A black sign with white text

Description automatically generated

**Unlocking Market Insights: A Predictive Framework for CSR/ESG News Impact on Investment Worthiness**

Chinmay Malhotra

2407168

*A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Master of Science in M.Sc. Business Analytics in the Faculty of Social Sciences and Law.*

**University of Bristol Business School**

September 2024

Word Count: 13603

**AUTHOR’S DECLARATION**

I declare that the work in this dissertation was carried out in accordance with the requirements of the University’s Regulations and Code of Practice for Taught Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, this work is my own work. Work done in collaboration with, or with the assistance of others, is indicated as such. I have identified all material in this dissertation which is not my own work through appropriate referencing and acknowledgement. Where I have quoted or otherwise incorporated material which is the work of others, I have included the source in the references. Any views expressed in the dissertation, other than referenced material, are those of the author.

I have read and understand the [University’s Academic Integrity Policy](https://www.bristol.ac.uk/media-library/sites/academic-quality/documents/policy/academic-integrity-policy.pdf) including plagiarism, contract cheating, falsifying data and breaching ethical research protocols



SIGNED: DATE: 04/09/2024

**Executive Summary**

This study develops a predictive framework using advanced machine learning (ML) and deep learning (DL) models to examine the impact of Corporate Social Responsibility (CSR) and Environmental, Social, and Governance (ESG) news on the investment worthiness of companies, as measured by abnormal stock returns. The research utilizes a comprehensive dataset, including financial data from CRSP and Compustat, market factors from the Fama-French model, and CSR/ESG news and ratings from Refinitiv’s LSEG platform, covering a broad spectrum of firms and events.

The empirical results indicate that CSR/ESG news has a limited but significant influence on abnormal stock returns, with varying effects depending on the nature and timing of the news. Specifically, environmental and human capital-related news tends to impact stock prices immediately, while corporate governance-related news shows a more delayed yet long lasting effect.

This research contributes to the literature by bridging the gap in understanding the dynamic and temporal aspects of CSR/ESG news on financial performance. The findings offer practical implications for improving investment decision-making and corporate governance through data-driven approaches, underscoring the importance of sustainability in financial markets.

**Table of Contents**

**Chapter I - Introduction and Background ……………………………………………….(**[5](#Chapter_1)**)**

1.1 Research Background **…………………………………………………………………...(**[5](#Chapter_11)**)**

1.2 Research Significance **…………………………………………………………………...(**[6](#Chapter_12)**)**

1.3 Research Aim, Questions and Objectives ……………………………………………….([7](#Chapter_13))

1.4 Research Structure ………………………………………………………………………([8](#Chapter_14))

**Chapter II – Literature Review and Hypothesis Development …………………………**([9](#Chapter_2))

2.1 Conceptual Framework and Theoretical Foundations …………………………………([10](#Chapter_21))

2.1.1 CSR/ESG ……………………………………………………………………([10](#Chapter_211))  
2.1.2 Technology in Finance ………………………………………………………([11](#Chapter_212))  
2.1.3 Event Studies in Finance ……………………………………………………([12](#Chapter_213))

2.2 Scope and Limitations of Existing Literature …………………………………………([13](#Chapter_22))

2.2.1 Initial Era: Simple Methodologies, Limited Scope and Linear Assumptions ([13](#Chapter_221))  
2.2.2 Evolving Era: Enhanced Data, Methods and Focus on Specific CFP Metrics ([14](#Chapter_222))  
2.2.3 Modern Era: Advanced Machine Learning and Dynamic Approaches ……([17](#Chapter_223))  
2.2.4 Research Gap ………………………………………………………………([18](#Chapter_224))

2.3 Bridging Research Limitations: An Analytical Framework …………………………..([19](#Chapter_23))

2.3.1 Hypothesis Formulation …………………………………………………….([20](#Chapter_231))

**Chapter III – Data and Methodology ………………………………………………**([21](#Chapter_3))

3.1 Methodology **…………………………………………………………………………**([21](#Chapter_31))

3.1.1 Research Methods …………………………………………………………([21](#Chapter_311))

3.2 Data Collection **………………………………………………………………………..**([22](#Chapter_32))

3.2.1 Stock and Financial Data from CRSP and Compustat ………………………([22](#Chapter_321))  
3.2.2 Market Factors Data from Fama-French ……………………………………([23](#Chapter_322))  
3.2.3 CSR/ESG News and Ratings Data from LSEG …………………………….([24](#Chapter_323))

3.3 Variable Selection **…………………………………………………………………..**([25](#Chapter_33))

3.3.1 Dependent Variables **……………………………………………………...**([25](#Chapter_331))  
3.3.2 Independent Variables **……………………………………………………**([27](#Chapter_332))  
3.3.3 Control Variables **…………………………………………………………**([28](#Chapter_333))

3.3..3.1 Fixed Control Variables **……………………………………….**([28](#Chapter_3331))  
3.3.3.2 Selected Control Variables …………………………………….([28](#Chapter_3332))

3.4 Model Comparison Setup and Framework ………………………………………...([29](#Chapter_34))

3.4.1 Final Predictive Model Framework ………………………………………([33](#Chapter_341))

**Chapter IV – Results and Findings ………………………………………………….**([34](#Chapter_4))

4.1 Initial Control Variable Selection Analysis **………………………………………..**([34](#Chapter_41))

4.1.1 Descriptive Statistics Analysis **…………………………………………..**([34](#Chapter_411))  
4.1.2 Variable Correlation Analysis **…………………………………………...**([36](#Chapter_412))  
4.1.3 Analysis Results **…………………………………………………………**([37](#Chapter_413))

4.2 Final Model Comparison Analysis and Framework Development ……………….([41](#Chapter_42))

4.2.1 Descriptive Statistics Analysis **………………………………………….**([41](#Chapter_421))  
4.2.2 Variable Correlation Analysis **…………………………………………..**([44](#Chapter_422))  
4.2.3 Hypothesis Testing using OLS Regression Results …………………….([45](#Chapter_423))  
4.2.4 Model Comparison Framework Results ………………………………..([49](#Chapter_424))

**Chapter V – Discussion and Conclusion …………………………………………...**([54](#Chapter_5))

5.1 Theoretical Contributions **…………………………………………………………**([55](#Chapter_51))

5.2 Practical Implications **……………………………………………………………..**([56](#Chapter_52))

5.3 Limitations and Recommendations for Future Research …………………………([57](#Chapter_53))

**References ……………………………………………………………………………**([59](#Chapter_6))

**Appendix ……………………………………………………………………………..** ([74](#Chapter_7))

Supervisor Agreement Form ………………………………………………… ([74](#supervisor_agreement))

Ethical Approval Form ……………………………………………………….. ([79](#ethics_approval))

**Chapter I – Introduction and Background**

#### *1.1 Research Background*

In the current business environment, integrating sustainability into corporate strategy is no longer just a trend but a necessity (Espahbodi et al., 2019). Environmental, Social, and Governance (ESG) factors, alongside Corporate Social Responsibility (CSR) initiatives, are increasingly vital in financial decision-making and strategic planning (Rumyantseva, 2022). This shift toward sustainability is driven by regulatory pressures as well as broader global trends, including the push towards net-zero emissions, the rise of stakeholder capitalism, and increased demands for social equity, making ESG and CSR considerations vital for assessing a company’s long-term viability and ethical standing.

The influence of CSR and ESG factors extends beyond internal corporate policies, significantly impacting financial markets by shaping investor behaviour and managerial decisions (Chang et al., 2022). As a result, there has been a growing body of research aimed at understanding the impact of CSR and ESG activities on corporate financial performance (Almeyda, 2019). However, the relationship between CSR/ESG factors and market dynamics is complex and often yields inconclusive results. Most studies focus solely on understanding this relationship but fail to consider its dynamic nature.

Traditional analyses have also struggled to incorporate a comprehensive set of variables that fully represent the landscape of this relationship, and they often lack a robust predictive framework that links CSR/ESG news to a company’s investment worthiness in a theoretically sound and practically useful way.

#### This study seeks to bridge this gap by exploring the dynamic interplay between CSR/ESG news and a company's investment worthiness.

#### *1.2 Research Significance*

The significance of this research lies in its potential to fill critical gaps in the existing literature by offering a nuanced understanding of how CSR/ESG news impacts investment worthiness. While prior studies have explored the correlation between CSR/ESG activities and financial outcomes, they have often fallen short in providing a conclusive, predictive analytical framework. Specifically, research is yet to fully explore the comprehensiveness in data attributes necessary for capturing the multifaceted nature of the CSR/ESG news – investment worthiness relationship, particularly its evolving and dynamic aspects over time.

#### This study will address these limitations by developing a comprehensive predictive modelling framework that encapsulates advanced machine learning (ML) and deep learning (DL) models, which are better suited to handle the complexities of such time-sensitive data and non-linear relationships. By using a more comprehensive set of variables, including those related to financial fundamentals, market factors and CSR/ESG news, this research seeks to improve decision-making for both investors and managers. In conclusion, by examining the effect of CSR/ESG-related news on stock abnormal returns, this study will contribute to a clearer and more actionable understanding of the CSR/ESG - investment worthiness dynamic. The ultimate goal is to empower all company stakeholders—both internal and external—with precise, data-driven insights, thereby contributing to more informed and sustainable financial practices.

#### *1.3 Research Aim, Questions, and Objectives*

The primary aim of this research is to explore and predict the impact of CSR/ESG-related news on a company’s investment worthiness (proxied by abnormal stock returns using event study methodology) over a dynamic period of up to one month following the release of such news. The study addresses the following central research question:

**Central Research Question**: *How does Corporate Social Responsibility (CSR) and Environmental, Social, and Governance (ESG) news impact a company’s investment worthiness, and how effectively can advanced machine learning models predict these effects?*

To further explore this, the research seeks to answer:

1. How does CSR/ESG news influence a company’s abnormal stock returns?
2. How do the effects of CSR/ESG news on abnormal stock returns evolve over time?
3. What are the key components within CSR/ESG news that drive changes in a company’s stock performance over time?
4. How effectively can advanced ML and DL models translate qualitative CSR/ESG news into quantitative financial outcomes, and predict these impacts to support strategic decision-making?

The research objectives that follow these questions are:

* To determine whether CSR/ESG-related news influences abnormal stock returns.
* To identify the key components within CSR/ESG news that significantly affect stock performance over the specified time frame.
* To compare traditional and advanced predictive models in terms of their ability to capture the complexities of the CSR/ESG-news relationship.
* To develop a predictive framework using the most effective model, aimed at enhancing decision-making for both investors and managers.

*1.4 Research Structure*

The structure of this dissertation is designed to systematically explore the aforementioned dynamic relationship. **The first chapter** sets the stage by discussing the increasing importance of sustainability in corporate strategy, driven by global trends such as net-zero emissions and stakeholder capitalism. It outlines the research's significance and aims to address gaps in the existing literature by developing a predictive framework for forecasting the impact of CSR/ESG news on a stock’s abnormal returns.

**The second chapter** delves into existing studies on the relationship between CSR/ESG factors and financial performance, highlighting the gaps present in them. This chapter also formulates the hypotheses that will guide this research.

**Meanwhile, the third chapter** details the research design, including data collection, variable selection, and the use of advanced machine learning (ML) and deep learning (DL) models. These methodologies are key to the predictive framework that this study thrives to design.

**Furthermore, the fourth chapter** presents the analysis results, testing the formulated hypothesis, comparing the performance of different models and examining how CSR/ESG news influences abnormal stock returns. This chapter provides insights into the factors driving investment worthiness.

Finally, **the last chapter** discusses the study's contributions, practical implications, and limitations, and offers recommendations for future research. This chapter emphasizes how the developed framework can enhance decision-making for investors and managers, contributing to more sustainable financial practices and concludes this study.

**Chapter II – Literature Review and Hypotheses Development**

The relationship between Corporate Social Responsibility (CSR), Environmental, Social, and Governance (ESG) factors, and Corporate Financial Performance (CFP) has been extensively studied in academic research (Klassen et al., 1996; Godfrey et al., 2009; Zhang et al., 2018; Guo et al., 2020). CFP is an umbrella term that includes a variety of outcomes related to a company’s financial health and investment attractiveness. Scholars have utilized diverse theoretical and practical frameworks, employing different data sources and analytical methodologies (Tsai et al., 2021; Xu et al., 2022; Pizzutilo, 2023). The presence of extensive variations in target variables and research approaches underpin the complexity of capturing the multifaceted and complex nature of this relationship.

This section aims to provide a comprehensive review of these studies, discussing their methodologies, findings, and limitations. The primary objective is to identify and present research gaps in existing literature that this study seeks to address.

This chapter is systematically organized to ensure a thorough exploration of the research topic. Section 2.1 introduces essential concepts, theories, and definitions, providing the necessary foundation for an in-depth review of existing literature. Section 2.2 follows with a comprehensive analysis of the scope of current literature, summarizing key studies, their contributions, and the limitations they reveal. In addition, this section identifies critical gaps in the research, highlighting areas that require further exploration. Finally, Section 2.4 explains how this study seeks to bridge these gaps by formulating and testing a hypothesis. This involves leveraging advanced methodologies to better represent data attributes and developing a predictive framework, both of which form the foundation of the academic and practical analysis and contributions in this research.

*2.1 Conceptual Framework and Theoretical Foundations*

*2.1.1 CSR/ESG*

Corporate Social Responsibility (CSR) encompasses a model by which corporations make a proactive and concerted effort to manage operations in ways that enhance, rather than degrade, the society and environment in which they operate. In contrast, Environmental (E), Social (S), and Governance (G) (ESG) are criteria used to measure a corporation’s overall environmental management, social improvement, and governance efforts. As Carroll (1999) explains: “There are four kinds of social responsibilities that constitute total CSR: economic, legal, ethical, and philanthropic. The CSR firm should strive to make a profit, obey the law, be ethical, and be a good corporate citizen”.

Despite the extensive increase in importance given to these terms, academicians still struggle with their precise definitions, making them elusive. They are often used interchangeably. For instance, ESG is often linked with CSR as a measure of its fulfilment by a corporation. However, both terms are considered constructs and lack a universally accepted definition (Trahan, 2023; Buniakova, 2021).

These terms have been integral to corporate assessment discussions since the early 1900s (Cochran, 1971). They have been connected and studied alongside the world of finance since the very beginning, as the name itself suggests: “Corporate” social responsibility. This study continues to use the same lineage and predict the impact of CSR/ESG-based news on the investment worthiness of a company.

*2.1.2 Technology in Finance*

The integration of technology in the field of finance began as early as the 1800s (Garbade, 1978). However, it gained significant traction only post the mid-1900s, particularly with the evolution of the data processing industry in the 1980s (Santarelli, 1995).

Today, technological advancements, particularly Artificial Intelligence (AI), play a significant role in the financial technology sector. AI, an umbrella term, is now being used or researched in every conceivable field and industry (Ergen, 2019; Abbass, 2021). One of the major fields where AI applications are widely used is finance. Specific applications include Machine Learning (ML) and Deep Learning (DL), which mimic human-like cognition but perform tasks faster inside machines. These technologies are used for tasks such as risk prediction, stock price forecasting, and volatility estimation (Arora, 2022; Singh, 2023; Yu, 2022; Nokeri, 2021; Mishev, 2020; Cheng, 2024).

For example, AI-driven algorithms are employed in high-frequency trading to execute trades at speeds and frequencies impossible for human traders (Dai, 2022). Additionally, even credit scoring models use ML to assess the creditworthiness of borrowers more accurately than traditional methods (Ramakrishnan, 2024).

The intersection of AI with CSR and ESG metrics provides a robust framework for analysing the financial health of corporations. These technologies help in determining and predicting based on the complex and intricate relationships between internal and external factors affecting a corporation’s financial health (Yu, 2022; Lee, 2024; Vo, 2019; Day, 2023; Guo et al., 2020).

For instance, deep learning models like Transformers are used for Natural Language Processing (NLP) tasks to estimate market sentiment based on user opinions or views expressed on various platforms such as social media and Google (Kokab, 2022). Similarly, these models are also used for processing news data for event studies based on CSR/ESG/Ethics-related events (Boudoukh, 2013).

However, the integration of AI in finance also presents challenges. Issues such as data privacy, model interpretability, and the potential for algorithmic bias need to be addressed to fully leverage AI’s potential in this field (Crookes, 2020; Giudici, 2018).

Overall, the advancements in technology, specifically AI, have significantly enhanced the ability to analyse and predict financial outcomes. These tools enable a deeper understanding of the factors influencing a company’s financial health and support more informed financial decision-making.

*2.1.3 Event Studies in Finance*

Event studies in finance refer to methodologies used to measure the effect of observed events on specific measures of a firm’s financial well-being (Sallinger, 1992; Thompson, 1995; Jong et al., 1991).

A wide variety of techniques can be used to perform event studies, but the most common and used one involves calculating and analysing abnormal returns (AR) before (estimation window), during (event window), and after (prediction window) the event. This approach helps to study the event’s impact and predict future effects based on the studied information on AR (Armitage, 1995).

For example, cumulative abnormal returns (CAR) are calculated over the aforementioned windows to assess the impact of various events, ranging from pandemics like COVID-19 to changes in ranking scales of indices such as the FTSE4Good Index, on the stock performance of related corporations (Chen, 2007; Curran et al., 2007). Other commonly studied events include mergers and acquisitions (M&A), regulatory changes, earnings announcements, and changes in leadership announcements (Cook, 2011).

Such event studies are particularly relevant to this research paper, as it focuses on the effect of observed CSR/ESG-related news on a company’s investment worthiness. These methodologies are well-suited to analysing CSR/ESG events due to their ability to capture the immediate and long-term impacts on financial performance irrespective of the complexity observed in their relationship with investment metrics. Numerous scholars have already conducted similar studies in their academic research (Dorfleitner, 2024; Serafeim et al., 2022; Keleş, 2023).

*2.2 Scope and Limitations of Existing Literature*

The relationship between Corporate Social Responsibility (CSR), Environmental, Social, and Governance (ESG) factors, and the Corporate Financial Performance (CFP) of a company has been a focal point in academic research and undergone significant evolution (Klassen, 1996; Bheenick et al., 2023). This evolution can be categorized into three distinct phases, each reflecting advancements in research methodologies, scope, and outcomes ([Table 1](#table_eralitreview)). In this section, we present a thorough review of the progression of these studies, highlighting their key findings, limitations, and the advancements that have shaped the current understanding of this relationship.

*2.2.1 Initial Era: Simple Methodologies, Limited Scope, and Linear Assumptions*

The earliest phase of research, beginning in the mid-1990s (Klassen, 1996; Boyle, 1997), sought to understand how CSR and ESG activities influenced a firm’s financial performance (CFP), affecting both external stakeholders like investors and internal stakeholders, such as employees (Sen, 2006). Early findings consistently showed that companies with strong CSR and ESG practices often benefited financially. Godfrey (2009) emphasized that firms with solid CSR efforts accumulated reputational capital, which acted as a buffer during challenging times. CSR-focused companies, as noted by Du (2007), gained competitive advantages through higher brand value and customer loyalty. Similarly, Daub (2005) found that enhanced customer satisfaction stemming from CSR initiatives fostered long-term support and loyalty. Collectively, these studies highlighted the importance of CSR in maintaining a company’s market position and financial stability.

A plethora of other studies also confirmed such positive relationships between CSR/ESG and different parts of CFP (Allouche, 2005; Sen et al., 2006; Du et al., 2010). Krüger (2015) identified a positive relationship between CSR events and shareholder wealth, with socially responsible firms experiencing stock price increases after positive CSR events. Additionally, Luo (2015) explored and found that firms with higher CSP receive more favourable stock recommendations from analysts, which in turn positively affect future stock returns, highlighting the mediating role of analysts in the positive social and financial performance relationship.

While the majority of studies supported a positive and significant correlation, a few presented contrasting views. Research by Brammer (2006) and Boyle (1997), using similar methodologies to previous studies, reported negative correlations between CSR/ESG and CFP. Additionally, McWilliams (2001) found no statistically significant relationship.

These inconsistencies, alongside the reliance on simple methodologies like linear regression, small sample sizes, and limited geographic scope, underscored the need for more rigorous and comprehensive research to fully capture the complexities noted in the CSR/ESG-CFP relationship.

*2.2.2 Evolving Era: Enhanced Data, Methods, and Focus on Specific CFP Metrics*

As the limitations of earlier studies became apparent, researchers in the next era sought to address these issues by utilizing larger datasets, expanding geographic coverage, and employing more advanced methodologies. This era marked a shift from the broad, generalized analyses of CSR/ESG impacts on financial performance towards more targeted studies that focused on specific aspects of Corporate Financial Performance (CFP).

*Methodological Advancements*

One of the key advancements in this era was the adoption of sophisticated methodologies. Studies began to move away from the simple linear regression models typical of earlier research and introduced more advanced techniques, such as two-stage least squares (2SLS) regression, and Natural Language Processing (NLP). Researchers increasingly shifted to applying larger sample sizes, longer time periods, and industry-agnostic approaches, which broadened the scope of their analyses. These advancements allowed researchers to analyse larger and more complex datasets with greater precision, providing more nuanced insights into the CSR/ESG-CFP relationship.

For instance, Serafeim (2022) demonstrated how advancements like NLP helped researchers to account for non-linear dependencies, revealing how ESG ratings influenced future ESG-related news. This, in turn, shaped a company's investment appeal as investors increasingly viewed ESG performance as a key indicator of long-term financial success. Furthermore, Serafeim (2022) found that ESG assets under management (AUM) rose by $20 billion between 2017 and 2019, reflecting how managers, recognizing the growing importance of ESG factors, increasingly prioritized CSR initiatives and worked to improve ESG ratings due to their perceived impact on CFP.

*Focus on Specific CFP Variables*

A major shift during this period was also the focus on specific financial metrics within CFP rather than examining overall corporate performance. Researchers targeted financial variables such as stock volatility, stock returns, and dividend payments to assess the effects of CSR/ESG initiatives. Studies like those by Bheenick (2023) and Al-Shammari (2022) explored CSR/ESG impacts using metrics such as Tobin’s Q and reported positive results, claiming that CSR/ESG initiatives contributed to overall firm value.

One of the earliest examples of research focusing on specific variables within CFP came from Klassen (1996), which assessed the impact of CSR/ESG news on corporate stock performance using cumulative abnormal returns (CAR). This foundational study paved the way for more in-depth explorations by scholars like Xu (2022), Pizzutilo (2023), and Havlinova (2021). They investigated how ESG performance influenced stock prices, revealing that positive ESG actions often led to significant improvements in stock value.

Similarly, Niccolo (2020) conducted another notable study on stock performance using dividend payments, finding that companies with higher ESG ratings generally showed a positive relationship with stock prices but a negative relationship with dividend pay-outs. This finding suggested that companies focused on CSR/ESG initiatives preferred to reinvest surplus funds into sustainable practices rather than distributing them to shareholders, thus highlighting different managerial strategies linked to ESG performance.

Dahal (2024) further explored this area, demonstrating that while CSR/ESG events generally had a positive effect on CFP, the significance of this relationship varied depending on external factors and firm characteristics, suggesting that CSR/ESG impacts were not uniform across industries or regions. This era also saw significant growth in the use of high-frequency data to capture the short-term effects of CSR/ESG activities on specific financial outcomes. Studies like Tsai (2021) and Szocs (2020) demonstrated that short-term stock performance and volatility could be impacted by CSR/ESG news, but only when analysed with high-frequency data, showing the importance of more time-sensitive analyses.

*Inconsistencies, Mixed Results and The Complexity of the Relationship*

While the majority of studies during this period confirmed a positive relationship between CSR/ESG activities and specific financial performance metrics, some studies produced conflicting results. For example, Liu (2023) reported mixed outcomes, where positive correlations were observed in parts of the dataset, neutrality in others, and even negative correlations in some instances. These mixed results underscored the multifaceted nature of the CSR/ESG-CFP relationship, highlighting that the impact of CSR/ESG initiatives could vary significantly depending on the specific context, dataset, or financial metric under analysis.

Although advancements in methodologies, such as the use of larger datasets and more sophisticated techniques, helped gain better insights into the CSR/ESG-CFP relationship, the intricacy of this relationship remained a significant challenge. Researchers acknowledged that despite improvements in analytical approaches, CSR/ESG impacts on CFP were highly context-dependent, influenced by external factors such as industry conditions, regulatory environments, and geographic differences (Dahal, 2024).

Moreover, while larger datasets and more specific financial metrics provided clearer insights, the lack of consistent results across all studies pointed to the need for even more advanced models and holistic approaches. The isolated improvements in either sample size or methodology, but rarely both, often limited the ability to generalize findings across different sectors and regions. Thus, despite the progress made during this era, the CSR/ESG-CFP relationship continued to present complexities that were difficult to fully capture with the existing methodologies.

In summary, this second era marked significant advancements in both the scope and depth of CSR/ESG research, moving from generalized analyses to more focused studies on specific CFP variables. However, while the majority of studies supported the positive link between CSR/ESG initiatives and financial performance, the ongoing challenges highlighted the need for continued methodological innovation to better understand the nuances of this multifaceted relationship.

*2.2.3 Modern Era: Advanced Machine Learning and Dynamic Approaches*

In the most recent era, researchers have increasingly turned to advanced technologies such as machine learning (ML) and deep learning (DL) in order to overcome the limitations of traditional methodologies and better understand the evolving CSR/ESG-CFP relationship. This shift represents a substantial leap forward in how the impact of CSR/ESG events on financial performance is studied. Researchers are now able to leverage vast and diverse datasets, applying sophisticated models capable of capturing the complexities and nuances that previous methods often overlooked.

While the second era saw the use of some advanced tools and techniques, the modern era has seen their full integration. Rather than applying these methods in isolation, contemporary research combines traditional models with ML and DL frameworks to provide deeper, more comprehensive insights into the dynamic relationship between CSR/ESG activities and CFP. Techniques such as Ordinary Least Squares (OLS) regression, the Fama-French Momentum Model, and GARCH, which were commonly used in earlier research (Zhang, 2018; Wang, 2019; Yen, 2019; Klassen, 1996), are now supplemented by a range of cutting-edge machine and deep learning algorithms. These include models like Support Vector Machines (SVM), Gated Recurrent Units (GRUs), and K-nearest Neighbours (KNN), which have been increasingly employed for their ability to handle complex, non-linear data structures (Xu, 2022; Teoh, 2019; Yu, 2022; Lee, 2024; Heng, 2022; Vo, 2019).

Additionally, state-of-the-art neural network architectures, including Bi-directional Recurrent Neural Networks (Bi-RNN) and Long Short-Term Memory (LSTM) networks, have become central to time-series prediction and pattern recognition in CSR/ESG-based data. These models are particularly effective in generating more accurate forecasts by capturing dependencies over time, which are crucial in analysing the long-term financial impacts of CSR/ESG initiatives. Transformer models, like Bidirectional Encoder Representations from Transformers (BERT) and its fine-tuned variants (e.g., Fin-BERT and RoBERTa), are widely used for comprehensive text analysis, helping researchers assess the influence of textual data on stock market behaviour and corporate performance (Day, 2023; Guo et al., 2020).

These advancements allow researchers to analyse larger datasets across extended time periods and broader industries, generating more reliable and precise results. The integration of these models has led to significant breakthroughs, affirming the significant correlation between CSR/ESG events and various aspects of corporate financial performance. By overcoming the shortcomings of traditional approaches, modern studies have hence established a strong foundation for future research and deeper exploration (Refer [Table 1](#table_eralitreview)).

*2.2.4 Research Gap*

Despite the progress made in the application of advanced models, several critical research gaps remain:

1. Lack of Comprehensive Data Attribute Exploration and Integration: Current research depicts insufficient depth and breadth of data attributes used. Even with the employment advanced techniques, the critical attributes related to CSR/ESG news and investment worthiness are either not fully explored or are examined in isolation, limiting the comprehensiveness of the analysis (Guo et al., 2020, Lee, 2024).
2. Static Analyses: A considerable portion of current research still relies on static analysis, which fails to account for the evolving and dynamic nature of the CSR/ESG news-investment worthiness relationship over time (Bheenick et al., 2023; Al-Shammari, 2022).
3. Lack of Practical Predictive Models and Frameworks: While studies have explored the correlation between CSR/ESG activities and investment worthiness and predicted its impact, there is still a lack of practical frameworks that can aid and guide end-to-end investment and managerial decisions (Yu, 2022; Lee, 2024; Heng, 2022).

To address these gaps, this study adopts a dynamic approach to analyse the relationship between CSR/ESG news and the investment worthiness of a company by developing a predictive model framework to forecast financial impacts over time. By thoughtfully comparing advanced methodologies like ML and DL, this study will identify the most effective approach, to explore this relationship by leveraging a wider range of relevant data attributes and critical variables, essential for capturing the CSR/ESG news-investment worthiness dynamic. This approach aims at empowering investors to make informed investment decisions, while also guiding managers to take proactive steps to boost their company’s investment appeal and overall performance.

*2.3 Bridging Research Limitations: An Analytical Framework*

Building on the insights from the critical literature review and the identified limitations, this study aims to validate the commonly observed significantly positive relationship between CSR/ESG/Ethics-related events and corporate financial performance. Specifically, this study aims to assess this by exploring the interaction between CSR/ESG news and investment worthiness of a company. By advancing the analyses through the integration of advanced technologies, a dynamic and robust predictive model framework will be developed, aiming to offer precise and actionable insights. The use of more extensive and relevant data attributes in these powerful models will instil confidence in the results, furthering the support to managerial decision-making within corporations and aiding investors, both institutional and individual, in making more informed investment choices.

*2.3.1 Hypotheses Formulation*

The study aims to test the following hypothesis, justified by the gaps identified in the literature review:

**Hypothesis:** There is a significant relationship between CSR/ESG related news and a company’s investment worthiness, reflected by the abnormal stock returns (AR) following the news release.

This hypothesis is designed to validate the widely claimed and accepted belief that CSR/ESG-related news leads to changes in stock performance. It also aims to address the need for checking the dynamic and time-specific relationship between stock abnormal returns and CSR/ESG news.

It can be more intuitively represented as:

*∆ CSR, ESG, Ethics News → ∆ Abnormal Stock Returns (AR)*

This study will not only test this hypothesis but also develop a predictive model framework aimed at providing strategic insights. The model will utilize a quantitative methods approach, combining both quantitative data analysis and predictive techniques. Additionally, the predictive framework will be made to ensure ethical and sustainable data collection, making it practical for both managerial and investment decision-making.

Based on the hypothesis and the objectives of this study, the following analytical framework and process flow is developed:



Figure 1: Analytical Framework

**Chapter III - Data and Methodology**

*3.1 Methodology*

This chapter sets the foundation for this study by detailing the research methodology and data collection procedures critical to addressing the hypothesis formulated. It outlines the diverse and high-quality sources of financial and CSR/ESG news data used, and emphasizes the meticulous methods employed to ensure data integrity, precision, and relevance throughout all analyses. The chapter further delves into the selection of dependent, independent, and control variables, which are key to building an effective predictive model framework. By incorporating a comprehensive set of both traditional financial metrics and textual analysis of CSR/ESG news, the study develops a robust framework aimed at evaluating the impact of corporate events on investment worthiness, offering valuable insights for investors and managers in making informed, data-driven decisions.

*3.1.1 Research Methods*

This study adopts a multifaceted approach to evaluate the impact of CSR/ESG news on a company’s investment worthiness. By integrating both quantitative financial data and qualitative news data, the research design leverages multiple high-quality sources to offer a holistic view of financial performance, market factors, and corporate sustainability news. Event study methodologies are employed to assess the effects of news events on investment worthiness.

To strengthen the analysis, advanced textual techniques are combined with traditional financial variables, providing a thorough understanding of the data. A range of statistical and machine learning models are applied to predict investment performance, forming the foundation for the predictive framework developed. This approach ensures a rigorous, data-driven assessment of how CSR/ESG news influences financial markets, delivering valuable theoretical insights and practical decision-making.

*3.2 Data Collection*

To ensure the collection of reliable, comprehensive, and relevant data, this study utilises multiple high-quality sources that are well-established in financial and academic research. The data is carefully selected to support various aspects of the study, including hypothesis testing, variable selection, model comparison, and framework development. By drawing data from four trusted sources—Center for Research in Security Prices (CRSP), S&P Compustat, Fama-French, and Refinitiv’s Data Platform (LSEG)—the study guarantees the robustness and accuracy necessary for both theoretical insights and practical applications.

These datasets provide the critical financial information, market factors, and CSR/ESG news essential to test the hypothesis and develop a dynamic predictive model framework, addressing the limitations identified in previous research (Liao, 2016; Stotz, 2021; Serafeim, 2022; Alareeni, 2020).

*3.2.1 Stock and Financial Data from CRSP and Compustat*

This study uses financial data from two widely trusted databases: CRSP and S&P’s Compustat, both known for their extensive and reliable coverage of North American publicly traded companies (Tsai et al., 2021; Khan, 2019). The data spans from January 1, 2016, to December 31, 2023, ensuring that all required historical financial information is available for thorough analyses.

*CRSP Data*

The CRSP Stock Database offers comprehensive stock data, including daily returns, trading volumes, and market capitalization for North American exchanges such as NASDAQ, NYSE, and AMEX (NYSE MKT). Data was collected from January 1, 2016, to December 31, 2022, for preliminary analysis and control variable selection, with additional data from January 1, 2023, to December 29, 2023, used for hypothesis testing, model comparison, and framework development.

To maintain data quality, only stocks with complete daily return data and active trading status were included. The dataset exclusively features trading days to ensure consistency and reliability. This curated dataset establishes a strong foundation for capturing the dynamic investment trends necessary to inform the study’s broader academic and practical contributions.

*Compustat Data*

The Compustat Database provides quarterly financial data from publicly traded companies, complementing the stock performance data from CRSP. It includes key fundamental financial metrics such as income statements, balance sheets, cash flow statements, and indicators like size, profitability, leverage, and valuation. Covering the same time period as CRSP (January 1, 2016, to December 31, 2023), this data enables a deeper exploration of the interaction between market variables and CSR/ESG news, offering additional insight into how financial outcomes are shaped.

*3.2.2 Market Factors Data from Fama-French*

This study incorporates market factor data from the Fama-French 5 Factor Model, sourced from the Fama-French database, to capture broader market movements that influence stock performance. Renowned for its ability to account for market and arbitrary stock changes (Yan, 2022; Serafeim, 2022), this model forms an essential foundation for this study. Data was collected from January 1, 2023, to December 31, 2023, ensuring alignment with other financial datasets for the final model comparison and framework development. Including these factors ensures that the analysis reflects real-world market conditions, enhancing the robustness of the overall analysis.

*3.2.3 CSR/ESG News and Ratings Data from LSEG*

In addition to financial and market data, this study incorporates CSR/ESG news and ratings data from Refinitiv's Data Platform (LSEG), known for its comprehensive coverage of corporate social responsibility, environmental, governance, and ethical practices (Dorfleitner, 2024; De Vincentiis, 2024; Khanchel et al., 2023). The dataset includes both news articles and ESG ratings, providing a thorough foundation for analysing the relationship hypothesis and accurately predicting a company’s investment potential.

The news data, spanning from May 2, 2023, to November 30, 2023, aligns with the financial and market data for model comparison and framework development. ESG ratings for 2023 are also included, capturing key metrics in this ever-evolving corporate sustainability landscape. Only complete and detailed news stories and ratings were selected to ensure data integrity. This combination of textual data and ratings provides a solid foundation for assessing the financial impact of CSR/ESG events on investment trends.

The collected data undergoes meticulous processing and cleansing to ensure its accuracy, consistency, and usability for further analysis. This involves removing incomplete and duplicate records, and aligning stock, financial, market, and CSR/ESG news data timelines. After filtering and merging CRSP and Compustat, only high-quality and complete records were retained. The final dataset integrates financial data with daily news coverage, focusing on companies with aligned data across all sources, ensuring robustness and consistency in the subsequent analysis (Refer [Table 2](#Table_preprocessinit) and [Table 3](#Table_preprocessfinal)).

*3.3 Variable Selection*

*3.3.1 Dependent Variables*

The dependent variable in this study's hypothesis testing and model comparison framework is the investment worthiness of a company. This is evaluated in relation to CSR/ESG and Ethics-related news, with Daily Abnormal Returns (AR) being used as a proxy.

Event Study Methodology (ESM) has been widely used in financial research since the late 1900s to measure the impact of various events, from corporate disclosures to public news, on a company’s financial performance (Jarrell, 1985; Krüger, 2015; Capelle-Blancard, 2019; Mu et al., 2023). While both Cumulative Abnormal Returns (CAR) and AR are used commonly in different contexts, this study opts for AR to capture the dynamic nature of stock price movements in response to news.

In prior literature leveraging ESM, several models have been employed to estimate AR, including the Market Model (Capelle-Blancard, 2019; Jacobs et al., 2010; Dorfleitner, 2024), the Market-Adjusted Model (Krüger, 2015), and the Fama-French 3 and 5 Factor Models (Xu, 2022; Serafeim, 2022). For the purpose of this study, the Fama-French 5 Factor Model (FFM 5) is selected due to its superior ability to predict expected stock returns and account for broader market movements (Yan, 2022).

The FFM 5 model can be represented as:

Where:

* ARᵢₜ is the abnormal return of an asset *i* at time *t*.
* Rᵢₜ is the actual return of an asset *i* at time *t*.
* R𝑓ₜ is the risk-free return at time *t*.
* α is the intercept, capturing unexplained returns.
* βₘₖₜ, βsmb, βhml, βrmw, βcma are sensitivities (betas) to the following factors:
  + Market risk premium (MKT): The excess return of the market over the risk-free rate.
  + Size premium (SMB): Small Minus Big, capturing the difference in returns between small-cap and large-cap stocks.
  + Value premium (HML): High Minus Low, distinguishing between high and low book-to-market ratios.
  + Profitability premium (RMW): Robust Minus Weak, accounting for differences in profitability.
  + Investment premium (CMA): Conservative Minus Aggressive, capturing differences in investment patterns.
* **ϵᵢₜ** represents the error term, capturing any residual variance not explained by the model.

To effectively apply the Fama-French 5 Factor Model (FFM 5) in this study, various windows are employed, as commonly found in previous research (Xu et al., 2022; Jarrell, 1985; Krüger, 2015; Capelle-Blancard, 2019). These windows are used to estimate the model's five-factor coefficients and calculate abnormal returns (AR). Typical setups include estimation windows spanning 120-250 days before the event, event windows ranging from -10 to +10 days around the event, and prediction windows that extend up to 90 days post-event. Given the fast-spreading and complex nature of CSR/ESG and Ethics-related news, this study adopts a [-120, -7] day window for estimation, a [-3, +3] day event window, and a separate [+4, +30] day prediction window. This structure allows for a precise and thorough comparison between the model's predictions and the actual abnormal returns, ensuring a thorough analysis of how such news influences a company's investment worthiness (Refer [Table 4](#Table_eventstudydef)).

*3.3.2 Independent Variables*

To assess the impact of CSR/ESG/Ethics-related news on a company’s investment worthiness, this study utilizes a comprehensive set of independent variables. These include key features derived from the news content data collected, which are described below:

1. **News Text Embeddings**: Representing the semantic information of the news articles, these embeddings are multi-dimensional (768 dimensions) vectors. They are generated from the hidden layers of a fine-tuned version of the Fin-BERT model, specifically ‘FinBERT-esg-9-categories,’ by inputting the collected LSEG news text data (Guo et al., 2020; Huang et al., 2023). This model, trained on financial text, is well-suited for analysing CSR/ESG content.
2. **News Story Topic & Subtopic Classification**: News articles are categorized into 9 subtopics related to CSR, ESG, and Ethics, which can further be classified into three main topics: Environmental (E), Social (S), and Governance (G) (Refer [Figure 2](#figure_esgtopicsubtopic)) (Capelle-Blancard, 2019; Xu et al., 2022). This classification is performed using the same fine-tuned FinBERT-esg-9-categories model (Huang et al., 2023).
3. **News Sentiment**: The sentiment of the news articles is classified into Positive, Neutral, and Negative categories. This classification is performed using the ‘FinBERT-tone’ model (Huang et al., 2023; Yu et al., 2023; Kim et al., 2019), which is specifically designed to evaluate sentiment in financial news.

These variables collectively provide a numeric representation of the news data collected. Rather than focusing on one aspect, such as sentiment or topic alone (Guo et al., 2020; Xu et al., 2022, Yu et al., 2023; Kim et al., 2019), this study integrates all three variables—embeddings, topic classification, and sentiment—to create a more holistic and representative dataset (Refer [Figure 3](#figure_newsfeaturesengg)). This approach follows Deng et al. (2023), arguing that human interpretation of news is complex and requires multiple layers of engineered representation to fully capture its effect on investment worthiness.

*3.3.3 Control Variables*

There are various factors, beyond CSR/ESG news, that influence a company’s investment worthiness. At times, these factors may even outweigh the direct impact of the news itself. Key financial metrics such as Size, Liquidity, Profitability, Leverage, Valuation, and Volatility play a pivotal role in assessing a company’s performance. Each metric can be measured using different proxies, with each proxy offering its own advantages and limitations (Nguyen, 2020; Dahiyat, 2016).

To identify the most relevant variables for this study, a combination of preliminary analysis, exploratory data analysis (EDA), and modelling is employed. These methods enable confident selection of variables that play a crucial role in predicting a company’s daily return (dlyret), which directly impacts abnormal returns (AR). While certain control variables were drawn directly from existing literature (Refer [Section 3.3.3.1](#Chapter_3331)), others were empirically tested due to varying uses across studies, allowing for the selection of the most effective proxies (Refer [Section 3.3.3.2](#Chapter_3332)).

*3.3.3.1 Fixed Control Variables*

Based on the review of relevant literature, the following fixed empirical control variables were selected for this study (Refer [Table 5](#table_fixcontrolsdef) and [Table 6](#table_fixcontrolsformulas)):

*3.3.3.2 Selected Control Variables*

While most studies utilize the given fixed control variables (as discussed in [Section 3.3.3.1](#Chapter_3331)), there is significant variation in the specific variables used to measure metrics like Size and Profitability. In this study, several empirical control variables were tested to select the most suitable proxies for these metrics (Refer [Table 7](#table_selectcontrolsdef) and [Table 8](#table_selectcontrolsformulas)):

*Selection Methodology*

To determine the most appropriate variable for measuring the Size and Profitability of a firm for final hypothesis testing, model comparison, and framework development, an initial regression was conducted using an ensemble machine learning model, **Extreme Gradient Boosting (XGBoost)**. XGBoost was selected due to its robust ability to handle large datasets and its proven effectiveness in predicting financial returns (Vuong et al., 2022; Worasucheep, 2022; Deng et al., 2023; Peng, 2024). The model was trained using **Time Series Split Cross Validation**, a method particularly suited for time-series data, providing a more accurate estimate of model performance by reducing variability (Vamsikrishna, 2024).

As outlined earlier, the control variables were fed into the XGBoost model, with Daily Returns (dlyret) set as the target variable. Given its close relationship with the final control variable, Abnormal Returns (AR), this approach allowed for a detailed comparison of six different combinations of Size and Profitability (3 Size and 2 Profitability proxies) metrics. The model’s performance was evaluated based on various metrics, including:

1. Mean Squared Error (MSE)
2. Mean Absolute Error (MAE)
3. Root Mean Squared Error (RMSE)
4. Goodness of Fit (R²)

The control variable combination that produced the best results on the test set was selected for the final analysis (Botchkarev, 2018; Dwivedi et al., 2023). [Figure 4](#figure_sixcombinations) illustrates the control variable combinations used for selecting the target Daily Returns (dlyret). Additionally [Figure 5](#figure_controlselectframework) outlines the applied framework for this selection.

*Selection Outcomes*

The trained XGBoost model is applied to the test set of financial data, and the best-performing model is chosen based on the metrics mentioned earlier (MSE, RMSE, MAE, R²). The final results of this selection process are detailed in Chapter IV ([Click Here](#Chapter_413)).

*3.4 Model Comparison Setup and Framework*

To address the limitations and complexity of the CSR/ESG news-investment worthiness relationship ([outlined in Chapter II Literature Review](#Chapter_224)), this study employs both traditional and advanced models to empirically test the formulated hypothesis. Initially, a linear OLS regression model tests the hypothesis, followed by a comparison of three state-of-the-art machine learning (ML) and deep learning (DL) models: Random Forest Regressor (RF), Light Gradient Boosting Machine Regressor (LightGBM), and Multi-Layer Perceptron (MLP) (Refer [Figure 6](#figure_modelcompare)). These models are chosen for their ability to capture non-linear relationships and computational efficiency (Gong et al., 2024; Martín-Cervantes, 2023; Rajabi et al., 2022).

Figure 6: Model Comparison Framework

All models are focused on predicting cube-root transformed abnormal returns (AR) for three specific days in the prediction window: +4th, +15th, and +30th days. This allows for capturing temporal dependencies and dynamic behaviour, an aspect often overlooked in prior research.

To handle the temporal dependencies in the dataset and accommodate the fact that these models are not inherently designed for time-series data, specific pre-processing steps are implemented. Control financial variables (both fixed and selected), as well as the target variable AR, are converted into lagged values within the estimation and event windows. This step not only leverages the autoregressive nature of these variables to enhance predictive accuracy (Garg et al., 2022) but also prevents any potential information leakage from the past into future observations, ensuring model reliability (Wang, 2019; Ahmed, 2010).

Transformations and standardization are applied to ensure all variables are machine-readable and that skewness in financial data is addressed (Passalis et al., 2021). Highly skewed features, particularly those related to multiple companies, can distort model predictions, so addressing this in the applied predictive models is crucial for balanced feature importance (Manning et al., 2005). A log(1+x) transformation is applied to control variables to handle zeros and small positive numbers (Tuvadaratragool, 2023; Lütkepohl, 2012). Meanwhile, the target variable (AR) undergoes a cube-root transformation to handle negative values and reduce skewness (Frecka, 1983).

Though the independent variables are not inherently skewed, they still require conversion into machine-readable formats to ensure compatibility with the used models. For categorical variables like Subtopic and Sentiment, one-hot encoding is applied, leaving n-1 columns (where n represents the number of categories). Additionally, the 768-dimensional BERT embeddings are flattened into 768 distinct columns to facilitate proper model input (Ranjan et al., 2022; Guo et al., 2020).

The final pre-processing step involves converting the daily stock data, along with attached news data, into a structured dataset where each row represents a single news article. This restructuring includes the independent variables and lagged financial data for control variables at specifically the first day of the event window: the 3rd day before the news release (-3rd). The chosen day, which is the first day within the event window of [-3, +3] trading days, aims to capture both immediate and past temporal trends, improving prediction accuracy (Wang, 2019; Garg et al., 2022).

Additionally, due to the high dimensionality of the flattened BERT embeddings, Principal Component Analysis (PCA) is applied to reduce the 768 variables to 17 principal components, accounting for 90% of the variability in the data. This step enhances computational efficiency while maintaining predictive accuracy (Bruna et al., 2022).

The final dataset, containing 3119 articles across 1562 companies ([Click Here](#github)), is first used for hypothesis testing. Specifically, the outcomes of the OLS Regression are employed to empirically test the hypothesis related to limitations identified in Chapter II ([Click Here](#Chapter_2)).

Where:

* ​represents the predicted abnormal return for the 4th, 15th, and 30th days following a news event.
* ​ is the intercept, representing the baseline level of abnormal returns.
* captures the contribution of ***k*** lagged financial control variables (e.g., daily volume, price-to-book ratio), where each ​ is the coefficient of a specific variable.
* accounts for the influence of ***m*** news-related variables (e.g., sentiment, subtopic), where each ​ represents the coefficient for a specific news variable.
* models the effect of ***n*** principal components derived from BERT embeddings, where each ​ is the coefficient of a reduced-dimension text embedding.
* represents the error term, capturing the residual variance not explained by the model.

After hypothesis testing, the data is split into training and test sets with time-dependency considerations. The first 70% of the data according to release date of news (2183 articles) is used for training, while the remaining 30% (936 articles) is reserved for testing. All ML and DL models—LightGBM, Random Forest (RF), and Multilayer Perceptron (MLP)—are trained, validated, and tested using Time Series Cross-Validation, as applied in the control variable selection process (Vamsikrishna, 2024).

The performance metrics—Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Goodness of Fit (R²)—are consistently used, as in earlier sections, to monitor and evaluate the generalizability and accuracy of all ML and DL models (Botchkarev, 2018; Dwivedi et al., 2023).

*3.4.1 Final Predictive Model Framework*

In light of the methodology employed throughout this study, a comprehensive framework is developed to assist in investment and managerial decision-making across companies. This framework integrates the entire process—from data collection, pre-processing, as well as variable and model selection to the application of insights derived from the selected models—helping both investors and managers make informed, data-driven decisions. It enables users to leverage predictive analytics for evaluating CSR/ESG news-related impacts on investment worthiness (Refer [Figure 7](#figure_overallframework)).

**Chapter IV - Results and Findings**

*4.1 Initial Control Variable Selection Analysis*

*4.1.1 Descriptive Statistics Analysis*

In order to thoroughly assess the financial dataset used for control variable selection, key statistical measures were employed to examine central tendencies and data dispersion. This analysis was critical to understanding the reasons behind the pre-processing steps taken, as outlined in Chapter III ([Click Here](#preprocessing_results)). These measures ensure that the dataset is consistent, reliable, and ready for further analysis.

[Table 9](#table_destatinit) below summarizes these descriptive statistics:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Count** | **Mean** | **Std** | **Min** | **Max** |
| **Daily Volume (Dlyvol)** | *6,058,077* | *1,458,492* | *6,106,032* | *1.0* | *1,341,812,000* |
| **Daily Market Capitalisation (Dlycap)** | *6,058,077* | *8,493,776* | *49,098,180* | *29.7* | *2,974,939,000* |
| **Total Quarterly Assets (Atq)** | *6,058,077* | *6,655.95* | *24,269.03* | *0.009* | *577,195* |
| **Quarterly Market Value (Mkvaltq)** | *6,058,077* | *8,914.27* | *52,177.89* | *0.008* | *2,901,645* |
| **Price to Book Ratio (Pb\_ratio)** | *6,058,077* | *9.423* | *957.40* | *-73,557.42* | *222,432.0* |
| **Debt to Equity Ratio (Debt\_to\_equity)** | *6,058,077* | *2.3849* | *373.66* | *-1,739.13* | *110,579.7* |
| **Current Ratio (Current\_ratio)** | *6,058,077* | *inf* | *N/A* | *0.000185* | *inf* |
| **Return on Equity (ROE)** | *6,058,077* | *-0.1471* | *51.04* | *-14,926.00* | *4,476.18* |
| **Return on Assets (ROA)** | *6,058,077* | *-0.0329* | *0.922* | *-166.00* | *147.0350* |

Table 9: Descriptive Statistics for all Control Variables

As illustrated in Table 9, the collected dataset exhibits significant variance across all variables, compelling rigorous transformation and standardization to render it suitable for machine learning applications used in this study. The Current Ratio (Current\_ratio), in particular, shows ‘infinite’ values for both mean and maximum, indicating the presence of extreme outliers that are incompatible with such machine learning models.

To address these challenges, variables with extreme outliers such as the Price to Book Ratio (Pb\_ratio), Debt to Equity Ratio (Debt\_to\_equity), Current Ratio (Current\_ratio), Return on Equity (ROE), and Return on Assets (ROA) were truncated to the 1st and 99th percentile values. This approach effectively reduces skewness and better centres the data, thereby making it more flexible for further standardization.

Given the large inherent skewness observed in variables like Daily Volume (Dlyvol), Daily Market Capitalization (Dlycap), Quarterly Market Value (Mkvaltq), and Total Quarterly Assets (Atq), a log(1+x) transformation was applied to these variables. Following this transformation, all numerical variables were standardized to ensure a normal distribution and ease in modelling (Refer [Figure 8](#figure_destatinit)).

These pre-processing steps ensure that the dataset is appropriately prepared for the advanced predictive modelling techniques employed in this study, i.e., Extreme Gradient Boosting (XGBoost).

*4.1.2 Variable Correlation Analysis*

To further refine the dataset for modelling, Pearson Correlation Coefficients are calculated to assess the relationships between all control variable proxies (Huang et al., 2022). While XGBoost is relatively robust to multicollinearity due to its decision-tree-based nature (Xia et al., 2024), addressing extreme multicollinearity is necessary, especially in financial datasets, to avoid skewed model interpretations and suboptimal performance.

To assess this, the correlation matrix (Refer [Table 10](#table_correlationinit)) and pairwise scatter plots alongside Kernel Density Estimation (KDE) plots are leveraged to visualize and validate relationships between all control variables (Refer [Figure 9](#figure_correlationinit)).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Dlyret** | **Dlyvol** | **Dlycap** | **Atq** | **Mkvaltq** | **Pb\_ratio** | **Debt\_to**  **\_equity** | **Current**  **\_ratio** | **ROE** | **ROA** |
| **Dlyret** | *1* |  |  |  |  |  |  |  |  |  |
| **Dlyvol** | *0.06* | *1* |  |  |  |  |  |  |  |  |
| **Dlycap** | *0.04* | *0.06* | *1* |  |  |  |  |  |  |  |
| **Atq** | *-0.01* | *0.13* | *0.42* | *1* |  |  |  |  |  |  |
| **Mkvaltq** | *0.03* | *0.06* | *0.86* | *0.39* | *1* |  |  |  |  |  |
| **Pb\_ratio** | *0.03* | *-0.00* | *0.48* | *-0.08* | *0.61* | *1* |  |  |  |  |
| **Debt\_to**  **\_equity** | *0.00* | *0.05* | *-0.13* | *0.13* | *-0.14* | *0.10* | *1* |  |  |  |
| **Current**  **\_ratio** | *0.00* | *0.03* | *0.10* | *0.04* | *0.11* | *-0.03* | *-0.21* | *1* |  |  |
| **ROE** | *0.01* | *-0.00* | *0.19* | *0.11* | *0.21* | *0.09* | *-0.19* | *0.10* | *1* |  |
| **ROA** | *0.01* | *0.00* | *0.21* | *0.13* | *0.23* | *0.11* | *-0.17* | *0.11* | *0.87* | *1* |

Table 10: Correlation Matrix for all Control Variables

Based on both the correlation matrix and pairwise visualizations, it is clear that variables such as Mkvaltq**,** Atq, and Dlycap exhibit moderate to high correlations, which is expected since they represent different dimensions of firm size. Likewise, ROE and ROA demonstrate a strong correlation (0.87), reflecting the overlap in how these metrics capture profitability.

However, the high correlation among these variables does not pose a challenge to this study. Since XGBoost is employed to select control variables related to Size and Profitability by comparing the model performance across six different combinations of fixed and selected variables, these correlated metrics will not be used together in the same model. This approach ensures that the high correlation among certain proxies does not interfere with or distort the outcomes of the study.

*4.1.3 Analysis Results*

After conducting a thorough exploration of variable correlations and descriptive statistics, this study now turns to the empirical evaluation of the six distinct combinations of firm size and profitability proxies alongside fixed control variables (Refer [Figure 4](#figure_sixcombinations)). The purpose of these combinations is to identify the most effective variables for predicting Daily Returns (Dlyret) and to inform subsequent hypothesis testing and model comparison as well as framework development involving Abnormal Returns (AR) as the target variable. By evaluating these various combinations, the analysis identifies the optimal set of control variables for further stages of this study.

Each combination was subjected to a Time Series Split Cross-Validation, ensuring the models were tested in temporally dependent environments, reflective of real-world financial prediction challenges. [Table 11](#table_resultsinit) and [Figure 11](#figure_resultsinit) below present the comparative results across all six models:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | **CV MSE** | **CV MAE** | **CV RMSE** | **CV R²** | **Test MSE** | **Test MAE** | **Test RMSE** | **Test R²** |
| Atq & ROE | *0.002* | *0.025* | *0.047* | *0.005* | *0.002* | *0.024* | *0.045* | *0.018* |
| Atq & ROA | *0.002* | *0.025* | *0.047* | *0.005* | *0.002* | *0.024* | *0.045* | *0.015* |
| Dlycap & ROE | *0.002* | *0.025* | *0.047* | *0.004* | *0.002* | *0.024* | *0.045* | *0.007* |
| Dlycap & ROA | *0.002* | *0.025* | *0.047* | *-0.002* | *0.002* | *0.024* | *0.045* | *0.001* |
| Mkvaltq & ROE | *0.002* | *0.025* | *0.047* | *0.003* | *0.002* | *0.024* | *0.045* | *0.016* |
| Mkvaltq & ROA | *0.002* | *0.025* | *0.047* | *0.003* | *0.002* | *0.024* | *0.045* | *0.015* |

Table 11: Results for all Control Variable Combinations

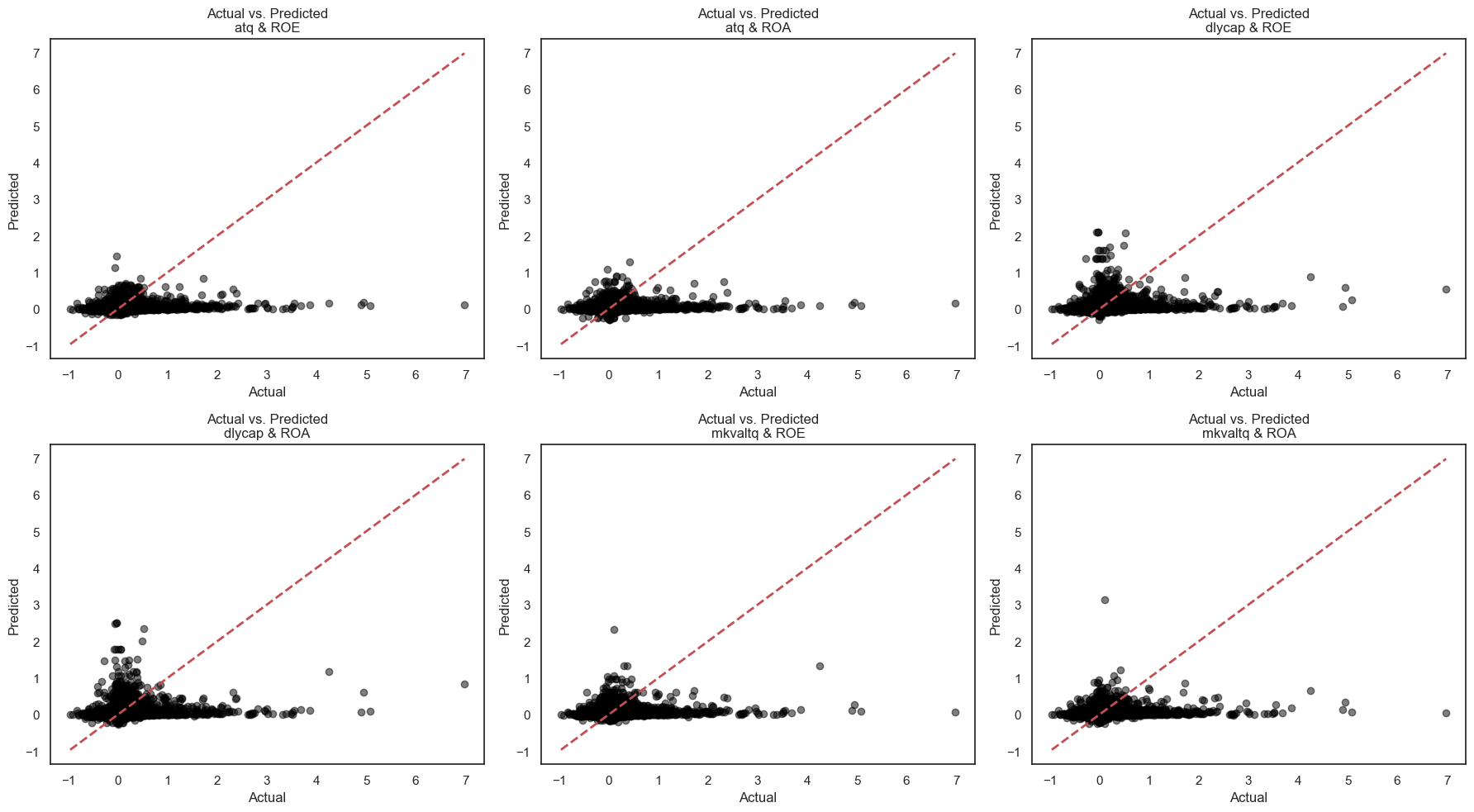


Figure 11: Results for all Control Variable Combinations

*Interpretation of Results*

The low R² values observed across all combinations are not unexpected, especially in financial modelling where predicting daily returns is notoriously complex. Financial markets are influenced by numerous unpredictable and external factors, which often result in models that capture only a small fraction of the variance in returns. This is consistent with prior research that shows low R² values are common in financial models (Kheradyar et al., 2011; Hussain et al., 2023; Wijesundera et al., 2016). Despite the low explanatory power of R², these models are still valuable, as even minor predictive gains can provide important insights for financial decision-making.

Where the models truly shine is in their performance based on the Mean Squared Error (MSE)**,** Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) metrics, which offer a clearer picture of the model's predictive accuracy. Across all combinations, the RMSE hovers around 0.045, indicating the model's predictions are quite close to the actual observed values, with only minimal error. The MSE values, ranging between 0.00204 and 0.00208, further confirm that the models are keeping prediction errors to a minimum, making them reliable for capturing even small movements in daily returns. Similarly, the MAE consistently stays between 0.02471 and 0.02476, showcasing that on average, the predictions deviate from the actual returns by a very small margin—approximately 2.5%. This further supports the model’s ability to predict returns with precision.

The scatter plots comparing "Actual vs. Predicted" values further reinforce these findings. In most combinations, the majority of data points cluster closely around the 45-degree line, indicating a strong alignment between the predicted and actual values. However, in combinations with higher error metrics—such as Dlycap & ROA—there are noticeable deviations from this line. These deviations highlight where the model faces difficulty in accurately predicting certain observations, particularly in cases where prediction errors are larger. These outliers correspond with the higher RMSE and MSE values, underscoring the variability in performance across different combinations.

When evaluating the six different combinations in parallel, the pairing of Total Quarterly Assets (Atq) and Return on Equity (ROE) stands out. This combination has the lowest Test MSEof0.00204 and the smallest Test RMSE of 0.04519, along with the lowest Test MAE of 0.02471. While the R² remains modest, it is the highest among the combinations, suggesting that this particular model has a slightly higher capacity to capture the variance in returns compared to the others.

In contrast, some combinations like Dlycap & ROA demonstrate a marginally weaker performance. Although the differences may seem minor, these subtle improvements in error minimization make Atq & ROE the more robust option for further modelling and hypothesis testing.

Overall, the combination of Atq & ROE offers the best balance between minimizing prediction errors, offering predictive accuracy, and generalizing well to new data. Its strong performance across all error metrics, coupled with minimal gaps between cross-validation and test set results, confirms that this model avoids overfitting and performs consistently. Thus, Total Quarterly Assets (Atq) & Return on Equity (ROE) are selected as the preferred control variables for the next phases of this study, including hypothesis testing, model comparison, and framework development.

*4.2 Final Model Comparison Analysis and Framework Development*

*4.2.1 Descriptive Statistics Analysis*

For the final data assessment utilized in hypothesis testing, model comparison, and framework development, similar statistical measures and techniques as those used in the control variable selection phase are applied. This ensures consistency and reliability throughout the analysis, while the inclusion of additional variables such as the ESG Score and Principal Components of the BERT Embeddings adds further depth to the final model. The objective is to ensure that the dataset remains robust for model comparisons across four different frameworks: Ordinary Least Squares (OLS) Regression, Random Forest Regressor (RF), Light Gradient Boosting Machine (LightGBM), and Multi-Layer Perceptron (MLP). Each of these models exhibits varying sensitivity to skewness in data—OLS being the most sensitive, MLP being moderately sensitive, and RF and LightGBM showing greater robustness (Sai et al., 2023; Peters, 1989). As a result, examining data distributions becomes essential to maintaining data reliability and ensuring accurate model performance.

*Lagged Financial Control Variables and ESG Score*

In this phase, the focus shifts to analysing the distributions of the lagged financial control variables and other independent variables like ESG Score. The same methods from the control variable selection analysis are employed, including descriptive statistics and Kernel Density Estimation (KDE) plots, to evaluate the data distributions. These visualizations and statistical summaries provide insight into the range, central tendencies, and dispersion of key financial metrics. The inclusion of both descriptive statistics and KDE plots offers a comprehensive understanding of the variables, enabling a deeper assessment of their readiness for model comparison. [Table 12](#table_destatfin) presents the key descriptive statistics for the dataset, followed by [Figure 12](#figure_destatfin) which further illustrates the distributional properties of the data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Count** | **Mean** | **Std** | **Min** | **Max** |
| **Dlyvol (-3)** | *3119* | *0.036* | *0.994* | *-3.138* | *6.121* |
| **Atq (-3)** | *3119* | *-0.029* | *0.992* | *-2.933* | *3.413* |
| **Pb\_ratio (-3)** | *3119* | *-0.073* | *0.971* | *-2.927* | *3.180* |
| **Debt\_to\_equity (-3)** | *3119* | *0.013* | *0.982* | *-3.537* | *3.571* |
| **Current\_ratio (-3)** | *3119* | *-0.165* | *0.964* | *-3.057* | *3.545* |
| **ROE (-3)** | *3119* | *-0.062* | *0.989* | *-3.292* | *3.517* |
| **Abnormal\_return (-3)** | *3119* | *-0.013* | *1.024* | *-2.696* | *3.926* |
| **ESG Score** | *3119* | *36.027* | *19.962* | *4.260* | *90.863* |

Table 12: Descriptive Statistics for Lagged Financial Variables

The descriptive statistics reveal that the dataset has been effectively transformed and standardized as done after noticing the significant skewness in financial variables (Refer [Section 4.1.1](#Chapter_411)), with most variables having means near zero and standard deviations close to one. However, despite this normalization, some variables exhibit broader ranges and some degree of skewness, highlighting the presence of outliers and long distribution tails, similar to the initial control variable selection analysis ([Click Here](#table_destatinit)). For instance, Daily Volume (Dlyvol) shows slight positive skewness, with longer tails indicating variability in trading volumes for certain firms. All other financial control variables exhibit near-symmetric KDE plots, reflecting a balanced mix of all kinds of company stocks within the dataset.

However, the ESG Score demonstrates a slightly right-skewed distribution, as seen in its KDE plot, with a mean value of 36.03. The wide range of values, from a range from 4.26 to 90.86, highlights the diversity of firms in terms of ESG performance. Firms with higher scores show greater commitment to sustainability, while those with lower scores reflect minimal ESG efforts. The presence of both extremes ensures that the dataset captures a broad spectrum of ESG activities, making this variable particularly useful for assessing the role of ESG in financial performance. The shape of the distribution, though not transformed like other variables, remains close to normal, offering a robust contribution to the analysis.

Overall, the dataset is well-prepared for further analyses, with standardized variables ensuring consistent means and deviations across time periods. While some skewness and outliers exist, typical in financial data, models employed in this study have proven to be robust to such minimal skewness (Sai et al., 2023). The broad representation of financial characteristics furthers the dataset’s suitability for supporting further analysis and hypothesis testing.

*Principal Components of BERT Embeddings*

Following the analysis of lagged financial variables, dimensionality reduction is applied to optimize the BERT embeddings used as independent variables ([As discussed in Chapter III Methodology](#bert_pca_reference)). A Scree Plot ([Figure 13](#figure_bertpca)) is then employed to determine how many Principal Components are necessary to retain most of the dataset’s variability while ensuring computational efficiency.

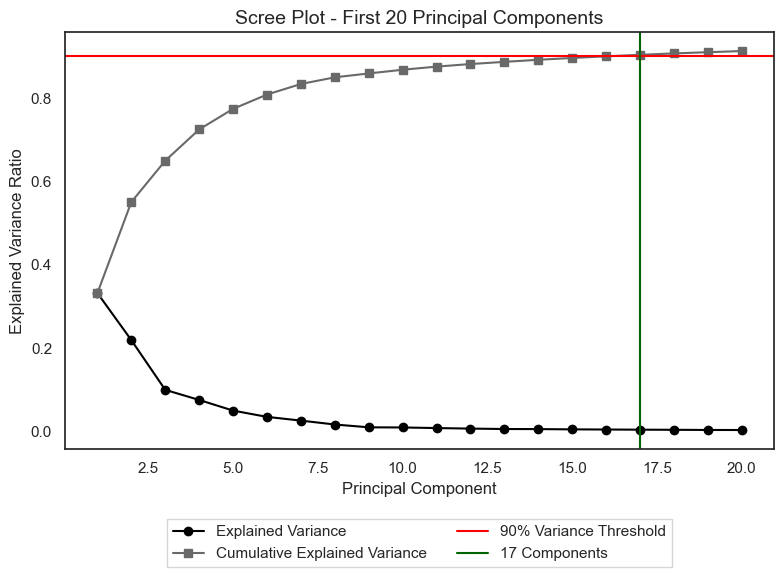


Figure 13: Scree Plot for BERT Embeddings PCA

The plot reveals that the top 17 components explain 90% of the variance, as highlighted by the green line. This sharp drop in explained variance after the first few components confirms that most of the essential information is concentrated within these initial components, making further reduction unnecessary.

By selecting these 17 principal components, the study achieves a balance between retaining critical information and reducing computational complexity, thus ensuring the efficiency and effectiveness of all predictive models (OLS, RF, LightGBM, MLP) without compromising on data quality (Bruna et al., 2022).

*4.2.2 Variable Correlation Analysis*

To assess relationships between all independent and control variables, Pearson Correlation Coefficients are calculated, excluding previously discussed principal components (Huang et al., 2022). Although the Machine Learning (ML) and Deep Learning (DL) approaches employed in this study, are robust at handling multicollinearity due to their various properties (Xia et al., 2024), addressing extreme multicollinearity is critical, especially in financial datasets, to ensure model efficiency.

The correlation table ([Table 13](#table_correlationfin)) displays the relationships between all variables used in final modelling except the principal components. The threshold chosen to focus on moderate to strong correlations that may indicate multicollinearity, particularly among lagged financial variables is 0.6. The table clearly indicates there are no strong correlations between time-lagged variables.

*4.2.3 Hypothesis Testing Using OLS Regression Results*

This study hypothesizes that there is a significant relationship between CSR/ESG/Ethics-related news and a company’s investment worthiness, as reflected in abnormal stock returns following the release of such news. To test this hypothesis, this study compares Ridge and OLS regression methods. This comparison allows to assess the statistical significance of both the independent and control variables in determining the dependent variable—Abnormal Returns (AR)—across three time windows: +4, +15, and +30 days post-news release. Additionally, through this, the analysis accounts for multicollinearity and the effect of regularization on the interpretation of coefficients. The results for all three regression models, corresponding to the 3 time windows, are presented in Table X (Refer to [Table 14](#table_olsresults) in the Appendix). In addition to the tables, Figure X (Refer to [Figure 14](#figure_olsresults) in the Appendix) visualises these results.

The results reveal that only a subset of the independent and control variables in this study are statistically significant, with significance tested at the 1%, 5%, and 10% levels across all three time periods. The R² values for abnormal returns (AR) range from 0.7% to 1.5%, indicating that the model explains only a small portion of the variation in AR. This limited explanatory power is further confirmed by the F-statistic values, where only the +4th day AR model exhibits statistical significance with an F-statistic p-value of 0.0992, while the +15th and +30th day models do not show statistical significance (p-values of 0.949 and 0.394, respectively). Although the +4th day model explains a small but meaningful proportion of changes in AR, the predictions for the +15th and +30th day windows are less robust. Furthermore, the independent and control variables that show statistical significance vary across the time periods, indicating temporal variation in their impact. This variation highlights the inherent complexities in financial data, where multiple factors contribute to market volatility that these models struggle to fully capture (Kheradyar et al., 2011; Hussain et al., 2023; Wijesundera et al., 2016).

In the first period, i.e., the +4th day AR post-news release, several variables stand out as statistically significant. Among the lagged financial control variables, the Pb ratio is significant at the 5% level (β = 0.0421, p-value = 0.028), suggesting that higher market valuations before the event lead to increased abnormal returns on the +4th day. This finding aligns with typical short-term market dynamics, where inflated valuations prompt investors to engage in profit-taking as the stock garners increased attention following a news release.

Turning to the news-related independent variables, the ESG Score is significant at the 1% level (β = 0.0025, p-value = 0.014), indicating that companies with strong ESG performance experience higher abnormal returns in the immediate aftermath of a news event. This finding highlights the importance investors take on ESG factors, supporting the broader notion of this study as well as past literature that ESG/CSR considerations do indeed matter, especially in the short term.

Additionally, several principal components derived from BERT embeddings of the news articles—namely PC7, PC12, and PC13—show statistical significance. PC7, significant at the 5% level (β = -0.0129, p-value = 0.025), suggests a negative impact on AR, likely tied to specific sentiments or themes within the news content. Meanwhile, PC12 and PC13, both significant at the 10% level (β = -0.0200, p-value = 0.082; β = -0.0221, p-value = 0.081, respectively), indicate that these components also capture underlying factors that negatively influence AR shortly after news is released, though with a slightly weaker effect.

Among the specific news subtopics, Natural Capital is particularly significant at the 5% level (β = -0.4765, p-value = 0.029), showing a strong negative impact on AR when environmental concerns related to a company come to light. This suggests that investors react adversely to news highlighting environmental issues. Similarly, the subtopic Non-ESG shows marginal significance at the 10% level (β = -0.1265, p-value = 0.062), indicating that non-ESG-related news can also have a negative impact on AR, though again to a lesser extent. For Non-ESG (Aside from the eight subtopics classified earlier using the Fin-BERT model (Refer to Chapter II)) news, the sentiment and content of these news articles contribute to their effect.

In the second period, i.e., the +15th day AR post-news release, the dynamics between the variables shift notably. At this stage, the Subtopic of Human Capital emerges as significant at the 5% level (β = 0.1684, p-value = 0.050). This suggests that positive news related to human capital—such as investments in the workforce or improvements in employee conditions—has a strong, positive influence on abnormal returns. This finding also underscores how qualitative factors, like company practices toward human capital, play a crucial role in shaping investor sentiment over a slightly longer time frame post-news release.

Interestingly, none of the financial control variables are statistically significant during this period, which marks notable differences from the earlier period analysis. This lack of significance suggests that the influence of prior financial performance on abnormal returns diminishes as time progresses. Instead, the impact of news content is continual, indicating that investors are continuously focused on the narrative and qualitative aspects of the news rather than just financial metrics. This shift highlights a growing tendency among investors to react to the broader implications of news, rather than solely relying on financial indicators.

In the third period, i.e., the +30th day AR post-news release, the trends from the second period become even more pronounced. The Subtopic of Corporate Governance emerges as significant at the 10% level (β = -0.0832, p-value = 0.080), indicating that news related to governance practices has a significant negative impact on abnormal returns in the medium to long term. This suggests that investors are particularly sensitive to corporate governance issues—such as changes in company leadership or management practices—which can have a more lasting effect on investor behaviour compared to other types of news. The persistent concern about governance reflects a deep-rooted caution, leading to a potential loss of confidence and a subsequent decline in stock prices.

This shift highlights a temporal progression in what drives abnormal returns. Initially, investors may react more strongly to environmental concerns and various types of news, but as time progresses, their focus shifts towards human capital and employee treatment. Ultimately, governance practices seem to have the most enduring impact on investor decisions, reflecting the significance of leadership and management integrity in sustaining investor confidence.

Interestingly, the ESG Score continues to show a lasting impact, reinforcing the idea that ESG practices do matter when predicting a company's investment worthiness—the central theme of this study. Additionally, principal components such as PC4 and PC13 exhibit varying levels of significance, with both being significant at the 10% level (β = 0.0057, p-value = 0.078; β = 0.0214, p-value = 0.083). This indicates that certain underlying factors captured by these components continue to influence abnormal returns over the medium term.

This study also observes the re-emergence of the influence of lagged financial control variables in this period. Specifically, the Abnormal Return control variable is significant at the 1% level (β = 0.0425, p-value = 0.016). This might suggest that as the impact of CSR/ESG-related news starts to fade, the market begins to stabilize, and investors gradually shift their focus back to past financial performance.

In summary, the key insight from these interpretations is that across all periods (+4th, +15th, and +30th day), abnormal returns (AR) are statistically significantly driven by CSR/ESG news released (but in a limited scope). Initially, investors place significant weight on the ESG Score and related performance, especially concerning the specific topic of the news, as the company comes into the spotlight. This heightened focus is natural as the news draws immediate attention to the company. As time progresses into the second period, the influence of the news persists but becomes less dominant. By the third period, while the news still plays a role, the market begins to normalize, and previous financial metrics regain importance. This shift suggests that as the immediate impact of the news fades, investors gradually shift their focus back to financial data, reflecting a return to a more rational assessment of the company’s stock performance over time, with less emphasis on market sentiment and more on financial fundamentals.

However, given the model's linearity assumption, the statistical significance (as reflected in the F-statistic), and the limited number of coefficients with significant explanatory power, it is clear that the relationship between CSR/ESG news and investment worthiness, as measured through AR, is highly complex. Linear regression alone cannot capture this complexity. **Thus, while the hypothesis that CSR/ESG news influences AR is supported to a limited extent, further validation using more sophisticated models is deemed necessary**. As discussed in Chapter II ([Click Here](#Chapter_223)), this study extends its investigation using advanced machine learning (ML) and deep learning (DL) models to capture these nuanced relationships more effectively.

*4.2.4 Model Comparison Framework Results*

The limitations of traditional linear regression models in capturing non-linear, complex relationships between CSR/ESG news and abnormal stock returns necessitate the use of more advanced models. Machine learning (ML) models such as Random Forest and LightGBM, along with deep learning (DL) models like the Multi-Layer Perceptron (MLP), are designed to handle larger feature spaces and non-linear interactions. Thus, these models are applied to the dataset to assess whether they can better predict abnormal returns by capturing intricate relationships between news content, financial performance, and stock reactions (as previously indicated in [Section 4.2.3](#Chapter_423)).

Upon analysing the results ([Table 15](#table_mldlresults)), we see that the R² values for all three models across different time periods (+4, +15, and +30 days) are negative. This indicates that none of the models outperform a simple mean-based prediction. While the ML models (Random Forest and LightGBM) slightly outperformed the MLP in R², all models show poor performance. For instance, the R² values for Random Forest and LightGBM across all time periods range from -0.011 to -0.079. In contrast, MLP performs worse, with R² values reaching -0.349, especially in the +15th day target.

The error metrics (MSE, MAE, and RMSE) across all time periods further emphasize the models' weak performance. The MSE for Random Forest and LightGBM hovers around 1.07, while for MLP, it is approximately 1.35. The MAE is consistently around 0.97 for all models, and the RMSE values range from 1 to 1.15, all of which indicate that the models fail to capture substantial deviations in abnormal returns. Although, these performance metrics do not show significant deviation in magnitude across different time periods, signalling the reliability of the results across different time frames, but also confirming the models' inability to generalize beyond simplistic patterns.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Name** | **Target Name** | **R2 Mean (CV)** | **MSE Mean (CV)** | **MAE Mean (CV)** | **RMSE Mean (CV)** |
| **Random Forest** | *Abnormal\_return (+4)* | *-0.011* | *1.054* | *0.959* | *1.026* |
| **LightGBM** | *Abnormal\_return (+4)* | *-0.073* | *1.118* | *0.957* | *1.057* |
| **MLP** | *Abnormal\_return (+4)* | *-0.304* | *1.359* | *1.007* | *1.165* |
| **Random Forest** | *Abnormal\_return (+15)* | *-0.014* | *1.019* | *0.946* | *1.009* |
| **LightGBM** | *Abnormal\_return (+15)* | *-0.076* | *1.081* | *0.946* | *1.039* |
| **MLP** | *Abnormal\_return (+15)* | *-0.349* | *1.355* | *1.012* | *1.164* |
| **Random Forest** | *Abnormal\_return (+30)* | *-0.010* | *1.005* | *0.940* | *1.002* |
| **LightGBM** | *Abnormal\_return (+30)* | *-0.079* | *1.073* | *0.943* | *1.036* |
| **MLP** | *Abnormal\_return (+30)* | *-0.333* | *1.326* | *0.998* | *1.151* |

Table 15: Model Comparison Results

Scatter plots of predicted versus actual values ([Figure 15](#figure_mldlscatter)) offer a visual confirmation of these numerical results. For Random Forest, the majority of predicted values are tightly clustered around zero, indicating that the models capture minor movements but fail to predict larger shifts in abnormal returns. LightGBM and MLP's scatter plots appear more dispersed, showcasing their struggle to make accurate predictions. This disparity in predictive power highlights that even with more complex models, significant market shifts remain abstract.

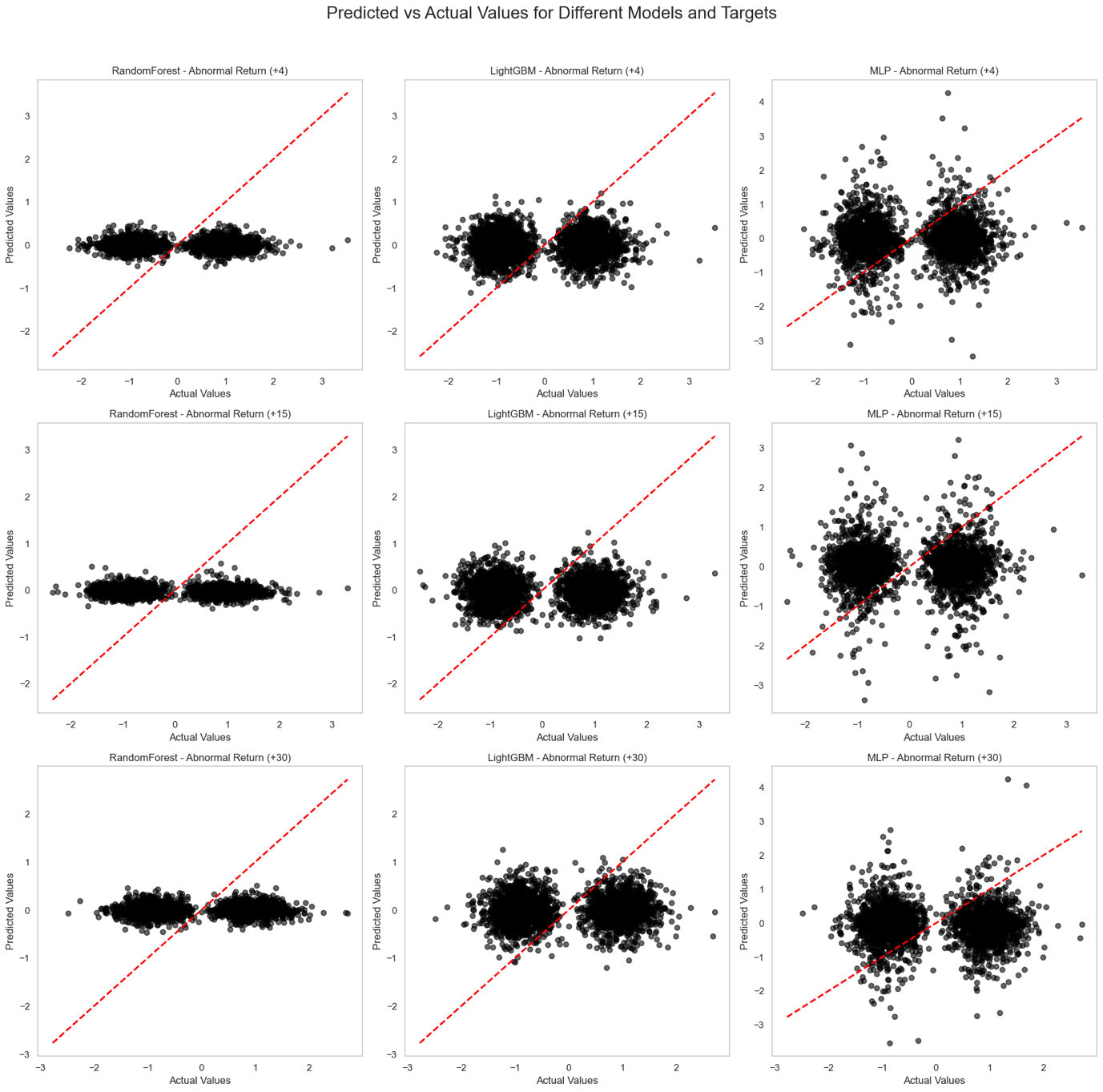


Figure 15: Model Comparison Results Visualised

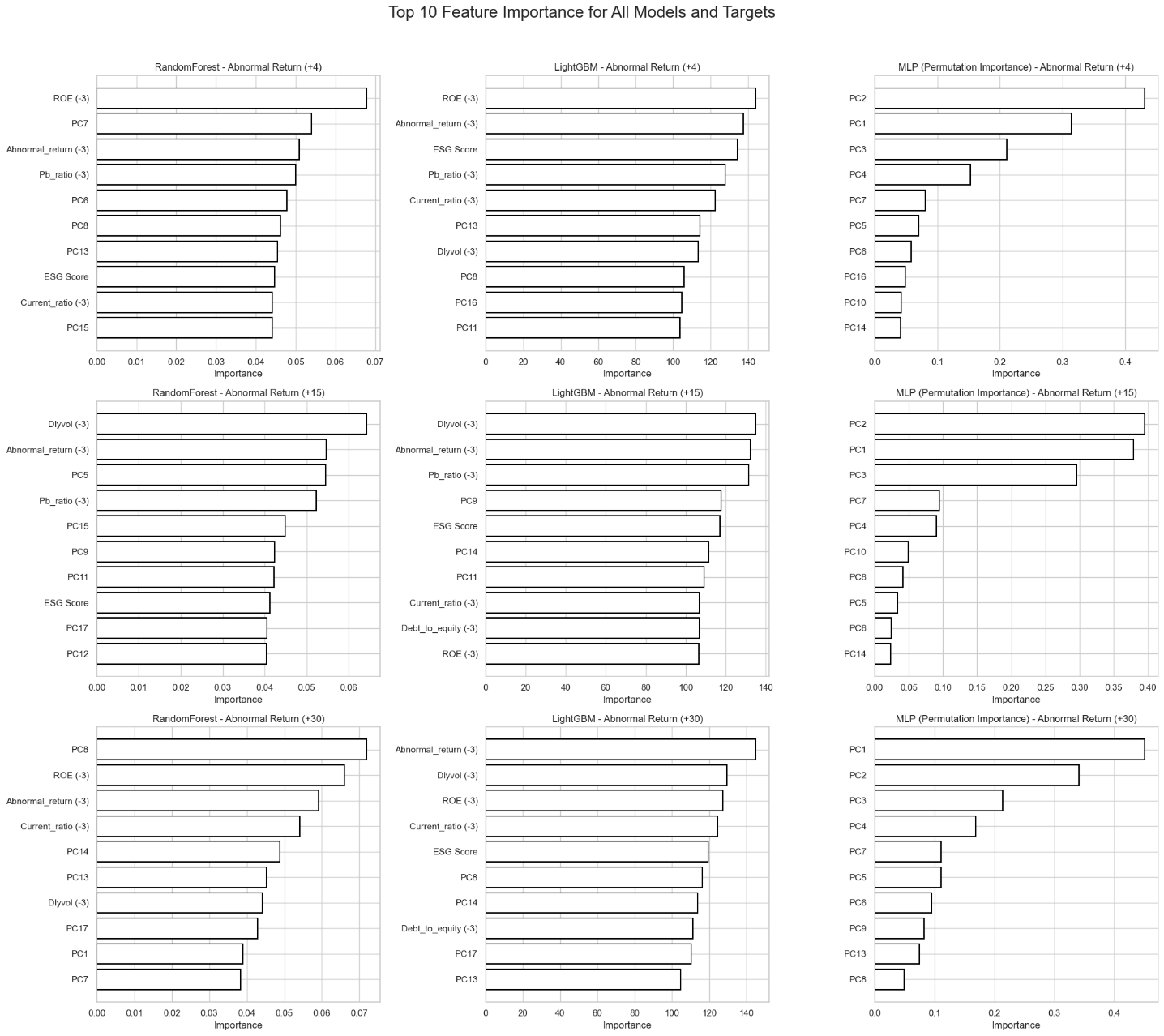
Feature importance analysis provides additional insights into the models' behaviours (Cheon et al., 2023) ([Figure 16](#figure_mldlfeatureimp)). For machine learning approaches, lagged financial control variables dominate the top spots in the feature rankings for all target values, with news-related independent variables like ESG Score and principal components (PCs) derived from BERT embeddings ranking lower or are fewer in quantity. This suggests that while financial controls such as past volatility, valuation and returns play a central role in predictions, the rich textual embeddings from news data offer limited value in these ML models. In contrast, the MLP model shows that all of its top-ranked features are the principal components, implying that while the model’s performance does not translate into meaningful predictive power, it is better equipped to handle nuanced, non-linear relationships such as the one in question. This also underlines the challenge of extracting actionable insights from news data, even when using sophisticated DL architectures.

Figure 16: Feature Importance for all Models

In conclusion, although advanced machine learning and deep learning models were employed in this study, the results did not show significant improvement compared to the initial findings generated by traditional linear models, such as OLS regression. This re-affirms the inherent complexity of the relationship between CSR/ESG news and a company's investment potential, as measured by abnormal returns (AR). In addition, these findings emphasize the need for further research to better understand and capture the nuances of this intricate relationship (Refer [Chapter V](#Chapter_5)).

**Chapter V – Discussion and Conclusion**

This study explores the relationship between CSR/ESG news and a company’s investment worthiness, with a focus on predicting abnormal stock returns (AR) following such news. To capture this relationship, the study employs a combination of traditional linear models, such as OLS regression for hypothesis testing, and more advanced machine learning (ML) and deep learning (DL) models like Random Forest, Light Gradient Boosting Machine (LightGBM), and Multi-Layer Perceptron (MLP) to further the understanding of this nuanced and complex relationship. The research is grounded in event study methodology, examining Abnormal Returns (AR) over three distinct time windows—+4th, +15th, and +30th day—after the news release to capture both immediate and delayed market reactions.

The results provide limited support for the hypothesis that CSR/ESG/Ethics-related news significantly influences a company’s investment worthiness, as reflected in abnormal stock returns. Traditional OLS regression models demonstrate low predictive power, with limited R² values and only a few statistically significant predictors of AR. In the short term (up to +4th day), stock returns are primarily influenced by CSR/ESG news, while the impact of financial variables like the Pb-ratio is minimal. Moving into the medium term (+15th day), the influence of financial metrics fades entirely, though the effect of news also starts to diminish. By the long term (+30th day), the impact of news fades further, and the influence of lagged financial controls resurfaces as the market normalizes and the initial reactions to the news dissolve. This suggests that while CSR/ESG news does impact stock returns, its effect is strongest in the short term and gradually weakens over time, with financial fundamentals eventually regaining importance.

Recognizing the limitations of traditional models, the study employs advanced ML and DL techniques to better capture the complex and non-linear nature of this relationship. However, these models do not show significant improvements in predictive performance. While ML models like Random Forest and LightGBM show marginally better generalization, their predictive power remains weak, especially for extreme AR values. The MLP model, which emphasizes the importance of Principal Components (PC) generated from BERT embeddings of the full news text, performed even worse, with a greater dispersion in predicted AR values. Nonetheless, MLP’s focus on news-derived features over financial variables highlights its potential to capture nuanced, highly non-linear relationships that simpler models may miss. This also reaffirms past literature's claim that the relationship between CSR/ESG news and financial performance is too complex and not easily quantifiable.

*5.1 Theoretical Contributions*

This study makes several important contributions to the theoretical understanding of the relationship between CSR/ESG news and investment worthiness:

Firstly, it enhances the scope of analysis by improving both data collection and feature engineering. The integration of textual analysis—using a combination of BERT embeddings, sentiment analysis, and topic classification—alongside all-inclusive, comprehensive financial data (including technical, fundamental, and market-related features) provides a novel approach for examining the market’s reaction to non-financial news. By combining natural language processing (NLP) with a robust set of financial metrics, this study underscores the growing importance of qualitative information in shaping investor sentiment and, ultimately, stock performance.

Secondly, the research affirms that CSR/ESG news does impact stock performance, but this effect is limited, often delayed and varies based on the type of news. For instance, governance-related news exerts a more significant influence over the long term, while short-term market responses are driven more by environment-related news. This provides a more nuanced understanding of how specific CSR/ESG factors—such as natural and human capital issues versus governance concerns—shape investor behaviour dynamically over time.

Lastly, the study highlights that while advanced ML and more specifically DL models have the potential to capture such non-linear relationships, not all such models can generalize effectively in the context of predicting abnormal stock returns from CSR/ESG news. The performance of these models can be limited by company-specific nuances, external market factors, and the inherent complexity of the CSR/ESG-news-investment worthiness relationship. This suggests that even with comprehensive datasets and sophisticated feature engineering as used in this study, certain factors—like sector-specific dynamics, macroeconomic conditions, and regulatory changes—are difficult to encapsulate fully. As a result, the generalizability of these models remains low. This emphasizes the need for continued exploration into more refined modelling approaches and the integration of additional variables that can better capture these complex and context-dependent relationships.

*5.2 Practical Implications*

In addition to the theoretical contributions, this study’s findings offer practical insights for both investors and corporate managers:

For investors, the nuanced relationship between CSR/ESG news and stock performance suggests that short-term market reactions tend to be highly driven by ESG performance and perception of the company in the headlines, while the long-term impacts of news are minimal and start to re-depend on lagged financial metrics and stock performance. This suggests that investors should look specifically at immediate stock price fluctuations following CSR/ESG news and incorporate non-financial information into their investment strategies, while in the long term should re-focus on financial information making better choices in the process.

For investors, the relationship between CSR/ESG news and stock performance indicates that short-term market reactions are largely influenced by how the company in the headlines is perceived by investors, particularly regarding its ESG performance. However, the long-term effects of such news are minimal, with the market gradually shifting its focus back to financial metrics and past stock performance during the news release period. This suggests that investors should pay close attention to immediate stock price movements following CSR/ESG news and consider integrating non-financial factors into their short-term investment strategies. In the long term, however, the emphasis should return to financial fundamentals, allowing for more informed and balanced investment decisions.

For corporate managers, the study highlights that CSR/ESG strategies, particularly those related to natural, human capital and governance, can have meaningful financial impacts. Positive CSR/ESG actions, especially those that are well-publicized and align with investor values, can boost stock performance over time. On the other hand, negative news, particularly regarding environmental and governance concerns, can adversely affect stock returns, as indicated in the short as well as long-term results. Therefore, managers should focus on activities like maintaining transparency, a cleaner carbon footprint, strong governance practices, and employee welfare programs to protect their company's reputation and financial standing.

Furthermore**,** this study introduces a new predictive framework—from data collection to advanced predictive modelling—that could help anticipate how markets might react to CSR/ESG-related news in the future. However, both investors and managers should be cautious when relying on advanced ML/DL models for stock prediction based on CSR/ESG news. The study indicates that even sophisticated models won’t necessarily improve predictive accuracy unless underlying data challenges are properly addressed. This serves as an important reminder for those leveraging cutting-edge technology to make informed decisions in the stock market.

*5.3 Limitations and Recommendations for Future Research*

### While this study provides valuable insights, several limitations remain that future research should address.

### Firstly, the relatively low predictive performance across all models suggests that key factors influencing stock returns may have been overlooked. Future studies should aim to incorporate additional variables such as macroeconomic indicators, investor sentiment, industry-specific dynamics, and time-sensitive factors like holiday effects or geopolitical events. This would create a more comprehensive understanding of how CSR/ESG news impacts financial outcomes and improve the models’ ability to capture market dynamics accurately.

### Secondly, the dataset used in this study was limited to a specific set of CSR/ESG news sources and covered a particular time period. To enhance the generalizability of findings, future research should expand the dataset to include a broader range of news outlets, real-time data, extended time periods, and even social media sentiment. Additionally, exploring the effects of CSR/ESG news on other financial metrics, such as volatility, trading volume, or dividend pay-outs, could offer a more holistic view of the relationship between sustainability initiatives and financial performance. Incorporating richer datasets that include global economic conditions, industry-specific shocks, and regulatory changes, alongside more granular data (High-Frequency or Ultra-High-Frequency), could further improve the accuracy of predictive models.

### Lastly, although this study utilized advanced ML and DL models, future research could explore newer AI approaches, such as state-of-the-art transformer-based models or reinforcement learning, which might offer better accuracy in predicting market reactions to CSR/ESG news. Emphasizing company-specific data rather than industry-wide aggregation could reduce ambiguity and provide clearer insights into the unique factors driving individual stock performance. Given the challenges posed by the qualitative nature of CSR/ESG news, future research should also focus on developing more sophisticated NLP techniques to better capture these nuances and provide more precise insights for investors.

In conclusion, this study presents limited but significant evidence that CSR/ESG news influences a company’s investment worthiness, with the impact varying based on the type of news and the timing horizon after its release. The complexity of this relationship makes it challenging to fully capture, even with the use of advanced ML and DL models. Therefore, enhancing data quality, refining feature selection, and improving model design is essential for future research to gain a deeper understanding of how CSR/ESG factors shape financial markets. Despite these challenges, the study offers valuable insights for investors and corporate managers, emphasizing the growing importance of CSR/ESG considerations in influencing market sentiment and driving long-term financial performance.

**References**

* A. A. V. I. Wijesundera et al. (2016) “Predictability of stock returns using financial ratios: empirical evidence from Colombo stock exchange,” Kelaniya Journal of Management, 4(2), pp. 44–55. Available at: <https://doi.org/10.4038/kjm.v4i2.7500>.
* Abbass, H. (2021) “Editorial: What is Artificial Intelligence?,” IEEE Transactions on Artificial Intelligence, 2(2). Available at: <https://doi.org/10.1109/TAI.2021.3096243>.
* Aguinis, H., & Glavas, A. (2012). What we know and don't know about corporate social responsibility: A review and research agenda. Journal of Management, 38(4), 932–968. <https://doi.org/10.1177/0149206311436079>
* Ajayi, O.A. and Mmutle, T. (2021), “Corporate reputation through strategic communication of corporate social responsibility”, Corporate Communications: An International Journal, Vol. 26 No. 5, pp. 1-15.
* Al-Shammari, M.A., Banerjee, S.N. and Rasheed, A.A. (2022) “Corporate social responsibility and firm performance: a theory of dual responsibility,” Management Decision, 60(6), pp. 1513–1540. Available at: <https://doi.org/10.1108/MD-12-2020-1584>.
* Alareeni, B.A. and Hamdan, A. (2020) “ESG impact on performance of US S&P 500-listed firms,” Corporate Governance: The International Journal of Business in Society, 20(7), pp. 1409–1428. Available at: <https://doi.org/10.1108/CG-06-2020-0258>.
* Allouche, J., & Laroche, P. (2005). A Meta-analytical investigation of the relationship between corporate social and financial performance. <https://hal.science/hal-00923906/>
* Almeyda, R. and Darmansya, A. (2019) “The Influence of Environmental, Social, and Governance (ESG) Disclosure on Firm Financial Performance,” IPTEK Journal of Proceedings Series, 0(5), p. 278. Available at: <https://doi.org/10.12962/j23546026.y2019i5.6340>.
* Anastasia V. Buniakova and Elena B. Zavyalova (2021) “ESG investment: a new word or a new world?,” RUDN Journal of Economics, 29(4), pp. 613–626. Available at: <https://doi.org/10.22363/2313-2329-2021-29-4-613-626>.
* Armitage, S. (1995) “EVENT STUDY METHODS AND EVIDENCE ON THEIR PERFORMANCE,” Journal of Economic Surveys, 9(1), pp. 25–52. Available at: <https://doi.org/10.1111/j.1467-6419.1995.tb00109.x>.
* Arora, G.D. et al. (2022) “Method of Predicting of Trend in the Stock Exchange using ML and DL Algorithms,” in 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), pp. 89–95. Available at: <https://doi.org/10.1109/IC3I56241.2022.10073453>.
* Atkinson, A.C., Riani, M. and Corbellini, A. (2020) “The analysis of transformations for profit-and-loss data,” Journal of the Royal Statistical Society: Series C (Applied Statistics), 69(2), pp. 251–275. Available at: <https://doi.org/10.1111/rssc.12389>.
  + Available at: <https://doi.org/10.1016/j.econmod.2023.106203>.
* Bharti, M. et al. (2022) “Cryptocurrency Stock Prediction using Deep Learning,” in 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC), pp. 612–617. Available at: <https://doi.org/10.1109/PDGC56933.2022.10053197>.
* Bhattacharya, A., Good, V., & Sardashti, H. (2020). Doing good when times are bad: The impact of CSR on brands during recessions. European Journal of Marketing, 54(9), 2049–2077. [https://doi.org/10.1108/EJM-01-2019- 0088](https://doi.org/10.1108/EJM-01-2019-%200088)
* Bissoondoyal-Bheenick, E., Brooks, R. and Do, H.X. (2023) “ESG and firm performance: The role of size and media channels,” Economic Modelling, 121.
* Boehmer, E., Masumeci, J. and Poulsen, A.B. (no date) “Event-study methodology under conditions of event-induced variance,” Journal of Financial Economics, 30(2), pp. 253–272. Available at: <https://doi.org/10.1016/0304-405X(91)90032-F>.
* Boudoukh, J., Feldman, R., Kogan, S., & Richardson, M. P. (2013). Which news moves stock prices? A textual analysis. Social Science Research Network. <https://autopapers.ssrn.com/sol3/papers.cfm?abstract_id=2207241>
* Boyle, E. J., Higgins, M. M., & Rhee, G. S. (1997). Stock market reaction to ethical initiatives of defense contractors: Theory and evidence. Critical Perspectives on Accounting, 8(6), 541–561. <https://doi.org/10.1006/cpac.1997.0124>
* Brammer, S., Brooks, C. and Pavelin, S. (2006) “Corporate Social Performance and Stock Returns: UK Evidence from Disaggregate Measures,” Financial Management, 35(3), pp. 97–116. Available at: <https://doi.org/10.1111/j.1755-053X.2006.tb00149.x>.
* Byun, S. K., & Oh, J.-M. (2018). Local corporate social responsibility, media coverage, and shareholder value. Journal of Banking & Finance, 87, 68–86. <https://doi.org/10.1016/j.jbankfin.2017.09.010>
* Capelle-Blancard, G. and Petit, A. (2019) “Every Little Helps? ESG News and Stock Market Reaction,” Journal of Business Ethics, 157(2), pp. 543–565. Available at: <https://doi.org/10.1007/s10551-017-3667-3>.
* Carhart, M.M. (1997) “On Persistence in Mutual Fund Performance,” The Journal of Finance, 52(1), pp. 57–82. Available at: <https://doi.org/10.2307/2329556>.
* Carroll, A.B. (1999) “Corporate Social Responsibility: Evolution of a Definitional Construct,” BUSINESS AND SOCIETY, 38(3), pp. 268–295.
* Chang, X. et al. (2022) “Sustainable Finance: ESG/CSR, Firm Value, and Investment Returns\*,” Asia-Pacific Journal of Financial Studies, 51(3), pp. 325–371. Available at: <https://doi.org/10.1111/ajfs.12379>.
* Chen, M.-H., Jang, S.(. and Kim, W.G. (2007) “The impact of the SARS outbreak on Taiwanese hotel stock performance: An event-study approach,” International Journal of Hospitality Management, 26(1), pp. 200–212. Available at: <https://doi.org/10.1016/j.ijhm.2005.11.004>.
* Cheng, J. et al. (2024) “Forecasting Bitcoin prices using artificial intelligence: Combination of ML, SARIMA, and Facebook Prophet models,” Technological Forecasting & Social Change, 198. Available at: <https://doi.org/10.1016/j.techfore.2023.122938>.
* Cheon, W. et al. (2023) “Feature Importance Analysis of a Deep Learning Model for Predicting Late Bladder Toxicity Occurrence in Uterine Cervical Cancer Patients,” Cancers, 15(13). Available at: <https://doi.org/10.3390/cancers15133463>.
* Cochran, T.C. (1971) “Book Review: The Social Responsibilities of Business: Company and Community, 1900-1960,” The Business History Review, 45(1), pp. 126–127.
* Connecting CSR and Stock Movements:
* Cook, A. and Glass, C. (2011) “Leadership change and shareholder value: How markets react to the appointments of women,” Human Resource Management, 50(4), pp. 501–519. Available at: <https://doi.org/10.1002/hrm.20438>.
* Crookes, L. and Conway, E. (2018) “Technology Challenges in Accounting and Finance,” in Contemporary Issues in Accounting : The Current Developments in Accounting Beyond the Numbers. Available at: <https://doi.org/10.1007/978-3-319-91113-7_4>.
* Dai, W., An, Y. and Long, W. (no date) “Price change prediction of Ultra high frequency financial data based on temporal convolutional network,” Procedia Computer Science, 199, pp. 1177–1183. Available at: <https://doi.org/10.1016/j.procs.2022.01.149>.
* Daub, C., & Ergenzinger, R. (2005). Enabling sustainable management through a new multi-disciplinary concept of customer satisfaction. European Journal of Marketing, 39(9/10), 998–1012. <https://doi.org/10.1108/03090560510610680>
* Day, M.-Y., Yang, C.-Y. and Ni, Y. (2023) “Portfolio dynamic trading strategies using deep reinforcement learning,” Soft Computing : A Fusion of Foundations, Methodologies and Applications, pp. 1–16. Available at: <https://doi.org/10.1007/s00500-023-08973-5>.
* De Jong, F., 2007. Event studies methodology. Lecture Notes.
* De Vincentiis, P. (2024) “ESG news, stock volatility and tactical disclosure,” Research in International Business and Finance, 68. Available at: <https://doi.org/10.1016/j.ribaf.2023.102187>.
* Di Tommaso, C. and Mazzuca, M. (2023) “The stock price of European insurance companies: What is the role of ESG factors?,” Finance Research Letters, 56. Available at: <https://doi.org/10.1016/j.frl.2023.104071>.
* Dirkx, P. and Peter, F.J. (2020) “The Fama-French Five-Factor Model Plus Momentum: Evidence for the German Market,” Schmalenbach Business Review, 72(4), pp. 661–684. Available at: <https://doi.org/10.1007/s41464-020-00105-y>.
* Dorfleitner, G. and Zhang, R. (2024) “ESG News Sentiment and Stock Price Reactions: A Comprehensive Investigation via BERT,” Schmalenbach Journal of Business Research, 76(2), pp. 197–244. Available at: <https://doi.org/10.1007/s41471-024-00185-3>.
* Du, S., Bhattacharya, C.B. and Sen, S. (2010) “Maximizing Business Returns to Corporate Social Responsibility (CSR): The Role of CSR Communication,” International Journal of Management Reviews, 12(1), pp. 8–19. Available at: <https://doi.org/10.1111/j.1468-2370.2009.00276.x>.
* El Ghoul, S. et al. (2022) “Event studies in international finance research,” Journal of International Business Studies, 54(2), pp. 344–364. Available at: <https://doi.org/10.1057/s41267-022-00534-6>.
* Ergen Keleş, F.H., Keleş, E. and Schlegelmilch, B.B. (2022) “Corporate social responsibility news and stock performance: A systematic literature review,” Borsa Istanbul Review [Preprint]. Available at: <https://doi.org/10.1016/j.bir.2022.11.017>.
* Ergen, M. (2019) “What is Artificial Intelligence? Technical Considerations and Future Perception,” Anatolian journal of cardiology, 22(Suppl 2), pp. 5–7. Available at: <https://doi.org/10.14744/AnatolJCardiol.2019.79091>.
* Espahbodi, L. et al. (2019) “Sustainability priorities, corporate strategy, and investor behavior,” Review of Financial Economics, 37(1), pp. 149–167. Available at: <https://doi.org/10.1002/rfe.1052>.
* Exterkate, P. et al. (2016) “Nonlinear forecasting with many predictors using kernel ridge regression,” International Journal of Forecasting, 32(3), pp. 736–753. Available at: <https://doi.org/10.1016/j.ijforecast.2015.11.017>.
* Fama, E.F. and French, K.R. (2015) “A five-factor asset pricing model,” Journal of Financial Economics, 116(1), pp. 1–22. Available at: <https://doi.org/10.1016/j.jfineco.2014.10.010>.
* FLAMMER, C. (2013) “CORPORATE SOCIAL RESPONSIBILITY AND SHAREHOLDER REACTION: THE ENVIRONMENTAL AWARENESS OF INVESTORS,” The Academy of Management Journal, 56(3), pp. 758–781.
* Galeone, G., Ranaldo, S. and Fusco, A. (2024) “ESG and FinTech: Are they connected?,” Research in International Business and Finance, 69. Available at: <https://doi.org/10.1016/j.ribaf.2024.102225>.
* Garbade, K.D. and Silber, W.L. (1978) “Technology, Communication and the Performance of Financial Markets: 1840-1975,” The Journal of Finance, 33(3), pp. 819–832. Available at: <https://doi.org/10.2307/2326479>.
* Giudici, P. (2018) “Fintech Risk Management: A Research Challenge for Artificial Intelligence in Finance,” Frontiers in Artificial Intelligence, 1. Available at: <https://doi.org/10.3389/frai.2018.00001>.
* Godfrey, P. C., Merrill, C. B., & Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. Strategic Management Journal, 30(4), 425–445. <https://doi.org/10.1002/smj.750>
* Gregor Dorfleitner and Rongxin Zhang (2024) “ESG News Sentiment and Stock Price Reactions: A Comprehensive Investigation via BERT,” Schmalenbach Journal of Business Research, 76(2), pp. 197–244. Available at: <https://doi.org/10.1007/s41471-024-00185-3>.
* Guo, T. *et al.* (2020) *ESG2Risk: A Deep Learning Framework from ESG News to Stock Volatility Prediction*.
* Haein Lee, Jang Hyun Kim and Hae Sun Jung (2024) “Deep-learning-based stock market prediction incorporating ESG sentiment and technical indicators,” Scientific Reports, 14(1), pp. 1–9. Available at: <https://doi.org/10.1038/s41598-024-61106-2>.
* Hamilton (no date) “Pollution as News: Media and Stock Market Reactions to the Toxics Release Inventory Data,” Journal of Environmental Economics and Management, 28(1), pp. 98–113. Available at: <https://doi.org/10.1006/jeem.1995.1007>.
* Harjoto, M.A. and Jo, H. (2015) “Legal vs. Normative CSR: Differential Impact on Analyst Dispersion, Stock Return Volatility, Cost of Capital, and Firm Value,” Journal of Business Ethics, 128(1), pp. 1–20. Available at: <https://doi.org/10.1007/s10551-014-2082-2>.
* Havlinova, A. and Kukacka, J. (2021) “Corporate Social Responsibility and Stock Prices After the Financial Crisis: The Role of Strategic CSR Activities,” Journal of Business Ethics, 182(1), pp. 223–242. Available at: <https://doi.org/10.1007/s10551-021-04935-9>.
* Huang, J. *et al.* (2022) ‘Analysis of Domestic Rebar Demand Based on Pearson Correlation Coefficient and XGBoost Model’, *Numerical Algebra and Scientific Computing by Clausius Scientific Press*, 1(1). Available at: <https://doi.org/10.23977/nasc.2022.010101>.
* Huanjing Wang et al. (2024) “Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods,” Journal of Big Data, 11(1), pp. 1–16. Available at: <https://doi.org/10.1186/s40537-024-00905-w>.
* Hui Do Jung and Beakcheol Jang (2024) “Enhancing Financial Sentiment Analysis Ability of Language Model via Targeted Numerical Change-Related Masking,” IEEE Access, 12, pp. 50809–50820. Available at: <https://doi.org/10.1109/ACCESS.2024.3385855>.
* Hussain, N. *et al.* (2023) ‘Stock Market Performance Analytics Using XGBoost’, in, pp. 3–16. doi: <https://doi.org/10.1007/978-3-031-47765-2_1>.
* Jacobs, B.W., Singhal, V.R. and Subramanian, R. (2010) “An empirical investigation of environmental performance and the market value of the firm,” Journal of Operations Management, 28(5), pp. 430–441. Available at: <https://doi.org/10.1016/j.jom.2010.01.001>.
* Jain, P. and A. Sunderman, M. (2014) “Stock price movement around the merger announcements: insider trading or market anticipation?,” Managerial Finance, 40(8), pp. 821–843. Available at: <https://doi.org/10.1108/MF-09-2013-0256>.
* Janina Witkowska (no date) “Corporate Social Responsability: Selected Theoretical and Empirical Aspects,” Comparative Economic Research, 19(1), pp. 27–43. Available at: <https://doi.org/10.1515/cer-2016-0002>.
* Jarrell, G. and Peltzman, S. (1985) “The Impact of Product Recalls on the Wealth of Sellers,” Journal of Political Economy, 93(3), pp. 512–536.
* Jong, F.d. et al. (1991) A contribution to event study methodology with an application to the Dutch stock market. Rotterdam: ECFR.
* Kamaliah (2020) “Disclosure of corporate social responsibility (CSR) and its implications on company value as a result of the impact of corporate governance and profitability,” International Journal of Law and Management, 62(4), pp. 339–354. Available at: <https://doi.org/10.1108/IJLMA-08-2017-0197>.
* Khadjeh Nassirtoussi, A. et al. (2014) “Text mining for market prediction: A systematic review,” Expert Systems With Applications, 41(16), pp. 7653–7670. Available at: https://doi.org/10.1016/j.eswa.2014.06.009.
* Khan, M. (2019) “Corporate Governance, ESG, and Stock Returns around the World,” Financial Analysts Journal, 75(4), pp. 103–123. Available at: <https://doi.org/10.1080/0015198X.2019.1654299>.
* Khan, M.A. et al. (2024) “Valuation effect of ESG and its impact on capital structure: Evidence from Europe,” International Review of Economics and Finance, 91, pp. 19–35. Available at: <https://doi.org/10.1016/j.iref.2024.01.002>.
* Khanchel, I., Lassoued, N. and Gargoury, R. (2023) “CSR and firm value: is CSR valuable during the COVID 19 crisis in the French market?,” Journal of Management and Governance, 27(2), pp. 575–601. Available at: <https://doi.org/10.1007/s10997-022-09662-5>.
* Kheradyar, S., Ibrahim, I. and Nor, F.M. (2011) “Stock Return Predictability with Financial Ratios,” International Journal of Trade, Economics and Finance, pp. 391–396. Available at: <https://doi.org/10.7763/IJTEF.2011.V2.137>.
* Kim, J. et al. (2019) “Stock Price Prediction Through the Sentimental Analysis of News Articles,” in 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), pp. 700–702. Available at: <https://doi.org/10.1109/ICUFN.2019.8806182>.
* King, B. G., & Soule, S. A. (2007). Social movements as extra-institutional entrepreneurs: The effect of protests on stock price returns. Administrative Science Quarterly, 52(3), 413–442. <https://doi.org/10.2189/asqu.52.3.413>
* Klassen, R. D., & McLaughlin, C. P. (1996). The impact of environmental management on firm performance. Management Science, 42(8), 1199–1214. <https://doi.org/10.1287/mnsc.42.8.1199>
* Koczar, J., Zakhmatov, D. and Vagizova, V. (2023) “Tools for considering ESG factors in business valuation,” Procedia Computer Science, 225, pp. 4245–4253. Available at: <https://doi.org/10.1016/j.procs.2023.10.421>.
* Krüger, P. (2015) “Corporate goodness and shareholder wealth,” Journal of Financial Economics, 115(2), pp. 304–329. Available at: <https://doi.org/10.1016/j.jfineco.2014.09.008>.
* Kuiper, C., & Adrián, G. (2020). The effect of ESG on stock prices : An event study on the S&amp;P 500. DIVA. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1438607&dswid=6229>
* Kulal, A., N, N. A., Dinesh, S., & MS, D. (2023). Impact of Environmental, Social, and Governance (ESG) factors on stock prices and investment performance. *Macro Management & Public Policies*, *5*(2), 14–26. https://doi.org/10.30564/mmpp.v5i2.5659
* Lamba, J., Jain, E. and 2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM) Dubai, United Arab Emirates 2022 Nov. 15 - 2022 Nov. 17 (2022) “Impact of Upsurge in Covid-19 Cases on Indian Stock Market: A Machine Learning Approach,” in 2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM), pp. 1–8. Available at: <https://doi.org/10.1109/ICCAKM54721.2022.9990082>.
* Lee, J., Koh, K.(. and Shim, E.D. (2024) “Managerial incentives for ESG in the financial services industry: direct and indirect association between ESG and executive compensation,” Managerial Finance, 50(1), pp. 10–27. Available at: <https://doi.org/10.1108/MF-03-2023-0149>.
* Li, H., Zhang, X. and Zhao, Y. (2022) “ESG and Firm’s Default Risk,” Finance Research Letters, 47. Available at: <https://doi.org/10.1016/j.frl.2022.102713>.
* Li, J. et al. (2022) “Stock Prediction Based on Deep Learning and its Application in Pairs Trading,” in 2022 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–7. Available at: <https://doi.org/10.1109/ISNCC55209.2022.9851776>.
* Li, Z. et al. (2019) “Event prediction via spatio-temporal sequence analysis,” in 2019 Chinese Automation Congress (CAC), pp. 1558–1563. Available at: <https://doi.org/10.1109/CAC48633.2019.8996903>.
* Li, Z. et al. (2022) “The impact of CSR and green investment on stock return of Chinese export industry,” Economic Research-Ekonomska Istraživanja, 35(1), pp. 4971–4987. Available at: <https://doi.org/10.1080/1331677X.2021.2019599>.
* Liao, Q. (2016) “The Stock Price Crash Risk Prediction by Neural Network,” Accounting and Finance Research, 5(2). Available at: <https://doi.org/10.5430/afr.v5n2p61>.
* Liu, X. and Yang, X. (2023) “ESG and the U.S. stock market equity returns – take the pharmaceutical industry as an example,” E3S Web of Conferences, 424. Available at: <https://doi.org/10.1051/e3sconf/202342404008>.
* Lu, J., Li, W. and Huang, W. (2024) “Corporate social responsibility and stock resilience to COVID-19: A contract theory perspective,” International Review of Economics and Finance, 89, pp. 12–29. Available at: <https://doi.org/10.1016/j.iref.2023.10.001>.
* Luo, X., Wang, H., Raithel, S., & Zheng, Q. (2015). Corporate social performance, analyst stock recommendations, and firm future returns. Strategic Management Journal, 36(1), 123–136. <https://doi.org/10.1002/smj.2219>
* Mahesh Dahal et al. (2024) “Mandatory CSR spending, ESG and firm performance: evidence from Bombay stock exchange,” Rajagiri Management Journal, 18(2), pp. 137–150. Available at: <https://doi.org/10.1108/RAMJ-08-2022-0132>.
* Martin Curran, M. and Moran, D. (no date) “Impact of the FTSE4Good Index on firm price: An event study,” Journal of Environmental Management, 82(4), pp. 529–537. Available at: <https://doi.org/10.1016/j.jenvman.2006.02.010>.
* McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. Academy of Management Review, 26(1), 117–127. <https://doi.org/10.5465/amr.2001.4011987>
* Mishev, K. et al. (2020) “Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers,” IEEE Access, 8. Available at: <https://doi.org/10.1109/ACCESS.2020.3009626>.
* Mu, X., Shi, Y. and International Academic Conference on Blockchain, Information Technology and Smart Finance Hangzhou, China February 17, 2023 - February 19, 2023 (2023) “An Event Study of the Impact of Negative ESG News on Stock Returns,” in Proceedings of the 2nd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2023), pp. 1384–1398. Available at: <https://doi.org/10.2991/978-94-6463-198-2_145>.
* Mücke, N.T. et al. (2023) “Markov chain generative adversarial neural networks for solving Bayesian inverse problems in physics applications,” Computers and Mathematics with Applications, 147, pp. 278–299. Available at: <https://doi.org/10.1016/j.camwa.2023.07.028>.
* Muller, N.Z. and National Bureau of Economic Research (2021) Measuring Firm Environmental Performance to Inform ESG Investing. Cambridge, Mass: National Bureau of Economic Research. Available at: https://www.nber.org/papers/w29454 (Accessed: June 13, 2024).
* Nakagawa, K., Masuda, T. and 2023 14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI) Koriyama, Japan 8 July 2023 - 13 July 2023 (2023) “Predicting Financial Asset Returns with Factor and Lead-Lag Relationships: Ridge Regression with Lagged Penalty,” in 2023 14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 534–539. Available at: <https://doi.org/10.1109/IIAI-AAI59060.2023.00107>.
* Niccolo, N. et al. (2020) “Shareholder value and dividend policy: the role of ESG strategies,” in 2020 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), pp. 1–5. Available at: <https://doi.org/10.1109/ICTMOD49425.2020.9380585>.
* Nokeri, T.C. (2021) “Market Trend Classification Using ML and DL,” in Implementing Machine Learning for Finance: A Systematic Approach to Predictive Risk and Performance Analysis for Investment Portfolios, pp. 143–165. Available at: <https://doi.org/10.1007/978-1-4842-7110-0_8>.
* Nsibande, L.M.Q. and Sebastian, A. (2023) “Is the environmental, social and corporate governance score the missing factor in the Fama-French five-factor model?,” South African Journal of Economic and Management Sciences, 26(1). Available at: <https://doi.org/10.4102/sajems.v26i1.4835>.
* Nyakurukwa, K. and Seetharam, Y. (2023) “Investor reaction to ESG news sentiment: evidence from South Africa,” EconomiA, 24(1), pp. 68–85. Available at: <https://doi.org/10.1108/ECON-09-2022-0126>.
* Oakley, P. (1997) “Finance and technological change: Theory and evidence,” Technovation, 17(1), p. 51. Available at: <https://doi.org/10.1016/S0166-4972(97)84222-8>.
* Patel D. and Oberai A.A. (2019) “Bayesian inference with generative adversarial network priors,” arXiv [Preprint].
* Peters, T.A. (1989) “The exact moments of ols in dynamic regression models with non-normal errors,” Journal of Econometrics, 40(2), pp. 279–305. Available at: <https://doi.org/10.1016/0304-4076(89)90086-9>.
* Pizzutilo (2023) “Is ESG-ness the vaccine?,” Applied Economics Letters, 30(4), pp. 484–487. Available at: <https://doi.org/10.1080/13504851.2021.1994124>.
* Qianyun Zheng et al. (2023) “TCN-GAT multivariate load forecasting model based on SHAP value selection strategy in integrated energy system,” Frontiers in Energy Research, 11. Available at: <https://doi.org/10.3389/fenrg.2023.1208502>.
* Ramakrishnan, R. et al. (2024) “Employing AI and ML in Risk Assessment for Lending for Assessing Credit Worthiness,” in 2024 2nd International Conference on Disruptive Technologies (ICDT), pp. 561–566. Available at: <https://doi.org/10.1109/ICDT61202.2024.10489313>.
* REVIEW OF ECONOMIC STUDIES (1991) “THE ECONOMETRICS OF FINANCIAL MARKETS,” The Review of economic studies [Preprint].
* Rumyantseva, A. and Tarutko, O. (2022) “Impact of the ESG Principles on the Corporate Financial Strategy,” in Challenges and Solutions in the Digital Economy and Finance: Proceedings of the 5th International Scientific Conference on Digital Economy and Finances (DEFIN 2022), St.Petersburg 2022, pp. 309–318. Available at: <https://doi.org/10.1007/978-3-031-14410-3_32>.
* Sai, M.J. et al. (2023) “An Ensemble of Light Gradient Boosting Machine and Adaptive Boosting for Prediction of Type-2 Diabetes,” International Journal of Computational Intelligence Systems, 16(1). Available at: <https://doi.org/10.1007/s44196-023-00184-y>.
* Salinger, M. (1992) “Value Event Studies,” The Review of Economics and Statistics, 74(4), pp. 671–677. Available at: <https://doi.org/10.2307/2109381>.
* Santarelli, E. and Santarelli, E. (no date) Finance and Technological Change: Theory and Evidence. New York, Secaucus: Palgrave Macmillan Springer [distributor]. Available at: https://link.springer.com/book/10.1007/978-1-349-39465-4 (Accessed: July 12, 2024).
* Scholes, M. and Williams, J. (no date) “Estimating betas from nonsynchronous data,” Journal of Financial Economics, 5(3), pp. 309–327. Available at: <https://doi.org/10.1016/0304-405X(77)90041-1>.
* Sen, S., Bhattacharya, C. B., & Korschun, D. (2006). The role of corporate social responsibility in strengthening multiple stakeholder relationships: A field experiment. Journal of the Academy of Marketing Science, 34(2), 158–166. <https://doi.org/10.1177/0092070305284978>
* Serafeim, G. and Yoon, A. (2022) “Stock price reactions to ESG news: the role of ESG ratings and disagreement,” Review of Accounting Studies, 28(3), pp. 1500–1530. Available at: <https://doi.org/10.1007/s11142-022-09675-3>.
* Singh, A. et al. (2023) “Harnessing ML and DL for Stock Market Analysis: A Tableau Forecasting Perspective,” in 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), pp. 1–5. Available at: <https://doi.org/10.1109/GCAT59970.2023.10353264>.
* Smith, K.T. and Huang, Y.-S.(. (2023) “A shift in corporate prioritization of CSR issues,” Corporate Communications: An International Journal, 28(1), pp. 68–85. Available at: <https://doi.org/10.1108/CCIJ-12-2021-0139>.
* Stotz, O. (2021) “Expected and realized returns on stocks with high- and low-ESG exposure,” Journal of Asset Management, 22(2), pp. 133–150. Available at: <https://doi.org/10.1057/s41260-020-00203-z>.
* Szocs, I., & Schlegelmilch, B. B. (2020). Embedding CSR in corporate strategies. In rethinking business Responsibility in a global context (pp. 45–60). Cham: Springer Nature.
* Tabinda Kokab, S., Asghar, S. and Naz, S. (2022) “Transformer-based deep learning models for the sentiment analysis of social media data,” Array, 14. Available at: <https://doi.org/10.1016/j.array.2022.100157>.
* Teoh, T.-T. et al. (2019) “From Technical Analysis to Text Analytics: Stock and Index Prediction with GRU,” in 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), pp. 496–500. Available at: <https://doi.org/10.1109/CIS-RAM47153.2019.9095772>.
* Theoretical definitions section:
* Thompson, R. (no date) “Chapter 29 Empirical methods of event studies in corporate finance,” in Handbooks in Operations Research and Management Science. Available at: <https://doi.org/10.1016/S0927-0507(05)80073-8>.
* Tiantian Meng, M. H. Yahya and Jingmin Chai (2022) “Deep Learning Model for Stock Excess Return Prediction Based on Nonlinear Random Matrix and Esg Factor,” 2022. Available at: <https://doi.org/10.1155/2022/5239493>.
* Trahan, R.T. and Jantz, B. (2023) “What is ESG? Rethinking the “E” pillar,” Business Strategy and the Environment [Preprint]. Available at: <https://doi.org/10.1002/bse.3371>.
* Trawinski, A. et al. (2022) “Sentiment Based LSTM for Stock Price Prediction: Congress vs General Public,” in 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 885–893. Available at: <https://doi.org/10.1109/ISMSIT56059.2022.9932752>.
* Tsai, H.-J. and Wu, Y. (2021) “Changes in Corporate Social Responsibility and Stock Performance,” Journal of Business Ethics, 178(3), pp. 735–755. Available at: <https://doi.org/10.1007/s10551-021-04772-w>.
* Vishwanathan, P., van Oosterhout, H. J., Heugens, P. P. M. A. R., Duran, P., & van Essen, M. (2020). Strategic CSR: A concept building meta-analysis. Journal of Management Studies, 57(2), 314–350. https://doi.org/10.1111/ joms.12514
* Vo, N.N.Y. et al. (2019) “Deep learning for decision making and the optimization of socially responsible investments and portfolio,” Decision Support Systems, 124. Available at: <https://doi.org/10.1016/j.dss.2019.113097>.
* Vu, T.N., Junttila, J.-P. and Lehkonen, H. (2024) “ESG news and long-run stock returns,” Finance Research Letters, 60. Available at: <https://doi.org/10.1016/j.frl.2023.104915>.
* Wang, H., Shen, H. and Li, S. (2023) “ESG performance and stock price fragility,” Finance Research Letters, 56. Available at: <https://doi.org/10.1016/j.frl.2023.104101>.
* Wang, Y., Delgado, M. S., Khanna, N., & Bogan, V. L. (2019). Good news for environmental self-regulation? Finding the right link. Journal of Environmental Economics and Management, 94, 217–235. <https://doi.org/10.1016/j.jeem.2019.01.009>
* Wen Long, T. et al. (2021) “News Event Prediction using Causality Approach on South China Sea Conflict,” in 2021 3rd International Cyber Resilience Conference (CRC), pp. 1–6. Available at: <https://doi.org/10.1109/CRC50527.2021.9392431>.
* Weng, X. et al. (2022) “Stock Price Prediction Based On Lstm And Bert,” in 2022 International Conference on Machine Learning and Cybernetics (ICMLC), pp. 12–17. Available at: <https://doi.org/10.1109/ICMLC56445.2022.9941293>.
* “What Is CSR?” (2014) in Corporate Social Responsibility: Definition, Core Issues, and Recent Developments. Available at: <https://doi.org/10.4135/9781483388014.n1>.
* Xia, Y. et al. (2024) “XGBoost-B-GHM: An Ensemble Model with Feature Selection and GHM Loss Function Optimization for Credit Scoring,” Systems, 12(7), p. 254. Available at: <https://doi.org/10.3390/systems12070254>.
* Xu, X. et al. (2022) “The Effect of ESG News on the Chinese Stock Market,” Journal of Global Information Management (JGIM), 30(1), pp. 1–21. Available at: <https://doi.org/10.4018/JGIM.313916>.
* Yadav, N. and Bhama, V. (2023) “Sustainability, Resilience, and Returns During COVID-19: Empirical Evidence from US and Indian Stock Markets,” Journal of Emerging Market Finance, 22(2), pp. 215–238. Available at: <https://doi.org/10.1177/09726527231158555>.
* Yang, X. and 2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID) Guangzhou, China 2020 Dec. 18 - 2020 Dec. 20 (2020) “CSR Disclosure and Stock Return Volatility in China: An Empirical Study Based on Panel Data Models,” in 2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID), pp. 55–58. Available at: <https://doi.org/10.1109/MSIEID52046.2020.00018>.
* Yen, M.-F., Shiu, Y.-M., & Wang, C.-F. (2019). Socially responsible investment returns and news: Evidence from Asia. Corporate Social Responsibility and Environmental Management, 26(6), 1565–1578. <https://doi.org/10.1002/csr.1833>
* Yu, G. et al. (2022) “Data Analysis of ESG Stocks in the Chinese Stock Market Based on Machine Learning,” in 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 486–493. Available at: <https://doi.org/10.1109/ICCECE54139.2022.9712837>.
* Yu, H. et al. (2023) “News-based ESG sentiment and stock price crash risk,” International Review of Financial Analysis, 88. Available at: <https://doi.org/10.1016/j.irfa.2023.102646>.
* Yu, X., Xiao, K. and Xu, T. (2023) “Does ESG profile depicted in CSR reports affect stock returns? Evidence from China,” Physica A: Statistical Mechanics and its Applications, 627. Available at: <https://doi.org/10.1016/j.physa.2023.129118>.
* Zanatto, C. et al. (2023) “The impact of ESG news on the volatility of the Portuguese stock market—Does it change during recessions?,” Business Strategy and the Environment, 32(8), pp. 5821–5832. Available at: <https://doi.org/10.1002/bse.3450>.
* Zhang, C. et al. (2024) “Do ESG ratings improve capital market trading activities?,” International Review of Economics and Finance, 93, pp. 195–210. Available at: <https://doi.org/10.1016/j.iref.2024.03.027>.
* Zhang, J., Djajadikerta, H., & Zhang, Z. (2018). Does sustainability engagement affect stock return volatility? Evidence from the Chinese financial market. Sustainability, 10(10), 3361. <https://doi.org/10.3390/su1010336>
* Zhang, Y. and Tang, X. (2023) “News event prediction by trigger evolution graph and event segment,” Journal of Systems Engineering and Electronics, 34(3), pp. 615–626. Available at: <https://doi.org/10.23919/JSEE.2023.000083>.
* Zhao, X., Fang, L. and Zhang, K. (2023) “Online search attention, firms’ ESG and operating performance,” International Review of Economics and Finance, 88, pp. 223–236. Available at: <https://doi.org/10.1016/j.iref.2023.06.025>.

**Appendix**

# MSc DISSERTATION / ARP / AEP

# STUDENT – SUPERVISOR AGREEMENT FORM

Instructions:

Students to complete this form, and both student and supervisor to sign it. The student saves it as a PDF and then uploads to the relevant Blackboard unit any time up to the deadline of Wednesday 21st June. Student and supervisor to both keep a copy for their records.

|  |  |
| --- | --- |
| MSc Programme  (including pathway if relevant) | Business Analytics |
| Name of student | Chinmay Malhotra |
| Student number | 2407168 |
| Name of supervisor | Tian Han |

|  |
| --- |
| Topic area [or AEP] |
| IBM Portal AEP |

|  |
| --- |
| Research Question |
| What is the most effective and ethical approach for developing a web-based tool to predict and recommend changes in a company's investment worthiness driven by ESG (Environmental, Social, and Governance) and ethics-based events? |

|  |
| --- |
| Proposed research outline |
| Justification  This project addresses the critical need for an integrated business analytics tool that enhances investment decisions by combining financial and non-financial (event based) data. Leveraging advancements in AI, such as ML, DL (specifically NLP), the tool aims to reduce the extensive time and resources currently required for such complex and often uninterpretable business analyses, ensuring efficient and comprehensive decision-making support. This tool also aims to provide investors with timely, actionable insights for real-time events, enhancing both academic research and practical investment strategies.  Methodology  The methodology for this project involves a structured 12-week timeline. Initial phases focus on recognizing reputable, reliable, and accessible sources for gathering multiple types of identified data, followed by the assessment of available business analytics tool and AI methodologies. The approach incorporates web scraping techniques. Identified and selected advanced ML and DL algorithms will then be applied to synthesize and analyse the collected data, culminating in the development of a prototype that demonstrates the tool’s capabilities.  **Timetable**  **Weeks 1-2**: Identifying data types and sources as well as ethical data gathering.  **Weeks 3-8**: in-depth research on selecting analyses and models, alongside addressing ethical concerns.  **Weeks 9-10**: Compilation of findings and report documentation.  **Weeks 11-12**: Prototype development and demonstration. |
| Agreed Scheduled Delivery of **Draft Chapters** [taking into account supervisor annual leave] |
| **Introduction and Research Background** – Week 5 (June 24th-30th)  **Literature Review and Research Design and Methodology** – Week 9 (July 22nd - 28th)  **Findings, Discussion and Conclusion**  – Week 12 (August 12th-18th) |

Terms and Conditions of the Student-Supervisor Agreement:

1. **Research proposal**

The proposed research to be conducted by the student is set out in this form. Any significant deviation from the plan for work described here should be discussed and agreed in a formal supervision meeting. It is the responsibility of the student to inform supervisors of any change to the plan of work.

1. **Supervision**

The maximum period of direct supervision provided to the student is five hours. The student is responsible for making the best use of this resource. A meeting schedule will be discussed at the first meeting. Supervisors will inform students when their annual leave will be taken over the summer period. Notes should be kept on the discussions and decisions made in supervision meetings. Best practice is for students to send notes on the discussions and decisions made in supervision meetings to supervisors by email immediately after meetings.

1. **Draft Chapters**

The supervisor should provide written feedback either within 10 working days for each submission of the draft - chapter by chapter - or within 15 working days for the draft of the whole dissertation. The protocol for receiving written feedback will be discussed during the first meeting. It is essential that students keep to the agreed schedule for delivery of drafts otherwise feedback and/or delivery within timeframes may be impacted due to supervisor leave or other commitments.

1. **Research Ethics**

The supervisor agrees to discuss the Ethics Form proposed by the student, and to grant approval when this is satisfactory. The form needs to be filled in and signed\* by both parties before any research is undertaken (including recruitment of participants). The student must upload the Ethics Form to the relevant Blackboard unit by the end of June. It is also the student’s responsibility to include the Ethics Approval Form (and any additional documentation) in the dissertation appendices.

1. **Agreement**

The Student and the Supervisor agree to work within this agreement.

This agreement must be signed \* by both parties and uploaded to Blackboard by the end of June. Copies should also be kept by the student and supervisor, so that the schedule as set out can be followed.

\*Any documents which require supervisor approval can be submitted to Blackboard and subsequently attached to the submitted dissertation using either of the following methods:

* Using digital signatures; or
* Using typed names plus attached confirmation email from the supervisor.

**We have read and agree to the Terms and Conditions of this agreement.**



Signature of Student: ………………………………………… Date: \_\_15\_\_\_ / \_\_5\_\_\_ / \_2024\_\_\_\_

Signature of Supervisor: ………………A black background with a black square

Description automatically generated with medium confidence………………… Date: \_28\_\_ / \_5\_\_\_ / \_2024

Notes on filling out the form

Subject Area

Your dissertation must be relevant to your named study programme and relate to the topic you have chosen (where relevant). You should not be writing a dissertation on a subject you haven’t studied. You can, however, draw on your first degree or past work experience.

**Title for the Proposed Research**

What is going to be researched? This might take the form of a research question or questions to be investigated, or a hypothesis to be tested.

Try to focus the title as much as possible and be succinct. Be specific, for instance: focus on an industry or a geographical area; name a company as an example or illustration or name a particular theoretical framework; or use a statement that you wish to challenge or a problem you want to address.

Note that the title can (and normally does) change over the course of your research, however a clear focus at the start makes the research process easier and the resulting dissertation is usually of a higher standard, normally because it is well structured. Indeed, you will be marked on the quality and appropriateness of your research question. Your supervisor will help you further with your research question and title during your first meeting. You should come to the first meeting with ideas you have developed.

Proposal Outline

Justification: explain why the research is important.   
This should take the form of a short critique of the main themes and controversies in the relevant literature as they relate to the research problem to be investigated. Identify gaps in the literature or problems eg with past methods of investigation that need to be addressed. You should also include your own reasons for wanting to study the area and why the results might be useful to you and others.

Outline the research methods you plan to use.

Briefly explain how you plan to carry out the research. How will you design any primary research? Will this overcome previous methodological weaknesses or follow examples of others in the field? What problems do you foresee and how do you plan to overcome them? You may have received methods training via a previous unit.

Conclude by giving a (draft) timetable of your research.

Include significant events (such as completion of on-line interviews or chapters) and set deadlines for yourself and in relation to planned delivery of draft chapters to your supervisor for comment. Take your supervisor’s annual leave into consideration as you plan your timetable.



**University of Bristol Research Ethics Application**



Title

Mr

First Name

Chinmay

Surname

Malhotra

Faculty

Faculty of Social Sciences and Law

Department

School of Management - Business School

School

School of Management - Business School

Telephone

Email

gs23170@bristol.ac.uk

Application Submitter Details

Preferred Name or Also Known As

**Investigator information**



Faculty



Social Sciences and Law

School / Department / Centre



University of Bristol Business School

Are you a student submitting this ethics application as part of your degree qualification?



Yes

Please declare your level of study



Taught Masters

Title

Dr

First Name

Tian

Surname

Han

Department

Management

Faculty

Faculty of Social Sciences and Law

Email

tian.han@bristol.ac.uk

Supervisor Contact Details

Are you an academic member of staff submitting an ethics application on behalf of a student(s) as part of their degree qualification?



No



Second Supervisor Details. If University of Bristol, please provide their full name and title.

If external to the University of Bristol, please provide their name, organisation details, email address and telephone number.

Please provide details of any other researchers/collaborators involved in the study.

Are you submitting this ethics application on behalf of another researcher?



No



Has or will your research be submitted to another research ethics committee for research involving human participants, their tissue

and or data?

Yes

No

**Important Information - Please note:**

It is extremely important that you select the

**correct Research Ethics Committee (REC)**

to review your research ethics application.

The REC selected, will determine the questions you are asked to complete on this online form and the research ethics committee that

will review your research ethics application.

**Please note**

, if you select the incorrect ethics committee, this may delay the review of your ethics application as your ethics

application will need to be returned to you so that you can select the correct REC and complete the relevant questions on the online

form.

If you are unsure of the correct research ethics committee to select please contact

research-ethics@bristol.ac.uk

Please select the Research Ethics Committee (REC) to review your research ethics application:



Business School Research Ethics Committee

To proceed to the next page select 'Next' in the Actions tiles.

To save your application for completion and submission at a later date please select 'Save' in the Actions tiles.

**Ethics Committee Review**





Brief Project Outline (up to approximately 300 words in plain English)

This project investigates the impact of ESG (Environmental, Social, and Governance), CSR (Corporate Social Responsibility), and

Ethics-related events on the investment worthiness of publicly traded companies. Utilizing comprehensive financial datasets from

Compustat and CRSP, alongside real-time ESG news from Bloomberg Terminal's API, the project analyzes and predicts the

relationship between these events and stock price and company valuation movements. The goal is to develop a predictive model

assessing both the immediate and medium-term effects of such news disclosures on a company's market value.

Advanced machine learning and data science techniques, including FinBERT and RoBERTa for NLP tasks, Principal Component

Analysis (PCA) for dimensionality reduction, and Generative Adversarial Networks (GANs) paired with Stochastic Gradient Markov

Chain Monte Carlo (SGMCMC) simulations for Bayesian Inference, are employed to train the model on historical data from the past

decade. The model estimates expected returns before ESG events, captures immediate returns after the events, and predicts future

stock performance over a 90-day horizon. This predictive tool aims to aid investors and financial analysts in making informed

decisions based on ESG factors, promoting sustainable and responsible investment practices. Additionally, the tool integrates IBM's

Watson Assistant to create a user-friendly interface for strategic suggestions using generative AI techniques.

Estimated dates of research data collection

Please note that you must not start data collection for your research project until you have received a formal favourable ethics

opinion from the appropriate Research Ethics Committee (REC). Please factor in the research ethics review timescales when

selecting the estimated dates for research data collection.

Anticipated Start Date

01/07/2024

Anticipated End Date

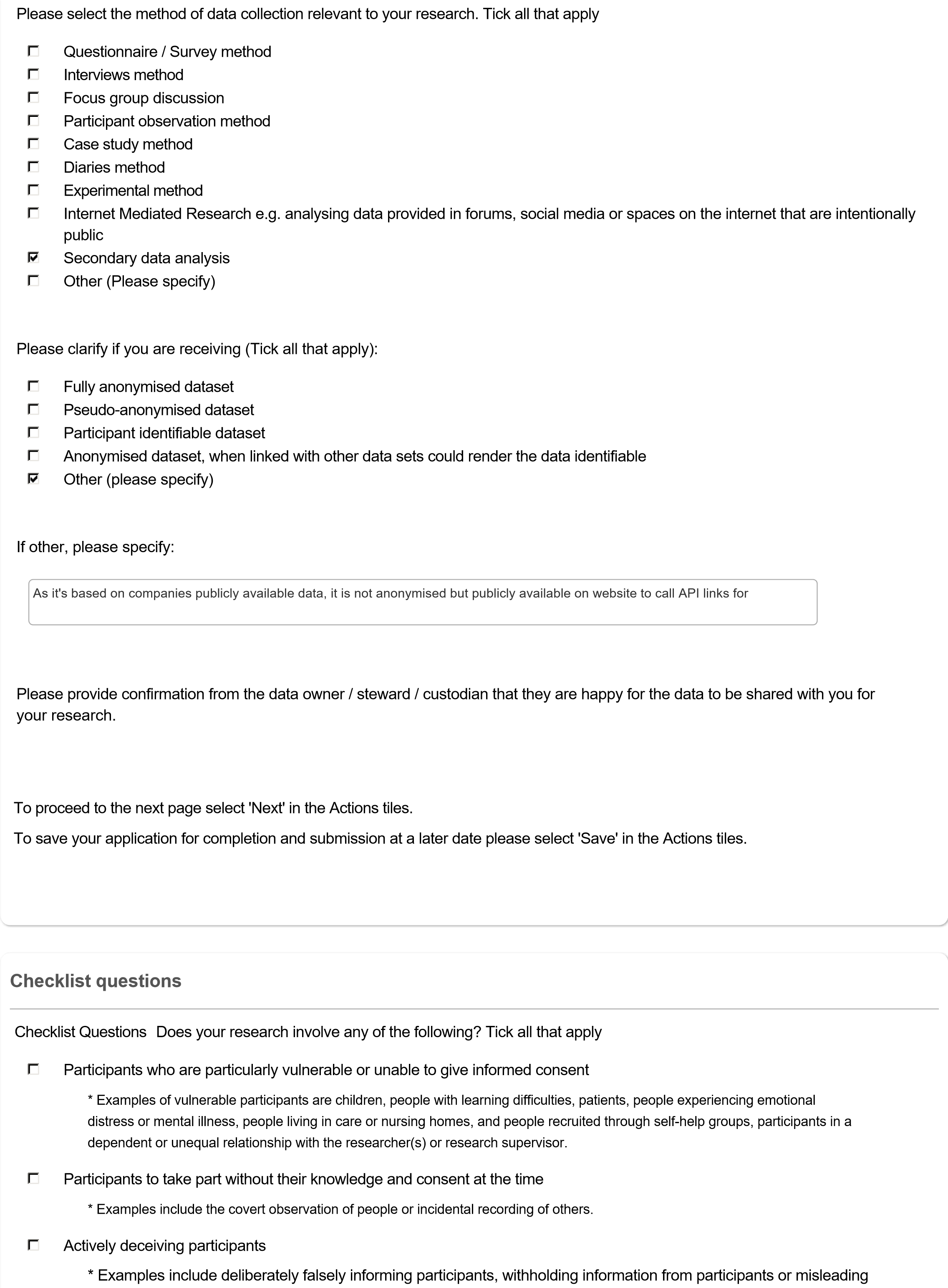
31/07/2024

To proceed to the next page select 'Next' in the Actions tiles.

To save your application for completion and submission at a later date please select 'Save' in the Actions tiles.



**Data collection method**





participants in such a way that they are likely to object or show unease when debriefed about the study.

Discussion or collection of information on sensitive topics or considered special category status under GDPR

\* Special Category Status under GDPR include:

personal data revealing racial or ethnic origin;

personal data revealing political opinions;

personal data revealing religious or philosophical beliefs;

personal data revealing trade union membership;

genetic data;

biometric data (where used for identification purposes);

data concerning health;

data concerning a person’s sex life;

and data concerning a person’s sexual orientation.

If the research is in relation to any of the sensitive topics listed then the legal issue requiring such scrutiny in such cases that

'explicit consent' must be obtained and the consenting process reviewed by the ethics committee

Invasive procedures

\* Invasive procedures may include:

Administration of drugs placebos:

Other substances (e.g., drinks, foods, food or drink constituents, dietary supplements) to study participants;

Biological samples from participants be obtained;

Pain or more than mild discomfort likely to result from the study.

Scans or x-rays of research participants

Photographs, videoing, audio recording or similar of research participants

Financial inducement (other than reasonable expenses and compensation for time)

The use or storage of information about living people whose personal identity could be discovered from that information

The risk of causing psychological stress or anxiety or other harm or negative consequences beyond that normally encountered

by the participants in their life outside research

Funds received from politically or culturally sensitive funding sources

\*Examples include the defence sector, projects with potential environmental effects and other internationally regulated or protected

industries. For more information, please follow the link to the '

[y](http://www.bris.ac.uk/red/support/governance/RGI.pdf)

[Research Governance and Integrity Polic](http://www.bris.ac.uk/red/support/governance/RGI.pdf)

'

Politically, culturally or socially sensitive topics

None of the above



**Study design and background**

Business School Research Ethics Application Form

Research involving humans by all academic and related Staff and Students in the University of Bristol Business School is subject to the standards set out in the University of Bristol Ethics of Research Policy and Procedure which can be found at: <http://www.bristol.ac.uk/red/research-governance/practice-training/researchethicspolicy.pdf>

It is a requirement prior to the commencement of all funded and non-funded research that this form be completed and submitted to the School’s Research Ethics Committee (REC). The REC will be responsible for issuing certification that the research meets acceptable ethical standards and will, if necessary, require changes to the research methodology or reporting strategy. It is a requirement that prior to the commencement of all funded and non-funded research that this form be completed and submitted to the School’s Research Ethics Committee (REC). The REC will be responsible for issuing certification that the research meets acceptable ethical standards and will, if necessary, require changes to the research methodology or reporting strategy.



1. Background and aims of the research.

The project investigates the impact of ESG, CSR, and ethics-related events on the investment worthiness of publicly traded companies. Leveraging comprehensive financial datasets from Compustat and CRSP, combined with real-time ESG news from Bloomberg Terminal's API, this study aims to develop a unique predictive model that analyzes the relationship between these events and stock price movements. Advanced machine learning techniques such as FinBERT and RoBERTa for NLP tasks, PCA for dimensionality reduction, and GANs paired with SGMCMC simulations for Bayesian Inference are employed.

What sets this project apart is its innovative approach to sample selection, feature development, and application. By utilizing historical data from the past decade, the model estimates expected returns before ESG events, captures immediate returns after the events, and predicts future stock performance over a 90-day horizon. Additionally, the project includes the creation of a user-friendly tool integrated with IBM’s Watson Assistant, providing strategic suggestions and enhancing decision-making through a generative AI interface. This applicative aspect, combined with the novel features developed for the machine learning models, ensures the project offers new, actionable insights for sustainable investment practices, an area not extensively covered in existing literature.

* + Guo T. et al. (2020) “Esg2risk: A deep learning framework from esg news to stock volatility prediction,” arXiv [Preprint].
  + Kuiper, C., & Adrián, G. (2020). The effect of ESG on stock prices : An event study on the S&amp;P 500. DIVA. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1438607&dswid=6229

1. Outline the design of the study and list the procedures/activities to which the participants will be subjected:

This study employs a mixed-methods design integrating quantitative data analysis and machine learning model development.

Participants, i.e., publicly traded companies, will not be directly involved but will be represented through their financial and ESG data.

Data Collection: Financial data will be sourced from Compustat and CRSP databases, while ESG event data will be gathered using Bloomberg Terminal's API.

Data Processing: The collected data will be preprocessed to handle missing values, outliers, and normalize the features. Gradient Boosting (XGB) and Principal Component Analysis (PCA) will be used for feature selection and dimensionality reduction to streamline the dataset respectively.

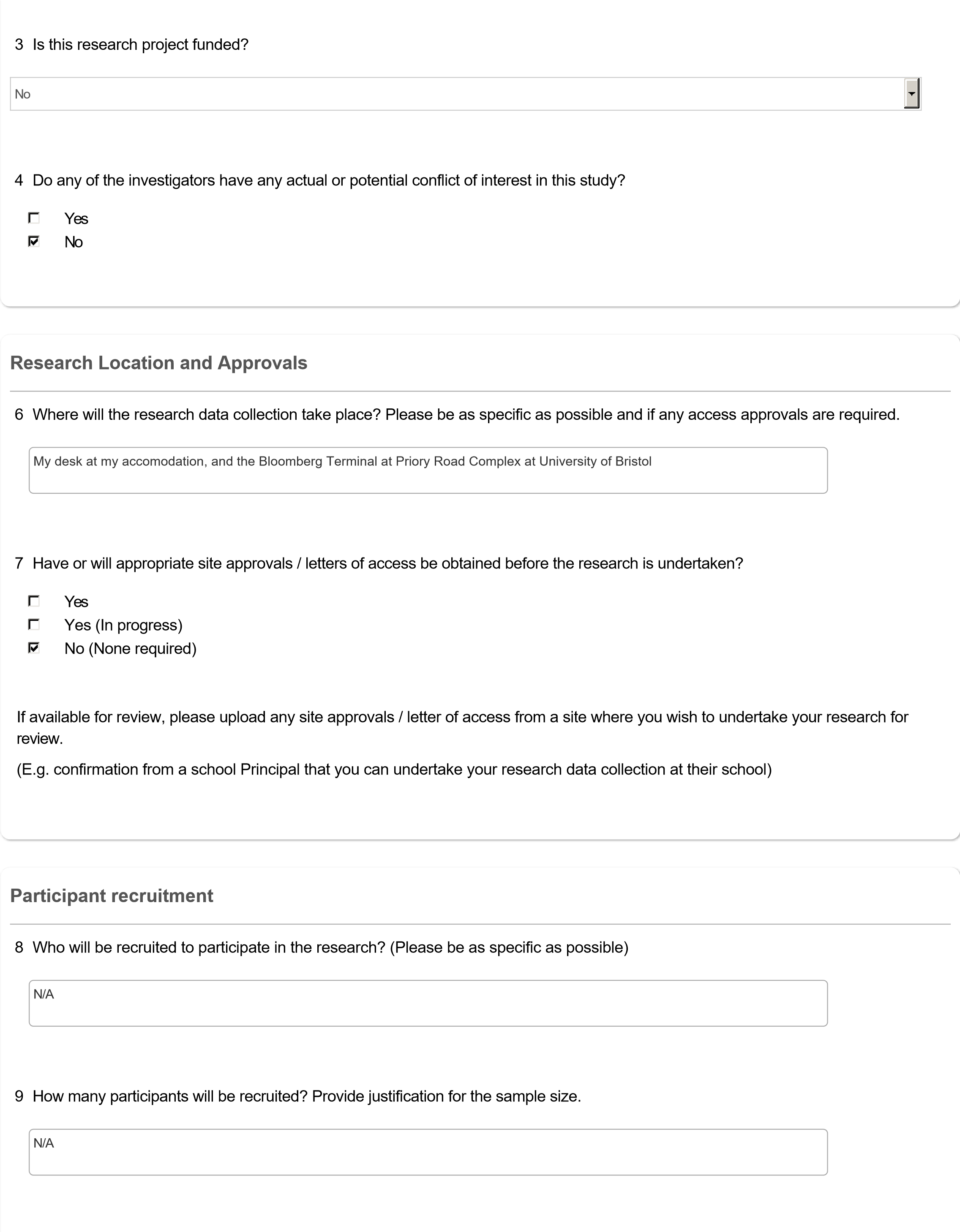
NLP Analysis: Natural Language Processing (NLP) models like FinBERT and RoBERTa will be applied to ESG news data to extract relevant features and sentiments that might affect stock prices.

Model Training: Generative Adversarial Networks (GANs) will be used in conjunction with Stochastic Gradient Markov Chain Monte Carlo (SGMCMC) to perform Bayesian Inference. The models will be trained on historical data to estimate expected returns, capture immediate effects of ESG events, and predict future stock performance over a 90-day period.

Evaluation: The predictive performance of the models will be evaluated using standard metrics such as Mean Squared Error (MSE) and R-squared values. Backtesting will be performed on a holdout sample to validate the model's effectiveness.

User Interface Development: A user-friendly interface will be created using IBM's Watson Assistant to provide real-time strategic suggestions based on model predictions.

The study aims to offer a novel tool for investors and financial analysts, enhancing decision-making through advanced analytics and real-time data integration. This comprehensive approach, from data processing to interface development, ensures practical applicability and innovation in the domain of ESG impact analysis.





10

How will the participants be identified and recruited to take part in your research?

N/A

11

Are there any potential participants who will be excluded? If so, what are the exclusion criteria?

N/A

13

Have you provided copies of all recruitment material with your ethics application for review?

Yes

No

13.1

If no, please provide justification as to why the recruitment material has not been provided for review.

N/A

Please provide any recruitment material used to recruit potential participants to take part in your research for review.



Please provide copies of any participant facing study documentation used to inform participants about the nature of the research for

review. Such as:

Participant Information Sheet (PIS)

Parental / Guardian Information Sheet (P/GIS)

Child Assent Information Sheet

15

Clearly outline how informed consent will be obtained from all participants and / or their parents / guardians prior to individuals

entering the research study?

N/A

16

How much time will participants be given to decide whether to give consent to participate after having been fully informed?

N/A

**Informed consent**



17

Has an independent named contact been provided if a participant wished to make a complaint or raise any concerns regarding

any ethical issues with the research?

Participants should be informed that they can contact the Research Governance Team (

research-governance@bristol.ac.uk

)

as an independent contact if they wish to make a complaint or raise any concerns with this research.

Yes

No

17.1

If 'No' please justify why an independent contact has not been provided to participants to raise a complaint or any issues with

the study.

N/A

19

Will participants be kept informed of new information that becomes available during the study which may influence their

continued participation?

N/A

20

Will participants be made aware they can withdraw their person or data from the research study at any time without having to

give a reason for doing so?

Yes

No

20.1.1

If no, please provide justification as to why participants will not be informed that they cannot withdraw their person and / or

data from the study.

N/A

Please provide copies of any documentation such as Consent Forms used to obtain consent from participants before taking part in

your research.



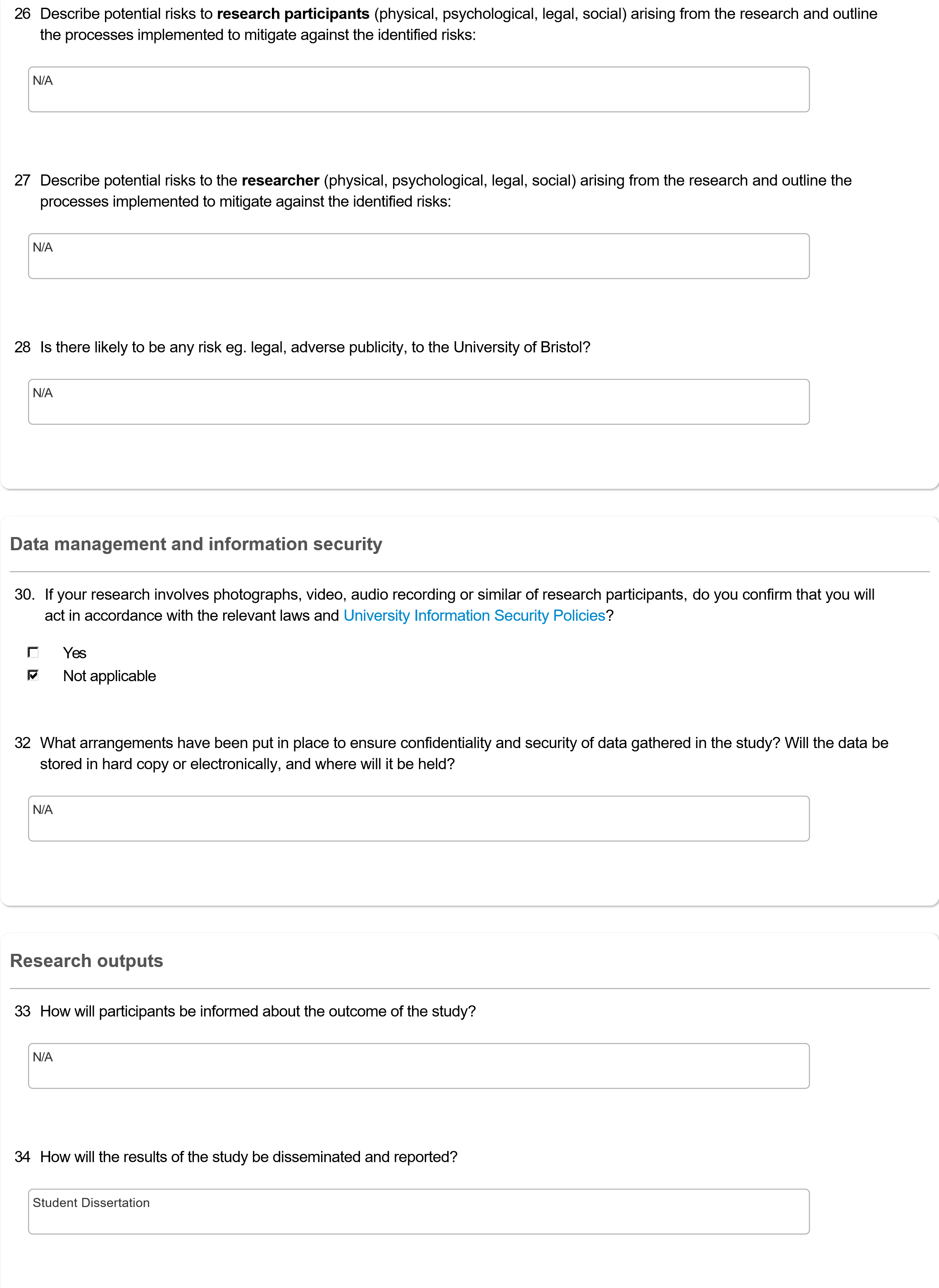
25

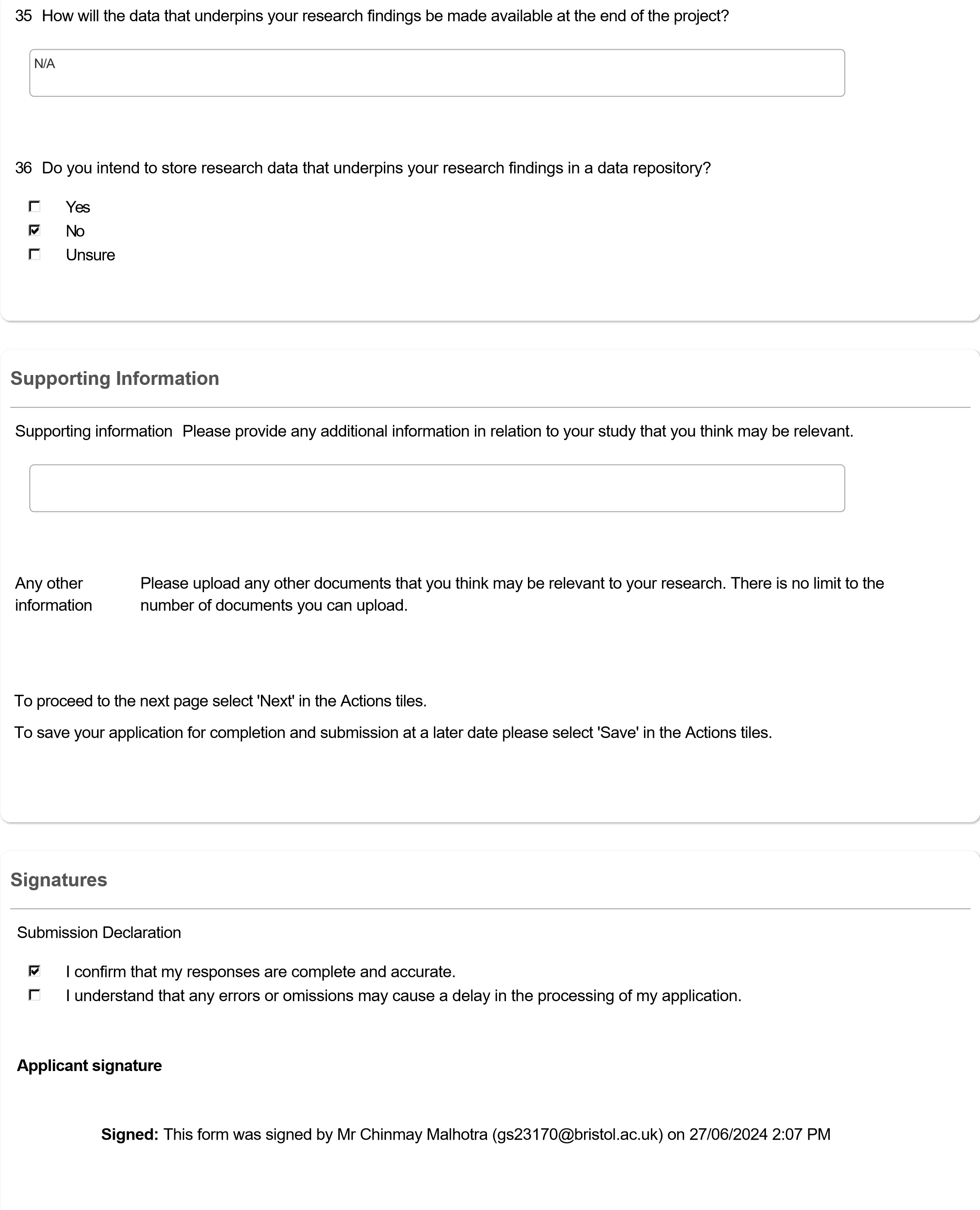
Are there any benefits to participants in taking part?

Be clear and realistic – if there are no direct benefits for the participant, then please state this.

N/A

**Participant and Researcher Safety**







Supervisor signature request

As the named supervisor please confirm:

1

.

I have read and reviewed this ethics application.

2

.

I am satisfied with the content and completeness of this ethics application.

**Signed:**

This form was signed by Dr Tian Han (tian.han@bristol.ac.uk)

on

27

/06/2024 5:21 PM

Submission Reminder

**Please note -**

Once all signatures have been obtained. You must ensure you select the

**Submit**

button on the left hand side of the

form.

If you have not received an email alert confirming submission of your research ethics application, then your research ethics

application has not been submitted for review.

If you have not received an OREMS submission confirmation email or if you are unsure whether your application has been

submitted, please contact

research-ethics@bristol.ac.uk

.

***\*\*\*Link to GitHub Repo Containing Python Framework from Data Collection to Model Comparison\*\*\****:

<https://github.com/chinmaylab/GenAI-Powered-Investment-Analytics-Tool-for-Strategic-Investment-Insights-based-on-ESG-News/blob/main/README.md>

|  |  |  |  |
| --- | --- | --- | --- |
| **Era** | **Characteristics**  **/Description** | **Key Findings**  **/Observations** | **Methodological Advancements** |
| **Initial Era** | Early research focused on the relationship between CSR/ESG and Corporate Financial Performance (CFP) using simple methodologies and linear assumptions. | Mostly positive correlation between CSR/ESG and financial stability, with some exceptions showing no or negative correlation (e.g., McWilliams, 2001; Brammer, 2006). | Basic methodologies like linear regression, small sample sizes, and limited geographic scope were used. |
| **Evolving Era** | More comprehensive studies with larger datasets, geographic expansion, and focus on specific financial metrics such as stock volatility and returns. | Shift in focus from overall CFP to specific metrics (e.g., Tobin’s Q, stock returns). Positive relationships between CSR/ESG and firm value (Bheenick et al., 2023; Al-Shammari, 2022). | Advanced techniques like two-stage least squares (2SLS) regression and Natural Language Processing (NLP) were introduced, along with larger sample sizes and longer time periods. |
| **Modern Era** | Application of machine learning (ML), deep learning (DL), and sophisticated models to better understand CSR/ESG's dynamic impact on financial performance over time. | Stronger correlations found between CSR/ESG activities and financial outcomes. Techniques like SVM, LSTM, and BERT are used for time-series and text-based analyses. | Integration of advanced models such as Support Vector Machines (SVM), Long Short-Term Memory (LSTM), Transformer models (e.g., BERT), and Gated Recurrent Units (GRUs) for complex data analysis. |

Table 1: 3 Eras of Literature

|  |  |  |
| --- | --- | --- |
| **Data Processing for CRSP and Compustat (Jan 1, 2016 - Dec 31, 2022)** | | |
| **Step** | **Description** | **Records Remaining** |
| Initial Data Collection | Data from CRSP and Compustat collected | ~13,500,000+ and 300,000+ rows |
| GVKEY Assignment & Filtering | Matched GVKEY with PERMNO and removed delisted/foreign entries | 9,500,000+ rows |
| CRSP-Compustat Matching | |  | | --- | | Kept companies with entries in both CRSP and Compustat |  |  | | --- | |  | | ~9,500,000+ and ~175,000 rows (8195 GVKEYs) |
| Data Merging | Merged CRSP and Compustat datasets, removing any null values | 6,000,000+ rows |
| Duplicate Removal | |  | | --- | | Removed duplicate rows from the merged dataset |  |  | | --- | |  | | 6,000,000+ rows |
| Financial Ratios and Trimming | Created financial ratios, dropped initial columns, and trimmed top/bottom 1% | ~5,500,000 rows and 5701 GVKEYs |

Table 2: Pre-processing Steps for Initial Control Variable Selection Data

|  |  |  |
| --- | --- | --- |
| **Data Processing for CRSP, Compustat and LSEG data (Jan 1, 2023 - Dec 31, 2023)** | | |
| **Step** | **Description** | **Records Remaining** |
| Initial Data Collection | Data from CRSP and Compustat collected | 2,000,000+ and ~100,000 rows |
| GVKEY Assignment & Filtering | Matched GVKEY with PERMNO and removed delisted/foreign entries | ~1,500,000 rows |
| CRSP-Compustat Matching | |  | | --- | | Kept companies with entries in both CRSP and Compustat |  |  | | --- | |  | | 1,350,000+ and ~50,000 rows (6066 GVKEYs) |
| Data Merging | Merged CRSP and Compustat datasets, removing any null values | 950,000+ rows |
| Duplicate Removal | |  | | --- | | Removed duplicate rows from the merged dataset |  |  | | --- | |  | | 950,000+ rows |
| Financial Ratios and Trimming | Created financial ratios, dropped initial columns, and trimmed top/bottom 1% | ~900,000 rows and 4381 GVKEYs |
| vii) Merging News Data | Combined the financial and news datasets, retaining only companies with daily news coverage. | ~600,000 rows and 1562 GVKEYs |

Table 3: Pre-processing Steps for Final Model Comparison Data

|  |  |  |
| --- | --- | --- |
| **Time Window** | **Duration** | **Purpose** |
| Estimation | [-120, -7] \*\* | Calculate coefficients for the Fama-French 5 Factor Model and compute Abnormal Returns (AR) for both Estimation and Event windows. |
| Event | [-3, +3] \*\* | Define the period to observe the event's impact. coefficients from the Estimation Window are used to calculate AR during this window. |
| Prediction | [+4, +30] \*\* | Predict the ongoing effect of CSR/ESG news using the same coefficients to estimate AR for the Prediction Window. |

\*\* - Represents trading Days

Table 4: Event Study Methodology Windows Description



Figure 2: Topic – Subtopic Matching Dictionary



Figure 3: CSR/ESG News Feature Engineering/Variable Selection

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Variable Name** | **Variable Code** | **Definition** | **Supporting Literature** |
| Liquidity | Current Ratio | current\_ratio | Evaluates a company's ability to meet short-term obligations using its current assets. A higher ratio indicates stronger liquidity. | Guo et al., 2020; Zhang et al., 2022 |
| Leverage | Debt to Equity Ratio | debt\_to\_equity | Shows the proportion of a company’s debt compared to its shareholders' equity, reflecting how much debt the company uses to finance its operations. | Sangkim, 2021;  Tsai, 2021;  Yadav, 2023;  Yang, 2020 |
| Valuation | Price to Book Ratio | pb\_ratio | Compares a company’s market value to its book value, showing how much investors are willing to pay for each dollar of the company's assets. | Yu, 2023;  Kamaliah, 2020; Khanchel et al., 2023 |
| Volatility | Daily Volume | dlyvol | Represents the total number of shares or contracts traded on a given trading day, serving as a key indicator of market activity and volatility. | Mupondo, 2022; Wang et al., 2005 |

Table 5: Fixed Control Variables Description and Supporting Literature

|  |
| --- |
| **Formulas** |
|  |
|  |
|  |
|  |

Table 6: Fixed Control Variables Formulas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Variable Name** | **Variable Code** | **Definition** | **Supporting Literature** |
| Size | Total Assets  (Quarterly) | atq | Measures the value of a company's assets, showing overall scale of operations. | Liao, 2016;  Xu et al., 2022; Dorfleitner, 2024; De Vincentiis, 2024 |
| Market Capitalization  (Daily) | dlycap | Reflects the total market value of a company’s outstanding shares, calculated by multiplying the stock price by the number of shares outstanding. (Reported at different frequencies) | Tsai, 2021; Serafeim, 2022; Khan 2019; Yadav, 2023 |
| Total Market Value  (Quarterly) | mkvaltq |
| Profitability | Return on Assets | ROA | Shows how efficiently a company uses its assets to generate profits. | De Vincentiis, 2024;  Tsai, 2021 |
| Return on Equity | ROE | Measures profitability relative to shareholders' equity investments. | Serafeim, 2022; Yadav, 2023 |

Table 7: Selected Control Variables Descriptions and Supporting Literature

|  |
| --- |
| **Formulas** |
|  |
|  |
|  |
|  |
|  |

Table 8: Selected Control Variables Formulas



Figure 4: Control Variable Selection Framework Model Input Combinations

****

Figure 5: Control Variable Selection Framework

****

Figure 7: Overall Final Predictive Modelling Framework

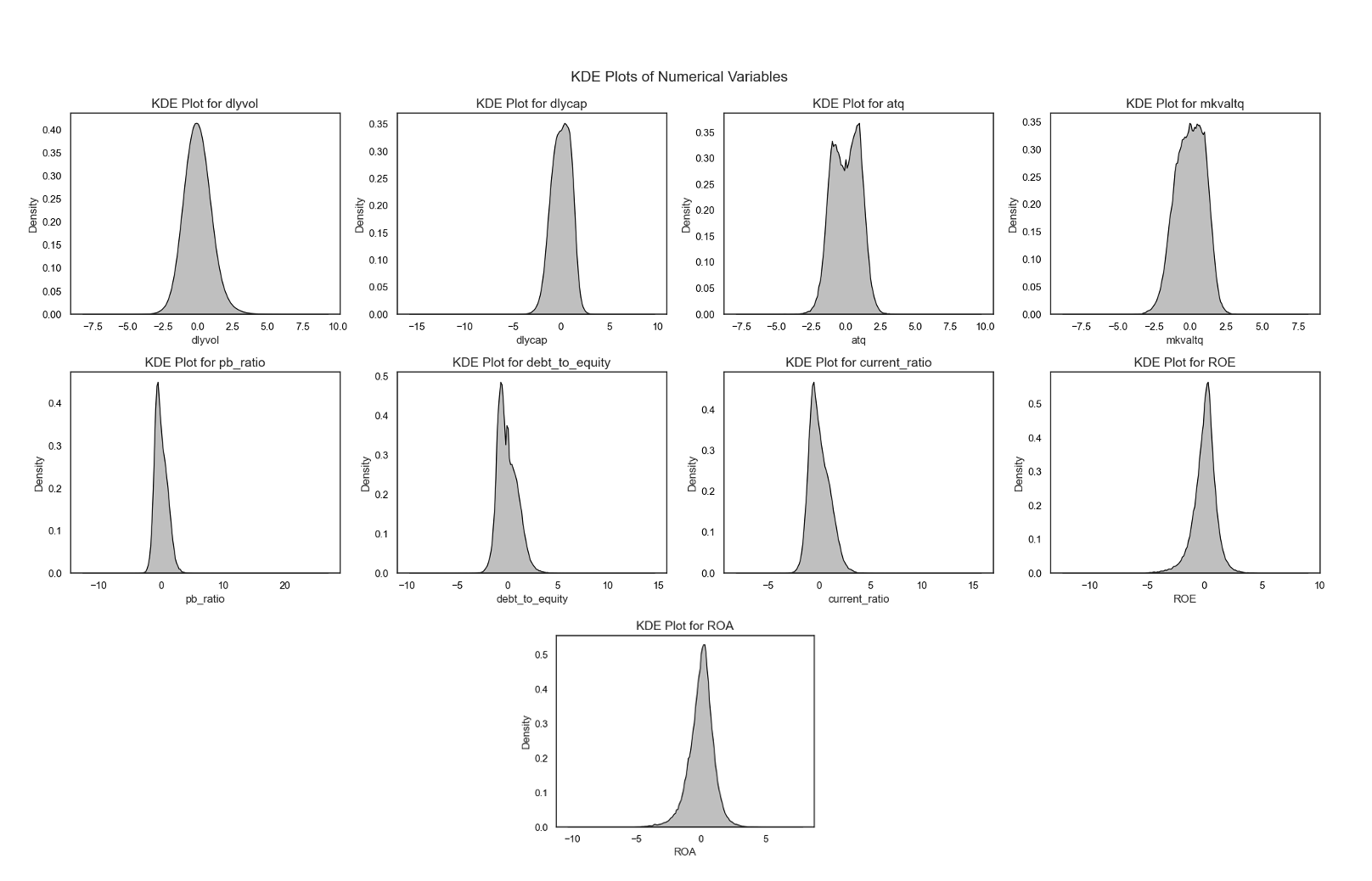
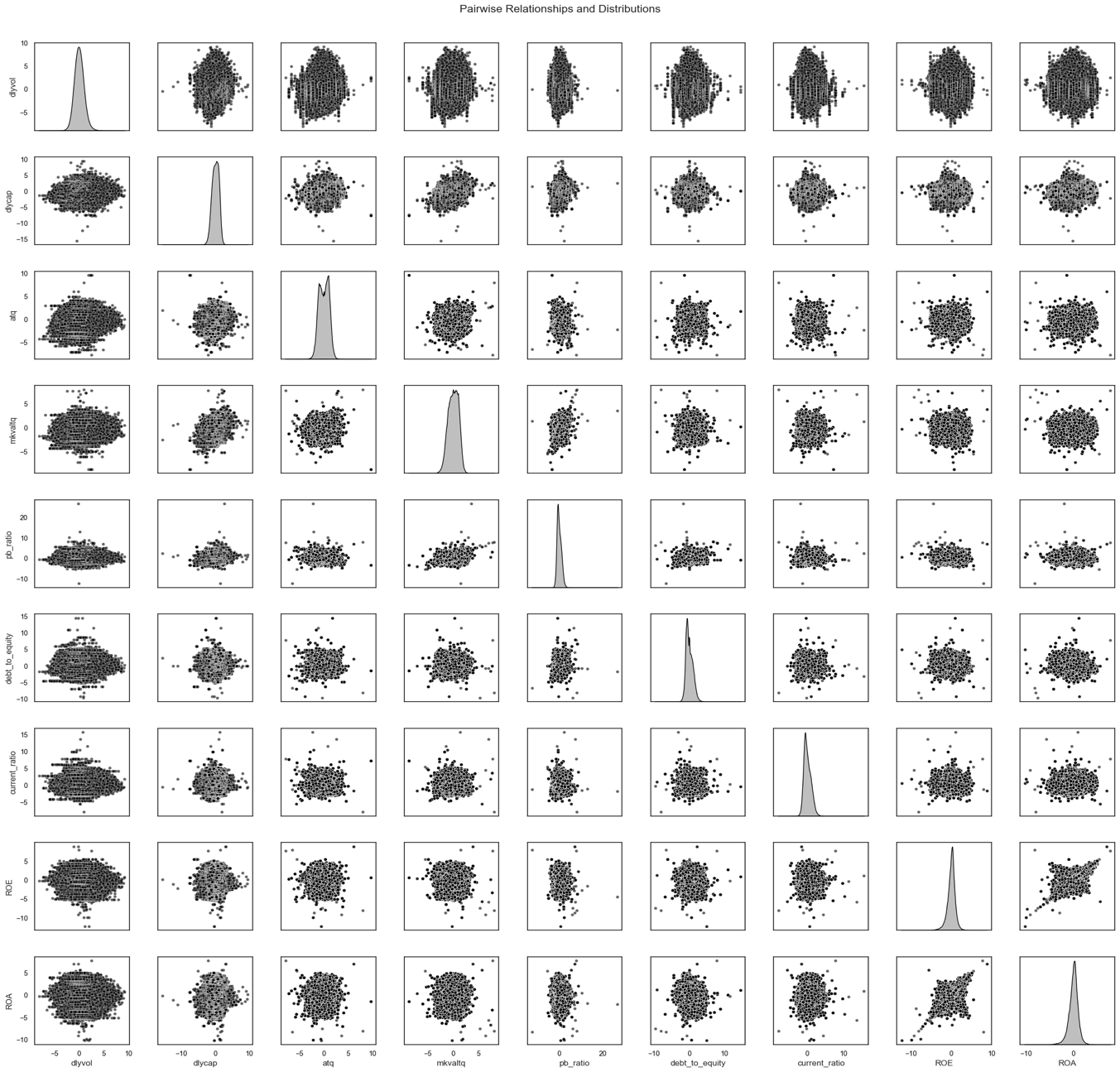


Figure 8: KDE Plots for all Control Variables

Figure 9: Pairwise Plots for all Control Variables

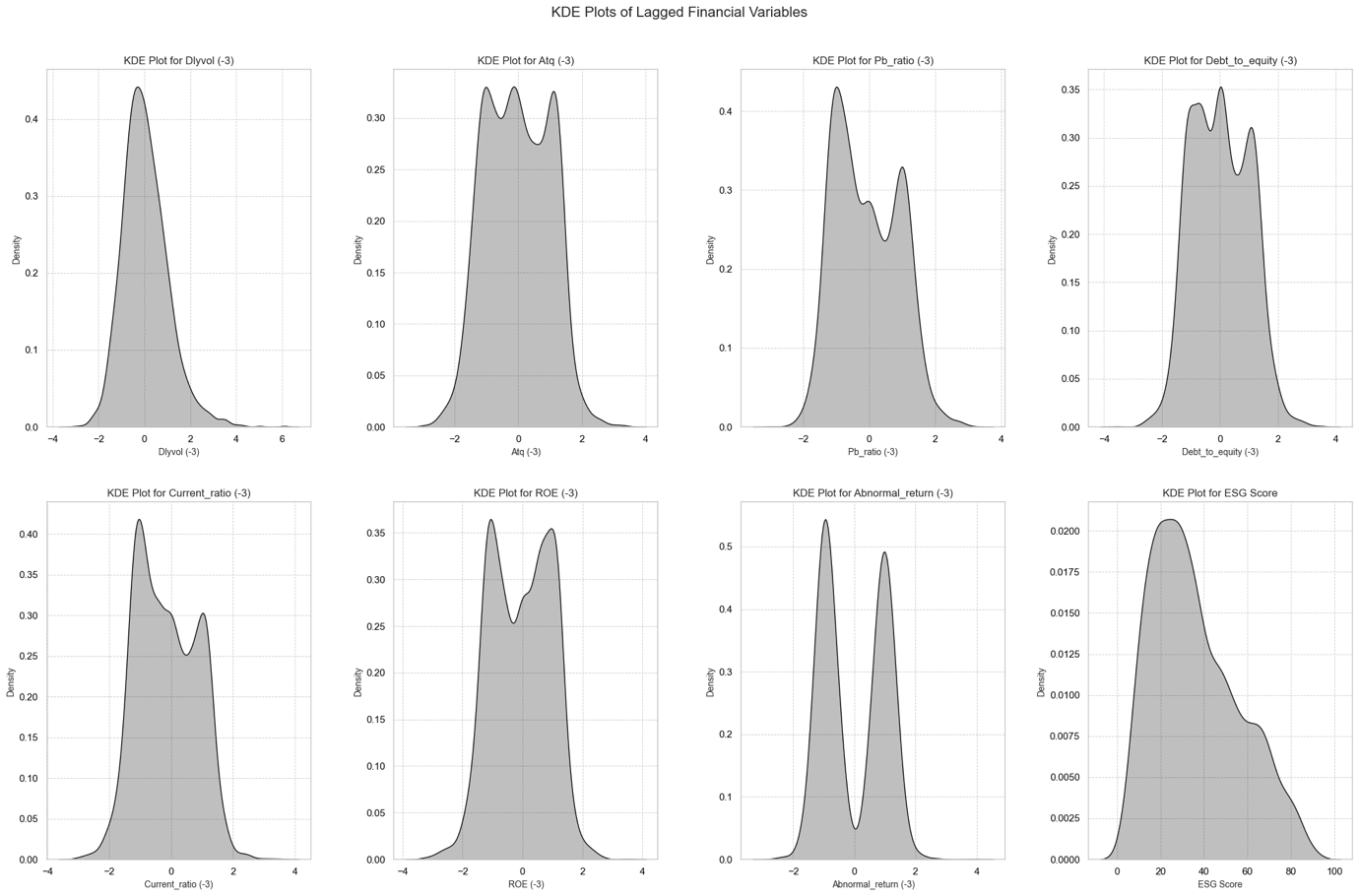
****

Figure 12: KDE Plots for Lagged Financial Variables

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Dlyvol (-3) | Atq (-3) | Pb\_ratio (-3) | Debt\_to\_equity (-3) | Current\_ratio (-3) | ROE (-3) | Abnormal\_return (-3) | ESG Score | Climate Change | Community Relations | Corporate Governance | Human Capital | Natural Capital | Non-ESG | Pollution & Waste | Product Liability | Sentiment Neutral | Sentiment Positive |
| Dlyvol (-3) | 1.0 | 0.02 | 0.01 | -0.04 | 0.06 | -0.01 | 0.03 | 0.07 | -0.02 | 0.01 | 0.02 | -0.03 | 0.0 | 0.01 | 0.03 | 0.02 | 0.02 | -0.02 |
| Atq (-3) | 0.02 | 1.0 | -0.01 | -0.27 | 0.19 | 0.28 | -0.03 | 0.08 | 0.05 | 0.06 | -0.05 | 0.07 | 0.02 | 0.02 | 0.01 | -0.04 | -0.06 | 0.07 |
| Pb\_ratio (-3) | 0.01 | -0.01 | 1.0 | 0.15 | -0.08 | -0.11 | 0.07 | 0.0 | 0.01 | 0.01 | 0.0 | 0.02 | 0.02 | 0.01 | 0.0 | 0.02 | -0.0 | 0.01 |
| Debt\_to\_equity (-3) | -0.04 | -0.27 | 0.15 | 1.0 | -0.3 | -0.3 | 0.05 | -0.1 | -0.05 | -0.05 | 0.03 | -0.02 | -0.02 | -0.01 | -0.02 | 0.05 | 0.06 | -0.06 |
| Current\_ratio (-3) | 0.06 | 0.19 | -0.08 | -0.3 | 1.0 | 0.22 | -0.01 | 0.03 | 0.01 | 0.03 | 0.02 | -0.02 | 0.03 | 0.02 | 0.02 | -0.04 | -0.02 | 0.02 |
| ROE (-3) | -0.01 | 0.28 | -0.11 | -0.3 | 0.22 | 1.0 | -0.04 | 0.03 | 0.0 | 0.02 | -0.0 | 0.04 | 0.02 | -0.01 | -0.01 | -0.02 | 0.0 | 0.01 |
| Abnormal\_return (-3) | 0.03 | -0.03 | 0.07 | 0.05 | -0.01 | -0.04 | 1.0 | 0.0 | -0.02 | 0.01 | 0.0 | -0.0 | 0.02 | 0.04 | 0.01 | -0.02 | -0.0 | -0.01 |
| ESG Score | 0.07 | 0.08 | 0.0 | -0.1 | 0.03 | 0.03 | 0.0 | 1.0 | 0.21 | 0.14 | -0.03 | 0.16 | 0.07 | 0.11 | 0.07 | -0.12 | -0.22 | 0.22 |
| Climate Change | -0.02 | 0.05 | 0.01 | -0.05 | 0.01 | 0.0 | -0.02 | 0.21 | 1.0 | -0.05 | -0.15 | -0.07 | -0.02 | -0.09 | -0.03 | -0.09 | -0.15 | 0.18 |
| Community Relations | 0.01 | 0.06 | 0.01 | -0.05 | 0.03 | 0.02 | 0.01 | 0.14 | -0.05 | 1.0 | -0.12 | -0.05 | -0.02 | -0.07 | -0.03 | -0.07 | -0.09 | 0.12 |
| Corporate Governance | 0.02 | -0.05 | 0.0 | 0.03 | 0.02 | -0.0 | 0.0 | -0.03 | -0.15 | -0.12 | 1.0 | -0.15 | -0.05 | -0.2 | -0.07 | -0.2 | 0.08 | -0.08 |
| Human Capital | -0.03 | 0.07 | 0.02 | -0.02 | -0.02 | 0.04 | -0.0 | 0.16 | -0.07 | -0.05 | -0.15 | 1.0 | -0.02 | -0.09 | -0.03 | -0.09 | -0.25 | 0.29 |
| Natural Capital | 0.0 | 0.02 | 0.02 | -0.02 | 0.03 | 0.02 | 0.02 | 0.07 | -0.02 | -0.02 | -0.05 | -0.02 | 1.0 | -0.03 | -0.01 | -0.03 | -0.05 | 0.05 |
| Non-ESG | 0.01 | 0.02 | 0.01 | -0.01 | 0.02 | -0.01 | 0.04 | 0.11 | -0.09 | -0.07 | -0.2 | -0.09 | -0.03 | 1.0 | -0.04 | -0.12 | -0.2 | 0.1 |
| Pollution & Waste | 0.03 | 0.01 | 0.0 | -0.02 | 0.02 | -0.01 | 0.01 | 0.07 | -0.03 | -0.03 | -0.07 | -0.03 | -0.01 | -0.04 | 1.0 | -0.04 | -0.08 | 0.08 |
| Product Liability | 0.02 | -0.04 | 0.02 | 0.05 | -0.04 | -0.02 | -0.02 | -0.12 | -0.09 | -0.07 | -0.2 | -0.09 | -0.03 | -0.12 | -0.04 | 1.0 | 0.11 | -0.1 |
| Sentiment Neutral | 0.02 | -0.06 | -0.0 | 0.06 | -0.02 | 0.0 | -0.0 | -0.22 | -0.15 | -0.09 | 0.08 | -0.25 | -0.05 | -0.2 | -0.08 | 0.11 | 1.0 | -0.84 |
| Sentiment Positive | -0.02 | 0.07 | 0.01 | -0.06 | 0.02 | 0.01 | -0.01 | 0.22 | 0.18 | 0.12 | -0.08 | 0.29 | 0.05 | 0.1 | 0.08 | -0.1 | -0.84 | 1.0 |

Table 13: Correlation Matrix for All Variables except Principal Components

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Abnormal Return (+4)** | **Abnormal Return (+15)** | **Abnormal Return (+30)** |
| **Const** | *-0.0308*  *(0.764)* | *0.012*  *(0.775)* | *0.007*  *(0.409)* |
| **Dlyvol (-3)** | *-0.0075*  *(0.688)* | *0.0221*  *(0.228)* | *-0.0075*  *(0.681)* |
| **Atq (-3)** | *0.0241*  *(0.225)* | *0.0055*  *(0.780)* | *0.003*  *(0.878)* |
| **Pb\_ratio (-3)** | *0.0421\*\*\**  *(0.028)* | *-0.0015*  *(0.936)* | *-0.0236*  *(0.210)* |
| **Debt\_to\_equity (-3)** | *0.0122*  *(0.555)* | *0.0062*  *(0.761)* | *0.0133*  *(0.514)* |
| **Current\_ratio (-3)** | *-0.0191*  *(0.348)* | *-0.0008*  *(0.970)* | *0.0118*  *(0.552)* |
| **ROE (-3)** | *0.0099*  *(0.623)* | *0.0136*  *(0.491)* | *0.0033*  *(0.868)* |
| **Abnormal\_return (-3)** | *0.0194*  *(0.281)* | *-0.0275*  *(0.119)* | *0.0425\*\*\**  *(0.016)* |
| **ESG Score** | *0.0025\*\*\**  *(0.014)* | *0.0013*  *(0.192)* | *0.0022\*\*\**  *(0.025)* |
| **Subtopic Climate Change** | *-0.0804*  *(0.346)* | *-0.0658*  *(0.434)* | *-0.0583*  *(0.485)* |
| **Subtopic Community Relations** | *0.0723*  *(0.460)* | *-0.0662*  *(0.492)* | *-0.0173*  *(0.857)* |
| **Subtopic Corporate Governance** | *0.0078*  *(0.872)* | *-0.0065*  *(0.892)* | *-0.0832\**  *(0.080)* |
| **Subtopic Human Capital** | *0.0025*  *(0.977)* | *0.1684\*\*\**  *(0.050)* | *-0.0808*  *(0.345)* |
| **Subtopic Natural Capital** | *-0.4765\*\*\**  *(0.029)* | *-0.1372*  *(0.522)* | *0.1754*  *(0.411)* |
| **Subtopic Non-ESG** | *-0.1265\**  *(0.062)* | *-0.0176*  *(0.792)* | *-0.0533*  *(0.421)* |
| **Subtopic Pollution & Waste** | *-0.0465*  *(0.772)* | *0.0347*  *(0.826)* | *-0.0831*  *(0.596)* |
| **Subtopic Product Liability** | *0.0473*  *(0.461)* | *-0.0099*  *(0.876)* | *-0.0521*  *(0.407)* |
| **Sentiment Neutral** | *-0.0589*  *(0.533)* | *-0.0165*  *(0.859)* | *0.0476*  *(0.607)* |
| **Sentiment Positive** | *0.0094*  *(0.930)* | *-0.1316*  *(0.215)* | *-0.0684*  *(0.517)* |
| **PC1** | *-0.0015*  *(0.355)* | *-0.0009*  *(0.546)* | *0.0004*  *(0.776)* |
| **PC2** | *-0.0016*  *(0.415)* | *-0.0030*  *(0.117)* | *0.0010*  *(0.611)* |
| **PC3** | *-0.0021*  *(0.474)* | *0.0011*  *(0.688)* | *0.0022*  *(0.435)* |
| **PC4** | *-0.0013*  *(0.696)* | *0.0017*  *(0.604)* | *0.0057\**  *(0.078)* |
| **PC5** | *-0.0024*  *(0.561)* | *0.0043*  *(0.295)* | *0.0015*  *(0.705)* |
| **PC6** | *0.0063*  *(0.203)* | *0.0020*  *(0.673)* | *0.0015*  *(0.751)* |
| **PC7** | *-0.0129\*\*\**  *(0.025)* | *0.0064*  *(0.257)* | *0.0088*  *(0.117)* |
| **PC8** | *-0.0094*  *(0.190)* | *0.0019*  *(0.786)* | *0.0012*  *(0.866)* |
| **PC9** | *0.0058*  *(0.539)* | *-0.0011*  *(0.907)* | *0.0120*  *(0.194)* |
| **PC10** | *-0.0003*  *(0.979)* | *-0.0057*  *(0.548)* | *0.0058*  *(0.541)* |
| **PC11** | *-0.0051*  *(0.633)* | *0.0053*  *(0.612)* | *-0.0073*  *(0.483)* |
| **PC12** | *-0.0200\**  *(0.082)* | *0.0021*  *(0.852)* | *-0.0069*  *(0.539)* |
| **PC13** | *-0.0221\**  *(0.081)* | *0.0001*  *(0.993)* | *0.0214\**  *(0.083)* |
| **PC14** | *0.0071*  *(0.583)* | *0.0022*  *(0.866)* | *-0.0147*  *(0.247)* |
| **PC15** | *-0.0138*  *(0.319)* | *0.0038*  *(0.781)* | *-0.0149*  *(0.273)* |
| **PC16** | *-0.0039*  *(0.784)* | *-0.0078*  *(0.583)* | *0.0034*  *(0.812)* |
| **PC17** | *0.0072*  *(0.638)* | *-0.0047*  *(0.754)* | *-0.0077*  *(0.608)* |
| **R2** | *0.015* | *0.007* | *0.012* |
| **F-Stat** | *1.320* | *0.6428* | *1.046* |
| **P(F-Stat)** | *0.0992* | *0.949* | *0.394* |

*Note: \*\*\*\*, \*\*\*, \* represent 1%, 5% and 10% significance levels respectively.*

*Values in brackets represent p-value of the coefficients*

Table 14: OLS Regression Results

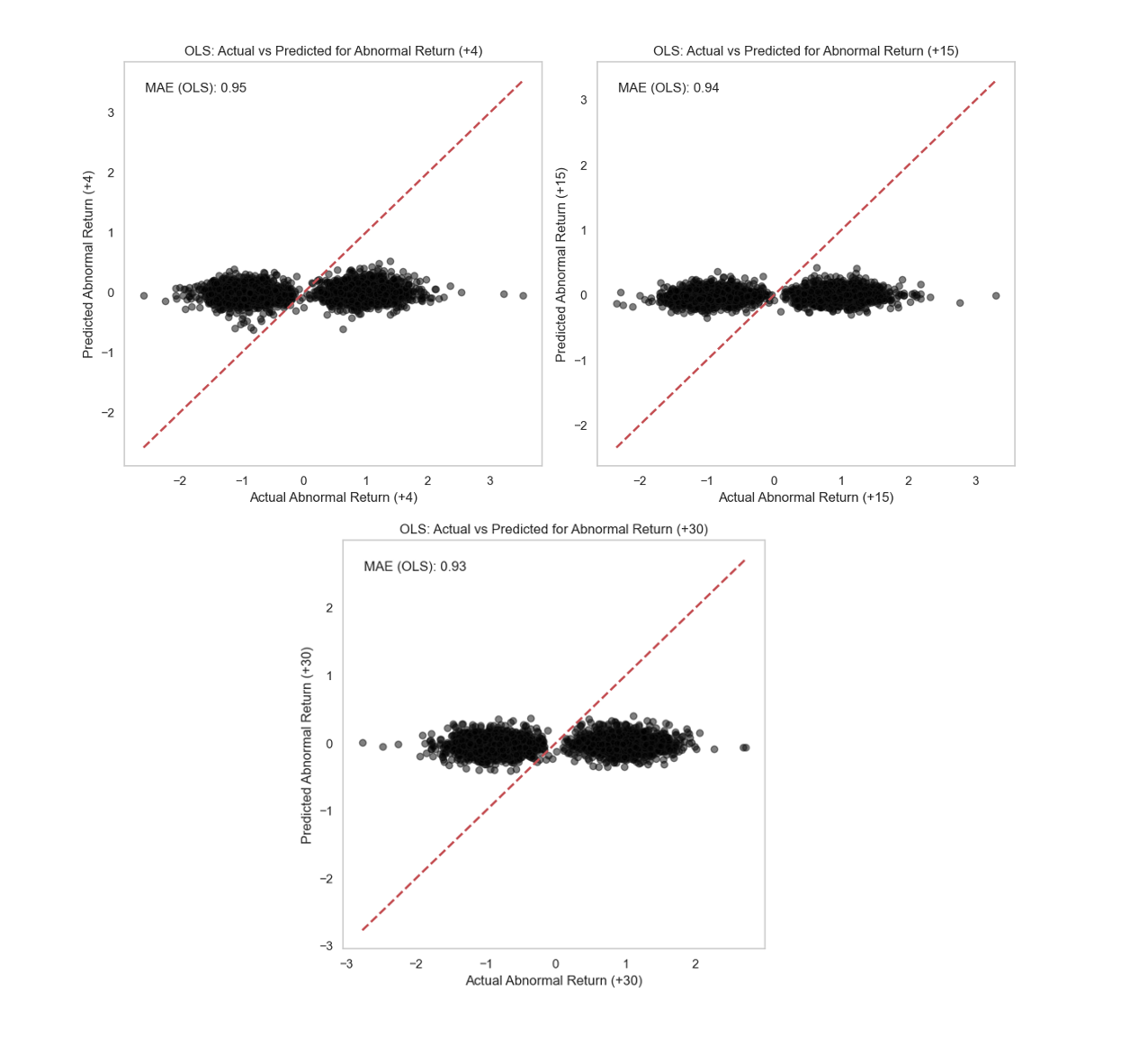


Figure 14: OLS Regression Results Visualised