Sustainability Performance and Financial Performance

Evidence from a Longitudinal Panel Study of European Firms

Fredrik Lichtenstein

Degree project/Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Faculty of Natural Resources and Agricultural Sciences/Department of Economics
Environmental Economics and Management. Master's Programme
Degree project/SLU, Department of Economics, 1733 • ISSN 1401-4084
Uppsala 2026

Sustainability Performance and Financial Performance: Evidence from a Longitudinal Panel Study of European Firms

Fredrik Lichtenstein

Supervisor:	**Per-Anders Langendahl, Swedish University of Agricultural Sciences, Department of Economics**
Assistant supervisor:	Sandra Kurniawati, Swedish University of Agricultural Sciences, Department of Economics
Examiner:	Per-Anders Langendahl, Swedish University of Agricultural Sciences, Department of Economics

Credits:	30 credits
Level:	Second cycle A2E
Course title:	Master thesis in Business Administration
Course code:	EX0904
Program/education:	Environmental Economics and Management
Course coordinating dept:	Department of Economics
Place of publication:	Uppsala
Year of publication:	2026
Title of series:	Degree project/SLU, Department of Economics
Part number:	1733
ISSN:	1401-4084

Keywords:	longitudinal panel study, sustainability, sustainable development, CSR, ESG, ESG scoring, sustainability performance, financial performance

Swedish University of Agricultural Sciences
Faculty of Natural Resources and Agricultural Sciences
Department of Economics

Abstract

The aim of this study is to contribute to a better understanding of the relationship between firm-level sustainability performance and financial performance among European firms.

The research has been conducted by collecting secondary financial and ESG data between 2014-2023 from a selection of 600 European firms across multiple industries and markets, that as of February 2026 make up the STOXX Europe 600 Index.

The data analysis was performed by conducting a panel regression with firm and year fixed effects, lagged ESG scores (t-1 and t-2), clustered standard errors, and winsorization of all continuous variables at the 1^{st} and 99^{th} percentile. The sample includes 508 firms and 4,833 firm-years after cleaning for missing financial and ESG data.

The analysis shows no significant association at t-1 and negative significant association at t-2 ($\beta = -0.0364$, $p = 0.005$). A joint Wald test confirms that the two ESG lag coefficients are jointly significant ($\chi^2(2) = 8.03$, $p = 0.018$), with the effect concentrated at the two-year horizon. The research null hypothesis (H_0), that the association between sustainability and financial performance is non-positive, cannot be rejected. The findings are consistent with recent European evidence of a negative association between sustainability and financial performance.

This is the first European study to combine lagged ESG, firm fixed effects and within-industry analysis simultaneously, adding temporal precision to the negative ESG-ROA pattern documented in prior European studies.

Keywords: longitudinal panel study, sustainability, sustainable development, CSR, ESG, ESG scoring, sustainability performance, financial performance

Table of contents

List of Tables

Abbreviations

Abbreviation	Description
Cap	Capitalization, i.e. value of stocks in the market, e.g. market cap, small/mid/large cap
CFP	Corporate Financial Performance
CO_2e	Carbon Dioxide Equivalents
CSP	Corporate Social Performance
CSR	Corporate Social Responsibility
CSRD	Corporate Sustainability Reporting Directive
ESG	Environmental, Social & Governance
ESRS	European Sustainability Reporting Standards
EU	European Union
FE	Fixed Effects
FP	Financial Performance
GenAI	Generative Artificial Intelligence
GHG	Greenhouse Gas/-es
ln	Natural logarithm
Log	Logarithm
LSEG	London Stock Exchange Group
NFRD	Non-Financial Reporting Directive
OLS	Ordinary Least Squares
ROI	Return on Investment
SDGs	United Nations Sustainable Development Goals
SLU	Swedish University of Agricultural Sciences
SP	Sustainability Performance
SP-FP	Sustainability Performance-Financial Performance (relationship)

1. Introduction

This chapter establishes the research context and motivation, reviews the state of current European evidence, and defines the scope, aim and contribution of the study. §1.1 and §1.2 provide background on sustainable development and corporate sustainability; §1.3 reviews the existing SP-FP literature; §1.4 describes the scope and approach; §1.5 defines the research aim and question; and §1.6 states the delimitations.

1.1 Sustainable Development

The concept of sustainable development was first introduced in the so-called Brundtland report, 'Our Common Future', that was published in 1987 by the United Nations World Commission on Environment and Development, led by the current Norwegian prime minister Gro Harlem Brundtland. Influenced by the 1980 'World Conservation Strategy' of the International Union for Conservation of Nature, IUCN, (Environment & Society Portal 2016), the report defined sustainable development as development that "… meets the needs of the present without compromising the ability of future generations to meet their own needs." (p. 16, WCED 1987).

The report argues that there is a need for integrating financial, environmental and social sustainability and that everyone in all levels of society needs to work together to create development that is sustainable (WCED 1987). The Brundtland report also laid the foundation for the United Nations Sustainable Development Goals (SDGs) that were adopted as part of the 2030 Agenda for Sustainable Development in 2015 and are now widely considered as the commonly accepted blueprint for global sustainable development action (UN SDGs n.d.).

In the years leading up to the mid-2020s, sustainable development was on most large companies agenda, driven by market expectations and policy decisions, driving up sustainability investments (Friede et al. 2015; European Commission 2018). As of the mid-2020s, in times of greater financial uncertainty (IMF n.d.), the mechanisms identified by Campello et al. (2010), where financially constrained firms cut long-term investments and prioritize short-term, quantifiable returns, offer a theoretical lens for understanding current corporate behavior. Under such conditions, sustainability investments, which often involve longer time horizons and less immediately quantifiable returns, risk being deprioritized in favor of investments with more certain near-term financial outcomes. This is further complicated by the challenge of objectively measuring sustainability performance – i.e. how well a firm performs within the ESG (Environmental, Social, Governance) framework on topics such as disclosure/reporting, quantifiable impact e.g. CO_2e emissions, investments and preemptive activities – leaving the relationship between sustainability performance and financial accounting performance contested in both practice and research (Khan et al. 2016).

1.2 Corporate Sustainability

The integration of sustainability into company strategy and operations is a vital part of reaching the Agenda 2030 SDGs (UN SDGs n.d.). Referred to as Corporate Social Responsibility (CSR) or, increasingly more common, ESG (Environmental, Social & Governance), corporate sustainability includes social, environmental and governance aspects of companies' operations (Gillan et al. 2021).

Sustainability integration refers to making sustainable development an integrated part of business operations (Khan et al. 2016). It is achieved by adopting and complying with sustainability policies and frameworks and by investing in sustainable development, i.e. investing and engaging in activities that make company operations more sustainable, e.g. increasing resource efficiency, reducing emissions, and improving working conditions.

Increasing regulatory requirements have compelled corporations to engage more in sustainable development (European Commission 2018). However, corporations engage in sustainable development for various other reasons beyond compliance, e.g. risk management, stakeholder management, and for gaining competitive advantages (Khan et al. 2016; Lins et al. 2017; Hartzmark & Sussman 2019; Gillan et al. 2021). Eccles et al. (2014) find that early sustainability policy adoption is associated with long-term financial outperformance, suggesting that corporate sustainability integration may itself be a source of financial value – a claim that has since motivated a substantial body of empirical research.

1.3 State of the Art on the Sustainability-Financial Performance Relationship

The study of the relationship between sustainability performance and financial performance asks the question whether shareholder value may be increased by optimizing value creation for other stakeholders that are identified by their sustainability impact, e.g. nature, employees, customers (Gillan et al. 2021). This is a field that has been studied empirically for decades. However, whether improved sustainability performance actually causes improved financial performance is widely contested in academic literature, as findings are not straightforward (Eccles et al. 2014; Khan et al. 2016; Berg et al. 2022).

In 2009, a study performing a meta-analysis of 251 other studies conducted during the preceding 35 years, showed a weak positive correlation between corporate social and financial performance (Margolis et al. 2009). By studying early adopters of sustainability policy among corporations through the years of 1993-2009, Eccles et al. found that those early adopters financially outperformed similar corporations with no early sustainability adoption (2014). A second-level review of 60 review studies, representing more than 2200 unique empirical studies, presented in 2015, argues that ESG investing has paid off financially, counter to the general perception among many investors (Friede et al. 2015).

Khan et al. nuances this conclusion, arguing that financial out-performance is dependent on whether the sustainability activity is financially material (2016), i.e. has a substantial financial risk/effect implication, derived from industry specific materiality guidelines from an investor perspective (IFRS n.d.). Other research claims that results are largely dependent on how sustainability performance is rated, since ESG scores are created by various private firms, constructed through combining a large set of selected indicators, using different scopes, proxies and weights, which consequently makes each ESG rating unique in its composition, creating a divergence between different ESG ratings (Berg et al. 2022). Moreover, increased ESG disclosure seems to increase the level of ESG rating divergence (Christensen et al. 2022). In a replication study of Eccles et al. (2014), King (2025) argues that the causal inferences in the original paper lack sufficient evidence. Furthermore, King (2025) is not able to replicate the results from the original paper, thus advising caution when drawing conclusions on the causal relationship between corporate sustainability and financial performance.

In a literature review of 504 articles published between year 2000-2021, Rahi et al. (2024), conclude that the corporate sustainability–financial performance relationship is largely dependent on factors such as institutional context, resource availability, time horizon and research method (Rahi et al. 2024). Their analysis shows that there is a positive association between corporate sustainability and financial performance when there is a time lag applied to the sustainability variable; and where institutions and legitimacy requirements support sustainability. Surprisingly capitalistic countries (defined as UK, Ireland, USA) show a negative relationship. Slack resources has a positive correlation with sustainability investment, which suggests bidirectional or reversed causality (Rahi et al. 2024). Inconclusive results across studies can also be partially attributed to methodological choices.

Recent European studies have found largely negative associations between sustainability and financial performance. Candio (Candio 2024), studying 600 firms between 2012 and 2022, using the STOXX Europe 600 index as a sampling frame, shows a negative relationship between ESG scores and ROA across all three regression frameworks. Bifulco et al. (Bifulco et al. 2023), using the same sampling strategy of 600 firms from the STOXX Europe 600 index, through 2014-2020, find a negative ESG score-share price (market value) association. Rostamicheri et al. (Rostamicheri et al. 2026), studying a large European panel of approximately 24,500 firm-years 2015-2024, find a negative ESG-ROA/ROE association, particularly in the non-Nordic subsample, which was interpreted as transitional adjustment costs and delayed financial realization in a period of evolving EU sustainability regulation. Thus, across all of these recent European studies, there is a consistent pattern of a negative association between ESG and financial performance expressed both in terms of accounting (ROA/ROE) and market value (share price).

To summarize, there is no apparent consensus in global literature about the relationship between sustainability and financial performance (Berg et al. 2022; Christensen et al. 2022; King 2025). The European evidence, however, has converged on a negative pattern (Bifulco et al. 2023; Candio 2024; Rostamicheri et al. 2026), but under methodological conditions that leave important questions unanswered. Thus, there is a need for further research on European firms, carefully addressing factors that the broader literature has identified as critical, including time horizon and heterogeneity, motivating the longitudinal time-lagged approach using firm fixed effects, undertaken in this study.

1.4 Scope and Approach of This Study

While recent European studies, covered in chapter 1.3, establish a negative SP-FP (Sustainability Performance – Financial Performance) association pattern, many questions remain unanswered and in need of further study, e.g. whether the negative pattern holds for variations in samples, performance metrics and methodological differences. Neither do any of these studies combine time lag, fixed effects and within-industry analysis, relevant aspects for isolating and understanding the SP-FP association over time.

This study addresses this methodological gap by analyzing financial performance, using firm-level pre-tax ROA as proxy, in relation to sustainability performance, measured by ESG scores, using a 10-year longitudinal panel of European firms. Using ESG scores and ROA, the analysis focuses on the association between firm-disclosed ESG and financial performance data, rather than metrics that are subject to market evaluation. Consistent with evidence that ESG effects on financial outcomes typically occur with a time lag (Rahi et al. 2024; Kazyte et al. 2026), and considering that none of the recent European studies apply a lagged ESG specification, this study analyses the relationship between lagged ESG scores and ROA. As evidence shows that heterogeneity among firms, industries and markets affects the SP-FP relationship (Khan et al. 2016; Rahi et al. 2024; Rostamicheri et al. 2026), the regression uses firm fixed effects to control for unobserved firm-specific characteristics that do not vary over time, isolating the within-firm effect of sustainability performance on financial performance over time. To capture heterogeneity between industries, the study also analyses within-industry effects.

Against this background, the study adopts the following research aim and question.

1.5 Research Aim and Question

The aim of this study is to contribute to a better understanding of the relationship between firm-level sustainability performance and financial performance. Thus, the research focuses on answering this question:

Is financial performance positively associated with sustainability performance among European firms?

The study intentionally focuses on association rather than causality, consistent with the conceptual framing developed further in chapter 2, which positions this study within the correlation stream of the SP-FP literature. The study is also limited to firms in developed countries in Europe, as European firms have been in the forefront of sustainability regulation during the past decade (European Commission 2018; European Commission n.d.), where mandatory sustainability reporting requirements have improved ESG data availability and comparability. Despite recent growth in European evidence, European firms are still underrepresented in existing literature that mostly study American (USA) and global samples (Eccles et al. 2014; Flammer 2015; Khan et al. 2016). The use of sustainability performance metrics is limited to those provided by LSEG, thus clearly delimited from sustainability performance metrics that may be accessed by other suppliers.

1.6 Delimitations

The research is delimited to studying European firms across a ten-year period (2014-2023). As proxy for financial performance, a single accounting-based measure (pretax ROA) is used. As proxy for sustainability performance a specific composite metric (ESG Score) provided by a specific supplier (LSEG) is used. These methodological decisions on scope and measurement methods are designed to provide a balance between generalizability and precision, offering a scope that is broad enough to draw conclusions beyond the sample, but methodologically sharp enough to keep it focused and enable replicability.

2. Conceptual Framework

This chapter develops the theoretical basis for the study. It defines the key constructs of sustainability performance and financial performance, reviews the three main theoretical streams in the SP-FP literature, and concludes with the conceptualization adopted in this study as well as its theoretical contribution.

2.1 Sustainability Performance

Sustainability performance, in the context of this study, refers to how well a firm manages its ESG impacts and obligations. It includes both current quantifiable impacts, e.g. emissions reduction, employee injury rate; and long-term potential impact related to strategic activities, e.g. policy adoption, goal setting and investments.

As opposed to financial performance that can relatively easily be quantified and measured objectively using firms' financial statements that follow a global reporting standard, there is no global standard for measuring sustainability performance. Some activities, such as measuring greenhouse gas (GHG) emissions, are possible to objectively quantify and use as a comparable metric if expressed in relative terms, e.g. CO_2e to total revenues. Other activities, such as board diversity and supply chain governance, are more difficult to objectively quantify, measure and combine into a comparable metric. Thus, out of necessity, there is a greater need for subjective assessments when aggregating sustainability activities into quantifiable metrics. This creates two related problems: Firstly, non-quantitative outcomes are subjectively assessed and aggregated using different proxies. Secondly, as shown in Berg et al. (2022) and Christensen et al. (2022), different rating providers use different scopes and methods in terms of weights, proxies and other criteria, resulting in divergence in ESG ratings, i.e. the same firm can receive widely different scores from different rating providers.

As a way of mitigating these measurement challenges, the European Union adopted its first standardized framework for mandatory non-financial reporting, the Non-Financial Reporting Directive, NFRD, in 2014, subsequently replaced by a stricter framework, the Corporate Sustainability Reporting Directive, CSRD, to take effect in 2024. CSRD requires covered firms to comply with the European Sustainability Reporting Standards, ESRS, a set of standards that aim to make sustainability reporting more similar to financial reporting, by being transparent, structured, and comparable (European Commission n.d.). In December 2025 the Omnibus I directive was adopted by the EU, pulling back on several of the principles stated in CSRD and ESRS, to ease the administrative burden and increase competitiveness of European firms (Ibid.). A study by Mahmood et al. (2025) also shows that average ESG *disclosure* scores decreased significantly in 2024, the first reporting year following CSRD issuance. This was interpreted as firms adjusting behavior in anticipation of the new stricter requirements, suggesting that regulatory transitions themselves introduce short-term instability in ESG reporting data.

The most widely used proxy for measuring sustainability performance in empirical research is ESG scores, which are composite metrics, meaning that they aggregate multiple indicators from all ESG dimensions into a single comparable score (Friede et al. 2015; Gillan et al. 2021). This addresses the challenge of subjectivity in aggregation and inter-provider divergence, which is why the choice of ESG score provider matters and is discussed in the following section.

This study uses the LSEG ESG Score as its primary sustainability performance metric. As one of the most comprehensive databases available, the LSEG database covers both financial and ESG data for over 11,500 companies globally (LSEG 2024), enabling consistent data collection. The LSEG ESG Score is a composite metric constructed by combining 186 individual ESG metrics based on more than 870 datapoints that rely on firms' own reported data and assessments, including both quantifiable impact and long-term strategic activities. The score is expressed in relative terms on a 0-100 scale in relation to the performance of industry peers, according to the Thomson Reuters Business Classification, TRBC (LSEG 2024). Since scores are expressed relative to industry peers, the ESG score captures within-industry variation in sustainability performance rather than absolute ESG levels across industries. Furthermore, with firm and year fixed effects included, β_1 is identified from within-firm changes over time in a firm's industry-relative ESG score, rather than from cross-sectional differences between firms.

2.2 Financial Performance

In previous studies examining the relationship between sustainability performance and financial performance, two proxies are generally used for measuring financial performance: market-based measures (e.g. Tobin's Q, share price) or accounting-based returns (e.g. ROA, ROE). Market-based measures represent, in broad terms, the market's expectations of future financial returns, while accounting-based returns are measured retrospectively, based on firms' financial statements, usually in relation to total assets or total equity (Hawn et al. 2018). Thus, the market-based measure is a forecasted measure of anticipated future returns, while the accounting-based returns measure is based on reported historical returns.

Since the aim of this study is to examine the relationship between *actual* sustainability performance and *actual* financial performance, the accounting-based measure is a better proxy for financial performance as it is based on historical, rather than anticipated, returns. Return on Assets (ROA) is a better proxy than ROE (Return on Equity), because it measures returns in relation to the total accounting value, independently of how the company is financed (debt-to-equity ratio), making it a cleaner measure of financial performance. ROA is therefore the more appropriate measure for this study's focus on operational financial performance.

2.3 The Sustainability Performance – Financial Performance Relationship

The SP-FP theoretical field comprises broadly three main theoretical streams: causality, correlation and framing/measurement.

2.3.1 The Causal Relationship

The causality stream is grounded in stakeholder theory, based on the argument that firms which more successfully manage environmental, social and governance risks, reduce costs, attract capital, and build long-term competitive advantage. Eccles et al. (Eccles et al. 2014) claim empirical evidence that corporate sustainability integration causes financial out-performance in relation to firms that do not integrate sustainability to the same extent. Khan et al. (Khan et al. 2016) argue that ESG materiality is the driver for financial performance and Flammer (Flammer 2015) uses a quasi-experiment to provide evidence for causality between sustainability (CSR) and financial performance.

2.3.2 The Correlational Relationship

Authors of the correlation stream argue that causality cannot be established since reverse causality and unobserved firm characteristics make it impossible to isolate direction. King (2025) fails to replicate Eccles et al. (2014) and also critiques their original causality claims due to identified methodological issues, specifically missing and miscalculated significance tests, a non-replicable matching process, and the disappearance of results when firm fixed effects are included. Rahi et al. (Rahi et al. 2024) conclude through a meta study that there is generally a positive correlation between sustainability performance and financial performance, but in a time lag, while also recognizing that bidirectionality (reversed causality) and self-selection biases create nonconclusive results. This study positions itself within this stream, studying association without claiming to prove causality.

2.3.3 Framing/Measurement as Predictor of SP-FP Relationship

The framing/measurement stream points to the lack of standardized measurements and ratings, creating rating divergence and construct validity issues. Several studies question the original assumption that all sustainability is the same, differentiating between ESG activities, concepts and measures, highlighting their unequal effects on financial performance (Berg et al. 2022; Christensen et al. 2022). This study addresses this issue by using a single standardized metric (LSEG ESG Score) consistently applied across all firms and years in the panel.

Against this theoretical background, this study adopts the following conceptualization.

2.4 My Conceptualization

As this study aims at answering the question of whether financial performance is positively *associated* with sustainability performance (among European firms), the SP-FP relationship is conceptualized as a correlation, not a causal relationship. Observational panel data without any experimental manipulation cannot establish causality, therefore this study belongs in the correlational stream by design.

Furthermore, since the explicit aim of this study is to contribute to a better understanding of the association between *firm-level sustainability performance* and *financial performance*, the choice of performance measure proxies is key to the validity of this relationship. To ensure that these proxies measure what they are intended to, they are selected both based on measurement decisions made in previous similar studies, and from what they measure in relation to the association of interest for this study. Sustainability performance is measured by the LSEG ESG Score as a proxy, because it measures performance without market perception/evaluation of performance and has been used in previous similar studies (Candio 2024; Rostamicheri et al. 2026). Pretax ROA is the proxy for financial performance, as it measures actual accounting performance in terms of financial returns (as opposed to market value of performance), and has widely been used in previous studies of financial performance (Eccles et al. 2014; Candio 2024).

Grounded in stakeholder theory – firms that manage stakeholder relationships well reduce risks, attract capital, and build competitive advantage – SP is hypothesized to be positively associated with FP, but with a time lag, consistent with evidence that ESG effects materialize with a time lag. The relationship is studied at firm level, within European small, mid and large cap listed firms, over a 10-year panel.

2.5 Contribution

This study contributes to the growing body of European empirical research on the association between sustainability performance and financial performance. Combining lagged ESG, firm fixed effects, and within-industry analysis – a combination not applied in the recent European evidence – this study tests whether the negative associations documented by Candio (2024), Bifulco et al. (2023) and Rostamicheri et al. (2026) hold under a methodologically stronger specification. The findings of this study add to knowledge on whether the negative SP-FP association persists when controlling for firm-level heterogeneity and applying a lagged ESG specification. The table below shows specifically how this study differentiates methodologically.

Table 1. Methodological Positioning.

Methodological differentiator	Candio (2024)	Bifulco et al. (2023)	Rostamicheri et al. (2026)	This study (2026)
Lagged ESG specification	✗	✗	✗	✓
t-2 lag as robustness check	✗	✗	✗	✓
Firm fixed effects (two-way FE)	FE tested, not primary	✓	FE (non-Nordic) RE (Nordic)	✓
Hausman test (FE vs RE)	✗	✓	✓	✓
Firm-clustered standard errors	✗	✗	✗	✓
Within-industry regressions (by GICS sector)	Multi-level (different method)	✗	Nordic vs non-Nordic only	✓
Accounting-based FP (ROA)	✓	✗	✓	✓
STOXX Europe 600 sampling frame	✓	✓	✗	✓
Includes 2020–2023 (post-COVID)	Up to 2022	✗	✓	✓
ESG variable type	ESG Score	ESG Score	ESG Combined Score	ESG Score
E, S, G pillar regressions	✓	✗	✗	✓
Winsorization threshold	Not applied	Not applied	5th/95th per firm-year	1st/99th (more conservative)

3. Method

This chapter describes the methodological approach used to examine the association between sustainability performance and financial performance. The chapter is structured as follows: §3.1 outlines the research design; §3.2 describes the data, sample and panel structure; §3.3 defines the variables; §3.4 presents descriptive statistics; §3.5 details the empirical framework including estimation and identification strategy; and §3.6 addresses ethical considerations.

3.1 Research Design

This study adopts a quantitative deductive approach, which is appropriate for testing a stated hypothesis about an observable relationship between measurable variables (Bryman et al. 2019). It uses a longitudinal panel design that enables studying change within firms over time, which is essential for capturing within-firm variation in the SP-FP relationship over time and reducing reverse causality risk.

To be able to study variations within firms independent of external shocks, the studied time period should be sufficiently long to capture variations over time while controlling for other variables that might also affect sustainability performance and/or financial performance (Bryman et al. 2019). The panel is constructed by observing a single sample of firms retrospectively over 10 years, allowing the panel to capture variations within firms over time, enabling the analysis of associations between firm-year variables.

Regression-based panel data methods are applied to examine the relationship between sustainability performance indicators and financial performance measures. Lagged independent variables are used to explore directional effects between variables.

This methodological approach enables the study to control for unobserved firm-specific characteristics that remain constant over time. Where appropriate, robustness checks and alternative model specifications are used to assess the stability of the results.

3.2 Data

The empirical analysis is based on secondary data retrieved from the LSEG database. The following subsections describe the data source, the sample of firms, the study period, the treatment of missing data and outliers, and the structure of the panel.

3.2.1 Data Source

The empirical material consists exclusively of secondary data from LSEG's database tool, LSEG Workspace, accessed and retrieved via Python and the LSEG Data Library API (LSEG n.d.).

In 2009, Thomson Reuters acquired ASSET4, a large dataset with 250 ESG performance metrics on approximately 3000 companies (Integrity Research Associates 2009). The ASSET4 dataset has been widely used in CSR research and has since the acquisition by Thomson Reuters been further expanded to cover more companies and datapoints (De Villiers et al. 2022). In 2019 LSEG acquired Refinitiv (Reuters 2026), part of Thomson Reuters, including the ASSET4 dataset. The ESG dataset is now integrated into the LSEG database alongside financial and company data covering 99% of the world's market cap.

The LSEG database was selected for this study due to its extensive coverage of both financial and ESG data within a single database, as discussed in §2.1.

3.2.2 Sample

The sample consists of the constituents of the STOXX Europe 600 index registered at a fixed point in time (per 2026-02-11). Thus, instead of studying 'companies included in the index' at each given point in time, the index is only used as a frame for sampling European small, mid and large cap listed firms across a wide range of industries and markets.

The index was created and launched in June 1998 by Stoxx Ltd. and covers 17 European countries including the euro zone, United Kingdom, Switzerland and the Nordics, i.e. all of developed Europe, across all industries (STOXX n.d.). The free-float market capitalization, i.e. the total value of stocks that are freely available for trade in the open stock market (Dhand 2025), of stocks included in the index constitutes over 90% of the total market equivalent, thus representing a large portion of listed companies in all of developed Europe (STOXX n.d.).

Firms that are based in United Kingdom, France, Germany and Switzerland together constitute the largest portion of the index, weighing about 65% of the total index composition. Among industries, the heaviest represented are financials, industrials and healthcare, weighing roughly 55% of the total index composition. The index composition is reviewed quarterly, on the third Friday of March, June, September and December (STOXX n.d.).

In order to identify firms and trace historical data for the constituting firms, a combination of the RIC number, which is an LSEG specific ID number, and the International Securities Identification Number (ISIN), is used, of which the latter ensures replicability and comparability since it is generally available in other databases and commonly used in financial research.

3.2.3 Period

The study focuses on European firms observed retrospectively over a 10-year period, 2014-2023, enabling a longitudinal analysis using secondary data collection. As Bryman et al. (Bryman et al. 2019) point out, observing a sample over a long period of time not only allows for studying long-term organizational effects, but also reduces the risk of reverse causality since there can be a time lag between dependent and independent variables.

The study period has been chosen to coincide with the implementation of the NFRD framework in 2014 and pre-implementation of the CSRD framework in 2024, meaning ESG reporting across the panel is governed by the same reporting policy framework throughout, ensuring comparability between firm-years. From the implementation of NFRD in 2014, ESG data availability has increasingly improved, making it a natural starting point for studying European firms. Starting in 2014 also avoids post-financial crisis effects that may distort the panel. The time period is intentionally as recent as possible, without risking missing potentially not yet reported data due to non-calendar fiscal years and long reporting cycles.

3.2.4 Missing Data and Outlier Treatment

Firms where ESG scoring was missing in 4 or more years in total were excluded from the panel, in order to have sufficient data to perform longitudinal analysis and firm fixed effects analysis. Note that consistently excluding firms with less historical ESG data constitutes a selection bias towards firms with more complete ESG reporting histories, which may affect the results.

Firm-years with any of ROA, ESG lag 1, Size, Leverage or Tobin’s Q missing were excluded from the panel.

After cleaning the data, the original sample of 600 firms and 6000 firm-year values was reduced to 508 firms and 4833 firm-year values. Thus, the attrition rate of firms is 15.33% and the attrition rate of firm-years is 19.45%.

Winsorization is applied to all continuous variables at the 1st and 99th percentile to mitigate the influence of outliers, which is standard practice in panel data research (Mahmood et al. 2025).

3.2.5 Panel Data Structure

The original panel constitutes the 600 firms that as of February 11, 2026, are included in the STOXX Europe 600 index, studied retrospectively over a period of 10 years (2014-2023). Note that this selection model implies a natural survivorship bias, including only firms that are still existing (as of 2026-02-11).

The panel is unbalanced due to missing firm-year observations. Firm-years with any of ROA, ESG lag 1, Size, Leverage or Tobin’s Q missing have been entirely excluded. Firms where ESG scoring was missing in 4 or more years in total were excluded from the panel.

Table 2. Sample Construction.

Step	Firms	Firm-year observations
Initial sample	600	6000
Excluding firms with insufficient ESG data	509	5090
Excluding firm-year observations with missing ROA, ESG lag 1, Size, Leverage or Tobin's Q	508	4833

3.3 Variable Definitions

This section defines the dependent, independent and control variables used in the regression model, as well as the fixed effects included in the specification.

3.3.1 Independent Variable

Firm Sustainability Performance

As a firm sustainability measure, the LSEG ESG Score is used, which is a composite metric expressed in a 0-100 number, representing a performance value relative to other firms within the same industry, according to the Thomson Reuters Business Classification (TRBC). The metric is calculated by combining 186 comparable ESG measures based on more than 870 datapoints that are weighted, categorized into Environmental, Social and Governance, and combined into an overall industry relative score of 0-100 (LSEG 2024).

Since the score is expressed relative to industry peers, it captures within-industry variation in sustainability performance rather than absolute ESG levels across industries. With firm and year fixed effects included, β_1 is identified from changes within the same firm over time in its industry-relative ESG score, rather than from differences between firms.

The measure is used in lagged form (baseline = t-1) to capture real lagged effects and to avoid reversed causality (Rahi et al. 2024; Kazyte et al. 2026).

3.3.2 Dependent Variable

Firm Financial Performance

ROA (Return on Assets) is used as proxy for firm financial performance, consistent with Eccles et al. (2014) and widely used in SP-FP research. In order to exclude national tax differences, pretax ROA is used. As argued in §2.2, ROA is chosen over ROE (Return on Equity) because it measures the financial performance ratio independent of the financial structure of the firm, which is instead controlled for using specific control variables.

3.3.3 Control Variables

Control variables include commonly used variables in ESG and financial performance research that are expected to have an effect on the dependent variable (Eccles et al. 2014; Khan et al. 2016; Hawn et al. 2018; Gillan et al. 2021):

Firm size

As a proxy for firm size Total Assets is used, measured on a logarithmic scale, log(Total Assets), to account for large differences in size between firms. Size is expected to affect ROA negatively, due to diseconomies of scale and bureaucratic inefficiency, while also recognizing that size is typically positively associated with more stable earnings.

Leverage

Leverage, measured as the debt over assets percentage ratio (Debt/Assets), controls for effects by the extent to which a firm is financed by debt. A firm that has more debt in relation to its total assets, typically has higher interest costs and increased financial risk, effectively decreasing ROA.

Growth Opportunities

Tobin's Q (Enterprise Value/Assets) captures the market's assessment of a firm's growth opportunities relative to its asset base. Fast growing firms typically have a lower current ROA due to growth and expansion costs and increased investments which delay earnings realization.

Firm Fixed Effects

Firm fixed effects are included to control for unobserved firm-specific characteristics that remain constant over time, allowing for the observation of in-firm changes in sustainability and financial performance over time (Angrist & Pischke 2008).

Year Fixed Effects

Year fixed effects are included to control for unobserved year-specific characteristics such as external shocks or other macroeconomic characteristics that affect all firms in a given year (Angrist & Pischke 2008).

Beta was also considered as a control variable but excluded from the baseline since it is a market-based risk measure with a weak theoretical link to accounting-based ROA and is instead included as a robustness check.

3.4 Descriptive Statistics

The table below shows descriptive statistics of the variables that are being used to analyze the sample of firms during 2014-2023.

Table 3. Descriptive Statistics.

Variable	Mean	SD	Min	Max	N
ROA	6.868	7.704	-12.316	38.699	4,833
ESG	67.116	16.553	16.301	92.708	4,831
ESG_{t-1}	65.485	17.662	12.980	92.672	4,833
ESG_{t-2}	64.142	18.271	11.515	92.610	4,755
ln(Tot. Assets)	23.763	1.827	20.001	28.553	4,833
Leverage	0.234	0.151	0.000	0.627	4,833
Tobin's Q	2.279	3.233	-0.045	20.641	4,833

The statistics show an average (mean) ROA of 6.868 % with a large variation between firms in the sample. The negative min value indicates firms that are run with a loss. The mean ESG score of 67.116 indicates an above average ESG performance relative to their respective industry peers. The variation in ESG score reflects the wide range of sustainability performance levels across firms in the sample. Two firm-years have a missing ESG score while still having a valid lagged ESG value, which is why N differs slightly between ESG and ESG (t-1), N = 4,831 vs 4,833. ESG (t-2) has fewer observations (N = 4,755) than ESG (t-1) as it requires two prior years of ESG data. ESG, ESG (t-1) and ESG (t-2) all show different values due to individual winsorization and different total firm-years. The size, proxied by the logarithm of total assets, shows a relatively narrow size span (min-max value) and a small standard deviation of 1.827, indicating a fairly homogenous sample in terms of size. The leverage (debt-to-assets ratio) min value of 0.000 reflects firms with no debt that will automatically receive a null value. The Tobin's Q negative min value of -0.045 shows that there are firms within the 1-99 percentile of the sample that effectively hold more cash than their market cap, typically seen in distressed or capital-light firms sitting on large cash reserves with depressed market valuations. All statistics are post-winsorization.

3.5 Empirical Framework

While Chapter 2 established the theoretical and conceptual basis for the study, this section describes the empirical framework — the statistical approach used to operationalize and test the hypothesized relationships between sustainability performance and financial performance.

This study explores the relationship between firm-level sustainability performance and financial performance over time using panel data regression methods. The empirical framework consists of two parts: The estimation strategy where methods for statistically estimating the association between variables is explained, and the identification strategy discussing implications for interpretations and conclusions enabled by the empirical design (Bellemare 2022).

The analysis uses firm-level panel data to examine whether sustainability performance variations associate with financial performance variations over time. Fixed effects panel regression is applied to control for unobserved firm-level variables and external year-specific variables that might affect sustainability and financial performance.

3.5.1 Estimation Strategy

The empirical analysis estimates the relationship between sustainability performance and financial performance using the following panel regression model that applies a fixed effects approach, as laid out by Angrist & Pischke (2008):

$$ROA_{it} = \alpha + \beta_1 ESG_{it-1} + \gamma X_{it} + \mu_i + \lambda_t + \epsilon_{it}$$

Where:

- ROA_{it} represents financial performance measured by return on assets for a given firm and year
- ESG_{it-1} represents lagged sustainability performance measured by the LSEG ESG score for a given firm and year -1
- α represents the intercept, the constant term of the regression model
- β_1 represents the coefficient of interest, capturing the relationship between sustainability performance and financial performance
- γ represents the coefficients associated with the control variables
- X represents all control variables including firm size, leverage, and growth opportunities
- μ_i represents firm fixed effects
- λ_t represents year fixed effects

- ϵ_{it} represents the error term
- i represents firm
- t represents time (year)

In addition to the baseline regression model, an extended equation is used to capture the joint lag effect of ESG_{t-1} and ESG_{t-2}:

$$ROA_{it} = \alpha + \beta_1 ESG_{it-1} + \beta_2 ESG_{it-2} + \gamma X_{it} + \mu_i + \lambda_t + \epsilon_{it}$$

Both ESG variables are included simultaneously in the extended model, to avoid omitted variable bias due to high autocorrelation of ESG scores across adjacent years within firms (Angrist & Pischke 2008, Ch. 3). That way the association of one variable can be isolated from the effect of the other variable. The joint Wald test is used to test whether $\beta_1 = \beta_2 = 0$, i.e. that both ESG lag coefficients are simultaneously equal to zero (no joint effect on ROA). A statistically significant result ($p \leq 0.05$) rejects the null hypothesis, indicating that at least one of the ESG lag coefficients is jointly different from zero.

Standard errors are clustered at the firm level to account for within-firm serial correlation in the error terms. This methodology is preferable to un-clustered standard errors, but typically produce larger standard errors, thus making it more difficult to achieve statistical significance (Cunningham 2021, Ch. 8.2.3).

The hypotheses tested in the empirical analysis are:

> Null hypothesis (H_0): The relationship between sustainability performance and financial performance is non-positive ($\beta \leq 0$).
>
> Alternative hypothesis (H_1): The relationship between sustainability performance and financial performance is positive ($\beta > 0$).

The analysis is implemented in Python using the linearmodels package for panel data estimation.

To assess the stability of the baseline results, the following robustness checks are performed, see table below.

Table 4. Robustness Checks Explained.

Check	Description	Motivation
E, S, G pillars	Re-run joint lag model separately for E, S and G pillar scores	Tests whether the aggregate ESG effect is driven by a specific pillar
ESG Combined Score	Re-run joint lag model (t-1 + t-2) with ESG Combined instead of ESG Score	Tests comparability with Rostamicheri et al. (2026), who use the combined score
Industry × Year FE	Re-run joint lag model adding industry × year interactions	Tests robustness to time-varying industry shocks beyond firm FE

Country × Year FE	Re-run joint lag model adding country × year interactions	Tests robustness to time-varying country shocks beyond firm FE
Year FE only	Re-run joint lag model with year FE only, no firm FE	Captures between-firm variation; tests sensitivity to fixed effects specification
Beta control	Re-run joint lag model adding Beta as control	Tests whether market-based risk measure affects baseline results
No winsorization	Re-run joint lag model on non-winsorized data	Tests sensitivity of results to outlier treatment

3.5.2 Identification Strategy

The empirical design does not allow strict causal identification but aims to estimate the relationship between sustainability performance and financial performance while mitigating several common sources of statistical bias (Angrist & Pischke 2008).

To capture potential effects on financial performance from sustainability performance, the relationship needs to be studied with a time lag. Firstly, this method reduces the risk of reversed causality, i.e. that a dependent variable is mistakenly treated as an independent variable – in this case, that sustainability performance is an effect of financial performance, rather than the anticipated reversed association. Secondly, the time lag will increase the chances of capturing actual financial performance effects from changes in sustainability performance, as some activities that aim at improving sustainability performance will naturally need some time to take effect. And lastly, as some investments in sustainability are not treated as financial investments that are capitalized over their economic lifespan, e.g. learning, R&D, and supply chain management, there will be an accounting mismatch, where costs are disproportionally high at the time of undertaking the investment (initially affecting financial performance negatively), and disproportionally low during the time of their expected returns (affecting financial performance positively with a time lag).

Reverse causality may arise if financially successful firms have greater resources available for sustainability investments (Rahi et al. 2024). The use of lagged sustainability performance (t-1) partially addresses this issue by introducing temporal separation between the explanatory and outcome variables. For robustness, t-2 will be used to check for effects that occur with an extended time lag.

Unobserved firm-specific characteristics may influence both sustainability performance and financial outcomes. Firm fixed effects are therefore included in the regression model to control for time-invariant firm characteristics such as:

- Managerial/organizational culture
- Corporate governance structures

- Long-term strategic orientation
- Brand, reputation and historical legacy

Unobserved time-specific effects on all firms within a sample may influence both sustainability performance and financial outcomes. Year fixed effects are therefore included in the regression model to control for firm-invariant time effects such as:

- Macroeconomic cycles and recessions
- Global financial market conditions
- Commodity price shocks (e.g. energy, raw materials)
- Pandemic and geopolitical shocks
- Shifts in ESG reporting norms and investor sentiment
- Changes in the regulatory environment affecting all firms

By controlling for these time-invariant and time-varying characteristics, firm fixed effects identify only within-firm changes over time, in this case within-firm changes in ROA in relation to within-firm changes in ESG score. This can be compared to cross-sectional (pooled OLS) models that capture between-firm changes, i.e. average ESG in relation to average ROA across the cross-section. A model that combines pooled OLS with year fixed effects captures a mix of both, i.e. cross-sectional variation between firms as well as time-varying shocks that affect the entire sample equally, whereas firm fixed effects isolate only within-firm variation over time. The distinction is important to make, since the choice of model can affect the direction of the association between ESG and ROA, as demonstrated in chapter 4.1.

Fixed and random effects differ in their assumption about the correlation between effects and independent variables. Random effects assume that effects are uncorrelated with the independent variable, while fixed effects assume correlation with the independent variable. The choice between fixed and random effects as the estimator for capturing this within-firm variation is determined by the Hausman test that formally tests whether the coefficients estimated by fixed and random effects differ systematically. If there is no significant variation, random effects should be used since it is a more efficient estimator; if there is significant variation between the two estimators, fixed effects should be used to avoid biased results. Since unobserved firm characteristics are expected to be correlated with both sustainability performance and financial performance, the random effects assumption of no correlation between effects and regressors is likely violated. The Hausman test formally confirms fixed effects over random effects in this sample, see result in chapter 4.1.

While industry fixed effects are commonly considered, they are intentionally excluded in the baseline model because the LSEG ESG Score is an industry relative score. Therefore, industry fixed effects are already absorbed in the variable, as effects on the whole industry will not change the score relative to its industry peers. Controlling for industry fixed effects would thus create a double control for industry. Industry–year interactions are instead included as a robustness check to test sensitivity to time-varying industry shocks.

Country fixed effects are also excluded in the baseline, as firm fixed effects already absorb time-invariant country-specific effects, since a firm's country of headquarters rarely changes over time. Thus, an additional country dummy in the model would be largely redundant for the time-invariant component. As for time-varying country shocks, those are partially captured by year fixed effects. However, country–year interactions are included as a robustness check to test sensitivity to residual time-varying country effects.

Measurement error may arise due to differences in ESG reporting practices and ESG rating methodologies. The use of the LSEG ESG Score provides a standardized metric across firms, but potential measurement limitations remain and should be considered when interpreting the results.

Overall, the empirical strategy allows the analysis to examine whether changes in sustainability performance within firms over time are associated with changes in financial performance, while controlling for observable and unobservable confounding factors where possible.

3.6 Ethical considerations

3.6.1 Data Collection

The study is based solely on secondary data derived from publicly available sources. No personal or sensitive data will be collected, and therefore no ethical approval is expected to be required.

3.6.2 AI Assistance

Throughout the work process generative artificial intelligence (GenAI) has been used for ideation on topics, literature search, work process and research method, critical reflection, disposition and language review as well as technical assistance on database software. The European Commission's guidelines on the use of generative AI in research (European Commission 2025) have been adhered to, with emphasis on responsibility, accountability and transparency. The GenAI tools that have been used for AI assistance are mainly ChatGPT and Claude. Analysis, discussion and conclusions are solely the author's own.

4. Results and Analysis

This chapter presents and analyses the empirical results. §4.1 presents the baseline regression results; §4.2 assesses robustness across alternative specifications; §4.3 examines within-industry heterogeneity; §4.4 analyses results by firm size; and §4.5 discusses the limitations of the study.

4.1 Baseline Results

Using the panel regression model,

$$ROA_{it} = \alpha + \beta_1 ESG_{it-1} + \gamma X_{it} + \mu_i + \lambda_t + \epsilon_{it}$$

the relationship between ESG and ROA has been estimated and presented in the table below, where X_{it} is a vector of time-varying control variables (Size, Leverage, Tobin's Q), μ_i denotes firm fixed effects, and λ_t denotes year fixed effects. Each value represents the percentage point change in ROA associated with a change in the corresponding variable, for ESG this corresponds to a one-point change on the 0-100 scoring scale; for control variables the unit of change corresponds to the variable definitions in chapter 3.3. Column (4) extends the baseline by adding ESG_{it-2} alongside ESG_{it-1} to form a joint lag model.

Table 5. Baseline Results.

Dependent variable: Pretax ROA (%). Standard errors clustered at firm level in parentheses. All continuous variables winsorized at 1st/99th percentile.

	(1) Controls only	(2) + Year FE	(3) + Firm FE	(4) Joint lag model
ESG (t-1)	0.0107 (0.0107)	0.0129 (0.0110)	-0.0122 (0.0121)	0.0100 (0.0113)
ESG (t-2)	-	-	-	-0.0364*** (0.0129)
Size ln(Total Assets)	0.1657*** (0.0345)	-1.0290*** (0.1390)	-1.4284* (0.8559)	-1.4470* (0.8662)
Leverage (Debt/Assets)	-3.7374** (1.4577)	-6.2264*** (1.3612)	-19.1045*** (2.4270)	-19.4270*** (2.4531)
Growth potential (Tobin's Q)	1.3042*** (0.1320)	0.9792*** (0.1402)	-0.1642 (0.1363)	-0.2216 (0.1416)

R² (within)	—	-0.0233	0.0946	0.0942
N	4,833	4,833	4,833	4,755
Firm FE	No	No	Yes	Yes
Year FE	No	Yes	Yes	Yes
Joint Wald test (ESG lags)	-	-	-	$\chi^2(2) = 8.03$ p = 0.018**

*Note: *** p < 0.01, ** p < 0.05, * p < 0.10. Column (3) is the baseline specification, selected as preferred estimator over random effects by the Hausman test ($\chi^2(4) = 301.21$, $p < 0.001$). Column (4) includes ESG (t-1) and ESG (t-2) simultaneously to avoid omitted variable bias from replacing one lag with the other. The joint Wald test in column (4) tests whether ESG_lag1 = ESG_lag2 = 0 jointly. Yellow highlight = baseline ESG coefficient. Green highlight = significant ESG (t-2) effect and joint test.*

As recommended by Bellemare (2020), results should be presented in a certain order, from most to least parsimonious. The table above thus presents the results accordingly from left to right, where the first column shows the results of a regression of pooled ordinary least squares (OLS), which is a method that treats the full panel as a cross-sectional data set without considering any firm or year specific characteristics. These results indicate a weak positive association between ESG (t-1) and ROA, after controlling for size, leverage and Tobin's Q, however the results are insignificant at a 5% significance level ($\beta = 0.0107$, $p = 0.315$). The association remains weakly but insignificantly positive when adding year fixed effects ($\beta = 0.0129$, $p = 0.241$) but flips to a negative association when adding also firm FE ($\beta = -0.0122$, $p = 0.314$). The negative association is insignificant ($p > 0.05$) with a confidence interval of -0.0358 (lower) to 0.0115 (upper), including zero, indicating that the real effects may be either negative or positive or none (zero). Thus, the research null hypothesis (H_0) cannot be rejected for ESG (t-1).

The Hausman test reveals that fixed effects is the preferred estimator over random effects, since firm characteristics are clearly correlated with the independent variables ($\chi^2(4) = 301.21$, $p < 0.001$). The sign reversal (plus to minus) illustrates how unobserved time-invariant firm characteristics confound the cross-sectional ESG-ROA relationship, and underscores why firm fixed effects is the appropriate specification.

Observing the association between ROA and the behavior of specific control variables across models, may indicate the quality of the baseline model used to analyze the association between the independent and the dependent variable. All control variables (size, leverage, Tobin's Q) show statistically significant associations with the dependent variable (ROA) across the first two columns (Controls only, Year FE). Size, proxied by ln(Total Assets), initially shows a positive association but then flips to negative when adding year FE and remains negative when also adding firm FE. This pattern indicates that there are large firms in the panel with inherently high ROA, showing as a strong positive correlation in the cross-sectional analysis, however when controlling for year FE and firm FE the association that appears is in fact negative. Leverage (Debt-to-Assets ratio) shows a statistically significant increasingly negative association across columns from left to right, where the biggest leap occurs between year FE (-6.2264***) and firm FE (-19.1045***). Tobin's Q shows an interesting pattern where the association is initially positive with high significance across the first two columns, but then flips to insignificantly negative, close to zero, as firm FE are added. This pattern implies that some firms are consistently highly valued and profitable, but when this fixed firm characteristic is controlled for there is no significant association between changes in Tobin's Q and profitability (ROA).

To observe whether patterns remain the same for additional time lag, a joint lag model is applied, adding ESG (t-2) alongside ESG (t-1). That way the model may simultaneously include both lags in the regression, to avoid omitted variable bias that would arise from replacing one lag with the other, given the high autocorrelation of ESG scores across adjacent years. The model shows an insignificant positive association between ESG (t-1) and ROA ($\beta = 0.0100$, $p = 0.377$), i.e. the one-year effect becomes insignificantly positive once ESG (t-2) is included in the model. The previously negative association is instead fully captured in ESG (t-2) and ROA, returning a statistically significant result at a 1% significance level ($\beta = -0.0364$, $p = 0.005$). The confidence interval is entirely negative (-0.0617, -0.0111), excluding zero, confirming the negative direction at the 5% significance level. Since the direction is negative rather than positive, the research null hypothesis (H_0: $\beta \leq 0$) cannot be rejected.

To test whether ESG (t-1) and ESG (t-2) are jointly statistically significant, a Joint Wald test is performed, which confirms a joint significance at a 5% significance level ($\chi^2(2) = 8.03$, $p = 0.018$), with the effect specifically showing at the two-year horizon (t-2).

The negative association between ESG and ROA at ESG (t-2) are consistent with the results of Bifulco et al. (2023), Candio (2024) and Rostamicheri et al. (2026), as they all find negative associations between ESG and financial performance. Noteworthy is that none of the other studies use lagged ESG, while this study shows the only statistically significant association at t-2, and none at t-1. This study applies firm-clustered standard errors (see §3.5.1), which are more conservative than the standard errors used in the comparator studies, and may partly explain why significance is achieved only at the two-year horizon rather than contemporaneously. Bifulco et al. (2023) also use the STOXX Europe 600 index as sampling frame, but use stock price as proxy for financial performance, suggesting the negative direction may not be sensitive to whether financial performance is measured on an accounting or market basis, though the measures capture different aspects of firm performance. Bifulco et al. observe a shorter time series, 2014-2020, excluding post-covid years, which may also affect the results. Candio's (2024) findings are very similar to this study's findings, since Candio also uses the STOXX Europe 600 index and to a very large extent overlapping time period (2012-2022). Candio does not apply a time lag but receives similar results for the contemporaneous ESG – ROA association ($\beta = -0.040$***) as this study's results for the ESG (t-2) – ROA association ($\beta = -0.0364$***), in terms of magnitude and significance, albeit different lag structures. Rostamicheri et al. (2026) finds a statistically significant negative association between ESG (no time lag applied) and ROA/ROE in their subsample of non-Nordic European firms, however ESG associations among Nordic firms show no statistical significance. Although the data is retrieved from the same database as this study's, the ESG *Combined* Score is being used instead, which is a metric similar to the ESG Score but also including controversies and perceived ESG performance based on media coverage. Therefore, they are essentially different metrics and direct comparisons of coefficients should thus be made with caution.

4.2 Robustness Checks

The purpose of the robustness checks is to assess whether the baseline t-2 findings hold under alternative specifications, addressing concerns about model sensitivity, control variable endogeneity and measurement choices. The joint lag specification is used in all checks to investigate individual and joint association under both t-1 and t-2.

Table 6. Robustness Checks Results.

Dependent variable: Pretax ROA (%). All specifications include ESG (t-1) and ESG (t-2) simultaneously unless noted. Firm FE + Year FE + clustered SE at firm level unless noted. Winsorization at 1st/99th percentile unless noted. The ESG Combined Score is based on the ESG Score with the addition of ESG controversies, a component that is also supposed to absorb some of the public perception of a firm's ESG performance.

Specification	β (t-1)	β (t-2)	p (t-1)	p (t-2)	Wald p	N
Baseline (joint lag)	+0.0100	-0.0364***	0.377	0.005	0.018**	4,755
1. No controls	+0.0178	**-0.0375*** **	0.137	0.003	0.010**	4,755
2. No winsorization	+0.0068	**-0.0470*** **	0.624	0.008	0.026**	4,755
3. ESG Combined Score	+0.0075	**-0.0349*** **	0.370	0.008	0.027**	4,755
4. Beta control added	+0.0068	**-0.0365*** **	0.551	0.006	0.024**	4,673
5. Year FE only (no firm FE)	+0.0252*	-0.0139	0.070	0.348	0.173	4,755
6. Industry × Year FE	+0.0142	**-0.0297** **	0.200	0.016	0.033**	4,755
7. Country × Year FE	+0.0044	**-0.0450*** **	0.706	0.001	0.003***	4,755

*Note: *** p < 0.01, ** p < 0.05, * p < 0.10. Clustered SE at firm level throughout. Green = ESG (t-2) significant and negative, consistent with baseline. Red = ESG (t-2) not significant (Year FE only, no firm FE).*

For all robustness checks that include firm fixed effects, the results are consistent with the baseline result, showing negative significant effects at 5% or better. Thus, the direction and significance of the robustness checks are consistent, concluding that the baseline result is robust.

When firm fixed effects are removed, keeping only year FE, ESG_{t-1} flips to weakly positive and ESG_{t-2} becomes insignificant (while still showing a negative direction). This pattern mirrors the sign reversal observed between columns (1) and (3) in Table 5 and confirms the Hausman test conclusion that firm characteristics are correlated with the ESG variable, making the within-firm estimator essential for unbiased inference.

As the controls used in the baseline regression may be partially endogenous – sustainability performance could very well affect Leverage, Tobin's Q and Size over time – check 1 tests this hypothesis specifically. The results with no controls show only a minor difference to the baseline results for t-2, while t-1 shows equally insignificant results in the same positive direction as the baseline results.

To investigate whether the independent variables are multicollinear (correlate with one another as well as the dependent variable), Pearson correlation coefficient (r) is calculated to show the pairwise strength and direction of correlation. The variance inflation factor (VIF) is then calculated for each variable to detect how much its coefficient variance is inflated due to multicollinearity with the other regressors.

Table 7. Within-Firm Correlation Matrix and Variance Inflation Factors.

Pearson correlations (r) computed on within-firm demeaned variables, consistent with the firm FE estimator. N = 4,833 firm-year observations. VIF threshold of concern: > 5.

Variable	ESG (t-1)	Size	Leverage	Tobin's Q	VIF	Verdict
ESG (t-1)	1.000	0.480	0.089	-0.395	1.309	No concern
Size	0.480	1.000	0.236	-0.724	2.345	No concern
Leverage	0.089	0.236	1.000	-0.236	1.070	No concern
Tobin's Q	-0.395	-0.724	-0.236	1.000	2.137	No concern

Note: Diagonal values (r = 1.000) are self-correlations. Yellow = highest pairwise correlation (Size–Tobin's Q, r = -0.724).

The highest pairwise correlation is found between Tobin's Q and Size (r = -0.724), however VIF values of 2.137 and 2.345 respectively show that this correlation does not meaningfully inflate ESG_{t-1}. The correlation between ESG_{t-1} and Size is also notable (r = 0.480), however VIF value of 1.309 confirms that this correlation does not inflate ESG_{t-1}. In conclusion, multicollinearity is not a concern in the baseline model.

As the ESG score is composed by combining and weighing three separate ESG pillar scores – Environmental Score, Social Score and Governance Score – it represents a weighted average score, without disclosing the effect of each pillar. Therefore, it is relevant as a robustness check, to investigate each pillar score's individual coefficient.

Table 8. Pillar Regressions.

Each pillar estimated separately using the joint lag specification (firm FE + year FE + clustered SE). ESG (t-1) and ESG (t-2) replaced by the respective pillar score at t-1 and t-2 simultaneously.

ESG Pillar	**β (t-1)**	**β (t-2)**	**p (t-1)**	**p (t-2)**	**Wald p**	**N**
Environmental Score	-0.0007	-0.0144	0.945	0.145	0.290	4,755
Social Score	+0.0131	-0.0164*	0.115	0.100	0.117	4,755
Governance Score	+0.0011	**-0.0147****	0.870	0.034	0.104	4,755

*Note: *** p < 0.01, ** p < 0.05, * p < 0.10. Green = statistically significant at 5%. Yellow = marginally significant at 10%.*

The pillar regressions reveal that the Environmental score shows no significance in either t-1 or t-2 and the Social score only marginal significance at 10% significance level (β = -0.0164*, p = 0.100) in a negative direction in t-2. The biggest driver of the aggregate ESG score coefficient seems to be the Governance score that shows a significant negative effect in t-2 (β = -0.0147**), which is consistent with the direction of the ESG score. None of the pillars show any significant effect in t-1. The joint Wald test shows no joint association, which indicates that the aggregate ESG effect is not strongly concentrated in one pillar.

In conclusion, the robustness checks collectively support the reliability of the baseline t-2 finding across measurement choices, model specifications and control configurations. The results from the individual ESG pillar regressions, offer clues to the mechanics of the SP-FP association, as governance seems to be the main driver of the negative effect over time.

4.3 Within-Industry Analysis

To receive a deeper understanding of the industry distribution of results of the panel regression, a within-industry analysis is performed. Industry heterogeneity is an expected feature of the ESG-FP relationship since the LSEG ESG Score is industry-relative, meaning within-industry variation in ESG is the relevant signal. The within-industry analysis uses the joint lag model, consistent with the extended baseline model (column 4, Table 5).

Table 9. Within-Industry Regression Results.

Dependent variable: Pretax ROA (%). Firm FE + Year FE, standard errors clustered at firm level. ESG (t-1) and ESG (t-2) included simultaneously. Each sector estimated separately.

GICS Sector	Firms	Obs.	β (t-1)	β (t-2)	p (t-1)	p (t-2)	Wald p
Communication Services	25	225	-	-	-	-	-
Consumer Discretionary	50	469	+0.0540	-0.0148	0.346	0.787	0.610
Consumer Staples	37	353	-0.0357	-0.0320	0.268	0.207	0.058 *
Energy	21	199	-	-	-	-	-
Financials	105	995	+0.0291	-0.0124	0.152	0.487	0.356
Health Care	37	331	+0.0554	-0.0174	0.189	0.626	0.376
Industrials	113	1,053	+0.0282	-0.0547**	0.173	0.023 **	0.066 *
Information Technology	22	193	-	-	-	-	-
Materials	41	404	-0.0099	-0.0288	0.851	0.522	0.812
Real Estate	26	235	-	-	-	-	-
Utilities	31	298	+0.0066	-0.0282	0.881	0.393	0.618
7 sectors estimated (4 excluded)	**433**	**4,098**					

*Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Sectors with fewer than 30 firms excluded (Communication Services, Energy, Information Technology, Real Estate). Green highlight = statistically significant result. Yellow highlight = marginally significant joint Wald test. Wald p reports the p-value of the joint test that $\beta(t-1) = \beta(t-2) = 0$. Industrials shows a significant negative ESG (t-2) coefficient ($\beta = -0.055$, $p = 0.023$) with a marginally significant joint Wald test ($p = 0.066$), consistent with the aggregate baseline finding.*

Consistent with the aggregated result, ESG_{t-1} shows mixed positive and negative associations across industries, none of which are statistically significant, while ESG_{t-2} shows a clearer negative pattern. Industrials, which constitutes the largest representation (113 firms, 1,053 observations = 27% of estimation sample) flips sign from insignificantly positive in ESG_{t-1} to significantly negative in ESG_{t-2} ($\beta = -0.0547$**), and the joint Wald test also reveals joint significance ($p = 0.066$*). Accounting for approximately 40% of the weighted sector contribution to the aggregate result, Industrials is the single largest contributor to the negative aggregate ESG_{t-2} coefficient. All other industries show negative ESG_{t-2} coefficients, albeit only one other, Consumer Staples, can show a statistically significant joint association result ($p = 0.058$*). The pattern across columns mirrors the same pattern on an aggregated level, indicating that the aggregated results are widely represented across industries. However, four industries (Communication Services, Energy, IT, Real Estate) are excluded due to their subsamples not reaching the 30-firm minimum threshold that is commonly applied for clustered standard errors to be reliable. This represents a limitation of the within-industry analysis, as the ESG-ROA relationship in these four sectors cannot be examined within this study's framework.

To further investigate the mechanism behind the Industrials finding, pillar-level regressions are run separately within each sector, replacing the aggregate ESG score with the Environmental, Social and Governance pillar scores at t-1 and t-2 simultaneously.[1] Within Industrials, the Environmental pillar shows the strongest and only statistically significant negative t-2 association ($\beta = -0.058$, $p = 0.011$), while the Social pillar is marginally significant ($\beta = -0.032$, $p = 0.066$) and the Governance pillar shows no association ($\beta = +0.003$, $p = 0.810$). This pattern is consistent with the transition cost hypothesis applied at the sector level — the costs depressing Industrials' ROA reflect environmental operational investments rather than governance restructuring. The marginally significant Social Pillar may indicate that workforce and supply chain-related social investments may contribute to the short-run cost effect alongside environmental compliance, though the evidence is not conclusive at conventional significance levels. Notably, Industrials is not a large cap-heavy sector — 50% of Industrials firms fall in the small cap tertile and the sector median total assets (€6.4bn) sit at the Small/Mid threshold — which rules out firm size or organizational complexity as the primary driver of the effects.

The same environmental pattern is found in Utilities, where the Environmental pillar also shows a significant negative t-2 association ($\beta = -0.040$, $p = 0.022$). Both sectors share capital-intensive environmental transition requirements,

[1] Sector-level pillar results are not tabulated separately. Full results are available from the author on request.

lending further support to the interpretation that the negative t-2 ESG-ROA association in operationally intensive sectors reflects the upfront costs of environmental compliance and energy transition investments. No other sector shows a statistically significant pillar effect at t-2, and no sector shows a dominant Governance effect at the sector level, suggesting that the aggregate Governance significance observed in the full-sample pillar regressions (§4.2) reflects a diffuse effect spread across sectors rather than concentration in any single industry.

4.4 Firm Size Analysis

To find out whether baseline t-2 findings hold across firm size segments, the sample is divided in three equal-sized groups (tertiles) of 143 firms each based on their ranked 2014 total assets (first panel year), to avoid endogeneity from classifying by size in outcome years. Then the baseline joint lag specification, including firm FE, year FE, clustered SE at firm level, is applied individually for each size segment to ensure size group membership is determined prior to the outcome period and remains fixed throughout the panel.

Note that the total sample has decreased with 79 firms, due to missing panel data in 2014 (their earliest observations appear between 2015 and 2018). A closer look at the data of the missing firms reveals that they would most likely have fallen into the small cap segment (median assets of excluded firms at first observation = €4.6bn vs full sample 2014 median = €15.1bn), consequently changing the thresholds, had they been included in the sample. Therefore the findings, especially for the small cap segment, should be interpreted with some caution.

Table 10. Firm Size Analysis.

Firms classified into tertiles by total assets in 2014. Baseline joint lag specification. Size thresholds: Small ≤ €6.4bn; Mid €6.4bn–€38.0bn; Large > €38.0bn total assets (2014).

Size group	2014 TA	Firms	Obs	β (t-1)	β (t-2)	p (t-1)	p (t-2)	Wald p
Full sample (baseline)		508	4,755	+0.0100	-0.0364***	0.377	0.005	0.018**
Small cap	≤ €6.4 bn	143	1,424	+0.0262	-0.0376	0.210	0.114	0.164
Mid cap	€6.4bn – €38.0 bn	143	1,423	-0.0014	**-0.0519****	0.942	0.016	0.038**
Large cap	> €38.0 bn	143	1,423	+0.0102	**-0.0606****	0.632	0.013	0.033**
Total		429	4270					

*Note: *** p < 0.01, ** p < 0.05, * p < 0.10. Firms with missing 2014 assets data are excluded (N reduces from 508 to 429; 79 firms excluded). Green = significant negative result at 5%, consistent with baseline. Yellow = negative direction but not significant.*

All size groups show negative directions at t-2, however only mid and large cap show significant results at the 5% threshold. The direction is consistent with the baseline full sample findings, which rules out the possibility that the aggregate result is driven by a single size segment. The joint Wald tests also show significance at the 5 % threshold for mid and large cap.

The large cap segment shows the strongest association among the size segments at t-2 (β = -0.0606**, p = 0.013), which indicates that transition costs are higher for larger firms, a reasoning that is intuitively sound given that large firms are expected to face more complex ESG transitions, larger absolute compliance investments and more capital-intensive restructuring requirements.

The small cap negative, but insignificant, result is likely due to a statistical power issue, as a smaller sample of 143 firms (1424 observations) makes it more difficult to achieve statistical significance even when a true effect exists, since standard errors increase as sample size decreases.

Overall, the size analysis suggests the negative ESG_{t-2}–ROA association is not confined to a specific size segment, as all three groups show a negative direction at t-2, with the effect most pronounced and most precisely estimated among mid and large cap firms.

4.5 Limitations

Several limitations of this study should be considered when interpreting the results.

First, the empirical design does not allow causal identification. While firm and year fixed effects control for time-invariant unobserved heterogeneity and the lagged ESG specification introduces temporal separation between the independent and dependent variables, time-varying confounders such as management changes or strategic pivots cannot be ruled out. The results should therefore be interpreted as associations rather than causal effects.

Second, the sample is subject to survivorship bias. By anchoring the sample to current STOXX Europe 600 constituents as of February 2026 and pulling historical data, firms that were delisted, acquired or dropped from the index during the study period are excluded. If underperforming firms also tended to have weaker ESG scores, this may result in a slight overstatement of the negative ESG-ROA association.

Third, results are specific to the LSEG ESG Score as the measure of sustainability performance. ESG ratings are known to diverge substantially across providers – Berg et al. (2022) document an average pairwise correlation of only 0.54 between major ESG rating agencies. Findings may therefore not generalize to studies using

Bloomberg, MSCI or Sustainalytics scores, and cross-study comparisons should be made with caution.

Fourth, financial variables are reported in local currencies. While firm fixed effects absorb the time-invariant component of currency-induced differences in firm size, the Size coefficient should be interpreted with caution in cross-country comparisons, as exchange rate variation introduces a degree of non-comparability across the 21 countries in the sample.

Fifth, four GICS sectors – Communication Services, Energy, Information Technology and Real Estate – could not be included in the within-industry analysis due to their subsamples falling below the 30-firm minimum threshold required for clustered standard errors to be reliable. The ESG-ROA relationship in these sectors therefore remains unexamined within this study.

Sixth, the LSEG ESG Score is constructed as an industry-relative metric, scoring firms relative to their industry peers rather than on an absolute scale. The results therefore capture the association between within-industry ESG variation and ROA and cannot be interpreted as reflecting the effect of absolute ESG performance levels across industries.

5. Discussion and Conclusion

This chapter interprets the empirical findings in the context of prior literature and theory, states the study's contribution, and identifies directions for future research.

5.1 Summary of Findings

This study investigates the association between sustainability performance and financial performance among European firms over the period 2014–2023, using a panel of 508 firms and 4,833 firm-year observations from the STOXX Europe 600 index. The baseline panel regression with firm and year fixed effects and clustered standard errors finds no statistically significant association between lagged ESG performance and pretax ROA at the one-year horizon ($\beta = -0.0122$, $p = 0.314$). However, when both one-year and two-year ESG lags are included simultaneously in a joint specification, a statistically significant negative association emerges at the two-year horizon ($\beta = -0.0364$, $p = 0.005$), with the joint Wald test confirming joint significance of the two lags ($\chi^2(2) = 8.03$, $p = 0.018$).

The within-industry analysis reveals that this negative aggregate result is concentrated in the Industrials sector, the largest sector in the sample, which alone accounts for approximately 40% of the weighted sector contribution to the aggregate effect. The negative direction is also consistent across all three size groups, with Mid and Large cap showing significance at 5% and Large cap showing the strongest effect ($\beta = -0.061$).

The t-2 negative finding holds across all robustness checks retaining firm fixed effects, with coefficients ranging from -0.030 to -0.047, confirming the baseline result is not sensitive to measurement or model choices. Thus, the research null hypothesis, that the association between sustainability performance and financial performance is non-positive, cannot be rejected.

5.2 Discussion

The negative association between ESG performance and ROA at a two-year horizon, while initially counterintuitive in the context of stakeholder theory, is consistent with a transition cost interpretation. This pattern follows the well-established cost-return logic of business transitions, where the initial phase is characterized by high costs and low returns, while the subsequent phase delivers lower costs and higher returns as transition-related capabilities mature.

The pillar regressions reveal that Governance is the dominant signal in the aggregate cross-sector analysis. This is consistent with the transition argument: EU sustainability regulation (CSRD, NFRD, and the EU Taxonomy) primarily imposes governance compliance requirements on firms regardless of whether the underlying regulation is environmental in nature. Board-level oversight, sustainability committee establishment, reporting infrastructure and audit processes are governance costs triggered by environmental and social policy. These are activities primarily associated with increased costs in the short run, contributing to a negative effect on ROA at the two-year horizon.

However, sector-level pillar analysis reveals that this aggregate Governance signal is not uniformly distributed across the sample. Within Industrials – the sector driving approximately 40% of the weighted aggregate t-2 effect – it is specifically the Environmental pillar that is significant ($\beta = -0.058$, $p = 0.011$), while the Governance pillar shows no association ($\beta = +0.003$, $p = 0.810$). The Social pillar is marginally significant ($\beta = -0.032$, $p = 0.066$), suggesting workforce and supply chain-related investments may contribute alongside environmental compliance, though the evidence is not conclusive at conventional significance levels. The same Environmental pattern is found in Utilities ($\beta = -0.040$, $p = 0.022$), which is the only other sector where a pillar-level effect reaches significance. Both sectors face particularly capital-intensive environmental transitions: decarbonization of production processes, major energy efficiency upgrades and restructuring of complex supply chains. The full-sample Governance signal therefore reflects a diffuse compliance cost effect spread across sectors, which is likely concentrated in large cap Financials that faces substantial board restructuring and reporting infrastructure costs, rather than any single industry driver. Together, these findings suggest that the mechanics behind the negative ESG-ROA association differ meaningfully across sectors: operationally intensive sectors bear environmental transition costs, while the broader cross-sector signal reflects governance compliance costs imposed by regulatory change.

Importantly, the Industrials effect is not explained by firm size. Fifty percent of Industrials firms fall in the small cap tertile and the sector median total assets (€6.4bn) sit at the Small/Mid threshold, ruling out organizational complexity or large firm governance restructuring as the primary driver. The capital-intensive environmental transition costs are sector-specific rather than size-driven.

The firm size analysis adds a further dimension. Large cap firms show the strongest negative t-2 association across all size groups ($\beta = -0.061$, $p = 0.013$), consistent with larger firms facing more substantial absolute compliance investments and more complex transition requirements. This connects directly to the regulatory argument: CSRD targets large firms first in its phased implementation timeline, and it is precisely these firms that show the most pronounced short-run ROA depression, which is consistent with the view that regulatory standardization should be most consequential for the firms currently bearing the largest transition burden.

Within this framework, the two-year window observed in this study likely captures only the cost side of the transition. The benefits typically associated with strong ESG performance, e.g. lower cost of capital, stronger stakeholder relationships, reduced regulatory risk and operational efficiencies, likely materialize beyond the t-2 horizon and remain unobserved within the empirical scope of this study. This interpretation is consistent with Rostamicheri et al. (2026), who find a negative ESG-ROA association in non-Nordic European firms and attribute it to transitional adjustment costs in a period of evolving EU sustainability regulation.

A natural implication is that the duration of the transition period is itself a key determinant of when ESG investments translate into financial returns. Shorter transitions allow firms to recover invested costs sooner and reach the benefit phase of the ESG-ROA relationship earlier. The EU Corporate Sustainability Reporting Directive (CSRD) and the associated European Sustainability Reporting Standards (ESRS) represent a significant standardization effort that, over time, should reduce the fragmented compliance burden firms have historically faced under voluntary frameworks such as GRI, SASB and TCFD. Reduced compliance complexity translates directly into lower transition costs and shorter adjustment periods.

However, the recent EU Omnibus proposal to scale back CSRD scope and delay implementation for many firms introduces regulatory uncertainty. Firms that have already invested in compliance may find their investments prematurely deployed, while firms that have delayed may now face longer adjustment horizons. In either case, regulatory uncertainty extends the transition period and delays the point at which ESG investments translate into improved financial performance. From this perspective, regulatory clarity and standardization are not merely governance instruments but financial performance enablers.

It is important to acknowledge that the transition cost hypothesis remains a hypothesis. The empirical design of this study cannot directly observe what happens to ROA at t-3, t-4 or beyond, and the inferred pattern of initial costs followed by delayed returns cannot be confirmed within the available data window. The most that can be claimed is that the observed pattern is consistent with this interpretation and that alternative explanations – such as ESG scores capturing risk exposure rather than firm quality, or measurement noise in the ESG metric itself – cannot be ruled out.

5.3 Contribution

This study contributes to the European empirical literature on the ESG-financial performance relationship in four specific ways. First, it is the first European study to combine a lagged ESG specification, two-way firm and year fixed effects, firm-clustered standard errors, and within-industry sector analysis simultaneously. Each of these elements is present individually in prior studies, but their combination provides a methodologically more conservative framework for examining the SP-FP association than has been applied previously in the European context.

Second, the introduction of a joint lag specification, including both ESG (t-1) and ESG (t-2) simultaneously rather than substituting one for the other, adds a temporal dimension to the negative ESG-ROA pattern documented by Candio (2024), Bifulco et al. (2023) and Rostamicheri et al. (2026). The finding that the negative association is concentrated at t-2 rather than t-1 is consistent with the transition cost interpretation and aligns with evidence from Kazyte et al. (2026), who find strongest ESG effects at a two-year lag in a credit risk context.

Third, the within-industry analysis identifies Industrials as the primary sector driving the aggregate negative result. This sector-level granularity is absent from the regional analysis in Rostamicheri et al. (2026) and the aggregate analyses in Candio (2024) and Bifulco et al. (2023) and provides a more specific basis for future research on sectoral heterogeneity in the ESG-FP relationship.

Fourth, sector-level pillar analysis reveals that the Environmental pillar drives the negative t-2 effect in operationally intensive sectors – Industrials and Utilities – while the aggregate Governance signal reflects compliance costs across sectors rather than concentration in any single industry. This granularity, connecting sector-specific ESG transition costs to specific pillar scores while ruling out firm size as an alternative explanation, provides a more precise basis for understanding the mechanics of the negative ESG-ROA association than prior European studies have offered.

5.4 Conclusion and Future Research

The findings of this study indicate that the negative association between ESG performance and financial performance documented in recent European evidence persists under a methodologically conservative specification with lagged ESG, firm fixed effects, and clustered standard errors. The effect is statistically significant only at the two-year horizon and is concentrated in the Industrials sector. The pattern is consistent with a transition cost interpretation, although the available data window does not allow direct observation of the longer-term benefits that stakeholder theory predicts.

Several directions for future research follow naturally from these findings. Most importantly, extending the time horizon to capture ESG effects at t-3, t-4 and beyond would directly test whether the negative short-run association reverses into a positive long-run association as transition costs are absorbed and benefits materialize. Such an extension requires longer panel data than was available within the scope of this study. Extending the time horizon introduces its own challenges, however. Reliable ESG data is sparse before the early 2010s, as systematic corporate ESG reporting only became widespread in the past decade. Additionally, extending the panel further back in time would span periods without a common regulatory framework (pre-CSRD, pre-NFRD, and pre-EU Taxonomy) making the transition period being studied less coherent across the panel. Future research designs would need to balance the value of a longer time horizon against these data and regulatory consistency constraints.

Second, sector-level analysis of the four GICS sectors excluded from the within-industry regressions (Communication Services, Energy, Information Technology and Real Estate) would require larger samples than the STOXX Europe 600 alone can provide. Combining multiple European indices or extending to a broader sample frame would enable a more complete picture of sectoral heterogeneity.

Third, the firm size analysis in this study finds a consistent negative direction across all three size groups, with significance concentrated in mid and large cap firms. Future research with larger samples — covering a broader range of European small cap firms, including those absent from the STOXX Europe 600 — would enable more precise estimation of the Small cap ESG-ROA relationship, which currently appears directionally consistent but statistically underpowered.

Fourth, the regulatory dimension of the ESG-ROA relationship represents an important research opportunity by itself. Empirical analysis of how CSRD/ESRS implementation timelines, the Omnibus revisions, and individual firms' regulatory exposure affect the magnitude and timing of the ESG-FP association would directly test the transition cost hypothesis from a policy perspective. The post-implementation period of CSRD will provide a natural empirical setting for such research in the coming years.

Fifth, comparison across ESG rating providers (Bloomberg, MSCI, Sustainalytics, alongside LSEG) would test the sensitivity of these findings to the choice of ESG measure. Given the documented divergence between major ESG ratings (Berg et al. 2022), such cross-provider analysis would help establish whether the negative ESG-ROA pattern is a general empirical regularity or a feature of specific scoring methodologies.

Finally, the sector-level pillar findings open a more targeted research direction: testing whether the Environmental pillar effect in Industrials and Utilities persists at longer time horizons, and whether a positive return materializes once environmental transition costs are absorbed. Similarly, understanding why the Governance signal is diffuse across sectors rather than concentrated would require firm-level data on specific governance activities, such as board restructuring, sustainability committee formation, reporting system investments, that are not captured in the aggregate pillar scores used here.

References

Angrist, J.D. & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricistís Companion*. Princeton University Press.

Bellemare, M.F. (2020). How to write applied papers in economics. Unpublished manuscript, Department of Applied Economics, University of Minnesota.

Bellemare, M.F. (2022). *Doing Economics: What You Should Have Learned in Grad School—But Didn't*. MIT Press.

Berg, F., Kölbel, J.F. & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, 26 (6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Bifulco, G.M., Savio, R., Paolone, F. & Tiscini, R. (2023). The CSR committee as moderator for the ESG score and market value. *Corporate Social Responsibility and Environmental Management*, 30 (6), 3231–3241. https://doi.org/10.1002/csr.2549

Bryman, A., Bell, E. & Harley, B. (2019). *Business research methods*. Fifth edition. Oxford University Press.

Campello, M., Graham, J.R. & Harvey, C.R. (2010). The real effects of financial constraints: Evidence from a financial crisis. *Journal of Financial Economics*, 97 (3), 470–487. https://doi.org/10.1016/j.jfineco.2010.02.009

Candio, P. (2024). The effect of ESG and CSR attitude on financial performance in Europe: A quantitative re-examination. *Journal of Environmental Management*, 354, 120390. https://doi.org/10.1016/j.jenvman.2024.120390

Christensen, D.M., Serafeim, G. & Sikochi, A. (2022). Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. *The Accounting Review*, 97 (1), 147–175. https://doi.org/10.2308/TAR-2019-0506

De Villiers, C.J., Jia, J. & Li, Z. (2022). Corporate social responsibility : a review of empirical research using Thomson Reuters Asset4 data. http://hdl.handle.net/2263/90210 [2026-02-20]

Dhand, A. (2025). Free Float Market Capitalization. *Scripbox*. https://scripbox.com/pf/free-float-market-capitalization/ [2026-02-19]

Eccles, R.G., Ioannou, I. & Serafeim, G. (2014). The Impact of Corporate Sustainability on Organizational Processes and Performance. SSRN Scholarly Paper, Social Science Research Network. https://doi.org/10.2139/ssrn.1964011

Environment & Society Portal (2016). *UN World Commission on Environment and Development, ed., Report of the World Commission on Environment and Development: Our Common Future*. *Environment & Society Portal*. https://www.environmentandsociety.org/mml/un-world-commission-environment-and-development-ed-report-world-commission-environment-and [2026-03-09]

European Commission (2018). *Action Plan: Financing Sustainable Growth*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0097 [2026-02-16]

European Commission (2025). *Living guidelines on the RESPONSIBLE USE OF GENERATIVE AI IN RESEARCH - ERA Forum Stakeholders' document, 2nd edition*. European Commission.

European Commission (n.d.). *Corporate sustainability reporting - Finance - European Commission*. https://finance.ec.europa.eu/financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en [2026-04-24]

Flammer, C. (2015). Does Corporate Social Responsibility Lead to Superior Financial Performance? A Regression Discontinuity Approach.

Management Science, 61 (11), 2549–2568. https://doi.org/10.1287/mnsc.2014.2038
Friede, G., Busch, T. & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5 (4), 210–233. https://doi.org/10.1080/20430795.2015.1118917
Gillan, S.L., Koch, A. & Starks, L.T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, 66, 101889. https://doi.org/10.1016/j.jcorpfin.2021.101889
Hartzmark, S.M. & Sussman, A.B. (2019). Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. *The Journal of Finance*, 74 (6), 2789–2837. https://doi.org/10.1111/jofi.12841
Hawn, O., Chatterji, A.K. & Mitchell, W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). *Strategic Management Journal*, 39 (4), 949–976. https://doi.org/10.1002/smj.2752
IFRS (n.d.). *SASB Standards*. https://www.ifrs.org/issued-standards/sasb-standards/#overview [2026-04-22]
IMF (n.d.). *World Economic Outlook - All Issues*. *IMF*. https://www.imf.org/en/publications/weo [2026-02-16]
Integrity Research Associates (2009). Thomson Reuters Acquires ASSET4. *Integrity Research Associates*. https://www.integrity-research.com/thomson-reuters-acquires-asset4/ [2026-02-20]
Kazyte, A., Krusna, R. & Sapkauskiene, A. (2026). From ESG disclosure to credit risk: evidence from European firms. *Journal of Financial Reporting and Accounting*, 1–23. https://doi.org/10.1108/JFRA-05-2025-0443
Khan, M., Serafeim, G. & Yoon, A. (2016). Corporate Sustainability: First Evidence on Materiality. SSRN Scholarly Paper, Social Science Research Network. https://doi.org/10.2139/ssrn.2575912
King, A.A. (2025). Do sustainable companies have better financial performance? Revisiting a seminal study. *Journal of Management Scientific Reports*, 3 (2), 88–118. https://doi.org/10.1177/27550311251332904
Lins, K.V., Servaes, H. & Tamayo, A. (2017). Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. *The Journal of Finance*, 72 (4), 1785–1824. https://doi.org/10.1111/jofi.12505
LSEG (2024). Environmental, Social and Governance scores from LSEG. LSEG.
LSEG (n.d.). *LSEG Workspace for students*. https://www.lseg.com/en/data-analytics/products/workspace/workspace-for-students [2026-02-20]
Mahmood, A., Mehmood, A., Terzani, S., De Luca, F. & Djajadikerta, H.G. (2025). The effect of ESG disclosure on firm value in the European context. *Management Decision*,. https://doi.org/10.1108/MD-10-2024-2480
Margolis, J.D., Elfenbein, H.A. & Walsh, J.P. (2009). Does it Pay to Be Good...And Does it Matter? A Meta-Analysis of the Relationship between Corporate Social and Financial Performance. SSRN Scholarly Paper, Social Science Research Network. https://doi.org/10.2139/ssrn.1866371
Rahi, A.F., Johansson, J., Blomkvist, M. & Hartwig, F. (2024). Corporate sustainability and financial performance: A hybrid literature review. *Corporate Social Responsibility and Environmental Management*, 31 (2), 801–815. https://doi.org/10.1002/csr.2600
Reuters (2026). LSEG's two decades of major deals, partnerships and strategic pivots. *Reuters*. https://www.reuters.com/sustainability/sustainable-

finance-reporting/lsegs-two-decades-major-deals-partnerships-strategic-pivots-2026-02-11/ [2026-02-20]
Rostamicheri, P., Popescu, V., Birau, R. & Bărbăcioru, I.C. (2026). ESG Score and Firm Performance: A Comparative Analysis of Nordic and European Companies. *Sustainability*, 18 (3), 1707. https://doi.org/10.3390/su18031707
STOXX (n.d.). *STOXX*. *STOXX*. https://stoxx.com/ [2026-02-19]
UN SDGs (n.d.). *THE 17 GOALS | Sustainable Development*. https://sdgs.un.org/goals [2026-03-09]
WCED (ed.) (1987). *Report of the World Commission on Environment and Development: Our Common Future*. UN. https://digitallibrary.un.org/record/139811 [2026-03-09]

Popular Science Summary

Every year, companies around the world invest billions of euros in sustainability – reducing carbon emissions, improving working conditions, strengthening governance structures. At the same time, investors and executives are asking an increasingly urgent question: does it pay off?

This study examines whether European companies that perform better on sustainability metrics, measured through ESG scores, which rate companies on Environmental, Social and Governance criteria, also tend to be more profitable. Using data on 508 large European companies over a ten-year period from 2014 to 2023, the study tests whether a company's ESG score in one year is associated with its financial returns (measured as pretax return on assets) in the following years.

The short answer is: not in the way stakeholder theory would predict. Companies that improve their ESG score do not show higher profitability the following year. In fact, the study finds a statistically significant negative association two years after an ESG improvement — meaning that companies that raised their sustainability performance tended to have slightly lower returns on assets two years later.

This finding is consistent with what researchers call the transition cost hypothesis. Improving sustainability performance is not free — it requires upfront investment in compliance systems, reporting infrastructure, supply chain restructuring and employee training. These costs hit the income statement immediately. The potential benefits, such as lower borrowing costs, stronger relationships with customers and suppliers, and reduced regulatory risk, take longer to materialize. The two-year window of this study likely captures the cost side of this transition before the returns have had time to appear.

A closer look at which sectors and types of sustainability investment drive this pattern reveals important nuances. In the Industrials sector, which includes manufacturing, engineering and transportation companies, the negative effect is driven specifically by environmental investments: decarbonization, energy efficiency and supply chain emissions reductions. In other sectors, governance-related investments appear to play a larger role, reflecting the compliance costs imposed by EU sustainability regulation such as the Corporate Sustainability Reporting Directive (CSRD). Larger companies show the strongest negative short-run effect, consistent with the fact that they face the most complex and costly transitions.

The policy implications are direct. If sustainability investments temporarily depress profitability before benefits materialize, then the length of the transition period matters — and the regulatory environment shapes that length. Clear, stable and standardized sustainability reporting requirements, such as CSRD and the European Sustainability Reporting Standards (ESRS), should over time reduce compliance complexity and shorten the transition period, bringing financial benefits forward. By contrast, the recent EU Omnibus proposal to scale back CSRD requirements introduces regulatory uncertainty, which risks extending transitions and delaying the financial payoff.

The results of this study do not suggest that sustainability investment destroys financial value. They suggest that it takes time – and that the process can be sped up by clear, consistent and well-implemented sustainability regulation.

Publishing and archiving

Approved students' theses at SLU can be published online. As a student you own the copyright to your work and in such cases, you need to approve the publication. In connection with your approval of publication, SLU will process your personal data (name) to make the work searchable on the internet. You can revoke your consent at any time by contacting the library.

Even if you choose not to publish the work or if you revoke your approval, the thesis will be archived digitally according to archive legislation.

You will find links to SLU's publication agreement and SLU's processing of personal data and your rights on this page:

- https://libanswers.slu.se/en/faq/228318

☒ YES, I, Fredrik Lichtenstein, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this