UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M. FANNO"

CORSO DI LAUREA MAGISTRALE IN
BUSINESS ADMINISTRATION- MANAGEMENT

TESI DI LAUREA

THE RELATIONSHIP BETWEEN PROFITABILITY INDEX AND ESG SCORES: A MACHINE LEARNING ANALYSIS OF THE STOXX EUROPE 600.

RELATORE:

CH.MO/A PROF. RETTORE ENRICO

LAUREANDO/A: LEZUO VERONICA

MATRICOLA N. 2034689

ANNO ACCADEMICO 2024 – 2025

Dichiaro di aver preso visione del “Regolamento antiplagio” approvato dal Consiglio del Dipartimento di Scienze Economiche e Aziendali e, consapevole delle conseguenze derivanti da dichiarazioni mendaci, dichiaro che il presente lavoro non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Dichiaro inoltre che tutte le fonti utilizzate per la realizzazione del presente lavoro, inclusi i materiali digitali, sono state correttamente citate nel corpo del testo e nella sezione ‘Riferimenti bibliografici’.

I hereby declare that I have read and understood the “Anti-plagiarism rules and regulations” approved by the Council of the Department of Economics and Management and I am aware of the consequences of making false statements. I declare that this piece of work has not been previously submitted – either fully or partially – for fulfilling the requirements of an academic degree, whether in Italy or abroad. Furthermore, I declare that the references used for this work – including the digital materials – have been appropriately cited and acknowledged in the text and in the section ‘References’.

Firma (signature) Veronica [illegible]

ABSTRACT

This study investogatea the relationship between environmental, social, and governance (ESG) performance and corporate profitability, in the context of increasing stakeholder attention to corporate sustainability practices. Drawing on the theoretical frameworks of agency theory and signalling theory, this research empirically evaluates the relationship between both overall ESG scores and their individual components (environmental, social, and governance) and firm operational performance, proxied by two accounting-based measures: Return on Assets (ROA) and EBIT Margin.

The empirical analysis employs a dual-method approach. First, fixed effects Panel OLS regression models are developed in Python using Jupyter Notebook and applied to a panel dataset of 402 European firms from the STOXX Europe 600 index, over a ten-year period from 2013 to 2022. Second, to address potential endogeneity and nonlinear relationships, the study incorporates Double Machine Learning (DML), that enhances causal inference in high-dimensional settings.

The results consistently indicate a positive and statistically significant relationship between overall ESG performance and profitability. However, when examining the ESG pillars separately, the effects vary in both magnitude and significance, highlighting the non-uniform contribution of each dimension. The robustness of the findings under both traditional econometric and machine learning frameworks underscores the importance of ESG integration in corporate strategy. These insights suggest that strategic investment in specific ESG areas can generate tangible improvements in operational financial outcomes.

INDEX

FIGURE INDEX

INDRODUCTION

In recent years, **Environmental, Social, and Governance (ESG)** considerations have transitioned from being marginal ethical concerns to becoming central components of corporate strategy and investment analysis. This evolution reflects a broader global awareness of environmental challenges, climate change, social inequality, and governance failures, issues that increasingly shape stakeholder expectations and regulatory interventions [Friede, et all. (2015)]. ESG metrics are now widely used as benchmarks for assessing non-financial risks and long-term value creation, prompting firms to integrate sustainability considerations into their core decision-making processes.

At the policy level, the European Union has played a leading role in institutionalizing ESG through a robust regulatory framework, including the **EU Taxonomy Regulation**, the **Corporate Sustainability Reporting Directive (CSRD)**, and the forthcoming **ESG Ratings Regulation**. These initiatives aim to enhance transparency, comparability, and accountability in sustainability reporting, thereby supporting more informed investment decisions and aligning capital flows with sustainability objectives [European Commission, (2021)].

Despite the proliferation of ESG-related disclosures and the increasing availability of ESG scores from data providers such as Refinitiv, **the academic literature remains largely focused on the relationship between ESG performance and financial market outcomes**, such as stock returns, cost of capital, and credit risk [Krüger, (2015)]. This emphasis reflects investor demand but leaves a gap regarding the internal operational implications of ESG adoption.

Empirical evidence on how ESG practices affect firms' internal operations, such as production efficiency, supply chain sustainability, and operating profitability, remains limited and ongoing. While some scholars argue that ESG investments enhance stakeholder engagement and generate reputational advantages that improve performance [Velte (2017)], others warn of potential overinvestment and misallocation of resources that may erode shareholder value [Brammer et al. (2006)].

This study aims to address this gap by empirically investigating the relationship between ESG performance and operational profitability among firms listed in the **STOXX Europe 600** index. Specifically, it examines whether ESG scores, both aggregated and disaggregated into their environmental, social, and governance components, are associated with short-term operating efficiency (measured by **EBIT margin**) and medium-to-long-term performance (measured by **Return on Assets, ROA**).

To ensure methodological robustness, the analysis employs a **dual empirical approach** combining traditional **panel data regression models** with advanced **machine learning techniques**, notably **Double Machine Learning (DML)**. This hybrid framework is designed to mitigate endogeneity concerns and address potential omitted variable bias, thus enabling more reliable causal inference [Chernozhukov et al., (2018)].

The analysis was conducted using **Python programming language within the Jupyter Notebook environment**, leveraging libraries such as **pandas**, **scikit-learn**, and **statsmodels**, which facilitated data cleaning, modeling, and interpretability.

The dataset is derived from **Refinitiv Eikon**, a leading ESG data provider widely used in academic and professional research, thereby ensuring both validity and comparability of the ESG metrics employed. The inclusion of firms from diverse sectors and European jurisdictions allows for the examination of ESG impacts across different institutional and regulatory contexts, enhancing the generalizability of the findings.

At the core of this inquiry lies a fundamental question: **Do higher ESG scores translate into superior operational profitability, and if so, through which dimensions?** By unpacking the aggregate ESG score into its constituent pillars, the study aims to reveal the heterogeneous effects that environmental, social, and governance factors exert on economic performance.

The findings aim to enrich the existing literature by providing a different perspective on the effects of ESG performance beyond financial markets, offering valuable insights for corporate managers, investors, and policymakers involved in advancing sustainable business practices.

1. THE ESG ISSUE

1.1 ESG Preliminary understanding

The term ESG is frequently employed in the field of financial economics to refer to responsible investment practices. This encompasses activities that, while aiming to achieve standard financial management goals, also take into account environmental, social, and governance factors.

According to some scholars [Gillian et al. (2021)], the term refers to the practice of companies incorporating ethical and social considerations into their business models. Others [Eccles et al. (2014)], argue that it can also be viewed as a framework for assessing a company not solely based on its financial performance, but rather through other factors such as those mentioned above.

1.1.1. ESG terminology

The term currently utilized in the economic domain and beyond has its origins in the early 2000s.

A key moment in the definition of ESG was the publication of the "Who Cares Wins" report in 2005 by the UN Global Compact in collaboration with several financial institutions.

This report highlighted the importance of integrating environmental, social, and governance considerations into investment decisions.

Additionally, the term gained prominence through academic studies such as those by Eccles et al. (2014), which explored the relationship between ESG performance and financial performance, demonstrating that companies with sustainable practices tend to achieve better long-term results.

The acronym is composed of three words which describe the fundamental pillars of social sensitivity:

- E: Environment
- S: Social
- G: Governance

The *environmental* pillar emphasizes the impact that a company exerts on the environment

[Hart, S. L. (1995)]. Key practices within this domain include:

- **Natural Resource Management:** This involves the sustainable use of resources, alongside efforts to minimize waste and pollution.
- **Climate Change:** Companies are expected to implement policies aimed at reducing greenhouse gas emissions and to develop adaptation strategies.
- **Biodiversity:** Initiatives designed to protect ecosystems and wildlife are also crucial components of this pillar.

The *social* pillar examines how a company manages its relationships with employees, suppliers, customers, and the communities in which it operates [Freeman, R. E. (1984)].
This includes:

- **Worker Rights:** Attention is given to labor conditions, human rights, and the promotion of diversity.
- **Community Impact:** Companies are assessed on their contributions to local communities and their social responsibility initiatives.
- **Health and Safety:** Policies aimed at ensuring the health and safety of employees are also fundamental in this area.

The *governance* pillar focuses on the governance structure and business practices of a company [Hart et al. (1976)]. Relevant aspects include:

- **Board Composition:** This encompasses diversity, independence, and the qualifications of board members.
- **Corporate Ethics:** Companies should pursue policies combat corruption and manage conflicts of interest.
- **Transparency:** Clear and open communication with investors and stakeholders is essential.

1.1.2 ESG and CSR: Key Differences and Overlaps

Corporate actions related to social aspects are sometimes interchangeably labeled as ESG and/or CSR practices [Gillian, Koch &Starks (2021)]. However, it is crucial to emphasize that these two terms do not cover the same concepts. Notably, the acronyms were developed during different historical periods, reflecting an evolution in the approach to sustainability within the economic sphere. In the following section, we will briefly examine the distinctions and similarities between Environmental, Social, and Governance (ESG) criteria and Corporate Social Responsibility (CSR) practices.

The term *ESG* enfolds the various practices by which companies and investors incorporate environmental and social considerations, along with sound governance, into their business models [Eccles et al. (2014)]. Conversely, *CSR* is a broader concept that refers to the initiatives a company engages to create a positive impact on society. CSR is often viewed as a voluntary approach, where companies strive to go beyond mere legal requirements and profit motives to contribute to social and environmental well-being; in other words, it describes the enterprise's commitment to being a better corporate citizen [Carroll et al. (1999)].

The primary distinctions between ESG and CSR lie in their purpose and application. ESG is mainly employed for investor assessment, concentrating on measurable and verifiable criteria. In contrast, CSR emphasizes philanthropic initiatives and social impact, which can vary widely across different organizations.

Additionally, the two frameworks differ in terms of assessability: ESG criteria are typically quantifiable and used to evaluate corporate performance over time. Conversely, CSR encompasses activities that are often difficult to measure, such as charitable donations and volunteer efforts.

Moreover, ESG is more focused on managing risks associated with environmental, social, and governance factors. In contrast, CSR tends to highlight opportunities for value creation through enhanced reputation and community engagement.

Finally, while ESG explicitly incorporates governance as a key component of its framework, CSR addresses governance issues only indirectly, in relation to their influence on environmental and social dimensions [Gillian et al. (2021)].

1.1.3. ESG Perception over time

The consideration of Environmental, Social, and Governance (ESG) factors has gradually become a central focus in the context of financial investments. Over the past decade, the significance of these dimensions has grown substantially, reflecting a shift in the way they are perceived by both corporations and investors [Silano (2016)].

Historically, sustainable finance was often viewed with skepticism, primarily regarded as a series of restrictive measures. However, contemporary attitudes have become more favorable, with ESG principles now seen as integral to proactive and constructive financial strategies.

From a corporate perspective, ESG considerations have become pivotal for fostering sustainable long-term impacts. Organizations are progressively embedding these criteria into their decision-making frameworks and resource allocation processes.

For investors, the integration of ESG factors within portfolio strategies is seen as a pathway to developing best practices, thereby enhancing not only financial outcomes but also contributions to environmental, social, and governance advancements [Silano (2016)].

1.2 European Union Regulation

Recent European Union (EU) regulations on Environmental, Social, and Governance (ESG) criteria have undergone substantial development, aiming to strengthen sustainability across corporate and investment practices. Notably, the EU Taxonomy Regulation, ratified by the EU Council and Parliament in 2022 [EU Regulation 2020/852, (2020)], represents a cornerstone of these efforts.

This regulatory framework seeks to standardize the definition and integration of ESG factors within the economic landscape.

The primary objective of the *EU Taxonomy* is to establish a comprehensive classification system for activities deemed sustainable. According to the "EU Taxonomy Guide" [Sustainalytics, (2020)], this system outlines a set of minimum requirements and securities that activities must satisfy to align with the EU’s environmental objectives. Specifically, these objectives encompass:

1. Mitigation of climate change.
2. Adaptation to climate change.
3. Sustainable use and conservation of water and marine resources.
4. Promotion of a circular economy, including waste prevention and recycling.
5. Prevention and control of pollution.
6. Protection and restoration of ecosystems and biodiversity.

This taxonomy provides a unified framework to guide businesses, investors, and policymakers in fostering environmentally sustainable practices while ensuring transparency and accountability in ESG reporting and implementation. Such measures represent a significant step toward achieving the EU’s broader sustainability goals and climate commitments [Sustainalytics, (2020)].

In February 2024, the European Union introduced the *Regulation on ESG Ratings*, aimed at enhancing the transparency and integrity of ESG ratings within financial markets. This regulation mandates that all ESG rating providers obtain authorization and be supervised by the European Securities and Markets Authority (ESMA). It establishes key standards for transparency, organizational structure, and conflict-of-interest management to bolster investor confidence in ESG-driven investments [Latham & Watkins. (2024)].

Key provisions of this regulation include:

- Mandatory licensing and oversight of ESG rating providers by ESMA.
- Transparency requirements for rating methodologies, allowing for the separate or combined reporting of Environmental, Social, and Governance (E, S, G) criteria, with clear disclosure of the weighting factors applied.
- Organizational safeguards to prevent conflicts of interest, ensuring the credibility of ESG rating activities.
- Provisions allowing non-EU rating providers to have their assessments recognized within the EU, contingent on meeting specific conditions.

In the context of the ongoing standardization of non-financial reporting, it is essential to highlight the *Non-Financial Reporting Directive (NFRD)* [European Union Directive 2013/34/EU, 2014].

This directive outlines the reporting obligations for large public interest entities, specifically those employing more than 500 workers. The NFRD applies to a subset of sectors that are deemed of public interest, including banking, insurance, and other sectors designated by national authorities.

The directive aims to ensure that these companies disclose relevant environmental, social, and governance (ESG) information, thus contributing to greater transparency and accountability in corporate sustainability practices. This initiative marks a significant step towards aligning corporate reporting with broader sustainability goals and ensuring that stakeholders have access to critical non-financial data [European Union Directive 2013/34/EU, 2014].

Furthermore, the *Corporate Sustainability Reporting Directive* (CSRD) came into force in February 2023, following an agreement between the European Union Council and Parliament on the framework for sustainability reporting. This regulation mandates that all large companies, as well as publicly listed small and medium-sized enterprises, provide detailed sustainability reports.

The purpose of this directive is to ensure that investors and other stakeholders can access relevant data to assess the potential risks associated with climate change, environmental impacts, and social factors [European Union Directive 2022/2464, 2022].

The significance of the CSRD lies not only in promoting a culture of transparency regarding the environmental and social impacts of corporate activities, but also in facilitating the digitalization of corporate reporting. This shift is designed to enhance the accessibility and comparability of sustainability information, thereby enabling more informed decision-making by stakeholders [European Union Directive 2022/2464, 2022].

In addition, the *Corporate Sustainability Due Diligence Directive* (CSDDD), finalized in 2024, aims to hold companies accountable for their impact on human rights and environmental standards across their entire value chains, both within and beyond the EU. This regulation is part of the EU's broader strategy to guide companies towards a climate-neutral, sustainable economy. The directive targets primarily larger businesses, with compliance timelines based on company size and sector. By 2026, EU member states must transpose this directive into national law, with full compliance expected by 2029. It includes enforcement mechanisms such as administrative supervision and penalties for non-compliance, as well as provisions for legal redress for victims of human rights and environmental harm [European Union Directive 2022/2464, 2022].

These regulations reflect the EU’s commitment to integrating ESG principles into business and financial practices, encouraging a shift toward a more sustainable and transparent economic landscape.

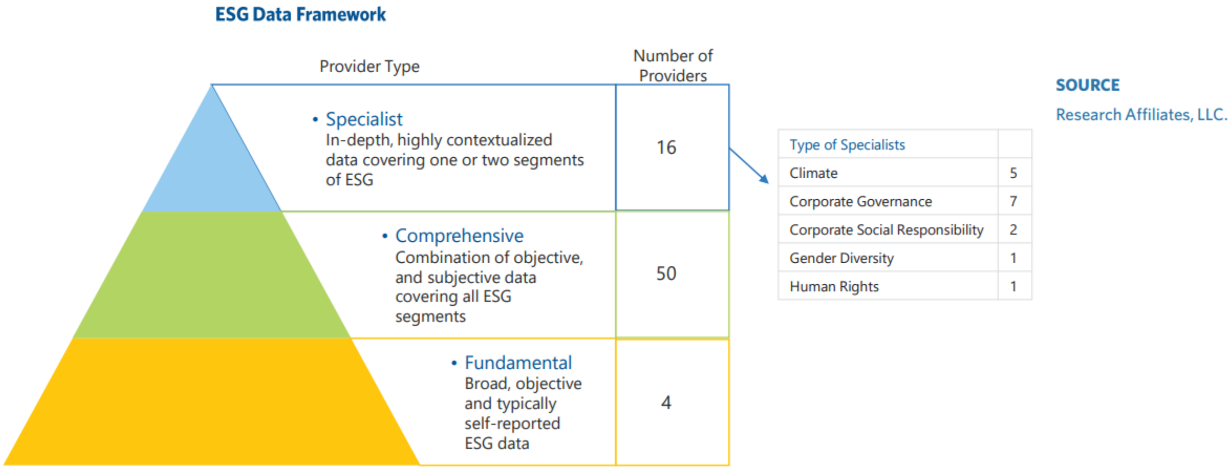
1.3 The ESG Ratings

Having established a definition of the term "ESG" and its foundational components, the discussion proceeds to examine the methodologies employed in measuring ESG performance, as well as the experts tasked with conducting these evaluations.

In the contemporary landscape, all publicly traded companies, alongside numerous private firms, undergo evaluation of their environmental, social, and governance (ESG) practices by various independent rating agencies and reports. These ratings enable stakeholders to track a company's ESG performance over time and compare it with that of its industry peers. With the growing prominence of ESG-focused investing, the reliability and accessibility of ESG assessments have become increasingly significant [Li & Polychronopoulos (2020)].

ESG databases have opened avenues for examining the advantages of non-financial evaluations of companies. Moreover, access to these databases facilitates comparative analyses of corporate opportunities, enabling evaluations of a specific entity relative to its peers, across sectors, or even using cross-country data Nonetheless, it is crucial to acknowledge that ESG assessment data often suffers from limitations, such as incomplete coverage and dependence on self-reported information. Furthermore, inconsistencies in the methodologies employed by different ESG data providers frequently result in significantly divergent outcomes [Li & Polychronopoulos (2020)].

The ESG sector is characterized by a multitude of rating providers offering a diverse array of data. These range from entities that focus on specific ESG dimensions to those that assess companies across several hundred ESG-related parameters. As noted by Li and Polychronopoulos (2020), it is essential to categorize this extensive range of material, enabling investors to focus on the specific information they seek. In the following sections, the conceptual framework proposed in the study by Li and Polychronopoulos (2020) will be resumed to classify the types of available data and their respective providers, facilitating the identification of the most effective sources for the subsequent analysis.



Type of Specialists	
Climate	5
Corporate Governance	7
Corporate Social Responsibility	2
Gender Diversity	1
Human Rights	1

Figure 1 ESG Data Framework [Li & Polychronopoulos, (2020)].

As illustrated in Figure 1, ESG data providers can be categorized into three distinct groups based on their approach and scope of analysis:

- **Fundamental Providers**: These organizations focus on collecting and aggregating publicly available information, typically sourced from corporate websites, non-governmental organizations, rating agencies, and other accessible platforms [Li & Polychronopoulos (2020)]. Their role is primarily to organize this data in a manner that is functional and accessible for end users. Consequently, the information offered by fundamental providers tends to be broad in scope and relatively objective.

  Notable examples in this category include Bloomberg [Blomberg (2022)] and Thomson Reuters (commonly referred to as Refinitiv) [Thomson Reuters (2017)].
- **Comprehensive Providers**: This category encompasses providers that utilize a combination of objective and subjective data to cover the entire spectrum of ESG considerations [Li & Polychronopoulos (2020)].

  These providers integrate publicly available data with proprietary insights generated through methodologies such as interviews and surveys conducted with the companies under review. By employing customized metrics and parameters (including assessments of controversies related to specific corporate activities) they translate the collected data into enterprise's comprehensive ESG evaluation.

  Prominent examples of comprehensive providers include MSCI [MSCI ESG Research LLC (2022)] and Sustainalytics [Sustainalytics (2022)].
- **Specialist Providers**: As the name suggests, specialist providers concentrate on specific ESG dimensions, such as carbon emissions, biodiversity management, human rights, or gender equity. Their specialized focus enables them to deliver highly detailed and expert analyses in their respective areas, offering valuable insights for users aiming to address or improve performance on targeted ESG issues.

  Examples of specialist providers include S&P Global [S&P Global (2023)], Equileap [Equileap (2023)] and Carbon Disclosure Project (CDP) [Carbon Disclosure Project (2022)].

1.3.1 ESG Ratings Disparities

The increasing emphasis on sustainability within the financial sector has catalyzed a proliferation of ESG rating providers. However, as noted by Capizzi et al. (2023), this multiplicity has concurrently given rise to significant divergences in ESG ratings. A single company can often receive markedly different ESG performance scores, depending on the rating provider.

Although rating agencies typically demonstrate a high degree of transparency in their scoring methodologies, these processes involve varying interpretations of underlying data, leading to differences in the relative weights and importance assigned to ESG components.

Berg et al. (2022) attribute these inconsistencies to four primary factors:

- *Scope*: Different agencies prioritize varying dimensions of ESG, such as environmental versus governance considerations.
- *Metrics*: The indicators used, and their respective weightings diverge across providers.
- *Data Sources*: Reliance on public disclosures, proprietary datasets, or external sources introduces variability.
- *Subjectivity*: The qualitative interpretation of data, such as corporate policies or public statements, differs among providers.

The discrepancies in ESG ratings arise not only from variations in data sources and quality but also from the measurement methodologies employed. Some divergences stem from differences in initial data inputs, while others result from subjective interpretations of the same data. Moreover, even when definitions align across providers, the application of differing weights to indicators often yields inconsistent outcomes [Berg, Kölbel, & Rigobon (2022)].

These inconsistencies are observable not only at the aggregate ESG score level but also across the individual Environmental, Social, and Governance (ESG) pillars. Governance scores exhibit the highest variance, attributable to the inherently subjective nature of governance-related metrics and evaluations. Conversely, divergences in the Environmental and Social pillars are somewhat less pronounced, as the associated indicators tend to be more objective, resulting in greater convergence in measurement and scoring [Capizzi et al. (2023)].

1.3.2 Thomson Reuters ESG Data

The Thomson Reuters EIKON database is classified within the category of Fundamental data providers. As previously discussed, it compiles and systematizes data that is publicly available.

This database offers access to an extensive repository of information sourced from over 400 stock exchanges and markets globally, with historical records dating back to 2002 [Silkacz &Wolczek (2018)].

Beyond Environmental, Social, and Governance (ESG) performance metrics, which are derived from over 6,000 companies, the database encompasses more than 70 key performance indicators (KPIs) and financial metrics. The information contained is recognized for its timeliness, accuracy, and reliability. The scope of the Thomson Reuters EIKON database includes comprehensive data on equities, bonds, trusts, investment funds, exchange rates, interest rates, financial derivatives, and commodities. Additionally, it provides macroeconomic data and projections [Thomson Reuters (2017)].

The Thomson Reuters EIKON database categorizes Environmental, Social, and Governance (ESG) indicators into three main classifications:

- **ESG Score**,
- **ESG Controversy Score (ESGC Score)**,
- **ESG Combined Score**.

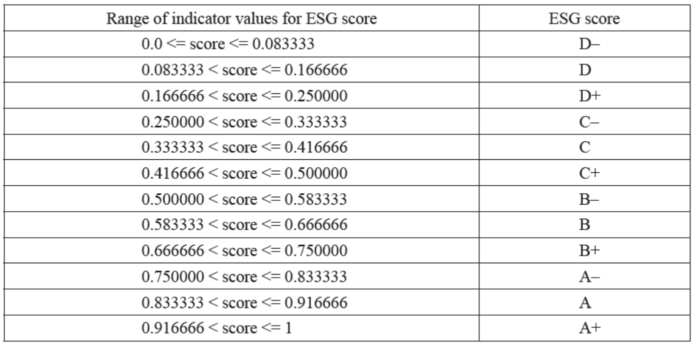
The **ESG Score** evaluates the companies' ESG performance based on publicly available information. The score is determined using the following formula:

$$\text{ESG Score} = \left(\frac{a + 0.5b}{c}\right) * 100$$

Where:

- *a* is the number of companies with worse performance than the company being assessed,
- *b* is the number of companies with equivalent performance to the company being assessed,
- *c* is the total number of companies evaluated in the dataset.

The computed score is then used to assign the company an ESG rating, as detailed in the accompanying table (Figure 2).

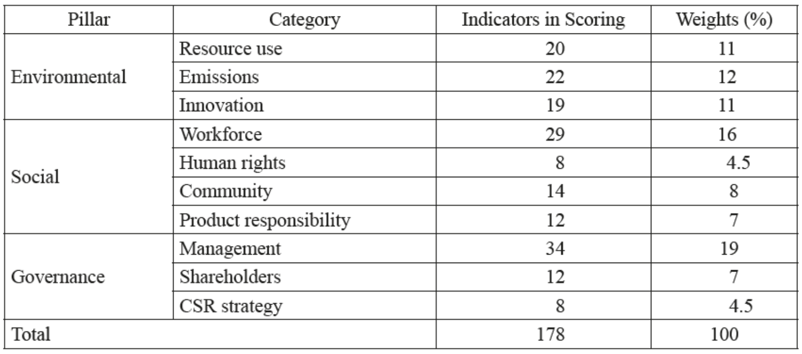


Range of indicator values for ESG score	ESG score
0.0 <= score <= 0.083333	D–
0.083333 < score <= 0.166666	D
0.166666 < score <= 0.250000	D+
0.250000 < score <= 0.333333	C–
0.333333 < score <= 0.416666	C
0.416666 < score <= 0.500000	C+
0.500000 < score <= 0.583333	B–
0.583333 < score <= 0.666666	B
0.666666 < score <= 0.750000	B+
0.750000 < score <= 0.833333	A–
0.833333 < score <= 0.916666	A
0.916666 < score <= 1	A+

Figure 2 Range ESG Score Indicator [Thomson Reuters (2017)].

The scoring process consists of several phases. First, data is collected based on its materiality, availability, and relevance to the sector. This information is then aggregated into three pillars—**Environment**, **Social**, and **Governance**—that collectively form the ESG score.

As demonstrated in Figure 3, these three pillars are further divided into 10 thematic areas, each with a corresponding score. The combined total of these scores amounts to a maximum of 178 points. The scoring process ensures appropriate weighting, recognizing that not all thematic areas contribute equally to the final ESG score.



Pillar	Category	Indicators in Scoring	Weights (%)
Environmental	Resource use	20	11
	Emissions	22	12
	Innovation	19	11
Social	Workforce	29	16
	Human rights	8	4.5
	Community	14	8
	Product responsibility	12	7
Governance	Management	34	19
	Shareholders	12	7
	CSR strategy	8	4.5
Total		178	100

Figure 3 Indicators' weight assigned to different ESG categories [Sikacz &Wolczek (2018)].

The **ESG Controversy Score (ESGC Score)** evaluates a company's exposure to controversies related to environmental, social, and governance issues, as noted by Thomson Reuters (2017). This score provides a holistic assessment of a company's ESG performance by integrating the baseline scores from the ESG pillars with additional insights derived from controversies reported in global media.

The primary objective of the ESGC Score is to adjust and rebalance a company's ESG performance to account for any adverse events or incidents that may negatively impact its overall evaluation.

The **ESG Combined Score** represents an overarching evaluation of a company's ESG performance. It combines the baseline ESG Score, derived from the environmental, social, and governance pillars, with adjustments for any ESG controversies, as described by Thomson Reuters (2017). The score is computed as a weighted average of the scores from the three ESG pillars and the evaluations of 23 specific ESG controversy topics.

Figure 4 presents a diagram that provides a summary of the concepts discussed in the preceding paragraph. The diagram illustrates the process by which ESG data is collected and organized within the Thomson Reuters database.

The data collection methodologies of other providers will not be examined, as this analysis is specifically focused on the EIKON database.

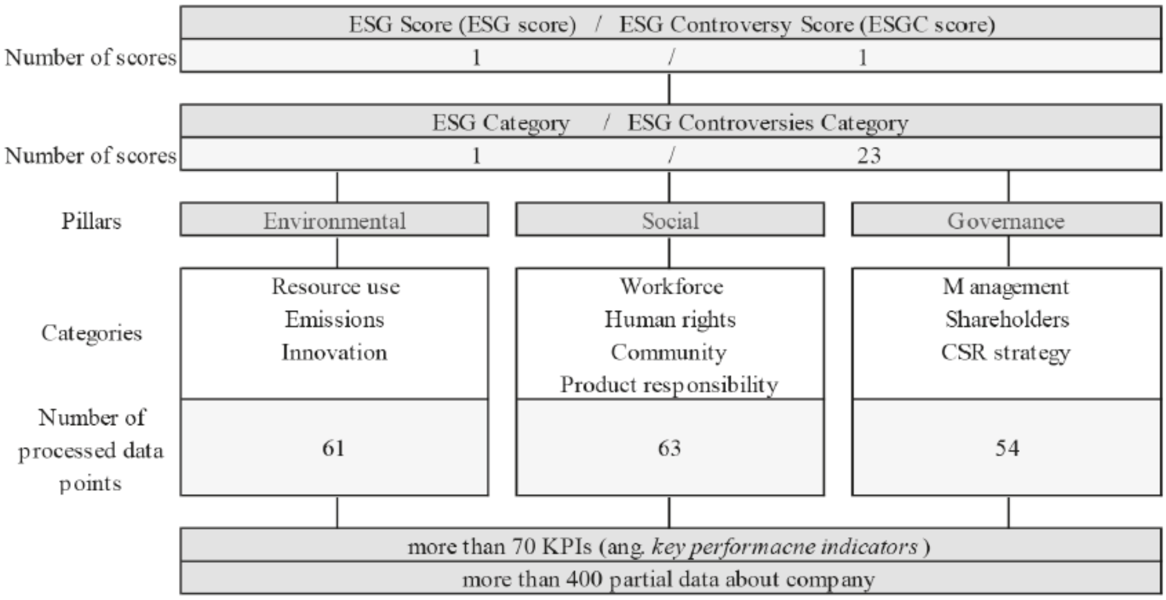


Figure 4 ESG Data Division in EIKON Database [Thomson Reuters (2017)].

2. MACHINE LEARNING

2.1 Introduction to Machine Learning

The concept of machine learning and, more generally, artificial intelligence has become increasingly relevant in contemporary discourse, exerting a profound influence on various aspects of daily life. Although these concepts have only recently entered common parlance, their theoretical foundations were established as early as the 1950s. Since the mid-20th century, mathematicians and researchers began to formulate the fundamental principles of machine learning, foreseeing its potential, long before computers had the processing capabilities necessary to implement complex algorithms effectively.

2.1.1 Definition Origins, and Early Developments

Arthur Samuel (1959), among the pioneers in the field of machine learning, provided an early definition of machine learning, describing it as:

"The field of study that provides computers with the ability to learn without being explicitly programmed".

This characterization underscores the fundamental premise of machine learning: the development of algorithms that enable computational systems to perform tasks autonomously by learning from data.

A more formal and widely accepted definition was later proposed by Tom Mitchell (1997), who articulated machine learning as follows:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance on the tasks in T, as measured by P, improves with experience E."

This definition emphasizes the iterative nature of machine learning, where models improve their predictive or decision-making abilities through the accumulation of experience, thereby refining their performance over time. As computing power has significantly advanced in recent decades, machine learning has evolved from a theoretical construct to a technological paradigm, driving innovations in numerous fields, including healthcare, finance, and artificial intelligence research.

In 1950, the British mathematician and cryptographer Alan Turing, following extensive research and the development of early computational device: Enigma; began to explore the question of whether machines could exhibit intelligence comparable to that of humans.
In his seminal paper, "Computing Machinery and Intelligence", Turing argued that the ability to learn from experience was a fundamental component of intelligence. His work laid the theoretical groundwork for what would later become the field of artificial intelligence (AI) and machine learning [Turing, A. M. (2021)].

Less than a decade later, in 1959, Arthur Samuel, a computer scientist at IBM, formally introduced the term machine learning while working on a checkers-playing program. His software demonstrated the ability to improve its strategy over time by analyzing past games, marking a groundbreaking advancement in the development of self-learning computational systems [Wang, W. (2024)].

During the same period, Frank Rosenblatt, drawing inspiration from the functioning of biological neurons, developed the perception, an early artificial neural network model capable of classifying inputs into distinct categories. Although the perception had inherent limitations, it represented a foundational step toward modern neural network architectures and served as a precursor to contemporary deep learning methodologies [Rosenblatt, F. (2021)].
These pioneering contributions collectively established the theoretical and practical foundations of machine learning, shaping its evolution into a central discipline in artificial intelligence research.

Despite the initial enthusiasm surrounding the development of this new technology, the field soon encountered a series of significant obstacles that slowed its early development.

- **Computational Limitations**: The hardware available at the time was insufficient to support large-scale learning. Computers lacked the necessary processing power and memory capacity to handle complex computations, thereby limiting the effectiveness of early machine learning models.

- **Mathematical Challenges**: The pioneers of artificial intelligence research faced considerable difficulties in formulating models capable of addressing intricate tasks beyond simple rule-based systems. The absence of well-established mathematical frameworks for handling learning processes further delayed progress.

These constraints resulted in a substantial decline in funding for machine learning research, with skepticism directed toward neural networks. Due to their restricted computational efficiency and inability to scale effectively, neural networks were criticized and ultimately deprioritized in favor of alternative approaches [Gogoi, A., & Kao, F. J. (2024)].

During the 1970s and 1980s, neural networks were gradually abandoned, and machine learning undertook a paradigm shift, transitioning from heuristic AI-driven techniques to more rigorous statistical methodologies. Among the approaches that gained notoriety during this period were Bayesian models for probabilistic modeling, decision trees, and logistic regression for supervised learning [Toosi, et al. (2021)].

2.1.2 The Path to Modern Machine Learning

By the 1990s, machine learning experienced rebirth, driven by advancements in hardware, including increased memory capacities and more compact storage solutions, alongside improvements in computational power and new algorithmic developments.
These technological breakthroughs enabled the revival of previously abandoned techniques.

A crucial moment occurred in 1986, when Geoffrey Hinton and Yann LeCun revisited neural networks and introduced significant optimizations. Their work on the backpropagation algorithm made neural networks viable for real-world applications [Zaras et al. (2022)].
In 1992, the introduction of Support Vector Machines (SVMs) and Kernel Methods provided a breakthrough in supervised learning. These models allowed for the mapping of data into higher-dimensional, facilitating the identification of complex relationships becoming linear [Cristianini, N. (2000)].
By 1996, the development of Random Forests, an ensemble method based on decision trees, further advanced the field. Unlike earlier models, Random Forests significantly enhanced predictive accuracy while reducing the risk of overfitting, making them a widely adopted technique in data science [Fawagreh, K. et al. (2014)].

With the advent of the new millennium, machine learning transitioned from a primarily theoretical discipline to a field with tangible real-world applications. More powerful predictive models emerged, including the AdaBoost algorithm, developed by Yoav Freund and Robert

Schapire, which laid the groundwork for subsequent innovations such as XGBoost and Gradient Boosting Machines (GBM) [Beja-Battais (2023)].

The 2010s saw an unprecedented rise in deep learning, fueled by the availability of big data, the proliferation of GPUs, and continuous algorithmic improvements.
During this period: Geoffrey Hinton's team introduced AlexNet, a deep neural network that revolutionized image recognition and demonstrated the superiority of deep learning over traditional methods [Krizhevsky et al. (2012)].
This era also saw the emergence of Generative AI, with landmark innovations such as: Google’s BERT (2018), which enabled contextual language understanding, improving natural language processing (NLP) capabilities.
OpenAI’s GPT models (2019–2023), which pushed the boundaries of AI-driven text generation, enabling computers to produce human-like language and, more recently, generate complex images and other forms of creative content [Dhamani, N. (2024)].

2.2 Machine Learning Architecture

Machine Learning (ML) encompasses a diverse array of algorithms and computational models designed to enable systems to infer patterns from empirical data and address complex tasks without explicit programming. These methodologies differ with respect to the characteristics of the input data and the underlying learning paradigms employed [Mitchell, T. M. (1997)].
The principal categories of machine learning approaches include:

- **Supervised Learning**: This paradigm involves training algorithms on labeled datasets, where each input is paired with a corresponding output. The objective is to infer a mapping function that can predict outputs for unseen inputs. It is predominantly applied in tasks such as classification and regression.

- **Unsupervised Learning**: In contrast, unsupervised learning deals with data that lacks explicit labels. Algorithms in this category aim to uncover hidden structures, patterns, or groupings within the data, making it particularly effective for clustering and association mining.

- **Semi-Supervised Learning**: This technique operates on a hybrid dataset comprising both labeled and unlabeled instances. It is especially beneficial in domains where labeled data is scarce or expensive to acquire, yet a high level of predictive accuracy is desired.

- **Instance-Based Learning**: Rather than constructing an explicit model, instance-based methods retain training examples and make predictions by comparing new inputs with stored instances. The k-Nearest Neighbors (k-NN) algorithm is a well-known exemplar of this approach.

- **Model-Based Learning**: This category involves the construction of an explicit predictive model based on training data. The model captures the underlying data distribution and is used to generalize new, unseen inputs [Subasi, A. (2020)].

2.2.1. Key Phases in the Development of Machine Learning Models

A machine learning framework typically comprises a sequence of interdependent components, encompassing data acquisition and extraction, data preparation, model development, evaluation, and deployment. Each phase plays a pivotal role in ensuring the accuracy, scalability, and overall efficacy of the resulting system [Harrington, (2012); Sarkar, Bali, & Sharma, (2018)].

To operationalize a machine learning algorithm effectively, practitioners generally adhere to a structured pipeline involving the following key stages:

1. **Data Collection**

   The initial phase involves the acquisition and recording of raw data, which constitutes the foundation of the machine learning process. The objective is to obtain high-quality, relevant data that reflects the problem domain and is suitable for subsequent analysis [Harrington, (2012); Sarkar et al., (2018)].

2. **Data Description and Exploration**

   A thorough descriptive analysis of the collected dataset is essential to comprehend its structural properties, distributional characteristics, and potential anomalies. This stage not only informs decisions about cleaning and transformation but also exposes inherent limitations and strengths within the data. Poor understanding at this stage often results in suboptimal models and misleading outcomes [Sarkar et al., (2018)].

3. **Feature Scaling and Extraction**

   At this juncture, data is normalized or standardized to ensure uniformity in scale, and relevant features are engineered or extracted. These transformations enhance model efficiency and performance by facilitating better pattern recognition and reducing computational complexity. The quality of features directly influences the model's ability to learn and generalize [Sarkar et al., (2018)].

4. **Modeling**

   Following preprocessing, the refined dataset is input into suitable algorithms tailored to the nature of the learning task. In supervised learning, models are trained using labeled data to predict outcomes, whereas in unsupervised learning, models aim to detect hidden patterns or structures in unlabeled data. Common algorithmic approaches include decision trees, support vector machines, neural networks, clustering algorithms, and ensemble methods [Sarkar et al., (2018)].

5. **Model Training**

   This stage involves fitting learning algorithms to the data, with the goal of constructing a model that generalizes well to unseen data while meeting defined performance criteria. The model's architecture and hyperparameters are often iteratively refined during this process [Sarkar et al., (2018)].

6. **Model Evaluation and Optimization**

   Once a model is trained, rigorous evaluation using metrics such as accuracy, precision, recall, or F1-score is essential. This phase may involve cross-validation, hyperparameter tuning, and error analysis to improve model robustness. A well-performing model must also be scalable and aligned with strategic objectives, whether in a commercial or research context [Harrington, (2012); Sarkar et al., (2018)].

7. **Deployment and Continuous Monitoring**

   The final step entails deploying the trained model into a real-world environment, where it begins generating predictions or classifications. Nevertheless, post-deployment evaluation remains vital. Over time, shifts in underlying data distributions—often termed “model drift”—necessitate ongoing performance monitoring and periodic retraining. This continuous feedback loop ensures the model remains relevant and valuable in dynamic settings [Sarkar et al., (2018)].

2.2.2 Supervised Machine Learning

Supervised machine learning represents a fundamental subset of the broader field of machine learning, wherein an algorithm is trained on a dataset composed of input-output pairs. Each input instance is associated with a corresponding, pre-defined output label, enabling the model to learn a functional mapping from feature space to target variables. The overarching objective of supervised learning is to develop predictive models capable of generalizing to previously unseen data [Subasi, (2020)].

Supervised learning tasks are generally categorized into two principal domains:

- **Classification**, where the target variable is categorical (e.g., distinguishing between spam and non-spam emails, or identifying employment status).
- **Regression**, where the target variable is continuous (e.g., estimating housing prices or predicting individual income based on educational attainment).

The development of a supervised learning model conventionally involves dividing the available dataset into distinct subsets, typically a **training set**, used to build the model, and a **testing set**, used to evaluate its performance. Key algorithms employed in this paradigm include decision trees, support vector machines (SVM), logistic regression, and artificial neural networks [Subasi, (2020); Sarkar, Bali, & Sharma, (2018)].

A critical determinant of model performance is the **quality and quantity of labeled data**, as well as the **relevance and informativeness of the input features**.

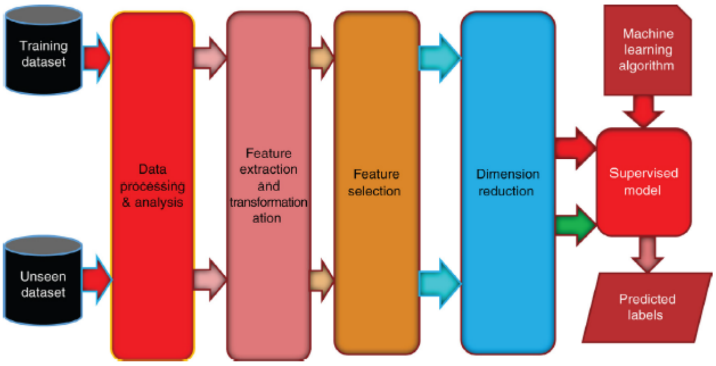


Figure 5 Supervised machine learning framework [Subasi, A. (2020)].

As illustrated in Figure 5 [Subasi, (2020)], the supervised learning workflow begins with data partitioning, followed by essential preprocessing steps such as normalization, feature extraction, and feature selection. It is imperative that these preprocessing operations are applied consistently across both the training and testing subsets to ensure methodological validity and avoid data leakage.

Subsequently, the training data, with its associated labels, is utilized to fit the supervised model. During the prediction or inference phase, the trained model receives input features from previously unseen data and outputs predicted labels.

This ability to infer accurate outputs from novel inputs is the hallmark of successful supervised learning [Sarkar et al., (2018)].

2.2.3 Unsupervised Machine Learning

Unsupervised machine learning constitutes a significant branch of machine learning methodologies in which algorithms are trained on datasets that lack predefined labels or target variables. The primary objective of these techniques is to uncover latent structures, patterns, or distributions inherent in the data without the guidance of explicit supervision or annotated outcomes.

In contrast to supervised learning, wherein models learn a functional mapping from inputs to known outputs, unsupervised learning methods aim to derive insights solely from the input data. Core objectives typically include:

- **Clustering**, where similar data instances are grouped together using algorithms such as k-means or hierarchical clustering.

- **Dimensionality Reduction**, where high-dimensional data is transformed into a lower-dimensional space while preserving the intrinsic relationships among features (e.g., via Principal Component Analysis, PCA).

- **Pattern Discovery**, which involves identifying underlying regularities, associations, or anomalies within the data.

Unsupervised learning is particularly valuable in exploration data analysis, data summarization, and preprocessing phases, including tasks such as feature extraction and anomaly detection. Due to the absence of ground truth labels, evaluation of model performance is inherently more complex and often relies on internal validation metrics (e.g., silhouette score) or qualitative assessments such as visual inspection [Sarkar, Bali, & Sharma, (2018)].

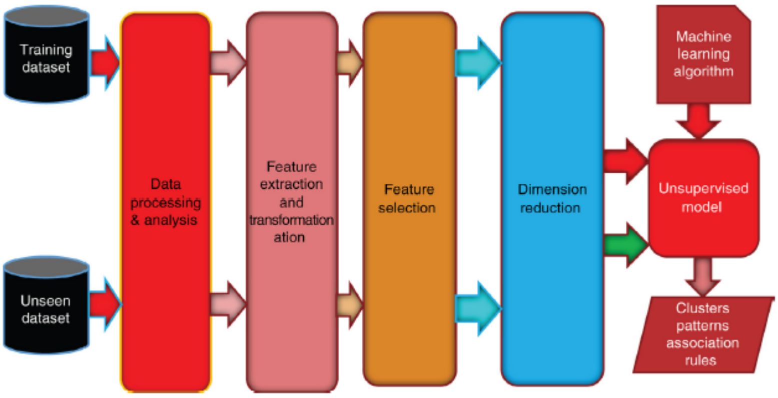


Figure 6 Unsupervised machine learning framework [Subasi, A. (2020)].

As illustrated in Figure 6 [Sarkar et al., (2018)], the unsupervised learning workflow mirrors that of supervised learning in its initial stages. Key preprocessing steps, including feature scaling, extraction, selection, and dimensionality reduction, are applied uniformly.

These transformations are critical for enhancing the model's ability to capture meaningful relationships within the data.

Subsequently, the unsupervised model is constructed using the prepared features derived from the input data. In the inference phase, features from new, unseen samples are fed into the trained model. Depending on the learning objective, the model then produces outputs in the form of cluster assignments, discovered patterns, association rules, or reduced feature representations [Subasi, A. (2020)].

2.3 Python and Jupyter Notebook in Empirical Econometrics.

Python is a high-level programming language developed by Guido van Rossum and first released in 1991. Conceived with the objective of balancing simplicity and expressive power, Python was designed from the outset to support extensibility, modularity, and object-oriented programming principles. Its clean syntax and flexibility have contributed to its widespread adoption, particularly within scientific and academic communities [Pérez et al., (2007)].

Over time, Python has evolved into one of the most extensively utilized languages in computational research and has become a standard in the field of data science and applied econometrics [Mckinney, (2010); Müller & Guido, (2016)].

This study has been conducted within a Python-based computational environment, selected primarily due to its suitability for statistical and econometric analysis. Python's syntactic clarity and internal consistency make it accessible to both learner users and experienced analysts, making it an ideal tool for a broad range of empirical research tasks.
Although the present analysis is designed with an introductory structure and methodological simplicity, its implementation remains sufficiently flexible to allow for seamless adaptation to more advanced and rigorous academic contexts.
Python’s ecosystem offers a coherent and integrated framework for data manipulation, numerical analysis, and visualization.
In particular, the pandas library provides high-level data structures analogous to those found in software such as R and Stata, facilitating efficient data preprocessing and management [Mckinney, (2010)]. For classical statistical modeling, the statsmodels package offers a comprehensive suite of econometric techniques, including ordinary least squares, time series analysis, and hypothesis testing. Moreover, for machine learning and predictive analytics, the scikit-learn library provides access to a broad array of algorithms and tools tailored for supervised and unsupervised learning [Pedregosa et al., (2011)].

Python code can be executed across a variety of interfaces. While the command-line interpreter represents the most basic execution environment, interactive notebook platforms, including Jupyter Notebook, JupyterLab, and Google Colaboratory, have become central to computational research workflows. These interfaces enable the seamless integration of executable code, output, and explanatory text, thereby facilitating reproducibility and interpretability in empirical analysis [Kluyver et al., (2016)].
Originally developed as a component of the IPython project, the Jupyter Notebook emerged as an independent platform in 2014. Its name, an acronym referencing Julia, Python, and R, reflects its foundational purpose as a unified interface for interactive scientific computing across multiple programming languages [Pérez et al., (2007)]. The notebook environment allows users to execute code in separate cells, view outputs immediately, and iteratively refine their analyses—features that are particularly advantageous for exploratory data analysis and model development.

One of the primary strengths of the Jupyter Notebook in empirical research lies in its support for fully reproducible workflows. By integrating narrative text, executable code, and output (including visualizations). Interactive visualization libraries such as Plotly further enhance the analytical process by enabling dynamic and responsive exploration of statistical results [Shirley, (2015)].

In both academic and applied econometric contexts, the combination of Python and Jupyter Notebooks offers several benefits. First, the programming paradigm inherent in Jupyter supports methodological transparency, encouraging reproducibility and peer validation [Rule et al., (2018)]. Second, Python's readable and compact syntax reduces the cognitive overhead often associated with statistical programming, facilitating both accessibility and efficiency. Third, the open-source nature of Python and its extensive ecosystem of libraries presents a cost-effective and extensible alternative to proprietary statistical software, such as Stata.
Moreover, the integration of explanatory text, computational logic, and graphical output within a single document renders Jupyter Notebooks particularly well-suited to instructional settings. This unified format supports pedagogically effective teaching and active learning, which are essential in contemporary data science and econometrics education [Rule et al., (2018)].

In summary, Python—when employed through Jupyter Notebooks—constitutes a powerful, flexible, and user-friendly platform for statistical analysis and machine learning. It enables a seamless transition between traditional econometric modeling and contemporary computational techniques. Supported by a vibrant development community and grounded in open science principles, Python continues to reshape the methodological landscape of data-intensive empirical research.

3. RESEARCH METODOLOGY

3.1 Literature Review

As discussed in the preceding chapters, the increasing prominence of issues such as climate change and environmental pollution has heightened the significance of sustainable investments. Attention to Environmental, Social, and Governance (ESG) factors has emerged as a critical consideration in the field of corporate finance [Wang, Liao, and Zhang (2022)].

Firstly, the integration of ESG principles can be considered a strategic decision for management.

A company's commitment to sustainability not only enhances corporate value but also fosters long-term sustainability [Freeman (1984)]. Moreover, sustainability initiatives contribute to competitive differentiation and cost efficiencies [Porter et al. (2019)], positively influence employee engagement, promote productive workplace behaviors, and strengthen customer loyalty [Kim and Park (2017)]. Secondly, the adoption of ESG frameworks supports more effective risk and opportunity management, bringing advantages key stakeholders, including management, employees, supply chain partners, and customers. Consequently, an increasing number of investors are prioritizing the long-term implications of a company’s environmental, social, and governance performance in their asset allocation decisions.

3.1.1 Theoretical Bases

There are two primary theories concerning the relationship between ESG (Environmental, Social, and Governance) performance and financial performance, namely:

1. **The Shareholder Theory** [Friedman (1970)].
2. **The Stakeholder Theory** [Freeman (1984)].

The Shareholder Theory

The shareholder theory posits a negative relationship between ESG practices and firm performance. According to Barnea and Rubin (2010), this perspective is grounded in the agency problem, wherein managers may overinvest in ESG initiatives at the expense of shareholders to advance their personal interests. Such overinvestment, particularly when it exceeds the optimal level, leads to diminishing returns, as the associated costs outweigh the derived benefits [Krüger (2015)].

The Stakeholder Theory

Conversely, the stakeholder theory emphasizes the potential advantages of ESG practices in enhancing a firm's financial performance. This view aligns with the conflict resolution hypothesis, which asserts that ESG initiatives can mitigate conflicts of interest between a firm's management and its non-investor stakeholders, thereby fostering alignment and improving overall performance [Freeman (1984)].

A pragmatic application of stakeholder theory is showed in the Hera Group’s sustainability balanced scorecard. This framework consolidates strategic objectives covering economic, competitive, and socio-environmental dimensions into a cohesive diagram [Merchant et al. (2014)]. As illustrated in Figure 7, the socio-environmental dimensions (depicted in blue) are not directly linked to short-term financial metrics, such as profitability and turnover growth (represented in white). However, these dimensions are indirectly connected through a network of cause-effect relationships that ultimately influence the financial outcomes over time.

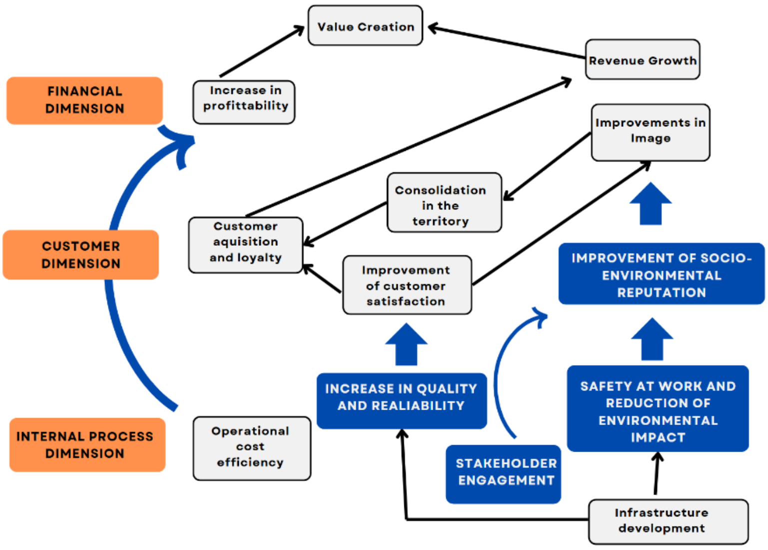


Figure 7 Hera Group's Balanced Scorecard [Merchant et al. (2014)].

Building on the foundations of stakeholder theory, Pulino et al. (2022) examine the limitations of traditional financial reporting in capturing the full scope of corporate activities. Their study highlights the inadequacy of such reports in providing a comprehensive view and highlights the role of integrating non-financial information. This inclusion, they argue, mitigates information asymmetries and agency costs, both negative outcomes predicted by agency theory.

Furthermore, the authors propose signaling theory, according to which the inclusion of sustainability disclosures serves as a positive signal to the market, reflecting the firm's commitment to sustainable practices. This signaling, in turn, has the potential to improve financial performance by promoting trust and aligning corporate activities with market expectations.

Agency theory is founded on two core principles:

1. The relationship between the principal and the agent
2. The separation of ownership and control

A conflict of interest often emerges between the principal, represented by shareholders, and the agent, represented by company managers, due to their divergent objectives. Shareholders typically adopt a long-term orientation, whereas managers may prioritize short-term goals. This divergence highlights the importance of transparency. Higher levels of non-financial information disclosure, particularly concerning ESG practices, tend to satisfy shareholders by enhancing transparency and mitigating agency conflicts.

Signaling theory emphasizes the critical role of information in business transactions.

Organizations strategically allocate financial resources to communicate favorable information regarding their sustainability commitments. By doing so, they provide stakeholders with valued insights while reducing information asymmetry between the company and external, non-financial stakeholders [Connelly et al. (2011)].

Both agency theory and signaling theory underscore the critical role of responsible corporate behavior in aligning with societal and stakeholder expectations. As noted by Ching et al. (2017), failure to meet these expectations can result in significant market sanctions, which are reflected in diminished corporate performance, lower sales growth rates, and declining EBIT. Consequently, transparency and responsible business practices transcend ethical obligations, emerging as pivotal components for ensuring financial stability and long-term success. Enhanced environmental, social, and governance (ESG) performance plays a vital role in reducing information asymmetries within the firm [Cui et al. (2018)], thereby lowering both the cost of equity and debt [Bhuiyan & Nguyen (2020)].

Furthermore, improved ESG performance strengthens corporate reputation, fostering greater stakeholder commitment [Turker (2019)]. In alignment with stakeholder theory, higher ESG performance is positively correlated with increased firm value, illustrating its importance in achieving sustainable corporate success [Lv et al. (2020)].

3.1.2 Literature supporting the positive impact of ESG on financial performance.

Since the 1970s, scholars have increasingly explored the relationship between corporate financial performance and Environmental, Social, and Governance (ESG) factors. Friede et al. (2015) highlights that this area of inquiry has given rise to over 2.000 empirical studies, supplemented by numerous review articles authored by both academics and business researchers.

As illustrated in Figure 8, the volume of research in this domain has expanded exponentially, with a particularly notable surge in the past decade.

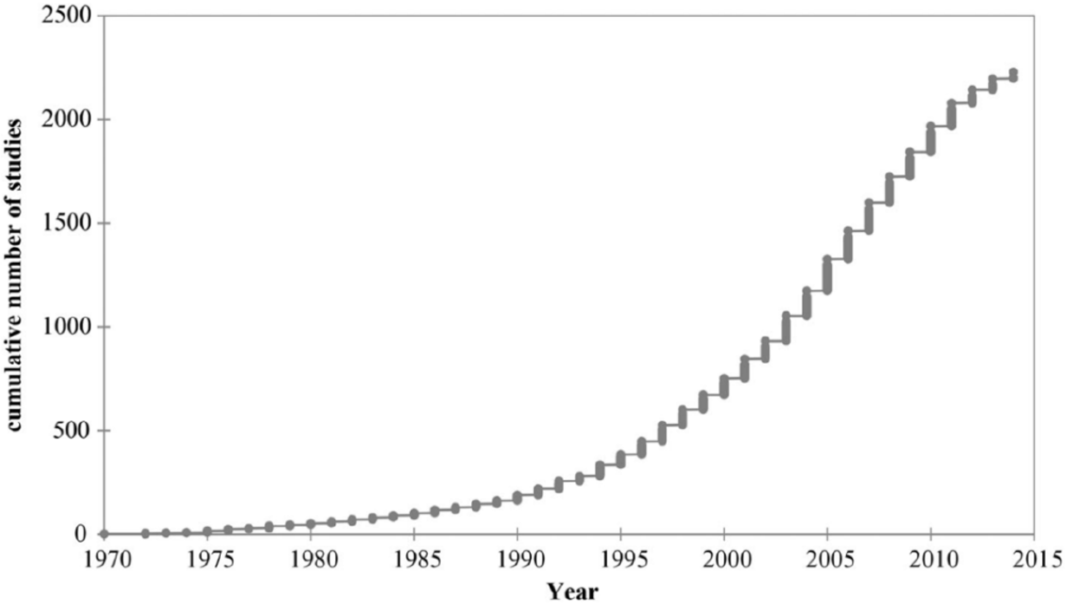


Figure 8 Estimated number of empirical studies on the ESG–CFP relation over time.

Among these more than 2.000 studies, the findings reveal varying results.

Specifically, a detailed analysis indicates that 58% of the studies report a positive relationship between ESG performance and financial performance, 8% identify a negative relationship, and 21% present mixed outcomes. Notably, 13% of the studies find no significant relationship between ESG and financial performance [Whelan et al. (2021)].

Examining the studies that support a positive relationship we find the paper by Naeem et al. (2022) conducted a statistical analysis of 1.042 companies in emerging markets over the period 2010–2019. Their findings indicate that both the overall ESG score, and its three individual dimensions (environmental, social, and governance) positively and significantly impact corporate profitability, measured by return on assets (ROA). Similarly, De Lucia et al. (2020)

analyzed 1.038 companies from 22 European Union countries between 2018 and 2019, concluding that ESG investments positively influence both return on equity (ROE) and ROA. Velte (2017) offers further evidence, arguing that ESG investments enhance corporate value, as measured by Tobin's Q index, and profitability, measured by ROA, in the context of companies listed in Germany.

These findings suggest a potential geographical bias, given the apparent homogeneity of the studied sample, which focuses on firms within specific and limited geographical regions. However, this concern is addressed by Bhaskaran et al. (2020), who conducted a global analysis of 4.887 companies spanning multiple regions during the 2014–2018 period. Using Tobin's Q as a measure of corporate value and ROA as an indicator of profitability, their study demonstrates that firms with stronger ESG performance consistently create greater market value and exhibit enhanced financial outcomes, regardless of geographical constraints.

3.1.3 Literature supporting the negative impact of ESG on financial performance

Some scholars contend that ESG investments may adversely affect profitability and corporate value. For instance, Brammer et al. (2006), in their analysis of the impact of social investments on market returns of publicly traded companies in the United Kingdom, found that firms with lower scores in the Social pillar exhibited higher market returns. Similarly, Landi and Sciarelli (2019), in their study of 54 Italian companies listed between 2007 and 2015, reported a negative association between ESG performance and financial outcomes. Marsat and Williams (2011) also identified a negative relationship between corporate social responsibility (CSR) ratings and corporate value, utilizing MSCI ESG ratings as a measure.

These findings suggest that ESG factors influence productivity and decisions regarding a firm's capital structure; however, the nature of these influences—whether positive or negative—remains a subject of ongoing debate. A key factor contributing to this debate is the lack of precision in assessing ESG pillars and the absence of a standardized measurement framework.

The variety of ESG indices available often leads to divergent results and substantial measurement errors in analyses examining the relationship between ESG performance and corporate outcomes.

Nollet et al. (2016) highlight these challenges by noting potential flaws in regression-based assessments. Their investigation into the relationship between social and financial performance of S&P 500 companies from 2007 to 2011 revealed contrasting results: a negative relationship in linear models and a positive relationship in non-linear models.

These discrepancies underscore the complexity of accurately measuring and interpreting the effects of ESG factors on corporate performance.

3.1.4 Literature supporting a mixed impact of ESG on financial performance

After examining the literature supporting both positive and negative relationships between ESG performance and corporate outcomes, this section explores studies that have yielded mixed results: findings that vary depending on the parameters analyzed.

A common approach in such studies is to assess the overall ESG score rather than focusing on the three ESG pillars (Environmental, Social, and Governance) individually. However, the interconnectedness of the pillars must be acknowledged, recognizing that analyzing them collectively or isolating just one of them can introduce biases or oversimplifications.

For instance, Almeyda and Darmansya (2019) conducted a study on Italian listed companies in the banking sector over a nine-year period, exploring the relationship between ESG disclosure and corporate performance, as measured by return on assets (ROA) and return on capital (ROC). Their findings indicated a significant overall correlation between ESG disclosure and corporate performance. However, when examining the individual pillars, the results diverged: the environmental pillar demonstrated a positive relationship with both ROC and share price, while the social and governance pillars showed no significant association.

Similarly, Han et al. (2016), in their analysis of Korean listed companies from 2008 to 2014, observed a positive correlation between the governance pillar and corporate performance, alongside a negative relationship for the environmental and social pillars.

These findings suggest that while the three pillars are inherently linked to the overall ESG score, analyzing them separately can reveal contrasting relationships. This underscores the complexity of ESG metrics and highlights the importance of a nuanced approach in assessing their impact on corporate performance.

Shifting the focus from ESG parameters to profitability metrics, the literature reveals contradictory findings depending on the indices employed. Much of the research has concentrated on the impact of ESG performance on aspects more closely tied to a firm's market value, while comparatively less attention has been given to operational performance, such as its influence on income statement data and investment returns.

For instance, studies by Atan et al. (2019) and Giannopulos et al. (2022) investigated the relationship between ESG scores and various financial measures, including the cost of capital, market value, and corporate profitability, using samples of companies from Malaysia and Norway, respectively.

Both studies identified a positive correlation between ESG performance and Tobin's Q, a commonly used measure of corporate value. However, they also observed a negative relationship between ESG performance and return on assets (ROA), a key indicator of profitability.

These findings highlight the nuanced and sometimes inconsistent nature of the relationship between ESG factors and profitability measures, suggesting the need for a more granular analysis of how ESG influences different dimensions of financial performance.

3.2 Hypothesis development

Building on the insights presented in the preceding chapters and the literature reviewed, this study aims to examine the relationship between Environmental, Social, and Governance (ESG) factors and key income statement indicators. Specifically, it seeks to assess the potential influence of ESG scores on corporate profitability.

Given the growing interest from investors, stakeholders, and the broader public in corporate reputation and sustainability, it is plausible to anticipate a significant relationship between a firm’s operating performance and its ESG evaluation.

This study analyzes a sample of Europeran publicly listed companies to determine whether the empirical results support the body of literature suggesting a **positive** correlation between ESG performance and profitability, or whether they align more closely with studies indicating a **negative** or negligible.

The following hypotheses will be tested:

- **Hypothesis 1**: Drawing on stakeholder theory and the rising importance of sustainable business practices, a correlation exists between a corporation’s operating performance and its ESG evaluation. ESG scores exert a positive and significant influence on corporate profitability.

- **Hypothesis 2**: The individual components of the ESG score (environmental, social, and governance) impact profitability differently, with each pillar exerting a distinct effect on operational profitability.

Numerous studies and observations from market practices have suggested a positive relationship between ESG performance and corporate financial performance. However, as discussed in earlier sections, prior research has also yielded a notable number of negative or mixed findings.

By testing these hypotheses, this study seeks to contribute to the ongoing discourse on ESG and corporate performance by utilizing a recent and comprehensive dataset.

3.3 Research Design and Data Collection

In this section, we will present and critically examine the dataset utilized in the analysis, along with the selection of variables. In the first part, we will elaborate on the process of constructing the dataset and delineate its principal features. Following this, the focus will shift to a brief discussion of the variables selected for the analysis, including their categorization into independent, dependent, and control variables.

3.3.1 Dataset Construction and Sample Selection

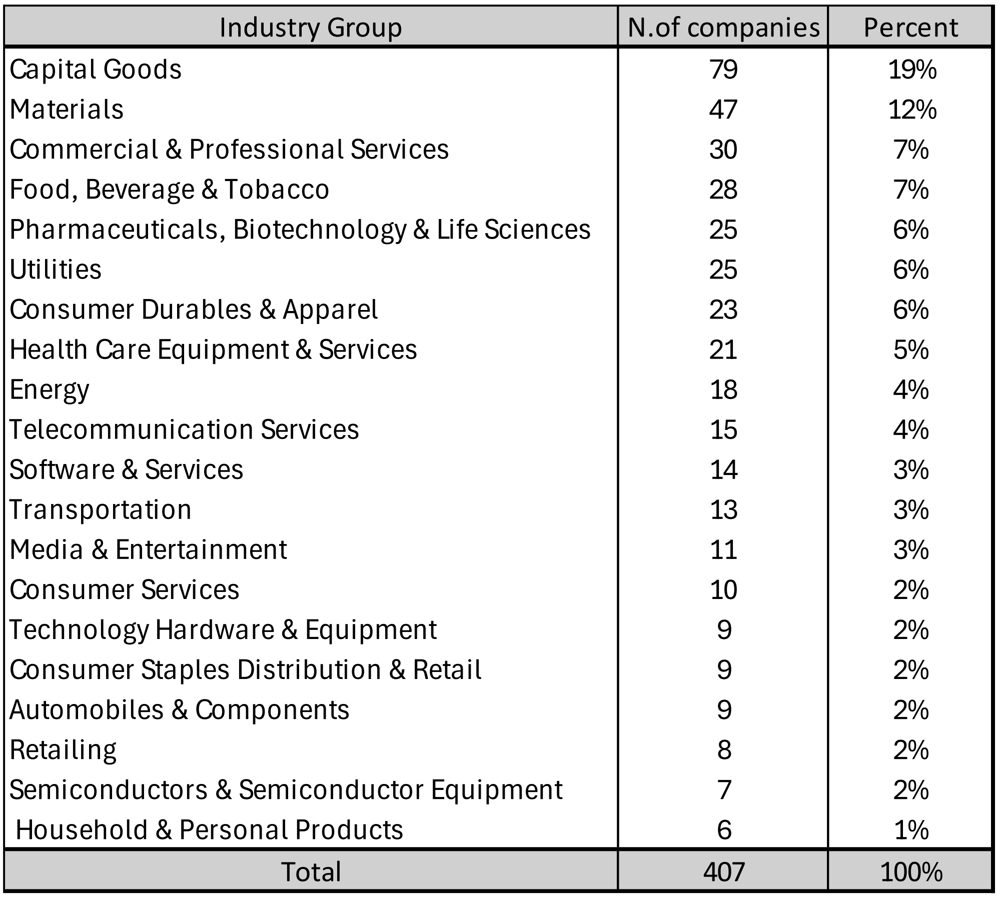
For the purposes of this analysis, financial, operational, and ESG data were sourced from Thomson Reuters. The final dataset was constructed using the Refinitiv Eikon database, a robust resource offering detailed information on market indices, macroeconomic trends, and financial metrics [Thomson Reuters (2017)].

Notably, Refinitiv hosts one of the most comprehensive ESG databases, which has been extensively utilized in recent studies on related topics [Demers et al. (2021); Mooneeapen et al. (2022)].

This study focuses on publicly listed European companies that comprise the Stoxx Europe 600 Index. These companies operate across various sectors and are geographically dispersed throughout Europe.

The Stoxx Europe 600 Index includes large-cap, mid-cap, and small-cap firms across 17 European countries, representing around 90% of the free-float market capitalization of the European stock market [Stoxx Ltd (2013)].

The index was not utilized in its entirety; instead, it was filtered to include only those companies with ESG scores available in the Refinitiv database for at least one year during the analysis period. Additionally, insurance, banking, and financial companies were excluded from the analysis due to their distinct financial and operational structures. This exclusion aimed to mitigate potential outlier effects and reduce bias in the results.



Industry Group	N.of companies	Percent
Capital Goods	79	19%
Materials	47	12%
Commercial & Professional Services	30	7%
Food, Beverage & Tobacco	28	7%
Pharmaceuticals, Biotechnology & Life Sciences	25	6%
Utilities	25	6%
Consumer Durables & Apparel	23	6%
Health Care Equipment & Services	21	5%
Energy	18	4%
Telecommunication Services	15	4%
Software & Services	14	3%
Transportation	13	3%
Media & Entertainment	11	3%
Consumer Services	10	2%
Technology Hardware & Equipment	9	2%
Consumer Staples Distribution & Retail	9	2%
Automobiles & Components	9	2%
Retailing	8	2%
Semiconductors & Semiconductor Equipment	7	2%
Household & Personal Products	6	1%
Total	407	100%

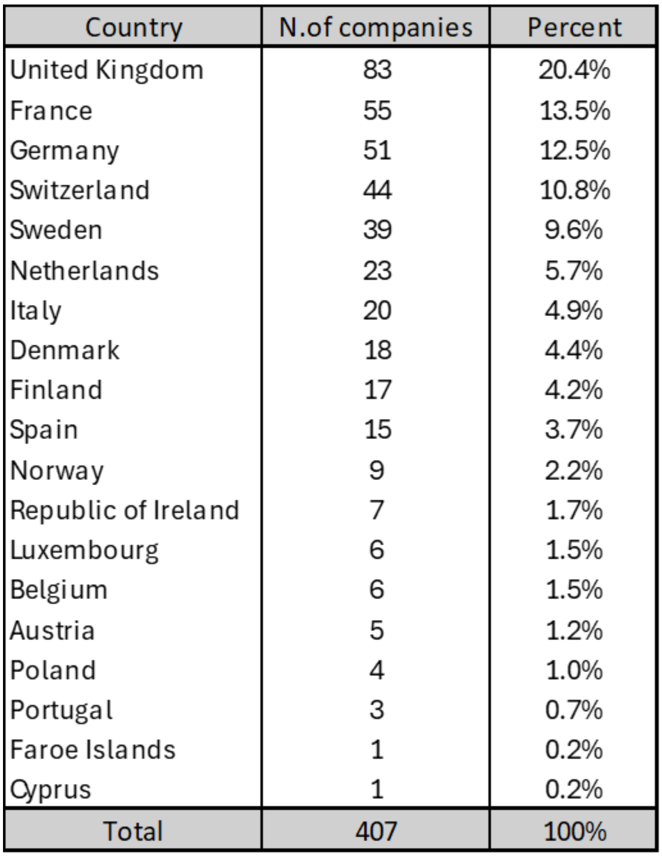
Figure 9 Sectoral Distribution of selected companies in the Stoxx Europe 600 Index [Personal Processing in Python].

It is significant to mention that all sector classifications were performed in accordance with the Global Industry Classification Standards (GICS), a widely recognized framework for industry analysis, investment research, portfolio management, and asset allocation.

The resulting dataset is a panel dataset encompassing 407 companies and approximately 3.705 company-year observations spanning a 10-year period from 2013 to 2022.

On average, each company contributes 8.80 observations, with the number of observations ranging from a minimum of 2 to a maximum of 10 per company. The median number of observations per company is 10.

Figure 10 presents the geographical distribution of the companies across Europe, while Figure 9 illustrates their sectoral distribution [Refinitiv (2023)].



Country	N.of companies	Percent
United Kingdom	83	20.4%
France	55	13.5%
Germany	51	12.5%
Switzerland	44	10.8%
Sweden	39	9.6%
Netherlands	23	5.7%
Italy	20	4.9%
Denmark	18	4.4%
Finland	17	4.2%
Spain	15	3.7%
Norway	9	2.2%
Republic of Ireland	7	1.7%
Luxembourg	6	1.5%
Belgium	6	1.5%
Austria	5	1.2%
Poland	4	1.0%
Portugal	3	0.7%
Faroe Islands	1	0.2%
Cyprus	1	0.2%
Total	407	100%

Figure 10 Geographical Distribution of selected companies in the Stoxx Europe 600 Index. Source: personal processing

3.3.2 Dependent Variable

To evaluate corporate performance, this study employs **Earnings Before Interest and Taxes Margin (EBIT Margin)** and **Return on Assets (ROA)** as the dependent variables for the companies included in the sample. Both metrics provide critical insights into the financial and operational efficiency of firms, with each offering distinct perspectives on performance.

Earnings Before Interest and Taxes Margin (EBIT Margin)

EBIT is a fundamental financial metric that reflects a company's operating performance by concentrating on profitability derived solely from core business operations.

By excluding the influence of financial structure (interest expenses) and tax obligations, EBIT offers a clearer representation of operating efficiency and earning capacity [Huber & Hirsch (2017)].

This metric is often referred to as "*operating income*" when derived from the income statement, although slight variations may arise based on the inclusion or exclusion of non-operating income.

EBIT can be calculated using two primary approaches:

- *From Revenue and Operating Expenses:*

$$\text{EBIT} = \text{Total Revenue} - \text{Operating Expenses}$$

- *From Net Income:*

$$\text{EBIT} = \text{Net Income} + \text{Interest Expense} + \text{Tax Expense}$$

In both methods, EBIT highlights a company’s ability to generate profits solely from its operations, irrespective of capital structure or tax environment. This feature makes EBIT particularly valuable for comparing companies with varying financing strategies or tax jurisdictions.

Furthermore, EBIT serves as an indicator of how effectively a company manages its operating expenses in relation to its revenues, shedding light on its core operational efficiency.

The EBIT margin (Earnings Before Interest and Taxes Margin) represents a widely recognized profitability metric that quantifies a firm's operating performance by expressing earnings before interest and taxes as a percentage of its total revenue. This metric provides insight into the firm's operational efficiency by evaluating its ability to generate profits from core business activities prior to accounting for financing and tax-related costs.

Given the heterogeneity of the firms included in our dataset—characterized by their geographic dispersion, membership in diverse industrial groups, and significantly varying sizes—the EBIT margin emerges as a particularly suitable metric for analysis.

It facilitates a standardized comparison of operating efficiency across companies, independent of variations in capital structure or financial leverage.

By employing the EBIT margin, we can objectively assess and benchmark the relative operational performance of firms, enabling meaningful comparisons even in the context of substantial differences in scale and industry characteristics. Higher EBIT margins are indicative of superior operational profitability [Greene, W. H. (2012)].

EBIT margin is calculated as:

$$\text{EBIT Margin} = \left(\frac{\text{EBIT}}{\text{Revenue}}\right) \cdot 100$$

Return on Assets (ROA)

ROA is a profitability ratio that measures a company's efficiency in generating profit relative to its total assets, providing insight into how effectively the firm utilizes its resources [Higgins (2016)].

It is calculated as follows:

$$\text{ROA} = \frac{\text{Net Income}}{\text{Average Total Asset}}$$

An alternative representation of ROA emphasizes its components:

$$\text{ROA} = \text{Profit Margin x Asset Turnover}$$

Where:

$$\text{Profit Margin} = \frac{\text{Net Income}}{\text{Net Sales}}$$

$$\text{Asset Turover} = \frac{\text{Net sales}}{\text{Average Total Asset}}$$

A higher ROA is desirable, as it indicates greater efficiency in utilizing the company's asset base to generate profits [Higgins (2016)].

Examining the two components of ROA provides further insights into its significance:

- *Operating Margin:* This metric evaluates a company's ability to convert sales into operating profit after covering all operating expenses. Expressed as a percentage, it isolates profitability derived from core operations while excluding financial structure and tax effects, like EBIT. Operating margin reflects how efficiently a company manages its costs relative to sales, enabling comparisons between firms across similar industries [Damodaran (2014)].

- *Asset Turnover:* This ratio captures a company's capacity to convert its investment in assets into sales, serving as a benchmark for operational efficiency. Asset turnover is particularly relevant for firms in asset-intensive sectors, where asset utilization significantly impacts profitability. A higher asset turnover ratio signals effective asset management, while a lower ratio may point to inefficiencies or underutilization [Damodaran (2014)].

In summary, EBIT and ROA, together with their respective components, provide a comprehensive view of corporate performance. EBIT emphasizes operational profitability, while ROA assesses the efficient use of resources to generate returns. Both metrics facilitate meaningful comparisons across firms, contributing to a nuanced understanding of financial and operational dynamics.

3.3.3 Independent Variables

The independent variables included in this analysis comprise the company's overall ESG score, and its three main pillars (environmental, social, and governance) considered individually.
These explanatory variables capture the company's commitment to non-financial objectives that extend beyond the traditional goal of maximizing shareholder profits [Pulino et al. (2022)].
For the ESG evaluation, the study utilizes the Refinitiv ESG score, a comprehensive metric derived from company-specific data across the environmental, social, and governance dimensions.
Each pillar provides insights into distinct aspects of the company's non-financial performance:

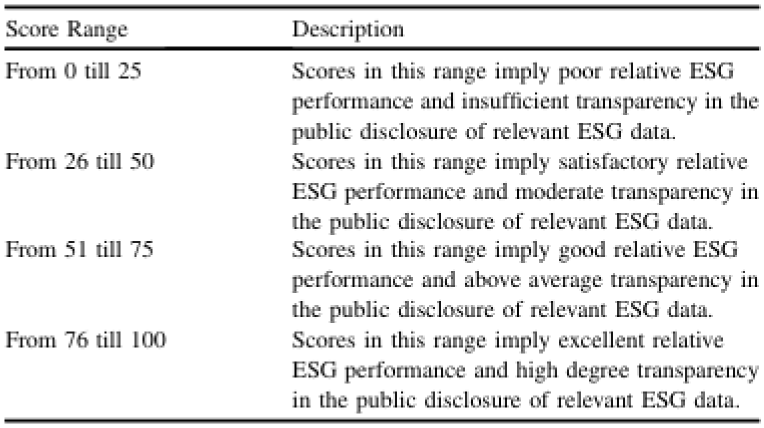
1. **ESG Score**: Reflects the company's approach to human rights, the integration of sustainability into core operations, emission reduction, and environmental protection practices.

2. **Environmental Score**: Indicates the company’s environmental impact, with a higher score denoting lower emissions and a stronger commitment to sustainability initiatives.

3. **Social Score**: Represents the organization’s efforts to enhance human resource management through various social and HR initiatives. A higher score reflects improved working conditions and employee well-being.

4. **Governance Score**: Assesses the company's dedication to social responsibility principles and the effective management of ethical and social issues. A high governance score signifies robust governance practices and alignment with corporate social responsibility standards [Refinitiv (2023)].

All independent variables were obtained from the Refinitiv Eikon database, a resource provided by Thomson Reuters. The ESG metrics are standardized on a scale ranging from 0 to 100, with a score of 100 denoting outstanding adherence to best practices and a strong capacity to generate sustained long-term value for shareholders [Aydoğmuş et al. (2022)].
Table 3 offers a detailed description of the Refinitiv ESG score ranges, providing context for the interpretation of these metrics.



Score Range	Description
From 0 till 25	Scores in this range imply poor relative ESG performance and insufficient transparency in the public disclosure of relevant ESG data.
From 26 till 50	Scores in this range imply satisfactory relative ESG performance and moderate transparency in the public disclosure of relevant ESG data.
From 51 till 75	Scores in this range imply good relative ESG performance and above average transparency in the public disclosure of relevant ESG data.
From 76 till 100	Scores in this range imply excellent relative ESG performance and high degree transparency in the public disclosure of relevant ESG data.

Figure 11 Refinitiv ESG Score range. [Refinitiv (2023)].

3.3.4 Control Variables

In alignment with the reviewed literature [Pulino et al. (2022)], a series of control variables related to profitability and ESG performance were incorporated into the analysis to enhance its robustness and reliability.

The first control variable introduced is **Firm Size**, as it can significantly affect the dependent variable in ways unrelated to the ESG variables. Accounting for firm size ensures that its effects are appropriately considered, yielding more precise and unbiased results.

The inclusion of firm size as a control variable is justified, as it accounts for heterogeneity in firm-specific characteristics. Larger firms tend to differ systematically from smaller firms in terms of their operational scale, financial structure, market influence, and strategic behavior.

Controlling for firm size enables the analysis to isolate the effects of the primary independent variables (ESG Score and its Pillars) by mitigating the confounding influence of these structural differences.

To measure firm size, the natural logarithm of total assets is utilized.

This measure is widely recognized for its robustness, as it simplifies large variations, captures proportional changes, reduces skewness, and aligns well with the nonlinear relationships commonly observed in economic models.

By adopting this approach, the analysis ensures that firm size is effectively accounted for, strengthening the overall validity of the results [Shen, Y. (2023)].

The second control variable incorporated into the analysis is **Leverage**, as provided by Thomson Reuters Eikon. This measure is calculated as the ratio of total debt to total assets, reflecting the proportion of a firm's assets financed through debt.

Leverage is included as a control variable due to its significant impact on firm performance, decision-making processes, and exposure to financial risk. By accounting for leverage, the analysis ensures that the effects of a firm's financial structure on its performance are adequately controlled, thereby enhancing the accuracy and reliability of the estimated relationships among the primary variables of interest [Shen, Y. (2023)].

The debt-to-total-assets ratio is utilized as the preferred measure of leverage because it is straightforward, easily interpretable, and broadly applicable across diverse industries, making it an effective tool for capturing variations in financial structure.

Another control variable included in the study is the **Asset Turnover**, sourced from Eikon.

The Asset Turnover Ratio is a key financial metric used to assess the efficiency with which a firm utilizes its assets to generate revenue. Specifically, it quantifies the amount of revenue generated per unit of assets employed and is calculated as total revenue divided by average total assets. A higher asset turnover indicates more effective use of assets in generating revenue, while lower values may suggest inefficiencies. As a dimension of firm performance, asset turnover complements other financial ratios and supports cross-firm comparisons in operational effectiveness [Xu et al., 2021].

Although asset turnover is a mathematical component of ROA via DuPont decomposition, it reflects a distinct aspect of firm performance. ESG practices may directly influence this efficiency (e.g., through sustainability initiatives or innovation), making it a potential confounder in the relationship between ESG performance and ROA. Controlling for asset turnover thus helps isolate the effect of ESG variables on profitability outcomes [Shan et al., 2024].

Given the study’s focus on non-operational determinants of profitability, the inclusion of asset turnover is justified. To avoid multicollinearity, particularly since ROA equals profit margin multiplied by asset turnover, profit margin is excluded from the model specification [Saygili et al., 2022].

In model specifications where EBIT margin is the dependent variable, the inclusion of asset turnover does not pose a multicollinearity concern, as asset turnover is not a direct mathematical component of EBIT margin.

4. THE ANALYSIS

4.1 Descriptive Analysis

In this section, the dataset's descriptive statistics are explored through a series of analytical steps, beginning with a panel data analysis spanning the last ten fiscal years. This is followed by a cross-sectional analysis and concludes with a correlation analysis.

The initial step involved analyzing the distribution of values for the selected variables. A box plot was employed to visually assess the dataset's spread and symmetry, providing insights into the variability of the data. The box plot in Appendix A[1] also facilitated the identification of potential outliers, extreme values that deviate significantly from the rest of the dataset. Such outliers could distort subsequent analyses and lead to misleading inferences [Tukey, J. W. (1977)].

To mitigate the impact of these anomalies, the interquartile range (IQR) method was applied to detect and remove outliers from the dataset. This robust statistical technique ensures that the analysis is not unduly influenced by extreme values [Tukey, J. W. (1977)].

Following the removal of outliers and additional data cleaning, the final dataset comprises a sample of 402 companies. Each company has an average of 8.21 observations, with the median number of observations per company being 10. This cleaned dataset serves as the foundation for the subsequent analyses.

4.1.1 Panel data Analysis

For the panel data analysis, Figure 12 provides an overview of the ESG Pillars trends over the past ten fiscal years, from 2013 to 2022. The figure illustrates the trends in the scores of the three ESG pillars—Environmental, Social, and Governance (ESG).

The analysis reveals a generally positive and growing trend across the three pillars. Among them, the social pillar demonstrates the highest average values, followed by the Environmental pillar.

---

[1] *Appendix A: Box Plot of the independent variables of the Regression Analysis*
Source: Personal Computation in Python

The Governance pillar, although displaying comparatively lower values, maintains average scores exceeding 50.

Notably, while the social pillar shows consistent and linear growth over the years, the Environmental and Governance pillars experience a significant surge between 2017 and 2018. This indicates distinct growth patterns among the pillars, underscoring the unique dynamics of each dimension within ESG performance.

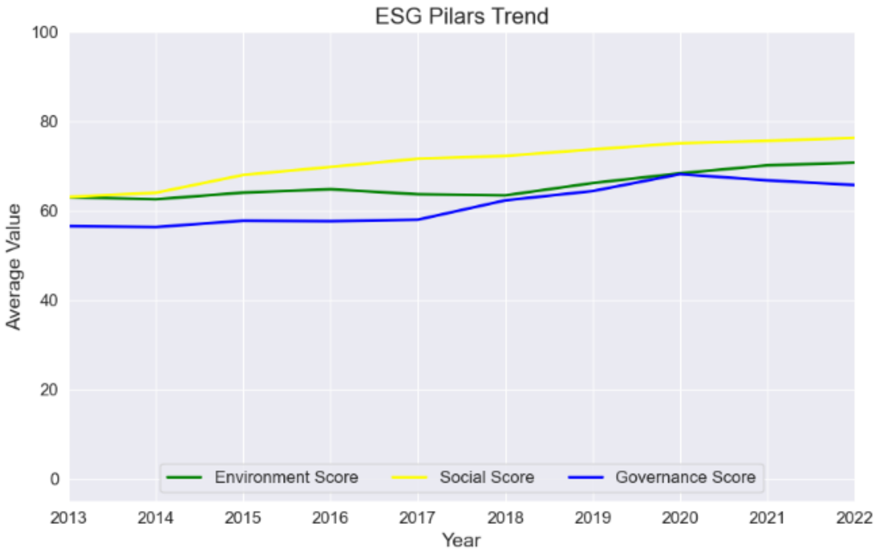


Figure 12 ESG Trends in the STOXX Europe 600 Over the Past Decade [Personal Processing in Python].

4.1.2 Cross-Sectional Analysis

As highlighted in the previous paragraphs a careful selection process was undertaken to minimize bias and eliminate potential outliers. This process resulted in a refined sample comprising 402 companies and 3.299 observations, upon which a detailed descriptive analysis was performed, categorized by geographical regions and GICS classifications.

Figure 13 illustrates the dataset's remarkable diversity and heterogeneity. Notably, the sectors with the highest concentration of observations are **Capital Goods** and **Materials**, accounting for 81 and 44 companies respectively, followed by **Commercial & Personal Services** with 29 firms. Conversely, sectors such as **Household & Personal Products**, **Semiconductors**, and **Consumer Staples Distribution & Retail** are underrepresented, each with fewer than 10 companies.

The geographical distribution of the dataset is well-balanced among major European countries. However, as demonstrated in Figure 14, there is a notable predominance of British companies, comprising 78 companies. This is followed by France and Germany, with 56 and 51 enterprises respectively.

Collectively, these three countries account for nearly half of the dataset, while the remaining observations are distributed more evenly across other nations.

It is worth noting that smaller countries typically exhibit a lower number of observations, which aligns with the expectation that larger nations tend to have a greater number of publicly listed companies.

However, an interesting observation emerges regarding relatively small states such as **Denmark**, **Luxembourg**, and **Ireland**, which contribute a significant number of observations. This anomaly may be attributed to their favorable governmental regulations and robust economic environments, which encourage corporate activity and market participation.

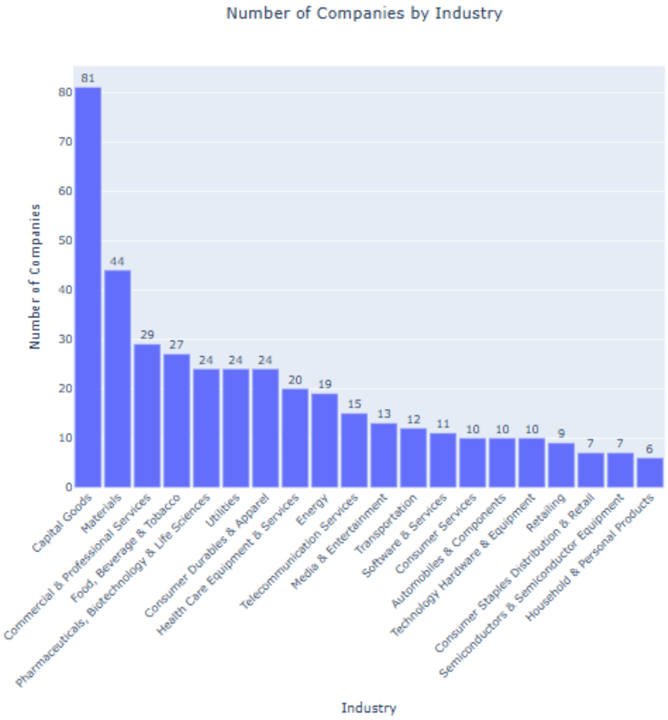


Figure 13 Industry Distribution of Sampled Companies from the STOXX Europe 600 [Personal Processing in Python]

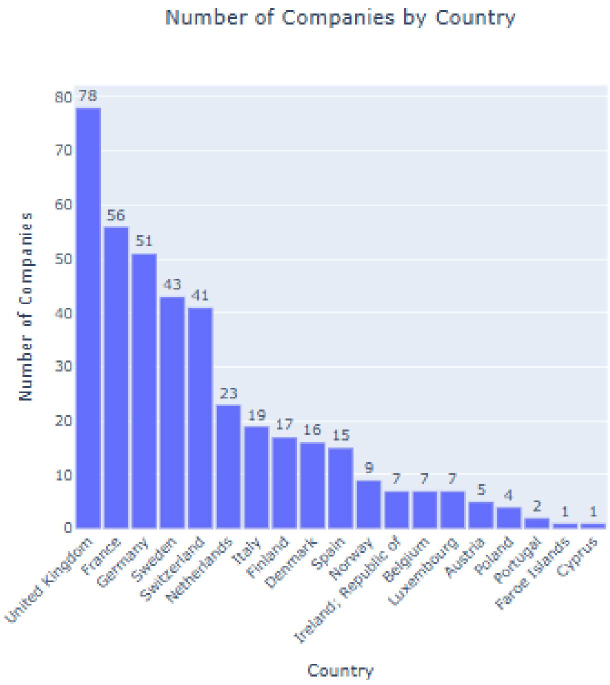


Figure 14 Geographical Distribution of Sampled Companies from the STOXX Europe 600 [Personal Processing in Python]

Table in Figure 15 presents the descriptive statistics utilized in cross-sectional analysis.

For this analysis, the average values of the last 10 fiscal years were considered for the economic indices **ROA** and **EBIT**, as well as for the total ESG indices, which were further broken down into the **three ESG pillars.**

As shown in the table, the **ROA** exhibits a maximum value of 22.87% and a minimum negative value of -5.13%, with an overall average of 8.62%. According to the guidelines provided by Borsa Italiana (2024), a ROA greater than 10% indicates very high efficiency, a ROA between 5% and 10% reflects good efficiency, while a ROA below 5% may suggest potential inefficiency or low margins.

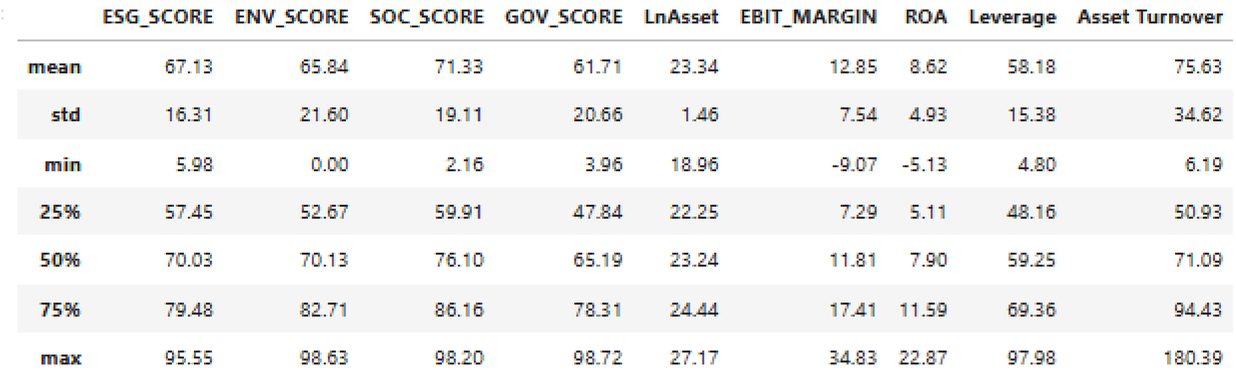
Based on these thresholds, the average ROA value across sectors can be considered satisfactory, reflecting good efficiency.

This suggests that, on average, companies in the index do not record critical situations in managing their resources.

Moving to EBIT Margin, it has an average value of 12.85%, a minimum value of -9.07% and maximum value of 34.83%.

Turning to the ESG pillars, the analysis reveals that the **Social pillar** has the highest average value (71.33), while the **Governance pillar** displays the lowest average value (61.71).

The **Environment pillar**, with an average score of 65.84, is relatively aligned with the total ESG and the other pillars. However, what stands out is its significant variability, with maximum values of 98.63 and minimum values of zero. This disparity could be attributed to the nature of the company's sector of operation. Companies in sectors with high environmental impacts such as chemicals or energy are likely to achieve lower environmental scores compared to those in less impactful sectors, such as services or communications. This discrepancy reflects varying levels of investment and focuses on the Environment pillar across industries.



	ESG_SCORE	ENV_SCORE	SOC_SCORE	GOV_SCORE	LnAsset	EBIT_MARGIN	ROA	Leverage	Asset Turnover
mean	67.13	65.84	71.33	61.71	23.34	12.85	8.62	58.18	75.63
std	16.31	21.60	19.11	20.66	1.46	7.54	4.93	15.38	34.62
min	5.98	0.00	2.16	3.96	18.96	-9.07	-5.13	4.80	6.19
25%	57.45	52.67	59.91	47.84	22.25	7.29	5.11	48.16	50.93
50%	70.03	70.13	76.10	65.19	23.24	11.81	7.90	59.25	71.09
75%	79.48	82.71	86.16	78.31	24.44	17.41	11.59	69.36	94.43
max	95.55	98.63	98.20	98.72	27.17	34.83	22.87	97.98	180.39

Figure 15 Descriptive Statistics of Sampled Companies from the STOXX Europe 600 [Personal Processing in Python].

Analyzing the sample over a 10-year period provides valuable insights into whether the results have evolved over time, allowing us to assess if companies have adapted different actions and strategies. As demonstrated in the Table in Figure 35 represented in Appendix B[2], the descriptive analysis conducted for each year of observation reveals how companies have progressively incorporated sustainable strategies into their business models.

Specifically, the ESG index itself exhibits a steady rise in its average value, from 61.39 in 2013 to 71.77 in 2022. This upward trend suggests that not only are more companies adopting ESG practices, but those already engaged in ESG efforts are intensifying their investments, incorporating greater attention to environmental sustainability, human rights, community well-being, alongside their economic and financial performance.

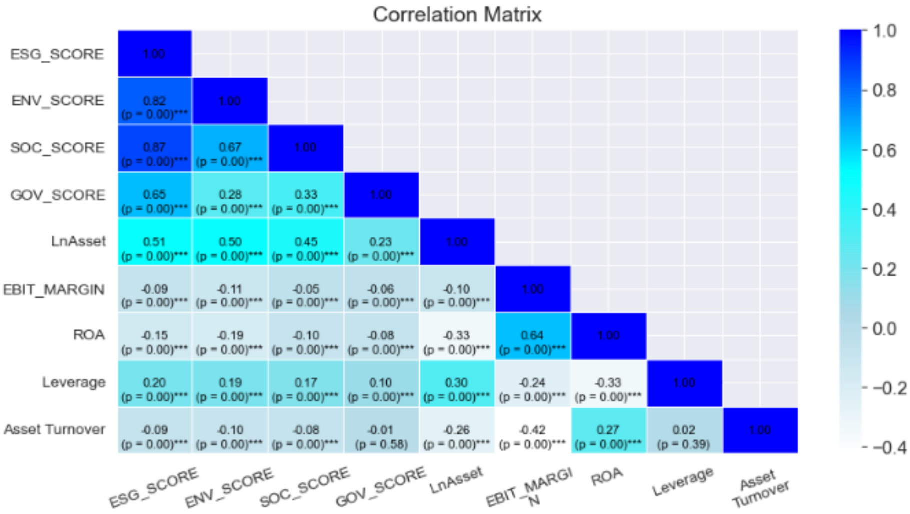
4.1.3 Correlation Analysis

The final stage of the descriptive analysis involves examining the correlation among all variables considered in the study. Table in Figure 16 presents the simple correlation matrix of these variables, with statistically significant correlations denoted using one star for a p-value < 0.1, two stars for a p-value < 0.05, and three stars for a p-value < 0.001.

At first glance, it is evident that all ESG variables exhibit a statistically significant correlation with both average ROA and EBIT.

---

[2] *Appendix B: Figure 35: Descriptive statistics per Year. Source: Personal Processing*



	ESG_SCORE	ENV_SCORE	SOC_SCORE	GOV_SCORE	LnAsset	EBIT_MARGIN	ROA	Leverage	Asset Turnover
ESG_SCORE	1.00								
ENV_SCORE	0.82 (p = 0.00)***	1.00							
SOC_SCORE	0.87 (p = 0.00)***	0.67 (p = 0.00)***	1.00						
GOV_SCORE	0.65 (p = 0.00)***	0.28 (p = 0.00)***	0.33 (p = 0.00)***	1.00					
LnAsset	0.51 (p = 0.00)***	0.50 (p = 0.00)***	0.45 (p = 0.00)***	0.23 (p = 0.00)***	1.00				
EBIT_MARGIN	-0.09 (p = 0.00)***	-0.11 (p = 0.00)***	-0.05 (p = 0.00)***	-0.06 (p = 0.00)***	-0.10 (p = 0.00)***	1.00			
ROA	-0.15 (p = 0.00)***	-0.19 (p = 0.00)***	-0.10 (p = 0.00)***	-0.08 (p = 0.00)***	-0.33 (p = 0.00)***	0.64 (p = 0.00)***	1.00		
Leverage	0.20 (p = 0.00)***	0.19 (p = 0.00)***	0.17 (p = 0.00)***	0.10 (p = 0.00)***	0.30 (p = 0.00)***	-0.24 (p = 0.00)***	-0.33 (p = 0.00)***	1.00	
Asset Turnover	-0.09 (p = 0.00)***	-0.10 (p = 0.00)***	-0.08 (p = 0.00)***	-0.01 (p = 0.58)	-0.26 (p = 0.00)***	-0.42 (p = 0.00)***	0.27 (p = 0.00)***	0.02 (p = 0.39)	1.00

Figure 16 Correlation Matrix of Sampled Companies from the STOXX Europe 600 [Personal Processing in Python].

Specifically, the ESG score demonstrates a negative correlation with both ROA (-0.15) and with EBIT Margin (-0.09).

Analyzing the individual ESG pillars reveals further nuances. Although all the pillars display a negative correlation with ROA, the magnitude of this relationship varies.

The **Governance pillar** shows the weakest negative correlation (-0.08), while the **Environmental pillar** exhibits the strongest negative correlation (-0.19).

When examining correlations with EBIT Margin, all ESG pillars display negative and statistically significant relationships. However, the **Environment pillar** shows a notably stronger correlation

(-0.11) compared to the **Governance pillar** (-0.06), nearly double in magnitude.

This finding underscores that the individual ESG pillars do not exert uniform effects on economic performance metrics.

In conclusion, these results highlight the differing correlations of ESG components on profitability indicators, confirming that not all ESG pillars carry equal weight in shaping economic outcomes.

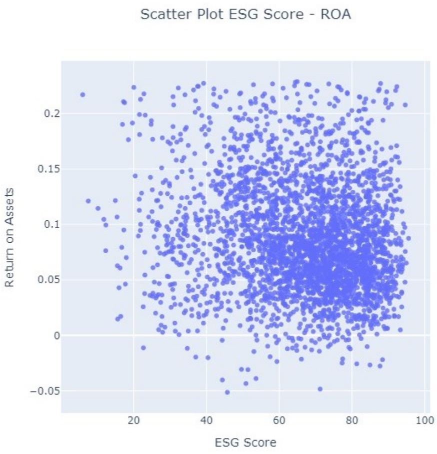


Figure 18 Scatter Plot between the ESG Score and ROA [Personal Processing in Python].

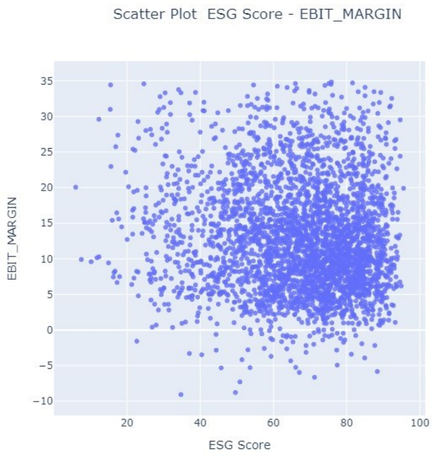


Figure 17 Scatter Plot between the ESG Score and Ebit Margin [Personal Processing in Python].

Figure 18 illustrates the scatterplot between ROA and the final ESG score.
The observations are predominantly concentrated in the upper right quadrant. However, the observed relationship does not appear to be particularly strong or well-defined, as evidenced by the dispersion of data points. The interpretation of this trend is further complicated by the presence of outliers, which may exaggerate patterns in the overall model and introduce additional variability.
A similar pattern is observed in Figure 17, which depicts the relationship between ESG scores and EBIT margin. Here, the observations are again concentrated in the upper right quadrant.
Nevertheless, like the previous case, the relationship between these variables does not exhibit a strong or distinct pattern through visual inspection alone.

To evaluate the presence of multicollinearity among the variables, a Variance Inflation Factor (VIF) test was employed. Two separate analyses were conducted: one using EBIT Margin and the other using ROA as dependent variables. Within each analysis, two scenarios were considered: one incorporating the overall ESG score and another disaggregating the ESG score into its three pillars Environmental, Social, and Governance.
Given the panel structure of the dataset, which consists of multiple firms observed over several years, it was deemed appropriate to control for unobserved, time-invariant firm-specific effects prior to the calculation of the VIF. This was accomplished by applying the **demeaning within transformation** to all independent variables: specifically, for each firm, the average value of each variable across all years was subtracted from the observed values for that firm [Gujarati, D. N., & Porter, D. C. (2009)].

$$x_{it}^* = x_{it} - \bar{x}_i$$

where: x_{it} is value of the variable x for company i at time t

$\bar{x}_i$ is average of the variable x for company i calculated over all available years for that company

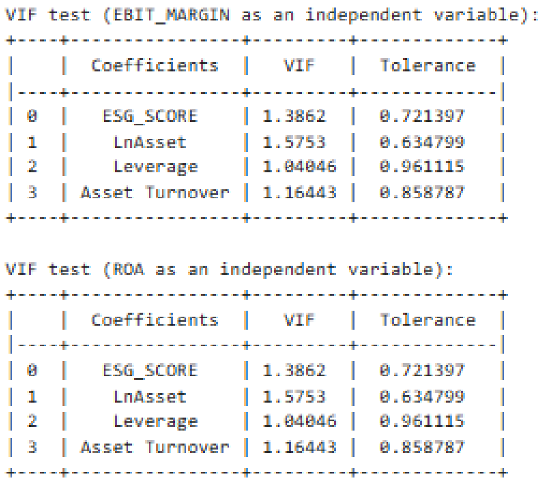
This procedure effectively removes all cross-sectional variations attributable to differences between firms, thereby isolating the within-firm (intra-company) variation over time.

As presented in Figures 19 and 20, the VIF values for the variables in the analysis with the overall ESG score ranged from 1.040 to 1.164. In the analysis with the separate ESG pillars, VIF values were slightly higher, ranging from 1.042 to 1.165.

Regarding tolerance levels, the overall ESG score analysis exhibited values between 0.72 and 0.96, indicating a high level of tolerance and an excellent situation for regression analysis.

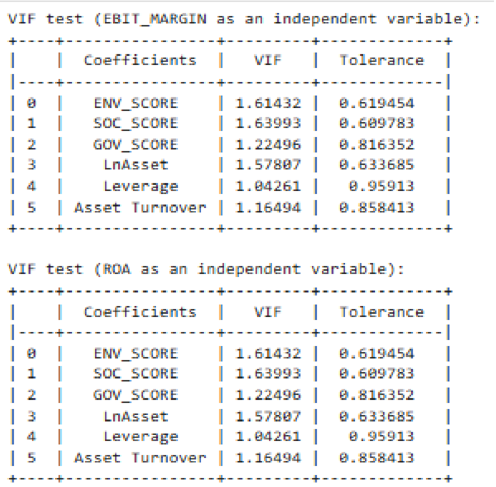
In the analysis with the individual ESG pillars, the tolerance levels were as follows: **ENV_SCORE** and **SOC_SCORE** demonstrated a moderate but acceptable level of 0.62 and 0.61, **GOV_SCORE** showed a high tolerance of 0.82, **LnAsset** was at 0.63, **Leverage** at 0.96, and **Asset Turnover** at 0.85, reflecting very favorable conditions.

The findings indicate that multicollinearity among the independent variables is negligible and does not pose a significant concern. This reinforces the robustness of the dataset for subsequent regression analyses, ensuring the reliability and validity of the results.



```
VIF test (EBIT_MARGIN as an independent variable):
+----+----------------+---------+------------+
|    |  Coefficients  |   VIF   |  Tolerance  |
|----+----------------+---------+------------|
| 0  |   ESG_SCORE    | 1.3862  |  0.721397   |
| 1  |    LnAsset     | 1.5753  |  0.634799   |
| 2  |    Leverage    | 1.04046 |  0.961115   |
| 3  | Asset Turnover | 1.16443 |  0.858787   |
+----+----------------+---------+------------+

VIF test (ROA as an independent variable):
+----+----------------+---------+------------+
|    |  Coefficients  |   VIF   |  Tolerance  |
|----+----------------+---------+------------|
| 0  |   ESG_SCORE    | 1.3862  |  0.721397   |
| 1  |    LnAsset     | 1.5753  |  0.634799   |
| 2  |    Leverage    | 1.04046 |  0.961115   |
| 3  | Asset Turnover | 1.16443 |  0.858787   |
+----+----------------+---------+------------+
```

Figure 20 VIF Values with Overall ESG Score [Personal Processing in Python].



```
VIF test (EBIT_MARGIN as an independent variable):
+----+----------------+---------+------------+
|    |  Coefficients  |   VIF   |  Tolerance  |
|----+----------------+---------+------------|
| 0  |   ENV_SCORE    | 1.61432 |  0.619454   |
| 1  |   SOC_SCORE    | 1.63993 |  0.609783   |
| 2  |   GOV_SCORE    | 1.22496 |  0.816352   |
| 3  |    LnAsset     | 1.57807 |  0.633685   |
| 4  |    Leverage    | 1.04261 |   0.95913   |
| 5  | Asset Turnover | 1.16494 |  0.858413   |
+----+----------------+---------+------------+

VIF test (ROA as an independent variable):
+----+----------------+---------+------------+
|    |  Coefficients  |   VIF   |  Tolerance  |
|----+----------------+---------+------------|
| 0  |   ENV_SCORE    | 1.61432 |  0.619454   |
| 1  |   SOC_SCORE    | 1.63993 |  0.609783   |
| 2  |   GOV_SCORE    | 1.22496 |  0.816352   |
| 3  |    LnAsset     | 1.57807 |  0.633685   |
| 4  |    Leverage    | 1.04261 |   0.95913   |
| 5  | Asset Turnover | 1.16494 |  0.858413   |
+----+----------------+---------+------------+
```

Figure 19 VIF Values with ESG Individual Pillars [Personal Processing in Python].

4.2 Regression Analysis

Before we conduct the panel data regression to estimate the potential relationship between profitability ratios (EBIT Margin and ROA) and ESG ratings, it is necessary to determine the most appropriate panel data model.

Two models have been considered for this purpose: the Random Effects Model and the Fixed Effects Model.

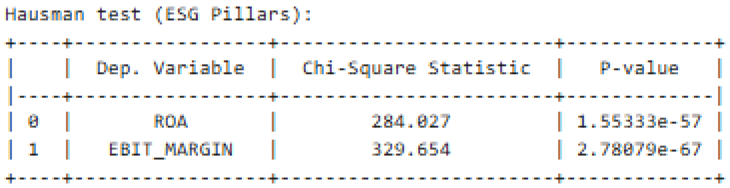
- The **Fixed Effects Model** assumes that unobservable differences across firms are constant over time and correlated with the independent variables of the model.
  This approach focuses exclusively on variations within the same unit over time, thereby excluding inter-unit differences.
  By doing so, it eliminates the risk of bias caused by omitted variables that are correlated with both the dependent and explanatory variables [Greene, W. H. (2008)].
- Conversely, the **Random Effects Model** assumes that unobservable differences between units are uncorrelated with the independent variables and can be treated as random components. This model enables the analysis within the same firm and across firms' variations, offering a broader perspective than the Fixed Effects Model [Greene, W. H. (2008)].

To determine the most suitable model for this analysis, a **Hausman Test [**Wooldridge, J. M. (2010)] was conducted.

The test evaluates the following hypotheses:

- **Null hypothesis (H_0):** No correlation exists between the unit-specific effects and the explanatory variables, supporting the use of the Random Effects Model.
- **Alternative hypothesis (H_1):** Correlation is present, indicating the preference for the Fixed Effects Model.

Based on the results of the Hausman Test (Figure 21), the Fixed Effects Model was selected, as strong correlation was detected between the specific effects and the explanatory variables.



```
Hausman test (ESG Pillars):
+----+-----------------+------------------------+-------------+
|    |  Dep. Variable  |  Chi-Square Statistic  |   P-value   |
|----+-----------------+------------------------+-------------|
|  0 |       ROA       |        284.027         | 1.55333e-57 |
|  1 |   EBIT_MARGIN   |        329.654         | 2.78079e-67 |
+----+-----------------+------------------------+-------------+
```

Figure 21 Hausman Test Results [Personal Processing in Python].

The Chi-Square statistic in the Hausman test quantifies the extent of the difference between the estimators derived under the random effects and fixed effects assumptions. The associated p-value indicates the probability of observing a result against the null hypothesis at least as the observed one, under the null hypothesis, i.e. no correlation between the firm specific effect and the explanatory variables. A higher Chi-Square value is evidence against the null hypothesis.
For the variable ROA, the test produces a Chi-Square statistic of 284.027, which is extremely large, and a corresponding p-value that is well below the conventional significance threshold of 0.05.
This provides strong evidence to reject the null hypothesis, indicating that the specific effects are correlated with the explanatory variables and that a fixed effects model is more suitable for this specification.
Also, for EBIT Margin, the Chi-Square statistic is 329.654, with a p-value well below the 0.05 threshold. In conclusion we opt to employ the fixed effects model for both regressions, (ROA and EBIT as dependent variables).

4.2.1 Fixed Effect Regression Models

After establishing that the fixed effects model is the most appropriate framework for our analysis, this section will explore the relationship between profitability metrics and ESG ratings.
Specifically, two distinct analyses will be conducted, with EBIT Margin and ROA serving as the dependent variables in each case.
For both profitability measures, the analysis will be structured into two separate models. Within these models, the dependent variable and the control variables will remain constant, while the ESG score will be tested in two configurations: the aggregate ESG score, and the three individual pillars (Environmental, Social, and Governance) analyzed separately.

This approach allows for a nuanced exploration of the impact of overall ESG performance and its individual dimensions on firm profitability.

1. $ROA_{it} = \alpha_i + \beta 1 * ESG\_SCORE_{it} + \beta 2 * LnAsset_{it} + \beta 3 * Leverage_{it} + \beta 4 * Asset\ Turnover_{it} + \mu_{it}$
2. $ROA_{it} = \alpha_i + \beta 1 * ENV\_SCORE_{it} + \beta 2 * SOC\_SCORE_{it} + \beta 3 * GOV\_SCORE_{it} + \beta 4 * LnAsset_{it} + \beta 5 * Leverage_{it} + \beta 6 * Asset\ Turnover_{it} + \mu_{it}$
3. $EBIT\_MARGIN_{it} = \alpha_i + \beta 1 * ESG\_SCORE_{it} + \beta 2 * LnAsset_{it} + \beta 3 * Leverage_{it} + \beta 4 * Asset\ Turnover_{it} + \mu_{it}$
4. $EBIT\_MARGIN_{it} = \alpha_i + \beta 1 * ENV\_SCORE_{it} + \beta 2 * SOC\_SCORE_{it} + \beta 3 * GOV\_SCORE_{it} + \beta 4 * LnAsset_{it} + \beta 5 * Leverage_{it} + \beta 6 * Asset\ Turnover_{it} + \mu_{it}$

Where:

- **Dependent Variable**:
  - ROA_{it}: Return on Assets for company *i* at time *t,* measuring the efficiency of asset utilization in generating returns.
  - $EBIT\_MARGIN_{it}$: Earnings Before Interest and Taxes Margin for company *i* at time *t* representing profitability before financing and tax considerations.

- **Fixed Effects (α_i):** Captures firm-specific, time-invariant characteristics, such as management style, industry sector, geographic location or organizational culture, which may influence the dependent variables.

- **Independent Variables**:
  - ESG_SCORE$_{it}$: Comprehensive Environmental, Social, and Governance (ESG).
  - ENV_SCORE$_{it}$, SOC_SCORE$_{it}$, GOV_SCORE$_{it}$: Individual components of the ESG score, representing the environmental, social, and governance dimensions, respectively.
  - LnAsset$_{it}$: The natural logarithm of total assets, serving as a proxy for firm size.
  - Leverage$_{it}$: Total Debt over Total Equity ratio, reflecting the extent of a firm's financial obligations relative to its equity.

    - Asset turnover$_{it:}$ The total Revenue over average Total asset quantifies the amount of revenue a company is able to generate per unit of asset employed. It represents the firm's operational efficiency.

- **Error Term (u_{it})**: Represents unobservable and idiosyncratic factors specific to company *i* at time *t.* Capturing variations not explained by the included independent variables.

4.2.2 Models results

The results of the linear regression analysis for the four models outlined above are presented below. The analysis was conducted on a sample of 402 companies drawn from the STOXX Europe 600 index. As detailed in the preceding chapters, the primary objective of this study is to assess the presence of a positive relationship between a firm's operating performance and its ESG rating.

Furthermore, the analysis seeks to investigate whether the individual components of the ESG score, the environmental, social, and governance dimensions, exert distinct effects on profitability, with a particular focus on their respective impacts on return on assets (ROA) and EBIT Margin.

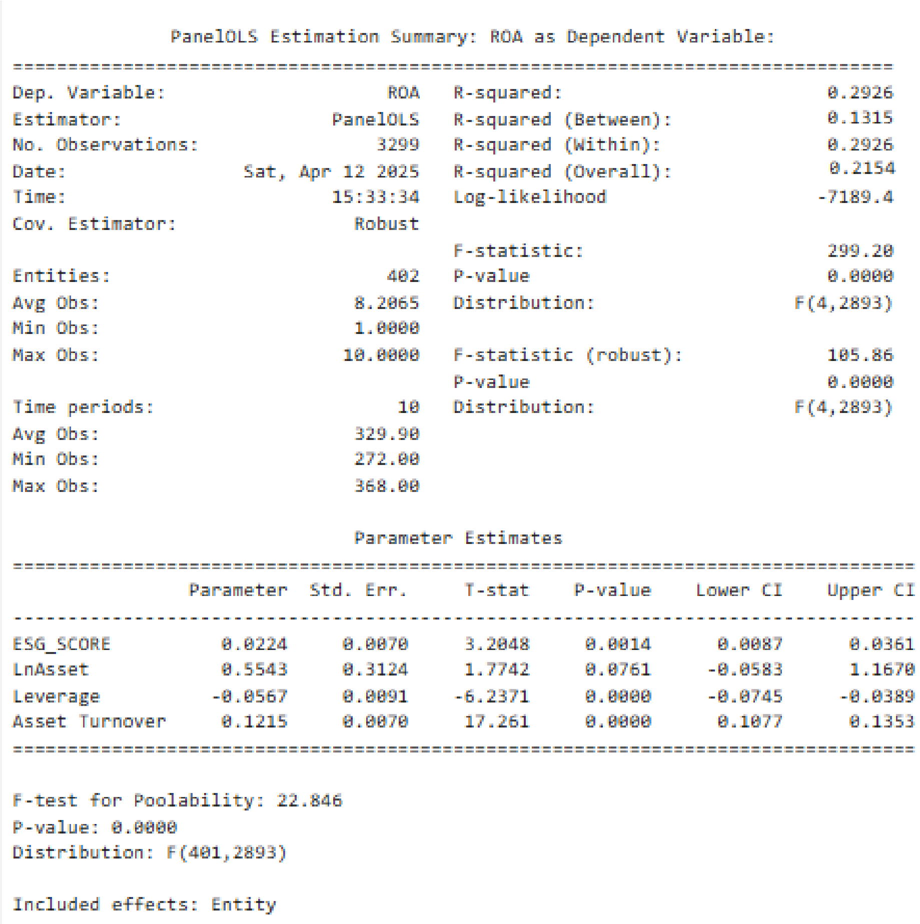
Table in Figure 22 provides a detailed summary of the first regression model outlined above, with **ROA as the dependent variable**.

The model yields a **F-statistic** of **299.20**, which indicates a high degree of joint statistical significance among the independent variables in explaining variation in ROA. This suggests that the explanatory variables collectively offer a meaningful explanation of the dependent variable within the context of the dataset.

Given the panel structure of the dataset, characterized by repeated observations of the same firms over time, it is reasonable to anticipate the presence of both **autocorrelation** (correlation within companies over time) and **heteroskedasticity** (non-constant variance across firms).

These features, if unaddressed, may compromise the reliability of classical F-statistics by inflating the apparent fit of the model and understating standard errors.

Therefore, to ensure the robustness of the findings, the **F-statistic (robust)** was also calculated.



```
                 PanelOLS Estimation Summary: ROA as Dependent Variable:
============================================================================
Dep. Variable:                    ROA   R-squared:                    0.2926
Estimator:                   PanelOLS   R-squared (Between):          0.1315
No. Observations:                3299   R-squared (Within):           0.2926
Date:                Sat, Apr 12 2025   R-squared (Overall):          0.2154
Time:                        15:33:34   Log-likelihood               -7189.4
Cov. Estimator:                Robust
                                        F-statistic:                  299.20
Entities:                         402   P-value                       0.0000
Avg Obs:                       8.2065   Distribution:               F(4,2893)
Min Obs:                       1.0000
Max Obs:                      10.0000   F-statistic (robust):         105.86
                                        P-value                       0.0000
Time periods:                      10   Distribution:               F(4,2893)
Avg Obs:                       329.90
Min Obs:                       272.00
Max Obs:                       368.00

                             Parameter Estimates
==============================================================================
               Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
ESG_SCORE         0.0224     0.0070     3.2048     0.0014      0.0087      0.0361
LnAsset           0.5543     0.3124     1.7742     0.0761     -0.0583      1.1670
Leverage         -0.0567     0.0091    -6.2371     0.0000     -0.0745     -0.0389
Asset Turnover    0.1215     0.0070     17.261     0.0000      0.1077      0.1353
==============================================================================

F-test for Poolability: 22.846
P-value: 0.0000
Distribution: F(401,2893)

Included effects: Entity
```

Figure 22 PanelOLS Estimation Results for the testing Model (1) [Personal Processing in Python].

This version of the test statistic adjusts for arbitrary forms of heteroskedasticity and intra-entity correlation, providing a more reliable assessment of the model's explanatory power under the likely violations of classical assumptions. The robust F-statistics yield a value of **105.86**, which, although lower than the conventional F-statistics, remains highly significant at the 1% level.

The consistency of significance across both versions of the F-test, alongside p-values approaching zero, provides compelling evidence against the null hypothesis of no joint significance.

This confirms that at least some of the independent variables have a significant effect on ROA, validating the relevance of the predictors included in the models.

Turning to the independent variable of principal interest, the estimated coefficient for **ESG_SCORE** is **0.0224**, with a **p-value of 0.0014**, indicating a **statistically significant and positive relationship** between ESG performance and corporate profitability as measured by ROA.

This implies that, ceteris paribus, a one-unit increase in a firm's ESG score is associated with an approximate **0.024 percentage point increase in ROA**. While the effect is statistically significant, its magnitude is economically negligible, suggesting that the ESG performance, although positively related to corporate profitability, contributes only marginally in practical terms.

Moreover, the **95% confidence interval** for ESG_SCORE (ranging from **0.0087 to 0.0361**) excludes zero and remains sufficiently narrow to suggest a high degree of precision in the estimated effect, further reinforcing the reliability of the result.

These findings are in line with the theoretical foundations laid out in the previous chapter, particularly **stakeholder theory**, which emphasizes the strategic value of aligning corporate interests with those of broader stakeholder groups. From this perspective, investment in ESG initiatives is not merely a reputational or ethical choice but constitutes a mechanism through which firms can enhance operational efficiency, risk management, and long-term value creation. Nonetheless, it is important to stress that despite the statistical significance, the economic impact of ESG performance on profitability remains limited. This finding supports the growing view in the literature that **sustainability and profitability are not mutually exclusive**, but it also highlights that the tangible **financial benefits** derived from ESG improvements may **be very modest in scale**.

In addition to the primary variable of interest, several control variables exhibit statistically significant associations with corporate profitability, as measured by ROA, offering further insights into the determinants of firm performance. In particular,

- The coefficient for **LnAsset**, which serves as a proxy for firm size, is positive (**β = 0.5543**) but only marginally significant at the 10% level (**p-value = 0.0761**). While not meeting the conventional 5% threshold for statistical significance, this result nonetheless suggests a potential positive relationship between firm size and profitability. This finding aligns with theoretical expectations that larger firms may benefit from economies of scale, enhanced market power, and superior resource allocation capabilities, which can contribute to improved operational efficiency and profitability.

- The **Leverage** variable displays a negative and highly statistically significant coefficient (**β = -0.0567**, **p-value < 0.001**), indicating that greater reliance on debt financing is associated with reduced operating performance. This result is consistent with capital structure theories which assess that excessive debt levels can lead to increased financial risk, higher interest expenses, and reduced managerial flexibility, all of which may adversely impact firm profitability [Myers, S. C. (1984)]. *The Capital Structure Puzzle*. **The Journal of Finance**, 39(3), 575–592.

- A**sset turnover** is positive and significantly related to ROA (**β = 0.1215**, **p-value < 0.001**), highlighting the importance of operational efficiency in driving profitability. This result underscores that firms that utilize their assets more effectively are better positioned to generate higher returns, reflecting prudent management of resources and streamlined operations.

Taken together, these control variables provide meaningful contextual understanding and reinforce the robustness of the model, capturing key firm-level characteristics known to influence financial performance in both theoretical and empirical literature.

Table in Figure 23 reports the results of the PanelOLS regression model designed to assess the individual contributions of the three ESG dimensions: **Environmental (ENV_SCORE), Social (SOC_SCORE), and Governance (GOV_SCORE)**, to corporate profitability, as measured by **Return on Assets (ROA)**. The specification accounts for firm-level fixed effects to control for unobserved heterogeneity across entities.

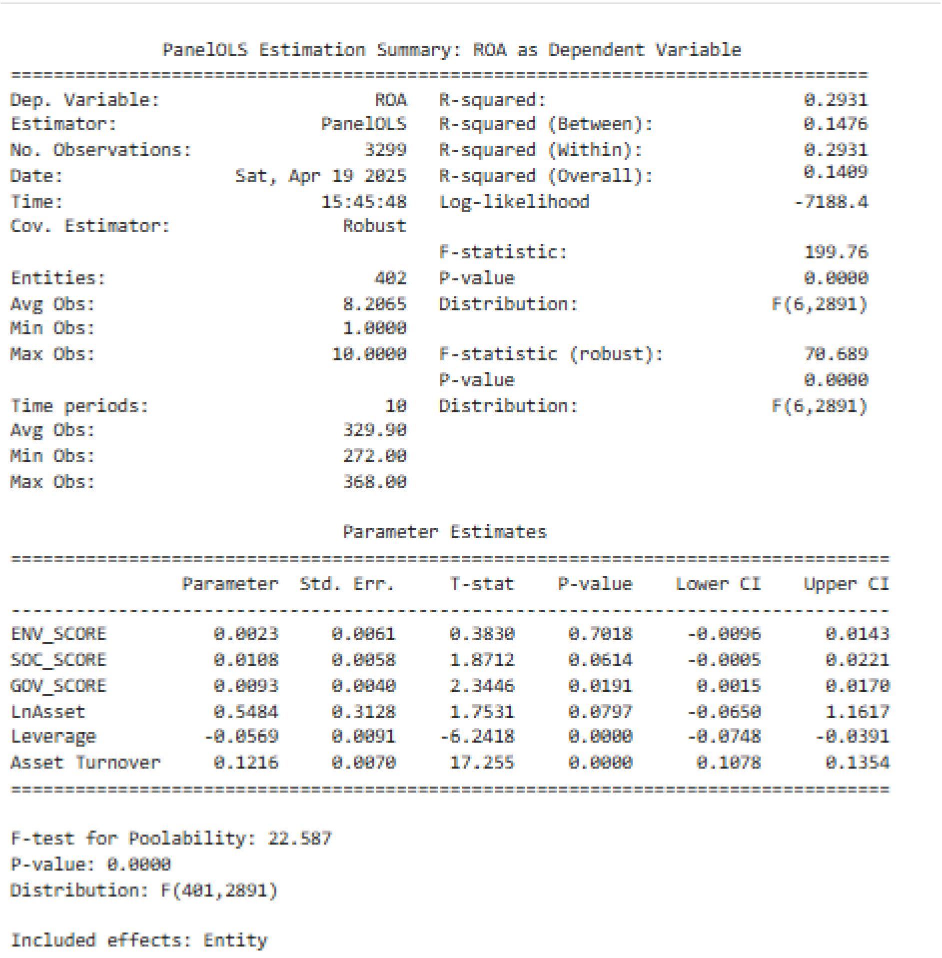
The **standard F-statistic** for the model is **199.76**, with an associated **p-value of 0.0000**, indicating that the explanatory variables, considered collectively, are statistically significant in explaining the variation in ROA. This result allows for the rejection of the null hypothesis that all slope coefficients are jointly equal to zero, thereby confirming the overall relevance of the model specification in capturing the determinants of corporate profitability within the panel.

To enhance the reliability of inference in the presence of potential **heteroskedasticity and intra-entity autocorrelation**, the model also reports a **robust F-statistic**. This adjusted statistic, calculated at **70.689** (p-value = 0.0000), remains highly significant, albeit lower than the conventional F-statistic. The use of robust standard errors offers a more conservative and dependable assessment of model significance under less restrictive econometric assumptions, thus reinforcing the robustness of the findings.

In sum, these diagnostic measures affirm the statistical validity and robustness of the regression model. The joint significance of the explanatory variables, confirmed through both standard and robust tests, supports the adequacy of the model in examining the differentiated effects of ESG components on firm-level profitability.

An examination of the estimated coefficients reveals important distinctions in the influence of each ESG pillar on firm profitability.

- The coefficient for **GOV_SCORE** is both **positive and statistically significant** (**β = 0.0093, p = 0.0191**), indicating that firms characterized by stronger governance structures tend to exhibit slightly higher levels of operational performance.
  This finding is consistent with the principles of **stakeholder theory** and corroborates prior empirical research suggesting that effective governance enhances **transparency, accountability, and managerial efficiency**, all of which are conducive to improved financial outcomes. Moreover, the corresponding **95% confidence interval excludes zero**, providing additional evidence of the precision and robustness of the estimate.



```
                 PanelOLS Estimation Summary: ROA as Dependent Variable
================================================================================
Dep. Variable:                    ROA   R-squared:                        0.2931
Estimator:                   PanelOLS   R-squared (Between):              0.1476
No. Observations:                3299   R-squared (Within):               0.2931
Date:                Sat, Apr 19 2025   R-squared (Overall):              0.1409
Time:                        15:45:48   Log-likelihood                   -7188.4
Cov. Estimator:                Robust
                                        F-statistic:                      199.76
Entities:                         402   P-value                           0.0000
Avg Obs:                       8.2065   Distribution:                  F(6,2891)
Min Obs:                       1.0000
Max Obs:                      10.0000   F-statistic (robust):             70.689
                                        P-value                           0.0000
Time periods:                      10   Distribution:                  F(6,2891)
Avg Obs:                       329.90
Min Obs:                       272.00
Max Obs:                       368.00

                                Parameter Estimates
==================================================================================
                Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
----------------------------------------------------------------------------------
ENV_SCORE          0.0023     0.0061     0.3830     0.7018     -0.0096      0.0143
SOC_SCORE          0.0108     0.0058     1.8712     0.0614     -0.0005      0.0221
GOV_SCORE          0.0093     0.0040     2.3446     0.0191      0.0015      0.0170
LnAsset            0.5484     0.3128     1.7531     0.0797     -0.0650      1.1617
Leverage          -0.0569     0.0091    -6.2418     0.0000     -0.0748     -0.0391
Asset Turnover     0.1216     0.0070     17.255     0.0000      0.1078      0.1354
==================================================================================

F-test for Poolability: 22.587
P-value: 0.0000
Distribution: F(401,2891)

Included effects: Entity
```

Figure 23 PanelOLS Estimation Results for the testing Model (2) [Personal Processing in Python].

- By contrast, **SOC_SCORE** demonstrates a **positive but only marginally significant** relationship with ROA (**β = 0.0108, p-value = 0.0614**). While this result does not meet the conventional threshold for statistical significance at the 5% level, it may nonetheless suggest a **positive association** between social responsibility initiatives, such as employee welfare, diversity, and community engagement, and firm profitability. However, caution is warranted in interpreting this result, as the confidence interval **marginally includes zero**, indicating a lower degree of certainty.

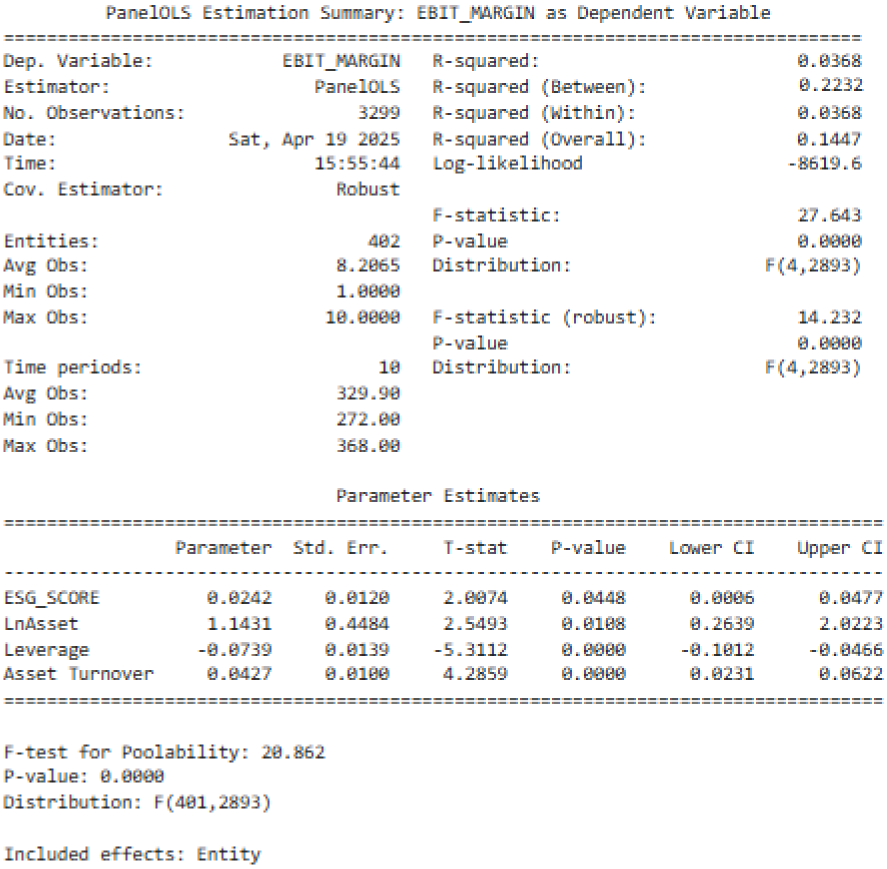
- Finally, the coefficient for **ENV_SCORE** is statistically insignificant (**β = 0.0023, p-value = 0.7018**), and its confidence interval clearly encompasses zero. This implies that, within the context of the present sample, **environmental performance does not exert a measurable impact on corporate profitability**. This may reflect the longer-term nature of environmental investments or a lack of market recognition for environmental initiatives in the short run.

Taken together, these findings suggest that the **governance dimension exerts the most substantial and reliable influence** on firm-level profitability among the three ESG components. They provide empirical support for the proposition that **ESG factors do not affect financial performance uniformly**, underscoring the need to disaggregate ESG scores when evaluating their economic implications.

The discussion now turns to Models 3 and 4, which mirror the structure of the previous regressions but differ in the choice of the dependent variable. Specifically, EBIT Margin is adopted here as a proxy for firm-level profitability, in contrast to ROA used in prior models. This measure allows for a more direct examination of operational efficiency, isolated from financial structure and taxation effects.

The standard **F-statistic** reported for the **Model 3** (Figure 24) is **27.643**, with an associated **p-value of 0.0000**, indicating that the set of explanatory variables is jointly significant in explaining variations in EBIT margin. As with Model 1, the null hypothesis that all regression coefficients are simultaneously equal to zero is firmly rejected, thereby affirming the overall adequacy of the model specification in capturing the determinants of profitability in the panel dataset.

Given concerns regarding the potential presence of heteroskedasticity and autocorrelation, a **robust version of the F-statistic** was also computed. The adjusted statistic, although lower in magnitude **(14.232), remains highly significant (p = 0.0000)**, reinforcing the robustness of the model's explanatory power even when standard assumptions are relaxed.



```
         PanelOLS Estimation Summary: EBIT_MARGIN as Dependent Variable
============================================================================
Dep. Variable:            EBIT_MARGIN   R-squared:                    0.0368
Estimator:                   PanelOLS   R-squared (Between):          0.2232
No. Observations:                3299   R-squared (Within):           0.0368
Date:                Sat, Apr 19 2025   R-squared (Overall):          0.1447
Time:                        15:55:44   Log-likelihood               -8619.6
Cov. Estimator:                Robust
                                        F-statistic:                  27.643
Entities:                         402   P-value                       0.0000
Avg Obs:                       8.2065   Distribution:              F(4,2893)
Min Obs:                       1.0000
Max Obs:                      10.0000   F-statistic (robust):         14.232
                                        P-value                       0.0000
Time periods:                      10   Distribution:              F(4,2893)
Avg Obs:                       329.90
Min Obs:                       272.00
Max Obs:                       368.00

                             Parameter Estimates
================================================================================
                Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
--------------------------------------------------------------------------------
ESG_SCORE          0.0242     0.0120     2.0074     0.0448      0.0006      0.0477
LnAsset            1.1431     0.4484     2.5493     0.0108      0.2639      2.0223
Leverage          -0.0739     0.0139    -5.3112     0.0000     -0.1012     -0.0466
Asset Turnover     0.0427     0.0100     4.2859     0.0000      0.0231      0.0622
================================================================================

F-test for Poolability: 20.862
P-value: 0.0000
Distribution: F(401,2893)

Included effects: Entity
```

Figure 24 PanelOLS Estimation Results for the testing Model (3) [Personal Processing in Python].

Turning to the interpretation of the estimated coefficients, the results are as follows:

- **LnAsset:** The coefficient for firm size is positive and statistically significant at the 1% level **(β = 1.1431; p = 0.0108).** While this variable exhibited only weak significance in Model 1 (ROA as the dependent variable), its impact is more pronounced when EBIT margin is employed. This finding suggests that larger firms, measured in terms of total assets, tend to achieve higher EBIT margins, likely due to economies of scale, operational efficiency, or greater market power, consistent with established theories of industrial organization [Schmalensee (1989)].

- **Leverage**: The leverage ratio exhibits a negative and highly statistically significant effect

(β = -0.0739; $p < 0.001$). This result aligns closely with that of Model 1 and supports the notion that increased debt levels are associated with reduced operating margins, potentially due to financial rigidity, increased risk exposure, or the burden of interest payments, which erode earnings before interest and taxes.

- **Asset Turnover:** This measure of operational efficiency shows a strongly positive and statistically significant association with EBIT margin **(β = 0.0427; p = 0.000).**
  The interpretation is intuitive: firms that utilize their asset base more effectively generate higher operating profits. Notably, this result is consistent across both specifications (ROA and EBIT margin), reinforcing the central role of asset efficiency in driving firm performance, as theorized in classical finance [Penman (2001)].

Overall, these findings underscore the stability and significance of the chosen control variables across different profitability metrics. The consistent direction and significance of the effects across models enhance the credibility of the results and suggest that the relationships observed are robust.

The regression estimates for the **ESG_SCORE** variable indicate a **positive and statistically significant association with firm profitability, as measured by the EBIT margin**.
Specifically, the coefficient estimate of **β = 0.0242**, with an associated **p-value of 0.0448**, implies that **a one-unit increase in the ESG score corresponds to a 0.024 percentage point increase in EBIT margin, ceteris paribus**. The result achieves statistical significance at the 5% level, suggesting that environmental, social, and governance performance exerts a modest but positive influence on firms' operational profitability.
In that specification, the ESG coefficient was β = 0.0224 (p-value = 0.0014), reinforcing the notion that ESG performance is positively correlated with firm profitability, regardless of the specific metric employed. The convergence of results across these two distinct performance indicators **strengthens the credibility of the observed relationship**.
Nonetheless, it is important to interpret the magnitude and implications of this effect with appropriate caution; the magnitude of the effect is economically negligible and of limited practical relevance. **The 95% confidence interval** for the ESG coefficient in the EBIT margin model is just above 0 on the lower limit **(0.0006 to 0.0477);** indicating that although the interval does not include zero, the effect size remains relatively small.

In sum, while ESG performance appears to have a statistically valid association with profitability, the actual economic benefit is so limited that it can be considered irrelevant in practical decision-making terms.

Figure 25 presents the results of **Model 4**, which investigates the relationship between firm-level profitability, proxied by the **EBIT margin**, and the **disaggregated components of ESG performance**: Environmental (ENV_SCORE), Social (SOC_SCORE), and Governance (GOV_SCORE). This model builds on prior specifications by decomposing the aggregate ESG measure to assess the relative explanatory power of each dimension.

The model exhibits an **F-statistic of 18.500** (p-value = 0.0000) and a **robust F-statistic of 9.5642** (p-value = 0.0000), indicating that the **set of explanatory variables is jointly statistically significant** at conventional levels. This confirms the **overall adequacy of the model specification**, even after adjusting for potential heteroskedasticity and serial correlation. However, when examining the **individual coefficient estimates** for each ESG sub-dimension, **none are statistically significant** at the 5% level. The findings are as follows:

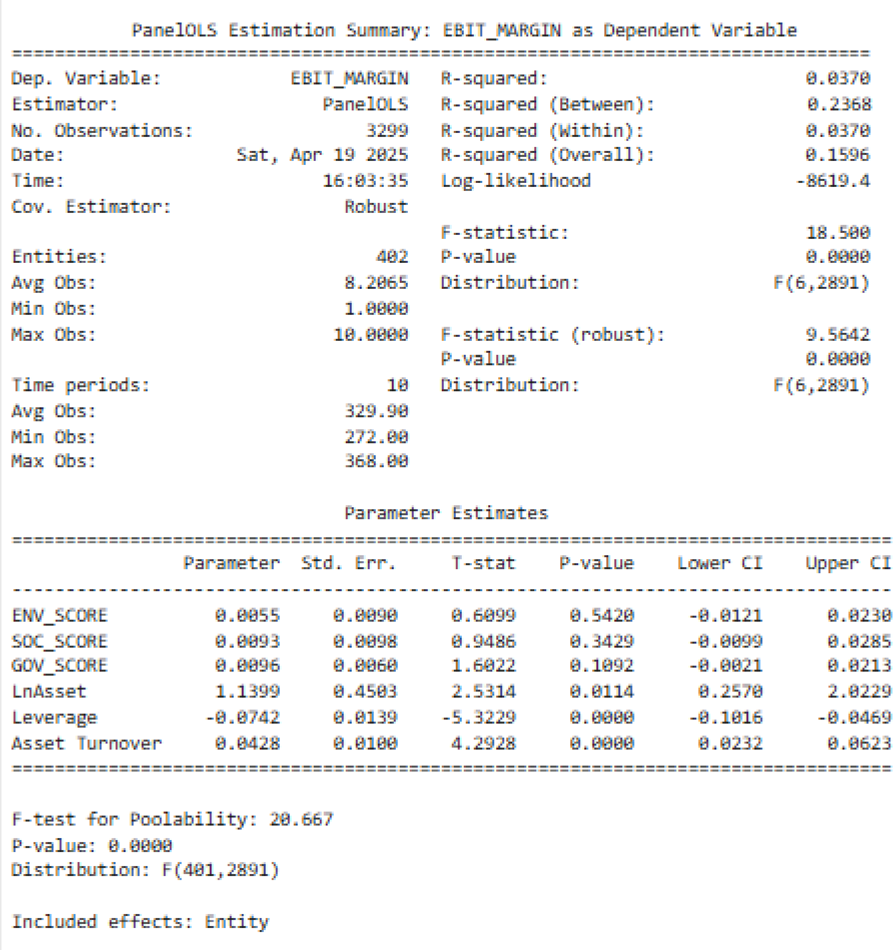
- **ENV_SCORE:** The estimated coefficient is **β = 0.0055** with a **p-value of 0.5420**, indicating a lack of statistical significance. The **95% confidence interval** [-0.0121, 0.0230] includes zero, suggesting that the effect of environmental performance on EBIT margin is statistically indistinguishable from zero. This may reflect the long-term nature of environmental investments, which often do not yield immediate financial returns or impact short-term operational profitability.

- **SOC_SCORE:** The coefficient for the social dimension is **β = 0.0093** (p-value = 0.3429), and the corresponding **95% confidence interval** [-0.0099, 0.0285] likewise contains zero.

  This suggests that improvements in social performance, such as employee well-being or community engagement, do not exert a statistically significant influence on firm profitability in the short run, consistent with existing literature emphasizing the delayed or intangible returns of social initiatives.

- **GOV_SCORE:** Although governance exhibits the largest coefficient (**β = 0.0096**), the **p-value of 0.1092** exceeds standard significance thresholds, and its **confidence interval** [-0.0021, 0.0213] also includes zero. While governance mechanisms are often cited as critical to financial performance, these results indicate that, within this specification, **no**

clear empirical relationship can be established between governance quality and EBIT margin.



```
                PanelOLS Estimation Summary: EBIT_MARGIN as Dependent Variable
================================================================================
Dep. Variable:             EBIT_MARGIN   R-squared:                        0.0370
Estimator:                    PanelOLS   R-squared (Between):              0.2368
No. Observations:                 3299   R-squared (Within):               0.0370
Date:                 Sat, Apr 19 2025   R-squared (Overall):              0.1596
Time:                         16:03:35   Log-likelihood                   -8619.4
Cov. Estimator:                 Robust
                                         F-statistic:                      18.500
Entities:                          402   P-value                           0.0000
Avg Obs:                        8.2065   Distribution:                  F(6,2891)
Min Obs:                        1.0000
Max Obs:                       10.0000   F-statistic (robust):             9.5642
                                         P-value                           0.0000
Time periods:                       10   Distribution:                  F(6,2891)
Avg Obs:                        329.90
Min Obs:                        272.00
Max Obs:                        368.00

                                 Parameter Estimates
==================================================================================
                Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
----------------------------------------------------------------------------------
ENV_SCORE          0.0055     0.0090     0.6099     0.5420     -0.0121      0.0230
SOC_SCORE          0.0093     0.0098     0.9486     0.3429     -0.0099      0.0285
GOV_SCORE          0.0096     0.0060     1.6022     0.1092     -0.0021      0.0213
LnAsset            1.1399     0.4503     2.5314     0.0114      0.2570      2.0229
Leverage          -0.0742     0.0139    -5.3229     0.0000     -0.1016     -0.0469
Asset Turnover     0.0428     0.0100     4.2928     0.0000      0.0232      0.0623
==================================================================================

F-test for Poolability: 20.667
P-value: 0.0000
Distribution: F(401,2891)

Included effects: Entity
```

Figure 25 PanelOLS Estimation Results for the testing Model (4) [Personal Processing in Python].

Although the magnitudes of the coefficients differ across dimensions, with the social and governance scores showing relatively larger estimates than the environmental score, all three indicators are **not statistically significant**.

The inclusion of zero within each 95% confidence interval implies that the **true population effect of each ESG component may be negative, null, or positive**.

Hence, the observed effects are **inconclusive** in both direction and magnitude.

Based on empirical findings, it can be inferred that while ESG (Environmental, Social, and Governance) practices are generally associated with enhanced operational profitability, this positive effect is not uniformly distributed across the individual pillars. Among the three dimensions, governance emerges as the most significant determinant of return on assets (ROA), underscoring the critical role that robust governance mechanisms play in fostering financial performance. Conversely, the environmental and social components, when analyzed independently, do not exhibit statistically significant relationships with profitability outcomes. Furthermore, in models utilizing EBIT margin as the dependent variable, the confidence intervals for the estimated coefficients of all three ESG pillars encompass zero. This suggests that the true effects of these variables in the population may be either positive, negative, or null, and thus no definitive conclusions can be drawn regarding their individual influence on profitability.

Overall, the analysis of the four statistical models does not provide conclusive evidence that the separate ESG components exert differential impacts on firm profitability.

Nevertheless, the aggregate ESG score demonstrates a positive association with corporate financial performance, supporting the notion that a comprehensive approach to ESG integration may yield greater benefits than consideration of the pillars in isolation.

This finding highlights the importance of evaluating ESG factors collectively when assessing their implications for financial outcomes.

4.2.3 Limitations and Potential Weaknesses of the Models

The regression analyses conducted so far appear to be well-structured, yielding statistically significant coefficient estimates within relatively narrow confidence intervals, at least for Models 1 and 3. Nevertheless, it remains crucial to further investigate the robustness of these models and undertake a more detailed evaluation of their overall performance.

Moreover, with respect to Models 2 and 4, it is necessary to consider the factors that may have contributed to the lack of statistical significance observed in some of the estimated coefficients. A series of diagnostic tests have already been addressed in the preceding sections.

In particular, the **Hausman test** has confirmed that a fixed effects model is preferable to a random effects model for the database at hand.

Potential multicollinearity among the independent variables was assessed by calculating the **Variance Inflation Factor (VIF)**, considering the within-firm demeaning transformation.

The analysis revealed no indication of problematic multicollinearity, thus supporting the reliability of the estimated coefficients.

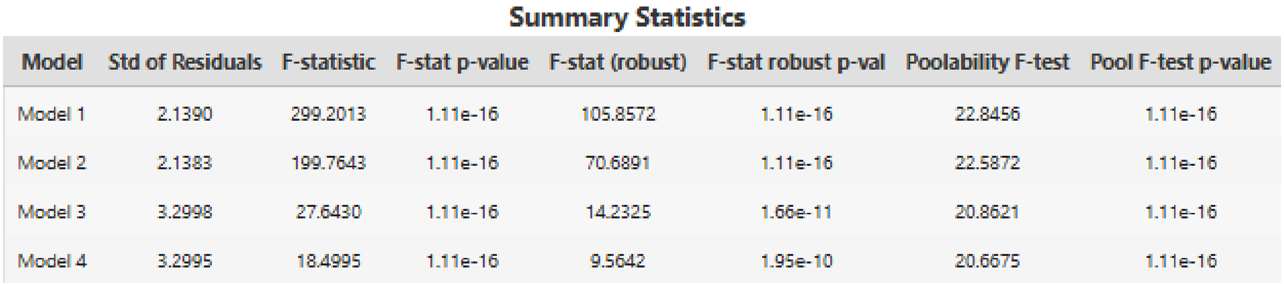
Additional validation for the adoption of a fixed effects approach is provided by the **F-test for Poolability**. This test assesses the suitability of a pooled model, which assumes homogeneity across all entities, versus a fixed effects model that explicitly accounts for unobserved individual heterogeneity [Baltagi, (2021)].

Specifically, the Poolability F-test is employed to determine the statistical relevance of entity-specific and time-invariant characteristics (α_i).

The null hypothesis theorises that all α_i are equal, implying the absence of significant individual differences, while the alternative hypothesis asserts that at least one α_i differs, indicating the presence of statistically significant unobserved heterogeneity.

The results of the Poolability F-test, as reported in Table in Figure 26, yield F-statistics ranging from 20.66 to 22.84, with associated p-values well below the 1% significance threshold.

These findings provide compelling evidence that firm-specific effects are statistically significant and should not be omitted. Accordingly, the fixed effects model is demonstrably preferable to the pooled OLS specification in this context.



Summary Statistics

Model	Std of Residuals	F-statistic	F-stat p-value	F-stat (robust)	F-stat robust p-val	Poolability F-test	Pool F-test p-value
Model 1	2.1390	299.2013	1.11e-16	105.8572	1.11e-16	22.8456	1.11e-16
Model 2	2.1383	199.7643	1.11e-16	70.6891	1.11e-16	22.5872	1.11e-16
Model 3	3.2998	27.6430	1.11e-16	14.2325	1.66e-11	20.8621	1.11e-16
Model 4	3.2995	18.4995	1.11e-16	9.5642	1.95e-10	20.6675	1.11e-16

Figure 26 PanelOLS Models Summary [Personal Processing in Python].

An essential final step in assessing the reliability of statistical inference involves examining the **distribution of the residuals**. This analysis offers a visual insight into the presence of autocorrelation and heteroskedasticity, which are assumed in the models under consideration; but it is also instrumental in detecting potential model misspecifications or omitted variable bias.

Figure 27 presents a summary of the Q-Q plots of the residuals for each model.

In **Model 1**, the residuals approximate a normal distribution with only minor deviations observable in the tails. Specifically, the residual distribution exhibits **modest leptokurtosis**, characterized by heavier tails alongside a central concentration. Given the relatively large panel dataset (comprising 402 firms over a ten-year period) such slight deviations from normality are generally acceptable for inference purposes. As Wooldridge (2010) and Baltagi (2005) note, in large samples, minor violations of normality do not materially affect the consistency or asymptotic properties of the estimators. Consequently, the inference drawn from Model 1 remains sufficiently robust.

In **Model 2**, which disaggregates the ESG score into its three constituent pillars (Environmental, Social, and Governance) the residual distribution closely takes after that of Model 1, though the **deviations from normality appear more pronounced**. A slightly elevated kurtosis is evident, alongside a **positive skewness**, which becomes clearer upon inspecting the histogram provided in Figure 36 Appendix C[3]. These distributional anomalies may pose a greater risk to the reliability of inferential procedures, particularly in the tails, where heavy-tailed behaviour can raise the variance of estimators. Moreover, the disaggregation of ESG into separate indicators might have introduced greater **instability in the explained component of the variance**, potentially contributing to the distortion observed in the residuals, despite the application of robust standard errors [Baltagi, (2005)].

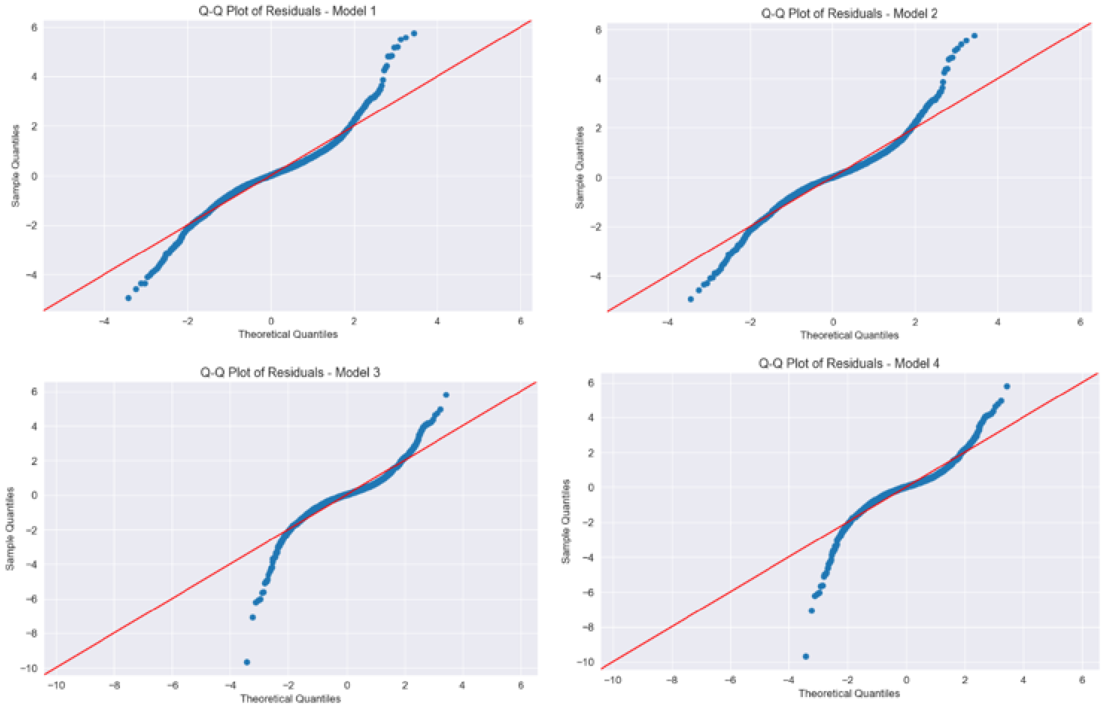


Figure 27 Q-Q Plots of Residuals for Models 1–4[Personal Processing in Python].

Turning to **Models 3 and 4**, where the dependent variable is **EBIT margin**, the departure from normality becomes more severe. In Model 3, the Q-Q plot reveals an "S"-shaped curvature, indicative of systematic deviations from theoretical quantiles. The histogram[4] confirms the

---

[3/4] Appendix C*: Figure 36 Probability distribution histogram. Source: Personal Processing* presence of **negative skewness**, suggesting that the distribution of residuals is asymmetric and biased toward lower values. Model 4 displays a similar distributional pattern, albeit with **even more pronounced tail behavior** and significant departures from normality.

These patterns suggest that both models may suffer from **misspecification**, potentially due to omitted variables, incorrect functional form, or the inherently volatile nature of EBIT margin, which is more sensitive to short- and medium-term fluctuations and **more exposed to outliers** compared to ROA [Greene, (2018)].

Model 4 combining EBIT margin with disaggregated ESG indicators appears to compound the distortion. While prior multicollinearity diagnostics indicated no severe collinearity between the three ESG pillars, **unexplained variance remains substantial**, likely absorbed by the residual component [Gujarati & Porter, (2009)]. This residual variance may reflect **non-linear relationships or omitted interactions** among the predictors.

In conclusion, **Model 1 appears to be the most statistically reliable and well-specified** among the four. The deviations from normality are minor and unlikely to undermine inferential validity. Conversely, **Models 2,3 and 4 demonstrate clear evidence of misspecification**, suggesting that a simple linear fixed effects framework may be insufficient to capture the complexity of the relationship between ESG performance and EBIT margin. In this case it may be better to better to explore non-linear transformations, include interaction terms or change the functional form of the model to correct for bias and improve fit.

Moreover, it is fundamental to conduct a more thorough investigation into the presence of heteroskedasticity and autocorrelation.

One of the key assumptions underlying the models employed in this analysis is that the error terms are not serially correlated over time within each cross-sectional firm.

This assumption is critical to ensure the validity of standard statistical inference.

However, in the context of panel data, particularly when applying fixed effects estimators, it is common to encounter serial correlation in the residuals [Baltagi, (2008)]. The presence of such autocorrelation can result in biased standard error estimates, ultimately compromising the reliability of hypothesis testing and confidence intervals.

To test the **autocorrelation of residuals** over time within the same company, this study uses the **Wooldridge test** [Wooldridge (2002)].Unlike the Durbin-Watson test, which is not well suited for panel data structures, the Wooldridge test accounts for fixed effects and remains robust in the presence of heteroskedasticity, a common characteristic in firm-level data [Drukker, (2003)]. This robustness makes it particularly appropriate for our dataset, which encompasses firms with potentially heterogeneous variance in their error terms across time and entities.

The Wooldridge test for first-order autocorrelation AR (1) is based on an auxiliary regression of the residuals from a fixed-effects model on their lagged values. Specifically, under the null hypothesis of no serial correlation, the residuals are assumed to be uncorrelated across time. The alternative hypothesis suggests the presence of first-order autocorrelation.

$$H0: \rho = 0$$

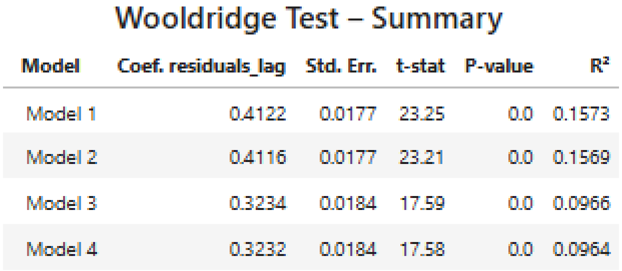
$$H1: \rho \neq 0$$

A statistically significant estimate of ρ provides evidence against the null, indicating serial correlation in the idiosyncratic errors [Wooldridge, (2002)].

The test statistic follows an F-distribution, and its implementation relies on estimating the following regression:

$$u_{it} = \rho * u_{it-1} + \varepsilon_{it}$$

where u_{it} are the residuals from the original fixed effects model.



Wooldridge Test – Summary

Model	Coef. residuals_lag	Std. Err.	t-stat	P-value	R^2
Model 1	0.4122	0.0177	23.25	0.0	0.1573
Model 2	0.4116	0.0177	23.21	0.0	0.1569
Model 3	0.3234	0.0184	17.59	0.0	0.0966
Model 4	0.3232	0.0184	17.58	0.0	0.0964

Figure 28 Wooldridge Test Results for Models 1-4 [Personal Processing in Python].

The outcomes of the Wooldridge test for the four models are presented in Table (Figure 28). In each model, the coefficient on the lagged residuals is significantly positive, providing strong statistical evidence of first order positive serial autocorrelation in the error terms.

This conclusion is reinforced by the high absolute values of the corresponding t-statistics and the associated p-values=0.00, leading to a clear rejection of the null hypothesis of no autocorrelation at conventional significance levels.

Furthermore, the estimated standard errors of the lagged residual coefficients are relatively small, indicating a high degree of precision in the ρ estimates. Interestingly, for models sharing the same dependent variable (ROA in Models 1 and 2, and EBIT Margin in Models 3 and 4) the standard errors are identical. This observation can likely be attributed to the similarity in model specification: regressing a corporate performance measure (ROA or EBIT Margin) on either the composite ESG score, or its disaggregated pillars yield highly comparable patterns in the residuals. Consequently, the auxiliary regression of the residuals on their lagged values produces nearly indistinguishable outcomes across such model pairs.

An inspection of the R^2 values from the auxiliary regressions suggests that a non-negligible proportion of the variance in the residuals is explained by their own temporal structure, further substantiating the presence of serial dependence. This form of autocorrelation is consistent with underlying economic dynamics: both ROA and EBIT Margin are firm-level operational indicators known to exhibit temporal continuity. The evidence therefore implies that the current model specifications may not adequately capture these dynamics, resulting in residuals that internalize part of these time patterns.

While such serial correlation is not uncommon in financial panel datasets, it may also point to potential model misspecification. For instance, the omission of relevant lagged dependent variables, dynamic adjustments, or unobserved permanent firm characteristics could contribute to the observed autocorrelation in the residuals [Baltagi, (2008); Wooldridge, (2010)].

Another essential diagnostic procedure in regression analysis is the evaluation of the homoskedasticity assumption. While the presence of heteroskedasticity does not make the coefficient estimates biased, it does lead to inefficient estimates and invalid standard errors, thereby compromising the reliability of hypothesis tests and confidence intervals [Wooldridge, (2010)].

Heteroskedasticity is particularly frequent in panel data structures, especially those involving economic or financial variables. In such datasets, error variance may differ across cross-sectional units as well as within the same enterprise over time, a phenomenon known respectively as cross-sectional and longitudinal heteroskedasticity [Baltagi, (2008)].

To account for this potential misspecification, the regression models in this study have been estimated using heteroskedasticity robust standard errors which produce consistent standard errors even in the presence of heteroskedasticity, thereby facilitating valid statistical inference [White, (1980)].

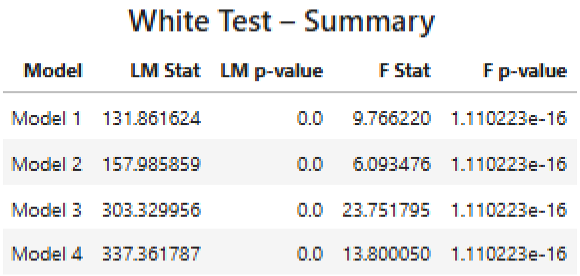
To effectively assess the **presence of heteroskedasticity**, the **White test** was applied to the within (after removing fixed effects) residuals of each mode. This diagnostic involves an auxiliary regression in which the squared residuals are regressed on the original regressors, their squares, and all possible cross-products.

$$u_{it}^{2} = \alpha_0 + \alpha_1 * x_{1i} + \ldots \alpha_k * x_{ki} + \alpha_{k+1} * x_{1i}^{2} + \ldots \alpha_{2k} * x_{ki}^{2} + \alpha_{2k} * x_{1i} x_{2i} + \ldots \varepsilon it$$

The null hypothesis (H_0) of the White test posits that the error variance is homoscedastic, whereas the alternative hypothesis (H_1) allows for error variance to be a function of one or more explanatory variables. Rejection of H_0 provides statistical evidence of heteroskedasticity in the residuals.

$$H0: \mathrm{Var}(ui|X) = \sigma 2$$
$$H1: \mathrm{Var}(ui|X) = \sigma 2(Xi)$$

The results of the heteroskedasticity tests are reported in Table in Figure 29. As illustrated, both the Lagrange Multiplier (LM) and F statistics uniformly reject the null hypothesis of homoskedasticity across all model specifications, with p-values effectively equal to zero. This provides strong statistical evidence of heteroskedasticity in the residuals of each fixed effects model, despite the application of robust covariance estimators.



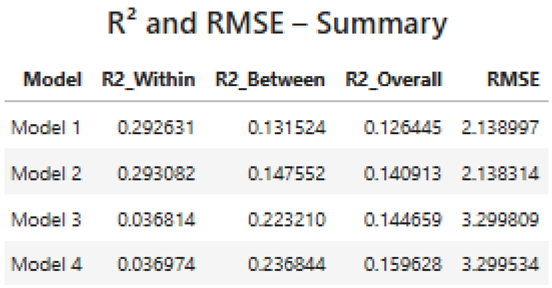
White Test – Summary

Model	LM Stat	LM p-value	F Stat	F p-value
Model 1	131.861624	0.0	9.766220	1.110223e-16
Model 2	157.985859	0.0	6.093476	1.110223e-16
Model 3	303.329956	0.0	23.751795	1.110223e-16
Model 4	337.361787	0.0	13.800050	1.110223e-16

Figure 29 White Test Results for Models 1-4 [Personal Processing in Python].

Notably, Models 3 and 4 (which employ EBIT margin as the dependent variable) exhibit substantially higher LM (303.33 - 337.36 vs 131.86 -157.98) and F statistics (23.75 – 13.80 vs 9.77 – 6.09) relative to Models 1 and 2, which are based on Return on Assets (ROA).

This indicates that the extent of heteroskedasticity is more pronounced in the EBIT margin models. These findings suggest that heteroskedasticity persists even after controlling for within entity fixed effects and robust standard errors, pointing to possible issues in model specification, such as omitted variables or inappropriate functional forms. While the use of robust standard errors preserves the validity of statistical inference, further model improvement is warranted, particularly concerning the modelling of the error variance structure and the pursuit of estimator efficiency in the panel data context.



R² and RMSE – Summary

Model	R2_Within	R2_Between	R2_Overall	RMSE
Model 1	0.292631	0.131524	0.126445	2.138997
Model 2	0.293082	0.147552	0.140913	2.138314
Model 3	0.036814	0.223210	0.144659	3.299809
Model 4	0.036974	0.236844	0.159628	3.299534

Figure 30 R^2 and RMSE Results for Models 1-4 [Personal Processing in Python].

To evaluate the overall performance and predictive power of the fixed effects models, key metrics such as the Root Mean Square Error (RMSE) and various forms of the R^2 statistic were examined [Greene, 2008; Wooldridge, 2010].

These metrics are summarized in Table in Figure 30.

In fixed effects models, the R^2_within is particularly relevant, as it reflects how much of the variation in the dependent variable is explained by the independent variables within firms over time. For Models 1 and 2, which use ROA as the dependent variable, the R^2_within is approximately 29% (0.2926 and 0.2930, respectively). While this may seem modest, such values are typical for panel data analysis, where a significant portion of variation tends to come from cross-sectional differences or unobserved firm-level heterogeneity. Interestingly, there is a minimal difference in explanatory power between using the aggregate ESG score (Model 1) and its disaggregated components (Model 2), implying that the three ESG pillars (Environmental, Social, and Governance) contribute similarly to firm performance as measured by ROA.

By contrast, Models 3 and 4, which use EBIT Margin as the dependent variable, exhibit considerably lower within-firm explanatory power, with R^2_within values of only about 3.7% (0.0368 and 0.0369). This suggests that ESG indicators are less effective at explaining short-term operational profitability than they are at explaining broader, longer-term performance like ROA. Even when ESG is broken down into its components in Model 4, the gain in explanatory power over the aggregate ESG score in Model 3 is negligible. This finding reinforces the conclusion that the added complexity of separating ESG into pillars does not significantly improve model performance in this context.

The R^2_between, which measures how well the models explain differences across firms, tells a slightly different story. Here, Models 3 and 4 outperform Models 1 and 2, with R^2_between values around 22% to 23.7% for EBIT Margin, compared to 13.2% and 14.8% for ROA. This indicates that ESG scores are more effective in capturing persistent firm-level differences in EBIT Margin than in ROA. These cross-sectional differences likely reflect more stable and structural attributes, such as corporate governance, strategic orientation, or organizational culture, that are better aligned with EBIT Margin than with longer-term profitability measures.

The R^2_overall, which combines both within and between effects, also improves modestly when disaggregated ESG components are used. For example, Model 2 shows an R^2_overall of 14.1% compared to 12.6% for Model 1, while Model 4 yields 16.0% compared to 14.5% for Model 3. However, these improvements are marginal, suggesting that the composite ESG score already captures most of the explanatory content of its components. Moreover, multicollinearity among the ESG dimensions may limit the marginal contribution of each pillar when they are included together in the model [Wooldridge, 2010].

The RMSE values further highlight the comparative strength of the ROA models. Models 1 and 2 report substantially lower RMSEs (around 2.14) than Models 3 and 4 (approximately 3.30), indicating that the ROA-based models offer more accurate in-sample predictions. This provides additional evidence that ROA is a more suitable financial performance indicator when investigating the effect of ESG factors.

In conclusion, Models 1 and 2 perform better overall, with stronger explanatory power and greater predictive accuracy than Models 3 and 4. The disaggregation of ESG scores contributes little to model performance, raising questions about their individual influence. The relatively low R^2 values across all models, however, may point to broader issues such as model misspecification, omitted variable bias, or the limitations of assuming linear relationships. Unobserved factors like macroeconomic shocks, managerial skill, or other latent firm-specific dynamics could be influencing the results, warranting further refinement of the modeling approach.

5. ALTERNATIVE APPROACH

5.1 Beyond Linearity in Panel Data Analysis

In the previous paragraph, it was emphasized that relationships between variables are not necessarily linear, and as such, a linear regression model may not always be the most appropriate methodological choice. Traditional regression approaches, including linear regression with fixed effects, often face limitations in handling nonlinearity, heterogeneity, and intricate variable interactions.

Moreover, in panel datasets, where observations are recorded over time for multiple firms in a cross-sectional framework, issues of residual heteroscedasticity frequently arise.

Firms within the dataset may exhibit significant differences in terms of size, industry, and financial structure, while variations in regulatory environments across different countries further contribute to heterogeneity. Additionally, firms do not behave uniformly across different time periods. Macroeconomic shocks, such as financial crises or global events like the COVID-19 pandemic, may exert asymmetric effects, impacting certain firms more than others.

These factors contribute to fluctuations in residuals, complicating traditional regression-based modelling.

5.1.1 Limitations of Classical Machine Learning Models

To address these challenges, machine learning techniques based on decision trees represent a promising alternative as they are robust to heteroskedasticity.

Nevertheless, classical machine learning (ML) models, such as Random Forest, XGBoost, or Support Vector Machines, are not well-suited for the objectives of the present study.

These algorithms are primarily designed for predictive tasks and are not inherently structured to address causal inference questions, namely, to establish whether and to what extent an explanatory variable exerts an influence on a dependent outcome.

While traditional machine learning models are well-versed in capturing complex nonlinear associations and optimizing predictive accuracy, they often fall short in distinguishing mere statistical correlations from genuine causal relationships [Pearl, 2009; Athey & Imbens, (2017)]. Specifically, they operate under the assumption that the relationship between input (X) and output (Y) is stable, observable, and free of endogeneity. Consequently, they do not adequately account for key issues such as confounding variables, reverse causality, or omitted variable bias, which are central to causal analysis.

Moreover, these models are ill-equipped to handle panel data structures, where observations are embedded across both cross-sectional units and time. Standard ML algorithms do not naturally incorporate temporal dependencies within entities, or control for unobserved heterogeneity, two critical dimensions of longitudinal data [Hsiao, (2014)].

As demonstrated in previous sections, the presence of residual autocorrelation suggests that outcomes at time *t* are systematically influenced by values observed at t_{-1}, t_{-2}, and so forth. Ignoring this structure may lead to biased and inconsistent estimates.

To address this, we adopt a **Double Machine Learning (DML)** approach, in which classical econometric techniques (i.e. OLS) are complemented by machine learning algorithms to estimate nuisance components, such as control variables, in a more flexible and data-driven manner.

5.1.2 Introducing the DML Framework

Double Machine Learning (DML), as introduced by Chernozhukov et al. (2018), represents a methodological framework that integrates machine learning techniques with traditional econometric tools to enable valid post-selection inference in the presence of high-dimensional nuisance parameters. DML provides estimators that are robust and asymptotically normal, even when the nuisance components are estimated via flexible machine learning methods [Chernozhukov et al., (2018); Clarke & Polselli, (2024)].

DML with his hybrid approach favors flexibility over analytical tractability, reflecting a broader trend in empirical research toward models that combine the predictive power of machine learning with the inferential rigor of econometrics [Knaus, (2022); Bach et al., (2024)].

The estimation strategy employed follows **a partially linear model** in the spirit of Robinson (1988), wherein the treatment effect is estimated linearly while the potentially complex relationship between covariates and outcomes is captured non-parametrically via machine learning methods.

Specifically, Random Forests [Breiman, (2001)] and Gradient Boosting Machines [Friedman, (2001)] are utilized to model the non-linear component, due to their capacity to flexibly capture high-order interactions and reduce bias in nonparametric estimation. For the linear component, standard OLS estimators are retained to identify the parameter of interest.

The empirical strategy adopted in this study is grounded in the theoretical underpinnings of the **Frisch–Waugh–Lovell (FWL) theorem** [Frisch & Waugh, (1933); Lovell, (1963)].

This theorem formalizes the intuition that, in a multiple linear regression model, it is possible to isolate the effect of a variable of interest by first removing the influence of control variables from both the dependent variable and the regressor of interest. In other words, the effect of a treatment variable D on an outcome Y, conditional on a set of covariates X, can be consistently estimated by first projecting both Y and D onto the space orthogonal to X, and then regressing the resulting residuals [Lovell (1963)].

Formally, consider the linear regression model:

$$Y_{it}=\beta 1+\beta 2\cdot D_{it}+\beta 3\cdot X_{it}+\mu_{it}$$

where:

- Y_{it} denotes the dependent variable for unit *i* at time *t*,
- D_{it} is the treatment or variable of interest,
- X_{it} is a vector of control covariates,
- μ_{it} is the idiosyncratic error term.

The FWL procedure entails the following steps:

1. **Regress Y_{it} on X_{it}** to obtain residuals $\mu_{it}^{Y}= Y_{it} - E\,[Y_{it} \mid X_{it}]$
2. **Regress D_{it} on X_{it}** to obtain residuals $\mu_{it}^{D}= D_{it} - E\,[D_{it} \mid X_{it}]$
3. **Regress μ_{it}^{Y} on μ_{it}^{D}** :

$$\mu_{it}^{Y} = \beta 2\cdot\mu_{it}^{D} + \varepsilon_{it}$$

The coefficient β2 obtained from this residual regression is algebraically identical to that estimated from the full model including D and X, thereby confirming the result of the FWL theorem [Tiffen (2025)].

While the FWL framework provides a foundational tool for partial regression analysis under linearity assumptions, more flexible approaches have been proposed to address the presence of **nonlinearities** and **model misspecification**.

In particular, Robinson (1988) extends the classical linear model by proposing a **partially linear model** of the form:

$$Y_{it}=\theta\cdot D_{it}+g(X_{it})+\mu_{it}$$

where:

- θ is the parameter of interest (e.g., the Average Partial Effect of D on Y),
- g $(\cdot)$ is an unspecified, potentially nonlinear function of covariates (X_{it})
- μ_{it} is an error term with $E[\mu it|Dit,Xit]=0$

In this setting, the effect of the treatment variable D_{it} is estimated parametrically, while the nuisance function $g(X_{it})$ is estimated nonparametrically.
The FWL principle is preserved in spirit, but instead of using OLS to partial out X_{it} a **nonparametric regression** technique is employed, thus allowing for greater flexibility in capturing complex interactions and nonlinear relationships between covariates and outcomes [Clarke et al. (2023)].

5.1.3 DML in a Panel Data Context

Building upon Robinson's contribution, Clarke and Polselli (2024) propose an extension of the partially linear model within the **Double Machine Learning (DML)** framework, specifically adapted to panel data with fixed effects. The baseline specification is given as:

$$Y_{it}= \alpha_i + \theta\cdot D_{it}+ g(X_{it})+\mu_{it}$$

Where:

- Y_{it} denotes the dependent variable for entity *i* at time *t*.
- α_i captures entity-specific fixed effects, accounting for unobserved, time-invariant heterogeneity.
- D_{it} represents the treatment variable.
- θ is the parameter of interest, interpreted as the **Average Partial Effect (APE)** of the treatment variable D_{it} on Y_{it}.
- $g(X_{it})$ is a potentially nonlinear and unknown function of covariates.
- X_{it} includes observed, time-varying control variables.
- μ_{It} is an idiosyncratic error term, assumed to satisfy $E\ [\mu_{it}|D_{it},X_{it},\alpha_i] =0$.

In this approach, machine learning algorithms are utilized to estimate the **nuisance components** $gY(X_{it})$ and $gD(X_{it})$ and while preserving the interpretability and inferential properties of the treatment effect estimate θ.
Following Clarke and Polselli, we estimate the nuisance components using flexible machine learning algorithms (e.g., LASSO, Random Forest, XGBoosting).
To mitigate overfitting and to avoid the estimation of the nuisance component to interfere with the estimation of treatment effects, we implement K-fold cross-fitting as proposed by Chernozhukov et al. (2018). This involves partitioning the sample into K folds, estimating the nuisance functions on $K-1$ folds, and evaluating the residuals on the held-out fold. This process is repeated such that each fold is used once for evaluation, enhancing robustness and ensuring valid inference.

Given the panel structure of the data and the need to address unobserved heterogeneity, we first estimate the nuisance functions using the full dataset (i.e., without fixed-effects transformation) and subsequently apply the within-group transformation (i.e., fixed effects estimator) to their predicted values.
This ordering is essential: when $g(\cdot)$ is nonlinear, the $Q(g(X_{it})) \neq g(Q(X_{it}))$ generally holds. Applying WG before estimation would introduce functional misspecification [Clarke & Polselli, (2024)].

Formally, we compute the residualized variables as:

$$\mu_{it}^{Y} = Q(Y_{it}) - Q(\widehat{gY}(X_{it}))$$

$$\mu_{it}^{D} = Q(D_{it}) - Q(\widehat{gD}(X_{it}))$$

Here, $Q(\cdot)$ denotes the **WG transformation** that demeans each variable over time within each cross-sectional unit:

$$Q(gY(X_{it})) = gY(X_{it}) - \frac{1}{T}\sum_{t=1} gY(X_{it})$$

$$Q(gD(X_{it})) = gD(X_{it}) - \frac{1}{T}\sum_{t=1} gD(X_{it})$$

This step removes unit-specific fixed effects α_i, addressing unobserved time-invariant heterogeneity while preserving the functional structure captured by the machine learning models.

While the **First Difference (FD)** estimator is theoretically more robust to nonlinear transformations, it was not employed here due to practical limitations: FD requires discarding an observation per unit, potentially reducing sample size and amplifying noise in the differenced variables. The WG approach, by contrast, retains the full sample and remains suitable when differencing is infeasible.

Once the fixed effects have been eliminated and the residualized variables μ_{it}^{Y} and μ_{it}^{D} are constructed, we proceed to the **second-stage regression**:

$$\mu_{it}^{Y}=\theta\cdot\mu_{it}^{D}+\varepsilon_{it}$$

The coefficient $\hat{\theta}$ which represents the **Average Partial Effect (APE)** of the treatment variable D_{it} on the outcome Y_{it}Y is then computed as:

$$\hat{\theta} = \frac{\sum_{it} \mu_{it}^{D} \cdot \mu_{it}^{Y}}{\sum_{it} {\mu_{it}^{D}}^{2}}$$

This estimator is consistent and asymptotically normal under regularity conditions outlined by Chernozhukov et al. (2018), even in the presence of high-dimensional or nonlinear covariates, due to the orthogonalization and sample-splitting properties intrinsic to the DML framework.

5.2 DML-Based Analysis of the STOXX Europe 600

5.2.1 DML Models specification

To further investigate the relationship between explanatory variables and firm profitability measured by Return on Assets (ROA) and EBIT margin, we specify a partially linear panel data model with fixed effects within the DML framework.

The model components are defined as follows:

$$Y_{it}= \alpha_{i} + \theta\cdot D_{it}+ g(X_{it})+\mu_{it}$$

- Yit denotes the profitability outcome (ROA or EBIT margin) for firm *i* at time *t*.
- α_i captures firm-specific fixed effects, accounting for time-invariant, unobserved heterogeneity.

- D_{it} represents ESG variables, either the overall ESG score or its components (Environmental, Social, and Governance).
- θ (or $\theta 1$, $\theta 2$, $\theta 3$) denotes the treatment parameter(s) of interest, interpreted as the Average Partial Effect (APE) of ESG performance on profitability.
- $g(X_{it})$ is a potentially nonlinear function of observed covariates.
- X_{it} includes time-varying control variables such as firm leverage, size, and asset turnover.
- μ_{it} is an idiosyncratic error term, assumed to satisfy $E\,[\mu_{it} \mid D_{it}, X_{it}, \alpha_i] = 0$.

The following four model specifications are estimated:

5) $ROA_{it} = \alpha_i + \theta^* \; ESG\_SCORE_{it} + g(X_{it}) + \mu_{it}$
6) $ROA_{it} = \alpha_i + \theta_1^* \; ENV\_SCORE_{it} + \theta_2^* \; SOC\_SCORE_{it} + \theta_3^* \; GOV\_SCORE_{it} + g(X_{it}) + \mu_{it}$
7) $EBIT\ MARGIN_{it} = \alpha_i + \theta^* \; ESG\_SCORE_{it} + g(X_{it}) + \mu_{it}$
8) $EBIT\ MARGIN_{it} = \alpha_i + \theta_1^* \; ENV\_SCORE_{it} + \theta_2^* \; SOC\_SCORE_{it} + \theta_3^* \; GOV\_SCORE_{it} + g(X_{it}) + \mu_{it}$

These specifications enable identification of the causal impact of ESG performance, both aggregated and disaggregated, on firm profitability, while flexibly controlling for nonlinearities in control variables and unobserved firm-level effects.

It is reasonable to assume that corporate profitability is influenced not only by the observed firm characteristics X_{it} but also by latent, time-invariant factors specific to each company α_i. Similarly, ESG engagement decisions may be shaped by these same unobserved characteristics. Therefore, the ESG scores themselves can be viewed as functions of both observable features (e.g., size, leverage, industry classification) and firm-specific attributes that are constant over time.

5.2.2 DML Model Results

The findings of the DML analysis for the models presented above are shown in the following section. This analysis was conducted using a consistent sample of 402 companies selected from the STOXX Europe 600 index.

Table in Figure 31 summarizes the estimation results of Double Machine Learning Model 5, where the dependent variable is the Return on Assets (ROA). The coefficient denoted as **Theta** captures the estimated causal effect of the ESG_SCORE on ROA, thus reflecting the impact of firms' ESG performance on their operating profitability.

In the first specification, where gD(X_{it}) is estimated using **Lasso** and gY(X_{it}) is learned via **XGBoost**, the estimated coefficient is **0.0181** with a **p-value of 0.0047**. In the second specification, where gD(X_{it}) is still estimated with **Lasso** but gY(X_{it}) is learned using **Random Forest**, the coefficient increases slightly to **0.0218** with a p-value of **0.0008**.

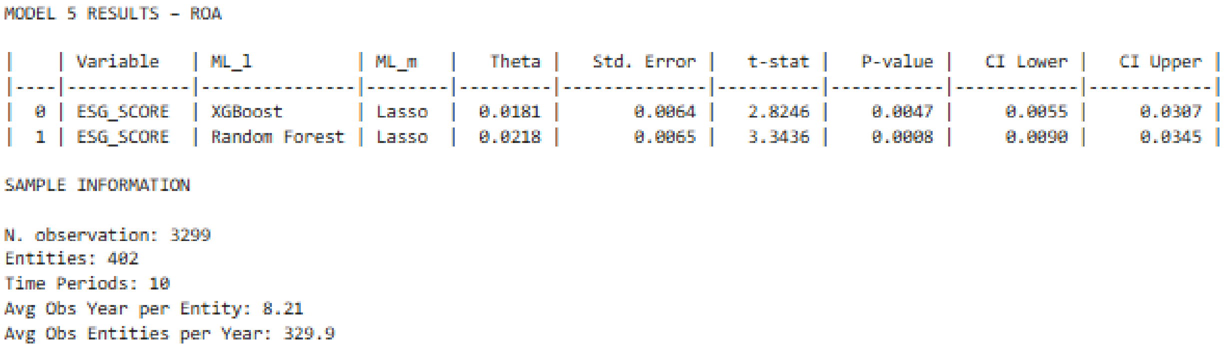
In both cases, **the estimated effect is modest in magnitude but positive and statistically significant at the 1% level.**

Accordingly, the corresponding 95% confidence intervals are entirely above zero, pointing to a statistically significant and robust positive association between ESG performance and firm profitability.

From an economic standpoint, the results indicate that a one-point increase in a firm's ESG_SCORE is associated with a relatively modest improvement in ROA ranging from 0.0181% to 0.0218%. Although the magnitude of this effect is small, it is statistically significant and suggests that ESG performance does contribute positively to firm profitability. In industries where average ROA typically ranges between 2% and 10%, such incremental gains may not be transformative individually but can accumulate over time or on a scale to produce meaningful financial benefits.

These findings imply that while ESG initiatives may not drastically boost profitability, they can still play a role in enhancing long-term operational performance potentially by improving risk mitigation, strengthening stakeholder relationships, or increasing organizational resilience.

Furthermore, the consistency of the estimated effects across two different machine learning methods (XGBoost and Random Forest) lends credibility to the results, suggesting that the observed relationship between ESG and ROA is robust and not driven by model-specific artifacts. For firms and investors, this underscores that ESG investments can offer small but stable economic benefits.



```
MODEL 5 RESULTS - ROA

|    | Variable   | ML_l          | ML_m   |   Theta |   Std. Error |   t-stat |   P-value |   CI Lower |   CI Upper |
|----|------------|---------------|--------|---------|--------------|----------|-----------|------------|------------|
|  0 | ESG_SCORE  | XGBoost       | Lasso  |  0.0181 |       0.0064 |   2.8246 |    0.0047 |     0.0055 |     0.0307 |
|  1 | ESG_SCORE  | Random Forest | Lasso  |  0.0218 |       0.0065 |   3.3436 |    0.0008 |     0.0090 |     0.0345 |

SAMPLE INFORMATION

N. observation: 3299
Entities: 402
Time Periods: 10
Avg Obs Year per Entity: 8.21
Avg Obs Entities per Year: 329.9
```

Figure 31 DML Estimation Results for the testing Model (5) [Personal Processing in Python].

Table in Figure 32 reports the results of Model 6, where ROA continues to serve as the dependent variable.

In this specification, the analysis focuses on assessing whether and to what extent the three distinct pillars of ESG environmental, social, and governance differentially affect firms' economic performance.

Among the three dimensions, **GOV_SCORE** exhibits the most pronounced and statistically robust association with ROA. The estimated coefficients are **0.0128 and 0.0126** when the treatment model is learned via XGBoost and Random Forest, respectively, **with both estimates significant at the 1% level (p-value = 0.0002).** Economically, this suggests that a one-point improvement in a firm's governance score is linked to a modest, yet consistently positive increase in profitability. While small in magnitude, such gains can become meaningful in aggregate, especially for large firms or over extended periods.

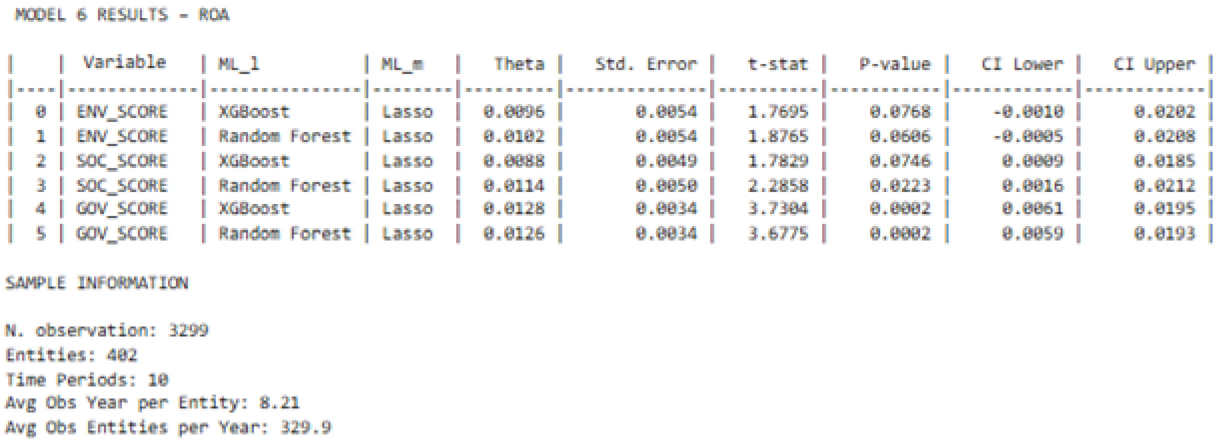
SOC_SCORE also demonstrates a positive relationship with ROA, albeit of smaller magnitude and weaker significance. **The estimated effect is 0.0083 (p = 0.0746) using XGBoost and 0.0114 (p = 0.0223) using Random Forest**. While the former is marginally significant at the 10% level, the latter achieves statistical significance at the conventional 5% threshold. In both cases, the estimated coefficients are positive, and the confidence intervals do not cross zero, suggesting a robust effect. From an economic perspective, these results imply that social initiatives may contribute to marginal profitability gains, potentially by improving brand reputation, employee satisfaction, or customer loyalty.

In contrast, the environmental component, represented by **ENV_SCORE**, appears to exert the weakest and least statistically reliable influence on ROA. Although the estimated coefficients are positive in both specifications **(0.0096 and 0.0102),** the corresponding p-values **(0.0768**

and 0.0606) exceed standard significance thresholds, and the 95% confidence intervals include zero even if only marginally.

Taken together, the results point to a robust and economically meaningful positive effect of corporate governance quality on firms' operating profitability. Social performance also appears to contribute positively, though to a lesser extent and with lower statistical confidence. Environmental performance, however, does not show a statistically significant impact within the current model specification and sample context.
This indicates a lack of statistical precision and suggests that, within this model and sample, environmental performance does not have a clear or measurable effect on operating profitability.

In summary, the findings point to a **modest but economically relevant role for corporate governance in enhancing firm profitability**, with social performance offering additional, though less robust, benefits. The environmental pillar, while important from a broader sustainability standpoint, does not appear to translate into immediate financial gains in the current context.



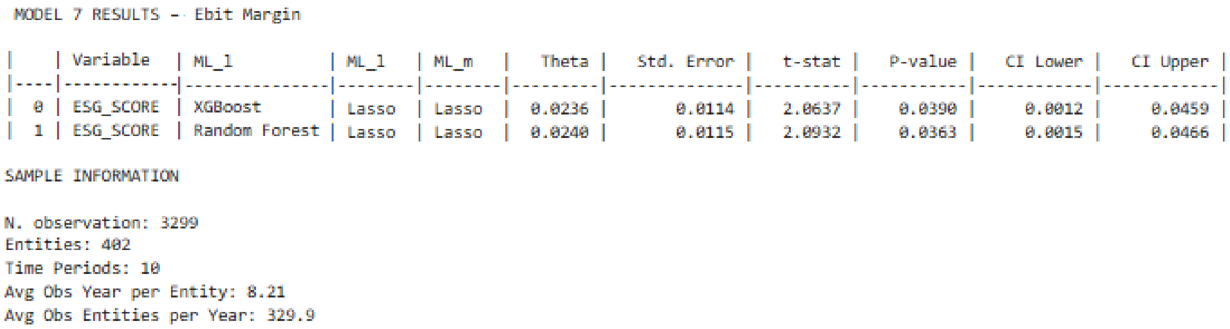
```
MODEL 6 RESULTS - ROA

|    | Variable   | ML_1          | ML_m   |   Theta |   Std. Error |   t-stat |   P-value |   CI Lower |   CI Upper |
|----|------------|---------------|--------|---------|--------------|----------|-----------|------------|------------|
|  0 | ENV_SCORE  | XGBoost       | Lasso  |  0.0096 |       0.0054 |   1.7695 |    0.0768 |    -0.0010 |     0.0202 |
|  1 | ENV_SCORE  | Random Forest | Lasso  |  0.0102 |       0.0054 |   1.8765 |    0.0606 |    -0.0005 |     0.0208 |
|  2 | SOC_SCORE  | XGBoost       | Lasso  |  0.0088 |       0.0049 |   1.7829 |    0.0746 |     0.0009 |     0.0185 |
|  3 | SOC_SCORE  | Random Forest | Lasso  |  0.0114 |       0.0050 |   2.2858 |    0.0223 |     0.0016 |     0.0212 |
|  4 | GOV_SCORE  | XGBoost       | Lasso  |  0.0128 |       0.0034 |   3.7304 |    0.0002 |     0.0061 |     0.0195 |
|  5 | GOV_SCORE  | Random Forest | Lasso  |  0.0126 |       0.0034 |   3.6775 |    0.0002 |     0.0059 |     0.0193 |

SAMPLE INFORMATION

N. observation: 3299
Entities: 402
Time Periods: 10
Avg Obs Year per Entity: 8.21
Avg Obs Entities per Year: 329.9
```

Figure 32 DML Estimation Results for the testing Model (6) [Personal Processing in Python].

Turning to the analysis of firm profitability as proxied by the **EBIT Margin**, Table in Figure 33 presents the results from Model 7, which aims to assess whether the ESG_SCORE exerts a positive causal impact on firms' operating margins. Consistent with the findings from Model 5 (ROA as dependent variable), the results in Model 7 indicate that the ESG_SCORE is positively and significantly associated with the EBIT Margin.

Specifically, the estimated Theta is **0.0236** when the gY(X_{it}) parameters are learned **via XGBoost (p-value = 0.0390), and 0.0240 when Random Forest is used (p-value = 0.0363)**. In both specifications, **the effect is statistically significant at the 5% level.**
Although the point estimates of Theta differ slightly between the two specifications, their close similarity enhances confidence in the robustness and reliability of the estimated effects.
Moreover, the 95% confidence intervals in both cases are entirely positive, reinforcing the statistical reliability and consistency of the results.
However, from an economic standpoint, the size of the estimated effect is extremely modest.
A one-point increase in ESG_SCORE is associated with only about 0.024 percentage of improvement in EBIT Margin; an effect so small that it is unlikely to have meaningful financial implications on its own. While statistically non-negligible, the economic significance of such a marginal increase is limited, especially when compared to typical fluctuations in operating margins driven by broader market, operational, or strategic factors.

These findings suggest that, although higher ESG performance implies slightly better operating margins, the financial impact is negligible in practical terms. The observed relationship may reflect minor efficiency gains or reduced risk exposure, but the magnitude of the effect implies that ESG engagement, while potentially beneficial in other dimensions, does not materially enhance operating margins in the short run.



```
 MODEL 7 RESULTS - Ebit Margin

|    | Variable   | ML_l          | ML_l   | ML_m   |   Theta |   Std. Error |   t-stat |   P-value |   CI Lower |   CI Upper |
|----|------------|---------------|--------|--------|---------|--------------|----------|-----------|------------|------------|
|  0 | ESG_SCORE  | XGBoost       | Lasso  | Lasso  |  0.0236 |       0.0114 |   2.0637 |    0.0390 |     0.0012 |     0.0459 |
|  1 | ESG_SCORE  | Random Forest | Lasso  | Lasso  |  0.0240 |       0.0115 |   2.0932 |    0.0363 |     0.0015 |     0.0466 |

SAMPLE INFORMATION

N. observation: 3299
Entities: 402
Time Periods: 10
Avg Obs Year per Entity: 8.21
Avg Obs Entities per Year: 329.9
```

Figure 33 DML Estimation Results for the testing Model (7) [Personal Processing in Python].

Lastly, Table in Figure 34 presents the results of Model 8, which investigates the potential differential effects of the three **ESG pillars** on short-term firms' operating performance, measured by the **EBIT Margin.**

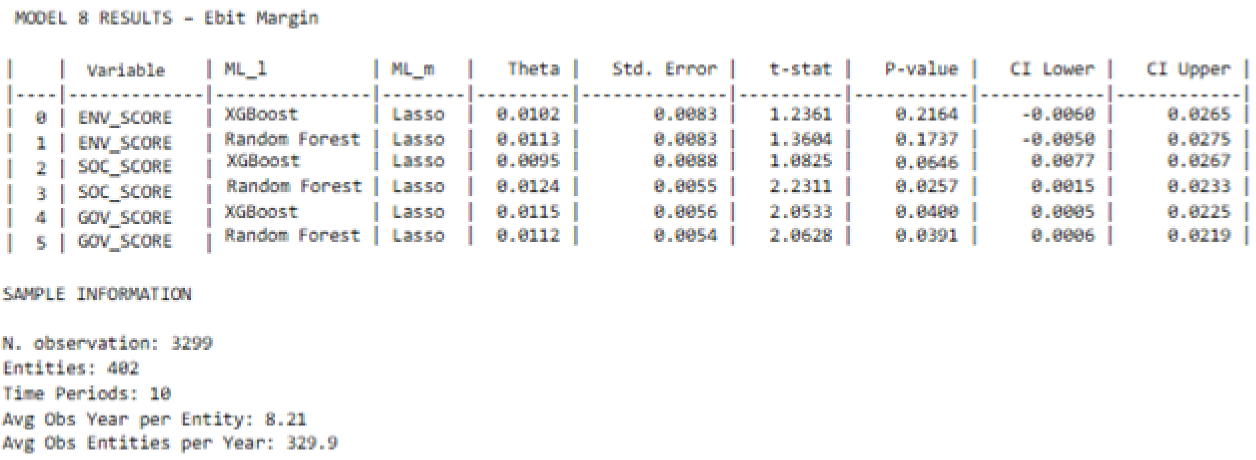
With respect to **ENV_SCORE**, the estimated coefficients (**Theta**) are positive across specifications **(0.0102 and 0.0113) but not statistically significant, with p-values ranging**

between 0.2164 and 0.1737. The resulting 95% confidence intervals include zero in both cases, indicating that the effect of environmental performance on EBIT Margin is not statistically distinguishable from zero.
Economically, the estimated impact is also minimal, suggesting that improvements in environmental performance are unlikely to yield any meaningful change in operating margins in the short term, at least within this sample.

The results for **SOC_SCORE** are mixed. In the specification using XGBoost, the estimated effect is not statistically significant at the 5% level **(Theta= 0.0095; p = 0.0646)**, whereas in the specification using Random Forest, the coefficient is positive and statistically significant **(Theta = 0.0124; p = 0.0257)**. Although both estimates are positive and their confidence intervals lie entirely above zero, the **economic impact remains a very small** gain of just under 1.25 basis points in EBIT Margin per one-point increase in the social score.
This implies that even where statistical evidence exists, financial **relevance is limited**, and the practical implications for firm profitability are likely negligible in isolation.

Finally, the estimates for **GOV_SCORE** display consistent and statistically significant positive effects across both specifications. The estimated coefficients are **0.0115 (p = 0.0400) using XGBoost and 0.0112 (p = 0.0391) using Random Forest**, with 95% confidence intervals that do not cross zero. However, like the other dimensions, the **economic effect is modest** just over one basis point increase in EBIT Margin per unit improvement in governance quality.
While statistically reliable, such a small effect is **unlikely to materially influence financial outcomes** at the firm level on its own.

In summary, while the analysis provides **statistical evidence** that corporate governance quality and to a lesser extent, social responsibility has a positive relationship with firms' operating margins, the **magnitude of the effects is modest across all ESG dimensions**. This indicates that, from a financial perspective, ESG-related improvements may contribute to slightly better operational performance, but the **short-term economic impact is slight**, particularly for the environmental component, which shows no significant effect.



```
MODEL 8 RESULTS - Ebit Margin

|    | Variable     | ML_l          | ML_m   |   Theta |   Std. Error |   t-stat |   P-value |   CI Lower |   CI Upper |
|----|-------------|---------------|--------|---------|--------------|----------|-----------|------------|------------|
|  0 | ENV_SCORE    | XGBoost       | Lasso  |  0.0102 |       0.0083 |   1.2361 |    0.2164 |    -0.0060 |     0.0265 |
|  1 | ENV_SCORE    | Random Forest | Lasso  |  0.0113 |       0.0083 |   1.3604 |    0.1737 |    -0.0050 |     0.0275 |
|  2 | SOC_SCORE    | XGBoost       | Lasso  |  0.0095 |       0.0088 |   1.0825 |    0.0646 |     0.0077 |     0.0267 |
|  3 | SOC_SCORE    | Random Forest | Lasso  |  0.0124 |       0.0055 |   2.2311 |    0.0257 |     0.0015 |     0.0233 |
|  4 | GOV_SCORE    | XGBoost       | Lasso  |  0.0115 |       0.0056 |   2.0533 |    0.0400 |     0.0005 |     0.0225 |
|  5 | GOV_SCORE    | Random Forest | Lasso  |  0.0112 |       0.0054 |   2.0628 |    0.0391 |     0.0006 |     0.0219 |

SAMPLE INFORMATION

N. observation: 3299
Entities: 402
Time Periods: 10
Avg Obs Year per Entity: 8.21
Avg Obs Entities per Year: 329.9
```

Figure 34 DML Estimation Results for the testing Model (8) [Personal Processing in Python].

5.2.3 Comparison of the Two Methodological Approaches and Concluding Remarks

In this study, two distinct statistical methodologies employed fixed-effects Panel OLS regressions and Double Machine Learning (DML) models, to investigate the relationship between Environmental, Social, and Governance (ESG) scores and corporate profitability, as measured by Return on Assets (ROA) and Ebit margin.

The aim was to assess the existence of an ESG-profitability nexus under both linear and flexible non-linear framework.

Following a thorough assessment of the respective strengths and limitations of each methodological approach, a comparative analysis was conducted to determine their suitability for examining the research question.

From a statistical standpoint, the fixed-effects Panel OLS models (Models 1–4) indicate a statistically **significant and positive association** between the aggregate ESG score and both ROA (Model 1) and EBIT margin (Model 3). These results suggest that firms demonstrating higher ESG engagement generally achieve better operational performance.

The DML models (Models 5–8), on the other hand, introduce non-linearity and allow for high-dimensional interactions among ESG dimensions and control variables. These models confirmed the key findings of the Panel OLS regressions: the ESG score positively and significantly affects both ROA (Model 5) and EBIT margin (Model 7). Notably, the **consistency of coefficient signs and statistical significance across methodologies supports the reliability of the linear model**, while also revealing **additional nuances** through the DML approach.

When disaggregating ESG components, the differences between methodologies become more evident. Panel OLS models (2 and 4) identified **social (SOC_SCORE)** and **governance (GOV_SCORE)** factors as significant drivers of ROA, while **no individual ESG component significantly impacted EBIT margin**. In contrast, DML models (6 and 8) not only confirmed the **statistical insignificance of the environmental pillar (ENV_SCORE)** but also highlighted **stronger and more precise effects of the social and governance scores on both profitability measures**, particularly EBIT margin. This reflects DML's ability to isolate the ESG variables from control variables enhances the precision of estimation by mitigating omitted variable bias, a limitation often presents in linear models.

Moreover, the DML methodology offers the advantage of accommodating intricate interactions between ESG dimensions and other covariates, and the use of cross-fitting further mitigates overfitting, thereby improving the reliability of causal inference.

Economically, the governance dimension consistently exerts the most robust positive effect on profitability. This underscores the value of corporate governance in enhancing transparency, risk management, and investor confidence, factors that translate into financial gains.

The social dimension also shows growing importance, especially in the DML framework, pointing to the profitability benefits of employee welfare, stakeholder engagement, and community investment. The environmental component, however, appears less impactful in both methodologies, potentially due to long-term cost structures or regulatory compliance costs that do not yield immediate financial returns.

From a methodological standpoint, the Double Machine Learning approach offers several distinct advantages over traditional econometric models.

First, DML is specifically designed to **capture complex, non-linear relationships and high-order interactions** among ESG components and firm-specific characteristics. This flexibility allows the model to reflect the true underlying structure of the data, which is often more intricate than what linear models can accommodate.

Second, and most importantly, DML enhances the **credibility of causal inference** in high-dimensional environments. By leveraging machine learning tools to control for a large set of covariates without overfitting, it isolates the effect of ESG on profitability more effectively than standard regression methods. This makes DML particularly powerful in contexts where traditional models may struggle to disentangle causal effects from mere associations.

These features make DML especially well-suited for contemporary corporate finance research, where **ESG-performance relationships are often nuanced, dynamic, and firm-specific**. By integrating robustness, flexibility, and causal clarity, DML provides a more reliable foundation for assessing the true economic impact of sustainability practices.

In conclusion, both methodologies confirm **Hypothesis 1** developed in Chapter 3: ESG performance has a **positive and significant effect on corporate profitability**. Moreover, the nuanced results from the DML approach support **Hypothesis 2**: **the individual ESG pillars impact profitability in distinct ways**, with governance and social dimensions playing more substantial roles than the environmental one.

This comparative analysis reaffirms the importance of combining traditional econometric rigor with modern machine learning tools. Fixed-effects models offer interpretability and transparency, while DML enhances robustness and granularity. Taken together, the findings contribute to the literature on ESG and firm performance, demonstrating that **strong ESG practices, particularly in governance and social areas, can be strategic assets in driving profitability**.

CONCLUSION AND FUTURE RESEARCH OPPORTUNIES

This thesis investigates the relationship between Environmental, Social, and Governance (ESG) performance and corporate profitability, employing both traditional econometric methodologies and machine learning techniques. The empirical analysis focuses on firms listed in the STOXX Europe 600 index and provides generally supportive evidence for the hypothesis that higher ESG scores are associated with improved operational performance, particularly in the long term, as measured by Return on Assets (ROA).

Although the results are statistically significant, their economic magnitude is relatively modest. This suggests that, while ESG integration may yield financial advantages when effectively embedded into corporate strategies, these benefits appear to be limited in scope. Moreover, the observed heterogeneity across ESG pillars and performance metrics indicates that the impact of ESG components on profitability is not uniform; certain dimensions may exert a more pronounced effect than others.

Despite the methodological rigor adopted, several limitations constrain the external validity and interpretative power of the findings:

- The dataset is restricted to companies included in the STOXX Europe 600 index; a benchmark primarily composed of large, well-established firms. This geographical and qualitative focus limits the generalizability of the results to firms operating in other regions, particularly emerging markets, and to small and medium-sized enterprises.
- The analysis is limited by the availability of ESG scores, which are reported for only a subset of firms. This introduces potential selection bias, as companies that disclose ESG information may systematically differ from those that do not.
- The study period (2013–2022) encompasses the COVID-19 pandemic, an economic shock that may have distorted typical firm performance. The pandemic's exceptional nature may have temporarily muted profitability metrics, thereby complicating the interpretation of causality.

Looking ahead, several avenues for future research emerge from this work:

- The increasing regulatory pressure in the European Union is expected to improve ESG data coverage across firm sizes and geographies. This will enable more comprehensive and representative empirical investigations.
- Future research could benefit from extending the temporal scope of analysis or treating periods of economic crisis, such as the COVID-19 pandemic, separately. This would facilitate the differentiation between structural and cyclical effects of ESG performance.

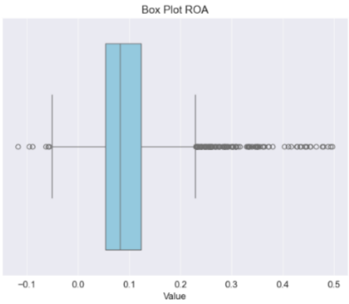
- Incorporating additional measures, such as free cash flow, cost of capital, or risk-adjusted returns, could offer a more nuanced understanding of how ESG affects various dimensions of corporate performance.
- Given the variability in ESG materiality across industries, disaggregated sectoral analyses may yield more granular and practically relevant insights. Customizing ESG indicators to reflect industry-specific sustainability challenges may further enhance analytical precision.
- Emerging methodologies such as Double Machine Learning (DML) offer gifted advances in addressing endogeneity and high-dimensionality challenges. Although still under development, these techniques are rapidly evolving and becoming increasingly powerful, accurate, and applicable to complex empirical settings. Their implementation may uncover dynamic and potentially non-linear relationships between ESG practices and firm performance.

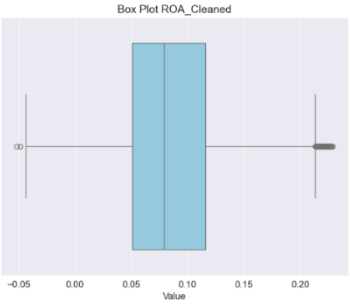
In conclusion, while the present study supports a generally positive association between ESG performance and corporate profitability among major European firms, caution must be exercised due to limitations in scope, data, and timeframe. Nevertheless, ongoing improvements in ESG reporting and empirical tools are likely to enhance future research capacity in disentangling the multifaceted link between corporate sustainability and financial outcomes.

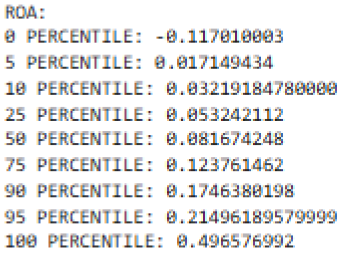
Appendix

A: Box Plot of the independent variables of the Regression Analysis.

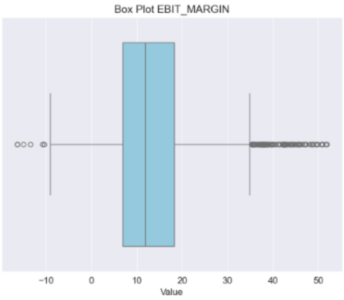
Source: Personal Computation in Python

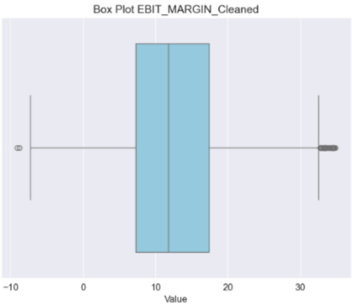


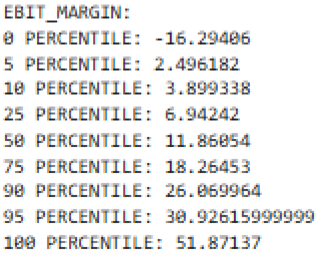




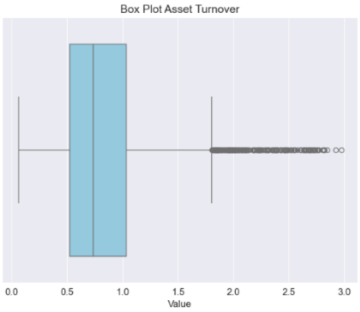
```
ROA:
0 PERCENTILE: -0.117010003
5 PERCENTILE: 0.017149434
10 PERCENTILE: 0.03219184780000
25 PERCENTILE: 0.053242112
50 PERCENTILE: 0.081674248
75 PERCENTILE: 0.123761462
90 PERCENTILE: 0.1746380198
95 PERCENTILE: 0.21496189579999
100 PERCENTILE: 0.496576992
```

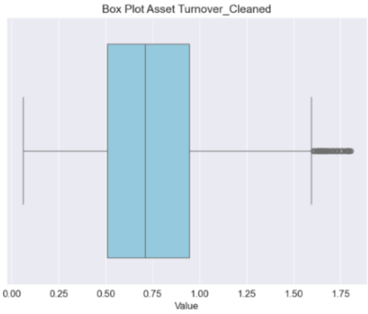


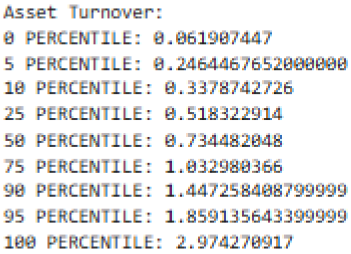




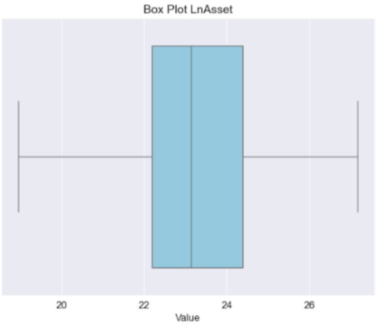
```
EBIT_MARGIN:
0 PERCENTILE: -16.29406
5 PERCENTILE: 2.496182
10 PERCENTILE: 3.899338
25 PERCENTILE: 6.94242
50 PERCENTILE: 11.86054
75 PERCENTILE: 18.26453
90 PERCENTILE: 26.069964
95 PERCENTILE: 30.92615999999
100 PERCENTILE: 51.87137
```

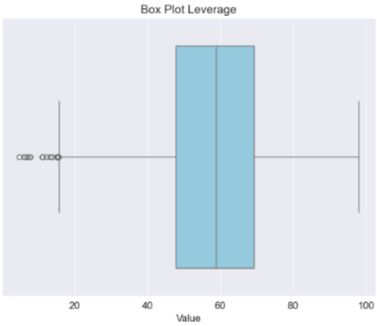


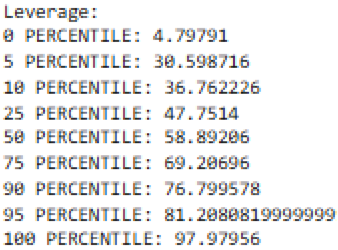




```
Asset Turnover:
0 PERCENTILE: 0.061907447
5 PERCENTILE: 0.2464467652000000
10 PERCENTILE: 0.3378742726
25 PERCENTILE: 0.518322914
50 PERCENTILE: 0.734482048
75 PERCENTILE: 1.032980366
90 PERCENTILE: 1.447258408799999
95 PERCENTILE: 1.859135643399999
100 PERCENTILE: 2.974270917
```

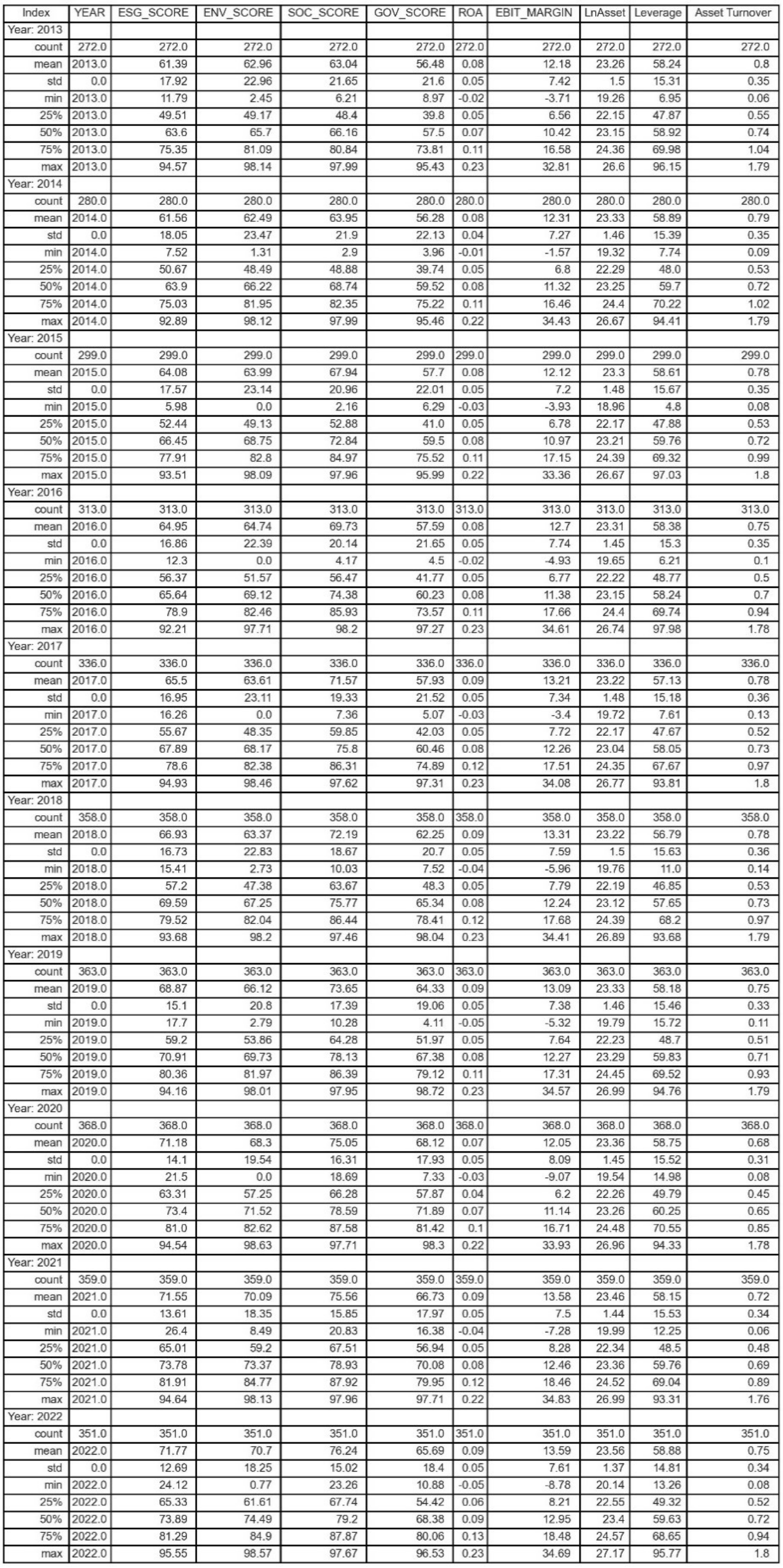






```
Leverage:
0 PERCENTILE: 4.79791
5 PERCENTILE: 30.598716
10 PERCENTILE: 36.762226
25 PERCENTILE: 47.7514
50 PERCENTILE: 58.89206
75 PERCENTILE: 69.20696
90 PERCENTILE: 76.799578
95 PERCENTILE: 81.2080819999999
100 PERCENTILE: 97.97956
```

B. Figure 35: Descriptive Statistics per Year. Source Personal Processing



Index	YEAR	ESG_SCORE	ENV_SCORE	SOC_SCORE	GOV_SCORE	ROA	EBIT_MARGIN	LnAsset	Leverage	Asset Turnover
Year: 2013										
count	272.0	272.0	272.0	272.0	272.0	272.0	272.0	272.0	272.0	272.0
mean	2013.0	61.39	62.96	63.04	56.48	0.08	12.18	23.26	58.24	0.8
std	0.0	17.92	22.96	21.65	21.6	0.05	7.42	1.5	15.31	0.35
min	2013.0	11.79	2.45	6.21	8.97	-0.02	-3.71	19.26	6.95	0.06
25%	2013.0	49.51	49.17	48.4	39.8	0.05	6.56	22.15	47.87	0.55
50%	2013.0	63.6	65.7	66.16	57.5	0.07	10.42	23.15	58.92	0.74
75%	2013.0	75.35	81.09	80.84	73.81	0.11	16.58	24.36	69.98	1.04
max	2013.0	94.57	98.14	97.99	95.43	0.23	32.81	26.6	96.15	1.79
Year: 2014										
count	280.0	280.0	280.0	280.0	280.0	280.0	280.0	280.0	280.0	280.0
mean	2014.0	61.56	62.49	63.95	56.28	0.08	12.31	23.33	58.89	0.79
std	0.0	18.05	23.47	21.9	22.13	0.04	7.27	1.46	15.39	0.35
min	2014.0	7.52	1.31	2.9	3.96	-0.01	-1.57	19.32	7.74	0.09
25%	2014.0	50.67	48.49	48.88	39.74	0.05	6.8	22.29	48.0	0.53
50%	2014.0	63.9	66.22	68.74	59.52	0.08	11.32	23.25	59.7	0.72
75%	2014.0	75.03	81.95	82.35	75.22	0.11	16.46	24.4	70.22	1.02
max	2014.0	92.89	98.12	97.99	95.46	0.22	34.43	26.67	94.41	1.79
Year: 2015										
count	299.0	299.0	299.0	299.0	299.0	299.0	299.0	299.0	299.0	299.0
mean	2015.0	64.08	63.99	67.94	57.7	0.08	12.12	23.3	58.61	0.78
std	0.0	17.57	23.14	20.96	22.01	0.05	7.2	1.48	15.67	0.35
min	2015.0	5.98	0.0	2.16	6.29	-0.03	-3.93	18.96	4.8	0.08
25%	2015.0	52.44	49.13	52.88	41.0	0.05	6.78	22.17	47.88	0.53
50%	2015.0	66.45	68.75	72.84	59.5	0.08	10.97	23.21	59.76	0.72
75%	2015.0	77.91	82.8	84.97	75.52	0.11	17.15	24.39	69.32	0.99
max	2015.0	93.51	98.09	97.96	95.99	0.22	33.36	26.67	97.03	1.8
Year: 2016										
count	313.0	313.0	313.0	313.0	313.0	313.0	313.0	313.0	313.0	313.0
mean	2016.0	64.95	64.74	69.73	57.59	0.08	12.7	23.31	58.38	0.75
std	0.0	16.86	22.39	20.14	21.65	0.05	7.74	1.45	15.3	0.35
min	2016.0	12.3	0.0	4.17	4.5	-0.02	-4.93	19.65	6.21	0.1
25%	2016.0	56.37	51.57	56.47	41.77	0.05	6.77	22.22	48.77	0.5
50%	2016.0	65.64	69.12	74.38	60.23	0.08	11.38	23.15	58.24	0.7
75%	2016.0	78.9	82.46	85.93	73.57	0.11	17.66	24.4	69.74	0.94
max	2016.0	92.21	97.71	98.2	97.27	0.23	34.61	26.74	97.98	1.78
Year: 2017										
count	336.0	336.0	336.0	336.0	336.0	336.0	336.0	336.0	336.0	336.0
mean	2017.0	65.5	63.61	71.57	57.93	0.09	13.21	23.22	57.13	0.78
std	0.0	16.95	23.11	19.33	21.52	0.05	7.34	1.48	15.18	0.36
min	2017.0	16.26	0.0	7.36	5.07	-0.03	-3.4	19.72	7.61	0.13
25%	2017.0	55.67	48.35	59.85	42.03	0.05	7.72	22.17	47.67	0.52
50%	2017.0	67.89	68.17	75.8	60.46	0.08	12.26	23.04	58.05	0.73
75%	2017.0	78.6	82.38	86.31	74.89	0.12	17.51	24.35	67.67	0.97
max	2017.0	94.93	98.46	97.62	97.31	0.23	34.08	26.77	93.81	1.8
Year: 2018										
count	358.0	358.0	358.0	358.0	358.0	358.0	358.0	358.0	358.0	358.0
mean	2018.0	66.93	63.37	72.19	62.25	0.09	13.31	23.22	56.79	0.78
std	0.0	16.73	22.83	18.67	20.7	0.05	7.59	1.5	15.63	0.36
min	2018.0	15.41	2.73	10.03	7.52	-0.04	-5.96	19.76	11.0	0.14
25%	2018.0	57.2	47.38	63.67	48.3	0.05	7.79	22.19	46.85	0.53
50%	2018.0	69.59	67.25	75.77	65.34	0.08	12.24	23.12	57.65	0.73
75%	2018.0	79.52	82.04	86.44	78.41	0.12	17.68	24.39	68.2	0.97
max	2018.0	93.68	98.2	97.46	98.04	0.23	34.41	26.89	93.68	1.79
Year: 2019										
count	363.0	363.0	363.0	363.0	363.0	363.0	363.0	363.0	363.0	363.0
mean	2019.0	68.87	66.12	73.65	64.33	0.09	13.09	23.33	58.18	0.75
std	0.0	15.1	20.8	17.39	19.06	0.05	7.38	1.46	15.46	0.33
min	2019.0	17.7	2.79	10.28	4.11	-0.05	-5.32	19.79	15.72	0.11
25%	2019.0	59.2	53.86	64.28	51.97	0.05	7.64	22.23	48.7	0.51
50%	2019.0	70.91	69.73	78.13	67.38	0.08	12.27	23.29	59.83	0.71
75%	2019.0	80.36	81.97	86.39	79.12	0.11	17.31	24.45	69.52	0.93
max	2019.0	94.16	98.01	97.95	98.72	0.23	34.57	26.99	94.76	1.79
Year: 2020										
count	368.0	368.0	368.0	368.0	368.0	368.0	368.0	368.0	368.0	368.0
mean	2020.0	71.18	68.3	75.05	68.12	0.07	12.05	23.36	58.75	0.68
std	0.0	14.1	19.54	16.31	17.93	0.05	8.09	1.45	15.52	0.31
min	2020.0	21.5	0.0	18.69	7.33	-0.03	-9.07	19.54	14.98	0.08
25%	2020.0	63.31	57.25	66.28	57.87	0.04	6.2	22.26	49.79	0.45
50%	2020.0	73.4	71.52	78.59	71.89	0.07	11.14	23.26	60.25	0.65
75%	2020.0	81.0	82.62	87.58	81.42	0.1	16.71	24.48	70.55	0.85
max	2020.0	94.54	98.63	97.71	98.3	0.22	33.93	26.96	94.33	1.78
Year: 2021										
count	359.0	359.0	359.0	359.0	359.0	359.0	359.0	359.0	359.0	359.0
mean	2021.0	71.55	70.09	75.56	66.73	0.09	13.58	23.46	58.15	0.72
std	0.0	13.61	18.35	15.85	17.97	0.05	7.5	1.44	15.53	0.34
min	2021.0	26.4	8.49	20.83	16.38	-0.04	-7.28	19.99	12.25	0.06
25%	2021.0	65.01	59.2	67.51	56.94	0.05	8.28	22.34	48.5	0.48
50%	2021.0	73.78	73.37	78.93	70.08	0.08	12.46	23.36	59.76	0.69
75%	2021.0	81.91	84.77	87.92	79.95	0.12	18.46	24.52	69.04	0.89
max	2021.0	94.64	98.13	97.96	97.71	0.22	34.83	26.99	93.31	1.76
Year: 2022										
count	351.0	351.0	351.0	351.0	351.0	351.0	351.0	351.0	351.0	351.0
mean	2022.0	71.77	70.7	76.24	65.69	0.09	13.59	23.56	58.88	0.75
std	0.0	12.69	18.25	15.02	18.4	0.05	7.61	1.37	14.81	0.34
min	2022.0	24.12	0.77	23.26	10.88	-0.05	-8.78	20.14	13.26	0.08
25%	2022.0	65.33	61.61	67.74	54.42	0.06	8.21	22.55	49.32	0.52
50%	2022.0	73.89	74.49	79.2	68.38	0.09	12.95	23.4	59.63	0.72
75%	2022.0	81.29	84.9	87.87	80.06	0.13	18.48	24.57	68.65	0.94
max	2022.0	95.55	98.57	97.67	96.53	0.23	34.69	27.17	95.77	1.8

Figure 35 Descriptive Statistics per Year. Source Personal Processing

C: Figure36: Probability distribution histogram. Source Personal Processing

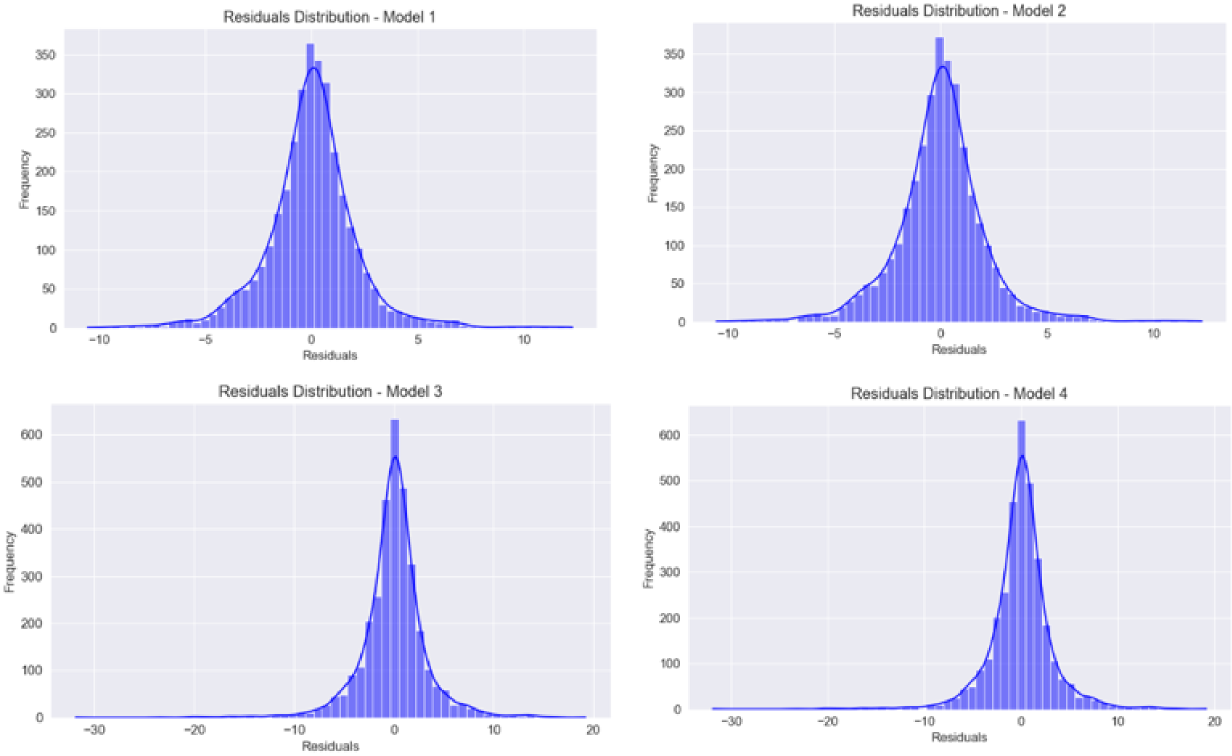


Figure 36 Probability distribution histogram. Source Personal Processing

Python Code

```
!pip install numpy
!pip install opendatasets
!pip install pandas
!pip install scikit-learn
!pip install plotly matlib seaborn
!pip install kaleido
!pip install missingno
!pip install doubleml

import numpy as np
import opendatasets as od
import pandas as pd
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
import scipy.stats as stats
import textwrap
import plotly.express as px
import os
import io
import missingno as msno
import statsmodels.api as sm
import scipy.stats as stats
from tabulate import tabulate
from scipy.stats import pearsonr
from matplotlib.colors import LinearSegmentedColormap
from matplotlib.lines import Line2D
from PIL import Image
from statsmodels.stats.outliers_influence import variance_inflation_factor
from linearmodels.panel import PanelOLS, RandomEffects
from scipy.stats import chi2
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from   sklearn.metrics   import   mean_absolute_error,   mean_squared_error,
r2_score
from sklearn.model_selection import GridSearchCV

pd.set_option('display.max_columns',None)
pd.set_option('display.max_rows',None)
sns.set_style ('darkgrid')
matplotlib.rcParams['font.size']=12
matplotlib.rcParams['figure.figsize']=(10,6)
matplotlib.rcParams['figure.facecolor']='#00000000'
database_df=pd.read_csv('DATABASE 1.csv',sep=';',
                      dtype={
                          'ESG SCORE':'float',
                          'ESG_ENVIRONMENT_SCORE':'float',
                          'ESG_SOCIAL_SCORE':'float',
                          'SG_GOVERNANCE_SCORE':'float',
                          'EBIT MARGIN':'float',
                          'ROA':'float',
                          'LN TOTAL ASSET':'float',
                          'TotDebt/TotAssets':'float',
                          'Asset Turnover':'float'
    }
)
database_df = database_df.rename(columns={
    'Asset Turover': 'Asset Turnover',
```

```
    'TotDebt/TotAssets': 'Leverage',
    'LN TOTAL ASSET': 'LnAsset',
    'ESG_ENVIRONMENT_SCORE': 'ENV_SCORE',
    'ESG_SOCIAL_SCORE':'SOC_SCORE',
    'ESG_GOVERNANCE_SCORE': 'GOV_SCORE',
    'EBIT MARGIN':'EBIT_MARGIN',
    'ESG SCORE': 'ESG_SCORE',
    'SECTOR/INDUSTRY':'INDUSTRY'
})
database_df.head()
database_df.shape
database_df.info()
database_df.isnull().sum()/len(database_df)*100
database_df.columns
num_of_companies = database_df['RIC'].nunique()
print(f"Numer of companies: {num_of_companies}")
database_df.shape
company_counts = database_df.groupby('RIC').size()
mean = company_counts.mean()
median = company_counts.median()
minimum = company_counts.min()
maximum = company_counts.max()
print(f"Mean of observations per company: {mean:.2f}")
print(f"Median of observations per company: {median}")
print(f"Minimum of observations per company: {minimum}")
print(f"Maximum of observations per company: {maximum}")
plt.figure(figsize=(8, 5))
sns.histplot(database_df['ROA'],    bins=40,    kde=True,    stat="density",
color="blue", alpha=0.6, label="ROA DISTRIBUTION")
plt.xlabel('Value')
plt.ylabel('Density')
plt.title('ROA DISTRIBUTION')
plt.legend()
plt.show()
#----------------------------------------------------
plt.figure(figsize=(8, 5))
sns.histplot(database_df['EBIT_MARGIN'], bins=40, kde=True, stat="density",
color="blue", alpha=0.6, label="EBIT_MARGIN DISTRIBUTION")
plt.xlabel('Value')
plt.ylabel('Density')
plt.title('EBIT_MARGIN DISTRIBUTION')
plt.legend()
plt.show()
#------------------------------------------------------------
plt.figure(figsize=(8, 5))
sns.histplot(database_df['Leverage'],  bins=40,  kde=True,  stat="density",
color="blue", alpha=0.6, label="Leverage DISTRIBUTION")
plt.xlabel('Value')
plt.ylabel('Density')
plt.title('Leverage DISTRIBUTION')
plt.legend()
plt.show()
#------------------------------------------------------------
plt.figure(figsize=(8, 5))
sns.histplot(database_df['LnAsset'],  bins=40,  kde=True,  stat="density",
color="blue", alpha=0.6, label="LnAsset DISTRIBUTION")
plt.xlabel('Value')
plt.ylabel('Density')
plt.title('LnAsset DISTRIBUTION')
plt.legend()
plt.show()
#------------------------------------------------------------
```

```
plt.figure(figsize=(8, 5))
sns.histplot(database_df['Asset     Turnover'],     bins=40,     kde=True,
stat="density",    color="blue",    alpha=0.6,    label="Asset    Turnover
DISTRIBUTION")
plt.xlabel('Value')
plt.ylabel('Density')
plt.title('Asset Turnover DISTRIBUTION')
plt.legend()
plt.show()
P0_ROA= database_df['ROA'].quantile(0)
P5_ROA= database_df['ROA'].quantile(0.05)
P10_ROA = database_df['ROA'].quantile(0.10)
Q1_ROA = database_df['ROA'].quantile(0.25)
Q2_ROA = database_df['ROA'].median()
Q3_ROA = database_df['ROA'].quantile(0.75)
P90_ROA= database_df['ROA'].quantile(0.90)
P95_ROA= database_df['ROA'].quantile(0.95)
P100_ROA= database_df['ROA'].quantile(1)
#---------------------------------------------------
P0_EBIT_MARGIN= database_df['EBIT_MARGIN'].quantile(0)
P5_EBIT_MARGIN= database_df['EBIT_MARGIN'].quantile(0.05)
P10_EBIT_MARGIN = database_df['EBIT_MARGIN'].quantile(0.10)
Q1_EBIT_MARGIN = database_df['EBIT_MARGIN'].quantile(0.25)
Q2_EBIT_MARGIN= database_df['EBIT_MARGIN'].median()
Q3_EBIT_MARGIN = database_df['EBIT_MARGIN'].quantile(0.75)
P90_EBIT_MARGIN= database_df['EBIT_MARGIN'].quantile(0.90)
P95_EBIT_MARGIN= database_df['EBIT_MARGIN'].quantile(0.95)
P100_EBIT_MARGIN= database_df['EBIT_MARGIN'].quantile(1)
#----------------------------------------------------------
P0_Leverage= database_df['Leverage'].quantile(0)
P5_Leverage= database_df['Leverage'].quantile(0.05)
P10_Leverage = database_df['Leverage'].quantile(0.10)
Q1_Leverage = database_df['Leverage'].quantile(0.25)
Q2_Leverage= database_df['Leverage'].median()
Q3_Leverage = database_df['Leverage'].quantile(0.75)
P90_Leverage= database_df['Leverage'].quantile(0.90)
P95_Leverage= database_df['Leverage'].quantile(0.95)
P100_Leverage= database_df['Leverage'].quantile(1)
#-----------------------------------------------------------
P0_AssetTurnover= database_df['Asset Turnover'].quantile(0)
P5_AssetTurnover= database_df['Asset Turnover'].quantile(0.05)
P10_AssetTurnover = database_df['Asset Turnover'].quantile(0.10)
Q1_AssetTurnover = database_df['Asset Turnover'].quantile(0.25)
Q2_AssetTurnover= database_df['Asset Turnover'].median()
Q3_AssetTurnover = database_df['Asset Turnover'].quantile(0.75)
P90_AssetTurnover= database_df['Asset Turnover'].quantile(0.90)
P95_AssetTurnover= database_df['Asset Turnover'].quantile(0.95)
P100_AssetTurnover= database_df['Asset Turnover'].quantile(1)
#-----------------------------------------
print('ROA:')
print('0 PERCENTILE:',P0_ROA)
print('5 PERCENTILE:',P5_ROA)
print('10 PERCENTILE:',P10_ROA)
print('25 PERCENTILE:',Q1_ROA)
print('50 PERCENTILE:',Q2_ROA)
print('75 PERCENTILE:',Q3_ROA)
print('90 PERCENTILE:',P90_ROA)
print('95 PERCENTILE:',P95_ROA)
print('100 PERCENTILE:',P100_ROA)
#-----------------------------------
print('EBIT_MARGIN:')
print('0 PERCENTILE:',P0_EBIT_MARGIN)
```

```
print('5 PERCENTILE:',P5_EBIT_MARGIN)
print('10 PERCENTILE:',P10_EBIT_MARGIN)
print('25 PERCENTILE:',Q1_EBIT_MARGIN)
print('50 PERCENTILE:',Q2_EBIT_MARGIN)
print('75 PERCENTILE:',Q3_EBIT_MARGIN)
print('90 PERCENTILE:',P90_EBIT_MARGIN)
print('95 PERCENTILE:',P95_EBIT_MARGIN)
print('100 PERCENTILE:',P100_EBIT_MARGIN)
#-------------------------------------
print('Leverage:')
print('0 PERCENTILE:',P0_Leverage)
print('5 PERCENTILE:',P5_Leverage)
print('10 PERCENTILE:',P10_Leverage)
print('25 PERCENTILE:',Q1_Leverage)
print('50 PERCENTILE:',Q2_Leverage)
print('75 PERCENTILE:',Q3_Leverage)
print('90 PERCENTILE:',P90_Leverage)
print('95 PERCENTILE:',P95_Leverage)
print('100 PERCENTILE:',P100_Leverage)
#------------------------------------
print('Asset Turnover:')
print('0 PERCENTILE:',P0_AssetTurnover)
print('5 PERCENTILE:',P5_AssetTurnover)
print('10 PERCENTILE:',P10_AssetTurnover)
print('25 PERCENTILE:',Q1_AssetTurnover)
print('50 PERCENTILE:',Q2_AssetTurnover)
print('75 PERCENTILE:',Q3_AssetTurnover)
print('90 PERCENTILE:',P90_AssetTurnover)
print('95 PERCENTILE:',P95_AssetTurnover)
print('100 PERCENTILE:',P100_AssetTurnover)
plt.figure(figsize=(8, 6))
sns.boxplot(x=database_df['ROA'], color="skyblue")
plt.title("Box Plot ROA", fontsize=14)
plt.xlabel("Value", fontsize=12)
plt.show()
#-------------------------
plt.figure(figsize=(8, 6))
sns.boxplot(x=filtered_df['ROA'], color="skyblue")
plt.title("Box Plot ROA_Cleaned", fontsize=14)
plt.xlabel("Value", fontsize=12)
plt.show()
#-------------------------
plt.figure(figsize=(8, 6))
sns.boxplot(x=database_df['EBIT_MARGIN'], color="skyblue")
plt.title("Box Plot EBIT_MARGIN", fontsize=14)
plt.xlabel("Value", fontsize=12)
plt.show()
#--------------------------------------------
plt.figure(figsize=(8, 6))
sns.boxplot(x=filtered_df['EBIT_MARGIN'], color="skyblue")
plt.title("Box Plot EBIT_MARGIN_Cleaned", fontsize=14)
plt.xlabel("Value", fontsize=12)
plt.show()
#--------------------------------------------
plt.figure(figsize=(8, 6))
sns.boxplot(x=database_df['Asset Turnover'], color="skyblue")
plt.title("Box Plot Asset Turnover", fontsize=14)
plt.xlabel("Value", fontsize=12)
plt.show()
#--------------------------------------------
plt.figure(figsize=(8, 6))
sns.boxplot(x=filtered_df['Asset Turnover'], color="skyblue")
```

```
plt.title("Box Plot Asset Turnover_Cleaned", fontsize=14)
plt.xlabel("Value", fontsize=12)
plt.show()
#---------------------------------------------
plt.figure(figsize=(8, 6))
sns.boxplot(x=database_df['Leverage'], color="skyblue")
plt.title("Box Plot Leverage", fontsize=14)
plt.xlabel("Value", fontsize=12)
plt.show()
#---------------------------------------------
plt.figure(figsize=(8, 6))
sns.boxplot(x=database_df['LnAsset'], color="skyblue")
plt.title("Box Plot LnAsset", fontsize=14)
plt.xlabel("Value", fontsize=12)
plt.show()
columns_to_analyze = ['EBIT_MARGIN', 'ROA', 'Asset Turnover']
filtered_df = database_df.copy()
for column in columns_to_analyze:

    Q1 = database_df[column].quantile(0.25)
    Q3 = database_df[column].quantile(0.75)

    IQR = Q3 - Q1

    lower_limit = Q1 - 1.5 * IQR
    upper_limit = Q3 + 1.5 * IQR

    print(f"\nAnalisi della colonna: {column}")
    print(f"Q1: {Q1:.2f}, Q3: {Q3:.2f}, IQR: {IQR:.2f}")
    print(f"Limite   Inferiore:   {lower_limit:.2f},   Limite   Superiore:
{upper_limit:.2f}")

    filtered_df = filtered_df[
        (filtered_df[column]  >=  lower_limit)  &  (filtered_df[column]  <=
upper_limit)
    ]
print("\nDataset pulito (filtered_df):")
desc_stats =  filtered_df.describe().drop(columns=['YEAR'])
desc_stats_without_count = desc_stats.drop(index='count').round(2)
desc_stats_without_count
count_by_Country = filtered_df.groupby('COUNTRY')['RIC'].nunique()
Companies_count_by_Country = count_by_Country.sort_values(ascending=False)
counts = Companies_count_by_Country.reset_index()
counts.columns = ['COUNTRY', 'COUNT']
fig = px.bar(counts,
             x='COUNTRY',
             y='COUNT',
             title='Number of Companies by Country')
fig.update_traces(
    text=counts['COUNT'],
    texttemplate='%{text}',
    textposition='outside'
)

fig.update_layout(
    bargap=0.1,
    autosize=False,
    width=600,   # Larghezza
    height=600,  # Altezza (uguale per forma quadrata)
     yaxis_title='Number of Companies',
    xaxis_title='Country',
    title=dict(x=0.5),
```

```
     xaxis=dict(
        tickangle=-45)# Centrare il titolo
)

fig.show()

fig.write_image("number_of_companies_by_country.jpeg", format="jpeg")
fig.write_image("C:\\Users\\Veronica\\Desktop\\Veruz\\number_of_companies_b
y_industry.jpeg", format="jpeg")
count_by_industry = filtered_df.groupby('INDUSTRY')['RIC'].nunique()
Companies_count_by_industry                                           =
count_by_industry.sort_values(ascending=False)
counts = Companies_count_by_industry.reset_index()
counts.columns = ['INDUSTRY', 'COUNT']
fig = px.bar(counts,
             x='INDUSTRY',
             y='COUNT',
             title='Number of Companies by Industry')
fig.update_traces(
    text=counts['COUNT'],
    texttemplate='%{text}',
    textposition='outside'
)

fig.update_layout(
    bargap=0.1,
    autosize=False,
    width=800,  # Aumenta la larghezza
    height=800,
    yaxis_title='Number of Companies',
    xaxis_title='Industry',
    title=dict(x=0.5),
    xaxis=dict(
        tickangle=-45,       # Ruota le etichette verticalmente
        tickfont=dict(size=12),  # Riduci la dimensione del font
    ),
    margin=dict(b=100)  # Aumenta il margine inferiore
)
fig.show()

fig.write_image("number_of_companies_by_industry.jpeg", format="jpeg")
fig.write_image("C:\\Users\\Veronica\\Desktop\\Veruz\\number_of_companies_b
y_industry.jpeg", format="jpeg")
mean_score   =   filtered_df.groupby('YEAR',   as_index=False)[['ENV_SCORE',
'SOC_SCORE', 'GOV_SCORE','ROA',

'EBIT_MARGIN']].mean()
print(mean_score)
colors = {
    "Environment Score": "green",
    "Social Score": "yellow",
    "Governance Score": "blue"
}

plt.figure(figsize=(10, 6))
plt.plot(mean_score['YEAR'],    mean_score['ENV_SCORE'],    linestyle='-',
linewidth=2, color=colors["Environment Score"], label='Environment Score')
plt.plot(mean_score['YEAR'],    mean_score['SOC_SCORE'],    linestyle='-',
linewidth=2, color=colors["Social Score"], label='Social Score')
plt.plot(mean_score['YEAR'],    mean_score['GOV_SCORE'],    linestyle='-',
linewidth=2, color=colors["Governance Score"], label='Governance Score')
plt.title('ESG Pilars Trend', fontsize=16)
```

```
plt.xlabel('Year', fontsize=14)
plt.ylabel('Average Value', fontsize=14)
plt.grid(True)
legend_lines  =  [Line2D([0],  [0],  color=colors[score_type],  lw=2)  for
score_type in colors]
legend = plt.legend(legend_lines, colors.keys(),

                    loc='lower center', ncol=len(colors))
legend.get_title().set_fontweight('bold')
plt.ylim(-5, 100)
plt.xlim(2013,2022)
plt.show()
grouped  =  filtered_df[['YEAR',  'ESG_SCORE',  'ENV_SCORE',  'SOC_SCORE',
'GOV_SCORE',
                       'ROA', 'EBIT_MARGIN', 'LnAsset', 'Leverage', 'Asset
Turnover']].groupby('YEAR')
tables_by_year = {year: group.describe().round(2) for year, group in grouped}
pd.set_option('display.width', 1000000)
pd.set_option('display.max_columns', None)
pd.set_option('display.max_colwidth', None)
def save_all_tables_as_image(tables, save_path):
    total_rows = sum(len(table) + 2 for table in tables.values())
    fig, ax = plt.subplots(figsize=(15, total_rows * 0.4))
    ax.axis('tight')
    ax.axis('off')
    rows = []
    col_labels                =                ["Index"]                +
list(tables_by_year[next(iter(tables_by_year))].columns)
    num_columns = len(col_labels)

    for year, table in tables.items():
        rows.append([f"Year: {year}"] + [""] * (num_columns - 1))
        rows.extend(table.reset_index().values.tolist())
    table = ax.table(cellText=rows, colLabels=col_labels, loc='center')
    table.auto_set_font_size(False)
    table.set_fontsize(8)
    table.auto_set_column_width(col=list(range(len(col_labels))))   # Auto-
resize colonne
    filename = f"{save_path}/all_tables_combined.jpeg"
    plt.savefig(filename, format="jpeg", bbox_inches="tight", dpi=300)
    plt.close()
    print(f"Tutte le tabelle sono state salvate come {filename}")
save_directory = r"C:/Users/Veronica/Desktop/Veruz"
save_all_tables_as_image(tables_by_year, save_directory)
grouped  =  filtered_df[['YEAR',  'ESG_SCORE',  'ENV_SCORE',  'SOC_SCORE',
'GOV_SCORE',
                       'ROA', 'EBIT_MARGIN', 'LnAsset', 'Leverage', 'Asset
Turnover']].groupby('YEAR')
tables_by_year = {year: group.describe().round(2) for year, group in grouped}
pd.set_option('display.width', 1000000)
pd.set_option('display.max_columns', None)
pd.set_option('display.max_colwidth', None)

for year, table in tables_by_year.items():
    print(f"{year}:\n")
    print(tabulate(table,        headers='keys',        stralign='center',
numalign='center'))
    print("\n" + "-" * 120 + "\n")
fig = px.scatter(filtered_df,
                 x='ESG_SCORE',
                 y='ROA',
                 opacity=0.8,
```

```
                 title='Scatter Plot ESG Score - ROA',
                labels={'ESG  SCORE':  'ESG  Score',  'ROA':  'Return  on
Assets'})
fig.update_layout(
    autosize=False,
    width=600,
    height=600,
    title=dict(x=0.5),
    xaxis_title="ESG Score",
    yaxis_title="Return on Assets",
)

fig.show()
fig.write_image("scatteresgros.jpeg", format="jpeg")
fig.write_image("C:\\Users\\Veronica\\Desktop\\Veruz\\scatteresgros.jpeg",
format="jpeg")
fig = px.scatter(filtered_df,
                 x='ESG_SCORE',
                 y='EBIT_MARGIN',
                 opacity=0.8,
                 title='Scatter Plot  ESG Score - EBIT_MARGIN',
                 labels={'ESG SCORE': 'ESG Score', 'EBIT': 'Ebit Margin'})
fig.update_layout(
    autosize=True,
    width=600,
    height=600,
    title=dict(x=0.25),
    xaxis_title="ESG Score",
    yaxis_title="EBIT_MARGIN",
)

fig.show()
fig.write_image("scatteresgebit.jpeg", format="jpeg")
fig.write_image("C:\\Users\\Veronica\\Desktop\\Veruz\\scatteresgebit.jpeg",
format="jpeg")
numeric_cols=filtered_df[['ESG_SCORE','ENV_SCORE','SOC_SCORE','GOV_SCORE','
LnAsset','EBIT_MARGIN','ROA','Leverage','Asset Turnover']]

corr = numeric_cols.corr()
mask = np.zeros_like(corr, dtype=bool)
mask[np.triu_indices_from(mask)] = True
np.fill_diagonal(mask, False)
fig, ax= plt.subplots(figsize=(10,5))
plt.title("Correlation Matrix", fontsize=14)
ax.tick_params(axis='y', labelsize=10)
Color_cmap    =    LinearSegmentedColormap.from_list("color",    ["white",
"lightblue", "Cyan","Blue"])
heatmap = sns.heatmap(corr,
                      annot= True,
                      annot_kws={"fontsize": 8, "color": "black"},
                      fmt='.2f',
                      linewidths=0.5,
                      cmap=Color_cmap,
                      mask=mask,
                      ax=ax)
x_labels   =   [textwrap.fill(label.get_text(),   width=10)   for   label   in
ax.get_xticklabels()]
ax.set_xticklabels(x_labels, rotation=20, ha="center", fontsize=10)
p_values = np.full((corr.shape[0], corr.shape[1]), np.nan)
for i in range(corr.shape[0]): #loop over rows
  for j in range(i+1, corr.shape[1]):
    x = numeric_cols.iloc[:, i]
```

```
    y =numeric_cols.iloc[:, j]
    mask = ~np.logical_or(np.isnan(x), np.isnan(y))
    if np.sum(mask) > 0:
      p_values[i, j] = pearsonr(x[mask], y[mask])[1]
p_values = pd.DataFrame(p_values, columns=corr.columns, index=corr.index)
mask_pvalues = np.triu(np.ones_like(p_values), k=1)
max_corr = np.max(corr.max())
min_corr = np.min(corr.min())
for i in range (p_values.shape[0]):
  for j in range(p_values.shape[1]):
    if mask_pvalues[i, j]:
      p_value = p_values.iloc[i, j]
      if not np.isnan(p_value):
        correlation_value = corr.iloc[i, j]
        text_color = 'black'
        if p_value <= 0.01:
            ax.text(i + 0.5, j + 0.8, f'(p = {p_value:.2f})***',
                    horizontalalignment='center',
                    verticalalignment='center',
                    fontsize=8,
                    color=text_color)
        elif p_value <= 0.05:
            ax.text(i + 0.5, j + 0.8, f'(p = {p_value:.2f})**',
                    horizontalalignment='center',
                    verticalalignment='center',
                    fontsize=8,
                    color=text_color)
        elif p_value <= 0.10:
            ax.text(i + 0.5, j + 0.8, f'(p = {p_value:.2f})*',
                    horizontalalignment='center',
                    verticalalignment='center',
                    fontsize=8,
                    color=text_color)
        else:
            ax.text(i + 0.5, j + 0.8, f'(p = {p_value:.2f})',
                    horizontalalignment='center',
                    verticalalignment='center',
                    fontsize=8,
                    color=text_color)
fig, ax = plt.subplots(figsize=(12, 8))
plt.show()
def calculate_vif(df):
    vif_data = pd.DataFrame()
    vif_data["Coefficients"] = df.columns
    vif_data["VIF"] = [variance_inflation_factor(df.values, i) for i in
range(df.shape[1])]
    vif_data["Tolerance"] = 1 / vif_data["VIF"]
    return vif_data

filtered_df = filtered_df.set_index(["RIC", "YEAR"])

multicollinearity_ROA = filtered_df[['ROA', 'ESG_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']]
multicollinearity_EBIT = filtered_df[['EBIT_MARGIN', 'ESG_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']]

multicollinearity_ROA =
multicollinearity_ROA.groupby(level="RIC").transform(lambda x: x - x.mean())
multicollinearity_EBIT =
multicollinearity_EBIT.groupby(level="RIC").transform(lambda x: x -
x.mean())
```

```
print("VIF test (EBIT_MARGIN as an independent variable):")
vif_test1 = calculate_vif(multicollinearity_ROA.drop(columns=['ROA']))
print(tabulate(vif_test1, headers='keys', tablefmt='psql',
stralign='center', numalign='center'))

print("\nVIF test (ROA as an independent variable):")
vif_test2 =
calculate_vif(multicollinearity_EBIT.drop(columns=['EBIT_MARGIN']))
print(tabulate(vif_test2, headers='keys', tablefmt='psql',
stralign='center', numalign='center'))
def calculate_vif(df):
    vif_data = pd.DataFrame()
    vif_data["Coefficients"] = df.columns
    vif_data["VIF"] = [variance_inflation_factor(df.values, i) for i in
range(df.shape[1])]
    vif_data["Tolerance"] = 1 / vif_data["VIF"]
    return vif_data

multicollinearity_ROA = filtered_df[[
'ROA','ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset', 'Leverage', 'Asset
Turnover']]
multicollinearity_EBIT =
filtered_df[['EBIT_MARGIN','ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']]
multicollinearity_ROA =
multicollinearity_ROA.groupby(level="RIC").transform(lambda x: x - x.mean())
multicollinearity_EBIT =
multicollinearity_EBIT.groupby(level="RIC").transform(lambda x: x -
x.mean())

print("VIF test (EBIT_MARGIN as an independent variable):")
vif_test1 = calculate_vif(multicollinearity_ROA.drop(columns=['ROA']))
print(tabulate(vif_test1, headers='keys', tablefmt='psql',
stralign='center', numalign='center'))

print("\nVIF test (ROA as an independent variable):")
vif_test2 =
calculate_vif(multicollinearity_EBIT.drop(columns=['EBIT_MARGIN']))
print(tabulate(vif_test2, headers='keys', tablefmt='psql',
stralign='center', numalign='center'))
Hausmandb_df = filtered_df[['COUNTRY', 'ENV_SCORE','SOC_SCORE','GOV_SCORE',
'EBIT_MARGIN', 'ROA', 'LnAsset', 'Leverage', 'Asset Turnover']].copy()
dependent_vars = ['ROA', 'EBIT_MARGIN']
independent_vars = [ 'ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']
def hausman_test(y, X):

    model_fe = PanelOLS(y, X, entity_effects=True)
    results_fe = model_fe.fit()

    model_re = RandomEffects(y, X)
    results_re = model_re.fit()

    b_diff = results_fe.params - results_re.params
    cov_diff = results_fe.cov - results_re.cov
    stat = np.dot(np.dot(b_diff.T, np.linalg.inv(cov_diff)), b_diff)
    df = b_diff.shape[0]
    p_value = chi2.sf(stat, df)
    return stat, p_value

hausman_results = []
```

```
for dep_var in dependent_vars:
    y = Hausmandb_df[dep_var]
    X = sm.add_constant(Hausmandb_df[independent_vars])
    stat, p_value = hausman_test(y, X)
    hausman_results.append({'Dep. Variable': dep_var,
                            'Chi-Square Statistic': stat,
                            'P-value': p_value})

hausman_df = pd.DataFrame(hausman_results)
print("\nHausman test (ESG Pillars):")
print(tabulate(hausman_df, headers='keys', tablefmt='psql',
stralign='center', numalign='center'))
dependent1 = filtered_df['ROA']
independent_vars1 = filtered_df[['ESG_SCORE', 'LnAsset', 'Leverage', 'Asset
Turnover']]
model1 = PanelOLS(dependent1, independent_vars1, entity_effects=True)
results1 = model1.fit(cov_type='robust')
y_pred_panel1 = results1.predict()
variables1 = ['ROA', 'ESG_SCORE', 'LnAsset', 'Leverage', 'Asset Turnover']
between_data1 = filtered_df.groupby("RIC")[variables1].mean()
y_between1 = between_data1['ROA']
X_between1 = between_data1[['ESG_SCORE', 'LnAsset', 'Leverage', 'Asset
Turnover']]
model_between1 = sm.OLS(y_between1, X_between1).fit()
y_pred_between1 = model_between1.predict(X_between1)
panel_data1 = filtered_df[variables1].dropna()
y_overall1 = panel_data1['ROA']
X_overall1 = panel_data1[['ESG_SCORE', 'LnAsset', 'Leverage', 'Asset
Turnover']]
model_overall1 = sm.OLS(y_overall1, X_overall1).fit()
y_pred_overall1 = model_overall1.predict(X_overall1)
SS_res1 = np.sum((y_between1 - y_pred_between1) ** 2)
SS_tot1 = np.sum((y_between1 - y_between1.mean()) ** 2)
R2_between1 = 1 - (SS_res1 / SS_tot1)
SS_res_overall1 = np.sum((y_overall1 - y_pred_overall1) ** 2)
SS_tot_overall1 = np.sum((y_overall1 - y_overall1.mean()) ** 2)
R2_overall1 = 1 - (SS_res_overall1 / SS_tot_overall1)
print(f"PanelOLS Estimation Summary: ROA as Dependent
Variable:\n{results1}\n")
print(f"R-squared (Between): {R2_between1:.4f}")
print(f"R-squared (Overall): {R2_overall1:.4f}")
dependent2 = filtered_df['ROA']
independent_vars2 = filtered_df[['ENV_SCORE','SOC_SCORE','GOV_SCORE',
'LnAsset', 'Leverage', 'Asset Turnover']]
model2 = PanelOLS(dependent2, independent_vars2, entity_effects=True)
results2 = model2.fit(cov_type='robust')
y_pred_panel2 = results2.predict()
variables2 = ['ROA', 'ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']
between_data2 = filtered_df.groupby("RIC")[variables2].mean()
y_between2 = between_data2['ROA']
X_between2 = between_data2[['ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']]
model_between2 = sm.OLS(y_between2, X_between2).fit()
y_pred_between2 = model_between2.predict(X_between2)
panel_data2 = filtered_df[variables2].dropna()
y_overall2 = panel_data2['ROA']
X_overall2 = panel_data2[['ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']]
model_overall2 = sm.OLS(y_overall2, X_overall2).fit()
y_pred_overall2 = model_overall2.predict(X_overall2)
SS_res2 = np.sum((y_between2 - y_pred_between2) ** 2)
```

```
SS_tot2 = np.sum((y_between2 - y_between2.mean()) ** 2
R2_between2 = 1 - (SS_res2 / SS_tot2)
SS_res_overall2 = np.sum((y_overall2 - y_pred_overall2) ** 2)
SS_tot_overall2 = np.sum((y_overall2 - y_overall2.mean()) ** 2)
R2_overall2 = 1 - (SS_res_overall2 / SS_tot_overall2)
print(f"PanelOLS Estimation Summary: ROA as Dependent
Variable:\n{results2}\n")
print(f"R-squared (Between): {R2_between2:.4f}")
print(f"R-squared (Overall): {R2_overall2:.4f}")
dependent3 = filtered_df['EBIT_MARGIN']
independent_vars3 = filtered_df[['ESG_SCORE', 'LnAsset', 'Leverage', 'Asset
Turnover']]
model3 = PanelOLS(dependent3, independent_vars3, entity_effects=True)
results3 = model3.fit(cov_type='robust')
y_pred_panel3 = results3.predict()
variables3 = ['EBIT_MARGIN', 'ESG_SCORE', 'LnAsset', 'Leverage', 'Asset
Turnover']
between_data3 = filtered_df.groupby("RIC")[variables3].mean()
y_between3 = between_data3['EBIT_MARGIN']
X_between3 = between_data3[['ESG_SCORE', 'LnAsset', 'Leverage', 'Asset
Turnover']]
model_between3 = sm.OLS(y_between3, X_between3).fit()
y_pred_between3 = model_between3.predict(X_between3)
panel_data3 = filtered_df[variables3].dropna()
y_overall3 = panel_data3['EBIT_MARGIN']
X_overall3 = panel_data3[['ESG_SCORE', 'LnAsset', 'Leverage', 'Asset
Turnover']]
model_overall3 = sm.OLS(y_overall3, X_overall3).fit()
y_pred_overall3 = model_overall3.predict(X_overall3)
SS_res3 = np.sum((y_between3 - y_pred_between3) ** 2)
SS_tot3 = np.sum((y_between3 - y_between3.mean()) ** 2)
R2_between3 = 1 - (SS_res3 / SS_tot3)
SS_res_overall3 = np.sum((y_overall3 - y_pred_overall3) ** 2)
SS_tot_overall3 = np.sum((y_overall3 - y_overall3.mean()) ** 2)
R2_overall3 = 1 - (SS_res_overall3 / SS_tot_overall3)
print(f"PanelOLS Estimation Summary: EBIT_MARGIN as Dependent
Variable\n{results3}\n")
print(f"R-squared (Between): {R2_between3:.4f}")
print(f"R-squared (Overall): {R2_overall3:.4f}")
dependent4 = filtered_df['EBIT_MARGIN']
independent_vars4 = filtered_df[['ENV_SCORE','SOC_SCORE','GOV_SCORE',
'LnAsset', 'Leverage', 'Asset Turnover']]
model4 = PanelOLS(dependent4, independent_vars4, entity_effects=True)
results4 = model4.fit(cov_type='robust')
y_pred_panel4 = results4.predict()
variables4 = ['EBIT_MARGIN', 'ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']
between_data4 = filtered_df.groupby("RIC")[variables4].mean()
y_between4 = between_data4['EBIT_MARGIN']
X_between4 = between_data4[['ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']]
model_between4 = sm.OLS(y_between4, X_between4).fit()
y_pred_between4 = model_between4.predict(X_between4)
panel_data4 = filtered_df[variables4].dropna()
y_overall4 = panel_data4['EBIT_MARGIN']
X_overall4 = panel_data4[['ENV_SCORE','SOC_SCORE','GOV_SCORE', 'LnAsset',
'Leverage', 'Asset Turnover']]
model_overall4 = sm.OLS(y_overall4, X_overall4).fit()
y_pred_overall4 = model_overall4.predict(X_overall4)
SS_res4 = np.sum((y_between4 - y_pred_between4) ** 2)
SS_tot4 = np.sum((y_between4 - y_between4.mean()) ** 2)
R2_between4 = 1 - (SS_res4 / SS_tot4)
```

```
SS_res_overall4 = np.sum((y_overall4 - y_pred_overall4) ** 2)
SS_tot_overall4 = np.sum((y_overall4 - y_overall4.mean()) ** 2)
R2_overall4 = 1 - (SS_res_overall4 / SS_tot_overall4)
print(f"PanelOLS Estimation Summary: EBIT_MARGIN as Dependent
Variable\n{results4}\n")
print(f"R-squared (Between): {R2_between4:.4f}")
print(f"R-squared (Overall): {R2_overall4:.4f}")
def summarize_model(results, model_name):
    """
    residuals = results.resids.to_numpy()
    std_residuals = np.std(residuals)
    r_squared = results.rsquared
    n = results.nobs
    k = results.params.shape[0]
    adj_r_squared = 1 - ((1 - r_squared) * (n - 1) / (n - k - 1))
    try:
        f_stat = results.f_statistic.stat
        f_stat_pval = results.f_statistic.pval
    except AttributeError:
        f_stat = np.nan
        f_stat_pval = np.nan

    try:
        f_stat_robust = results.f_statistic_robust.stat
        f_stat_robust_pval = results.f_statistic_robust.pval
    except AttributeError:
        f_stat_robust = np.nan
        f_stat_robust_pval = np.nan


    try:
        pool_f_stat = results.f_pooled.stat
        pool_f_pval = results.f_pooled.pval
    except AttributeError:
        pool_f_stat = np.nan
        pool_f_pval = np.nan

    return {
        "Model": model_name,
        "Std of Residuals": round(std_residuals, 4),
        "F-statistic": round(f_stat, 4) if not np.isnan(f_stat) else "-",
        "F-stat p-value": f"{f_stat_pval:.2e}" if not np.isnan(f_stat_pval)
else "-",
        "F-stat (robust)": round(f_stat_robust, 4) if not
np.isnan(f_stat_robust) else "-",
        "F-stat robust p-val": f"{f_stat_robust_pval:.2e}" if not
np.isnan(f_stat_robust_pval) else "-",
        "Poolability F-test": round(pool_f_stat, 4) if not
np.isnan(pool_f_stat) else "-",
        "Pool F-test p-value": f"{pool_f_pval:.2e}" if not
np.isnan(pool_f_pval) else "-"
    }

summary1 = summarize_model(results1, "Model 1")
summary2 = summarize_model(results2, "Model 2")
summary3 = summarize_model(results3, "Model 3")
summary4 = summarize_model(results4, "Model 4")

summary_table = pd.DataFrame([summary1, summary2, summary3, summary4])
styled_table = (
    summary_table.style
    .set_caption("Summary Statistics")
```

```
    .set_table_styles(
        [
            {"selector": "caption", "props": [("font-size", "18px"), ("font-
weight", "bold"), ("text-align", "center"), ("color", "#333")]},
            {"selector": "th", "props": [("font-size", "14px"), ("text-
align", "center"), ("background-color", "#e0e0e0"), ("color", "#333")]},
            {"selector": "td", "props": [("font-size", "13px"), ("text-
align", "center"), ("padding", "8px")]},
        ]
    )
    .set_properties(**{"text-align": "center"})
    .apply(lambda x: ['background: #f7f7f7' if i % 2 == 0 else '' for i in
range(len(x))], axis=0)
    .format({
        "Std of Residuals": "{:.4f}",
        "F-statistic": "{:.4f}",
        "F-stat p-value": lambda x: "{:.4e}".format(x) if isinstance(x,
float) else x,
        "F-stat (robust)": "{:.4f}",
        "F-stat robust p-val": lambda x: "{:.4e}".format(x) if isinstance(x,
float) else x,
        "Poolability F-test": "{:.4f}",
        "Pool F-test p-value": lambda x: "{:.4e}".format(x) if isinstance(x,
float) else x,
    })
)

display(styled_table)
df1 = filtered_df.copy()
df1['residuals'] = results1.resids
df1['residuals_lag'] = df1.groupby(level=0)['residuals'].shift(1)
df_aux1 = df1.dropna(subset=['residuals', 'residuals_lag'])

X_aux1 = add_constant(df_aux1['residuals_lag'])
y_aux1 = df_aux1['residuals']
aux_model1 = OLS(y_aux1, X_aux1).fit()

df2 = filtered_df.copy()
df2['residuals'] = results2.resids
df2['residuals_lag'] = df2.groupby(level=0)['residuals'].shift(1)
df_aux2 = df2.dropna(subset=['residuals', 'residuals_lag'])

X_aux2 = add_constant(df_aux2['residuals_lag'])
y_aux2 = df_aux2['residuals']
aux_model2 = OLS(y_aux2, X_aux2).fit()


df3 = filtered_df.copy()
df3['residuals'] = results3.resids
df3['residuals_lag'] = df3.groupby(level=0)['residuals'].shift(1)
df_aux3 = df3.dropna(subset=['residuals', 'residuals_lag'])

X_aux3 = add_constant(df_aux3['residuals_lag'])
y_aux3 = df_aux3['residuals']
aux_model3 = OLS(y_aux3, X_aux3).fit()

df4 = filtered_df.copy()
df4['residuals'] = results4.resids
df4['residuals_lag'] = df4.groupby(level=0)['residuals'].shift(1)
df_aux4 = df4.dropna(subset=['residuals', 'residuals_lag'])

X_aux4 = add_constant(df_aux4['residuals_lag'])
```

```
y_aux4 = df_aux4['residuals']
aux_model4 = OLS(y_aux4, X_aux4).fit()
import pandas as pd


models = [aux_model1, aux_model2, aux_model3, aux_model4]
model_names = ['Model 1', 'Model 2', 'Model 3', 'Model 4']
summary_data = []
for name, model in zip(model_names, models):
    coef = model.params['residuals_lag']
    std_err = model.bse['residuals_lag']
    t_stat = model.tvalues['residuals_lag']
    p_value = model.pvalues['residuals_lag']
    r_squared = model.rsquared
        summary_data.append({
        'Model': name,
        'Coef. residuals_lag': round(coef, 4),
        'Std. Err.': round(std_err, 4),
        't-stat': round(t_stat, 2),
        'P-value': round(p_value, 4),
        'R²': round(r_squared, 4)
    })


summary_df = pd.DataFrame(summary_data)
from IPython.display import display, Markdown
display(Markdown("<h3 style='text-align: center;'>Wooldridge Test –
Summary</h3>"))
display(summary_df)
def white_test_from_model(results, X_base, resids_within,
model_name="Modello"):
    """

    Args:
        results       → oggetto .fit() di PanelOLS
        X_base        → DataFrame delle variabili indipendenti originali
        resids_within → residui within del modello
        model_name    → nome descrittivo per stampa risultati
    """
    idx = resids_within.dropna().index
    X = X_base.loc[idx].copy()
    y = resids_within.loc[idx] ** 2

    X_const = sm.add_constant(X)
    X_white = X_const.copy()
    cols = X.columns.tolist()

    for i, col1 in enumerate(cols):
        X_white[f"{col1}^2"] = X[col1] ** 2
        for col2 in cols[i+1:]:
            X_white[f"{col1}*{col2}"] = X[col1] * X[col2]

    model_aux = sm.OLS(y, X_white).fit()
    n = X_white.shape[0]
    df = X_white.shape[1] - 1  # tolgo costante
    r2 = model_aux.rsquared
    lm_stat = n * r2
    lm_pval = 1 - chi2.cdf(lm_stat, df)
    f_stat = (r2 / df) / ((1 - r2) / (n - df - 1))
    f_pval = 1 - f.cdf(f_stat, df, n - df - 1)
      print(f"\n White Test – {model_name}")
```

```
    print(f"Lagrange Multiplier χ² stat: {lm_stat:.3f} (df={df}) — p =
{lm_pval:.4f}")
    print(f"F-statistic: {f_stat:.3f} — p = {f_pval:.4f}")

    return {
        "Model": model_name,
        "LM Stat": lm_stat,
        "LM p-value": lm_pval,
        "F Stat": f_stat,
        "F p-value": f_pval
    }
import statsmodels.api as sm
from statsmodels.stats.diagnostic import het_white
from scipy.stats import chi2, f
# Lista dei modelli
modelli = [
    {"results": results1, "X": independent_vars1, "residuals":
results1.resids, "name": "Model 1 "},
    {"results": results2, "X": independent_vars2, "residuals":
results2.resids, "name": "Model 2"},
    {"results": results3, "X": independent_vars3, "residuals":
results3.resids, "name": "Model 3"},
    {"results": results4, "X": independent_vars4, "residuals":
results4.resids, "name": "Model 4"}
]

white_test_results = []
for mod in modelli:
    result = white_test_from_model(
        results=mod["results"],
        X_base=mod["X"],
        resids_within=mod["residuals"],
        model_name=mod["name"]
    )
    white_test_results.append(result)
white_df = pd.DataFrame(white_test_results)
from IPython.display import display
display(Markdown("<h3 style='text-align: center;'>White Test -
Summary</h3>"))
display(white_df)
residuals1 = results1.resids
plt.figure(figsize=(8, 5))
sns.histplot(residuals1, kde=True, bins=60, color='blue')
plt.title('Residuals Distribution - Model 1')
plt.xlabel('Residuals')
plt.ylabel('Frequency')
plt.grid(True)
plt.tight_layout()
plt.show()
#---------------------------------
residuals2 = results2.resids
plt.figure(figsize=(8, 5))
sns.histplot(residuals2, kde=True, bins=60, color='blue')
plt.title('Residuals Distribution - Model 2')
plt.xlabel('Residuals')
plt.ylabel('Frequency')
plt.grid(True)
plt.tight_layout()
plt.show()
#------------------------------
residuals3 = results3.resids
plt.figure(figsize=(8, 5))
```

```
sns.histplot(residuals3, kde=True, bins=60, color='blue')
plt.title('Residuals Distribution - Model 3')
plt.xlabel('Residuals')
plt.ylabel('Frequency')
plt.grid(True)
plt.tight_layout()
plt.show()
#-----------------------
residuals4 = results4.resids
plt.figure(figsize=(8, 5))
sns.histplot(residuals4, kde=True, bins=60, color='blue')
plt.title('Residuals Distribution - Model 4')
plt.xlabel('Residuals')
plt.ylabel('Frequency')
plt.grid(True)
plt.tight_layout()
plt.show()
#----------------------------
sm.qqplot(residuals1, line='45', fit=True)
plt.title('Q-Q Plot of Residuals - Model 1')
plt.xlabel('Theoretical Quantiles')
plt.ylabel('Sample Quantiles')
plt.grid(True)
plt.show()
#----------------------
sm.qqplot(residuals2, line='45', fit=True)
plt.title('Q-Q Plot of Residuals - Model 2')
plt.xlabel('Theoretical Quantiles')
plt.ylabel('Sample Quantiles')
plt.grid(True)
plt.show()
#--------------------
sm.qqplot(residuals3, line='45', fit=True)
plt.title('Q-Q Plot of Residuals - Model 3')
plt.xlabel('Theoretical Quantiles')
plt.ylabel('Sample Quantiles')
plt.grid(True)
plt.show()
#-----------------------------
sm.qqplot(residuals4, line='45', fit=True)
plt.title('Q-Q Plot of Residuals - Model 4')
plt.xlabel('Theoretical Quantiles')
plt.ylabel('Sample Quantiles')
plt.grid(True)
plt.show()

rmse1 = np.sqrt(np.mean(results1.resids ** 2))
rmse2 = np.sqrt(np.mean(results2.resids ** 2))
rmse3 = np.sqrt(np.mean(results3.resids ** 2))
rmse4 = np.sqrt(np.mean(results4.resids ** 2))

data = {
    "Model": ["Model 1", "Model 2", "Model 3", "Model 4"],
    "R2_Within": [
        results1.rsquared,
        results2.rsquared,
        results3.rsquared,
        results4.rsquared
    ],
    "R2_Between": [
        R2_between1,
        R2_between2,
```

```
        R2_between3,
        R2_between4
    ],
    "R2_Overall": [
        R2_overall1,
        R2_overall2,
        R2_overall3,
        R2_overall4
    ],
    "RMSE": [
        rmse1,
        rmse2,
        rmse3,
        rmse4
    ]
}

summary_df = pd.DataFrame(data)

display(Markdown("<h3 style='text-align: center;'>R² and RMSE -
Summary</h3>"))
display(summary_df)
def extract_model_results(results, df):
    params = results.params
    p_values = results.pvalues
    r_squared_within = results.rsquared_within
    r_squared_between = results.rsquared_between
    r_squared_overall = results.rsquared_overall
    f_stat = results.f_statistic.stat
    robust_f_stat = results.f_statistic_robust.stat

    try:
        poolability_f_stat = results.f_pooled.stat
        poolability_f_pval = results.f_pooled.pval
    except AttributeError:
        poolability_f_stat = None
        poolability_f_pval = None

    n_entities = df.index.get_level_values('RIC').nunique()
    n_time_periods = df.index.get_level_values('YEAR').nunique()

    def add_stars(pval):
        if pval < 0.01:
            return "***"
        elif pval < 0.05:
            return "**"
        elif pval < 0.1:
            return "*"
        else:
            return ""

    coefficients_with_stars = {}
    for var in params.index:
        coefficients_with_stars[var] =
f"{params[var]:.4f}{add_stars(p_values[var])}"

    return [
        coefficients_with_stars.get("ESG_SCORE", None),
        coefficients_with_stars.get("ENV_SCORE", None),
        coefficients_with_stars.get("SOC_SCORE", None),
        coefficients_with_stars.get("GOV_SCORE", None),
        coefficients_with_stars.get("LnAsset", None),
```

```
        coefficients_with_stars.get("Leverage", None),
        coefficients_with_stars.get("Asset Turnover", None),
        r_squared_within if r_squared_within is not None else None,
        r_squared_between if r_squared_between is not None else None,
        r_squared_overall if r_squared_overall is not None else None,
        f_stat if f_stat is not None else None,
        robust_f_stat if robust_f_stat is not None else None,
        poolability_f_stat if poolability_f_stat is not None else None,
        poolability_f_pval if poolability_f_pval is not None else None,
        n_entities,
        n_time_periods
    ]

model1_results = extract_model_results(results1, filtered_df)
model2_results = extract_model_results(results2, filtered_df)
model3_results = extract_model_results(results3, filtered_df)
model4_results = extract_model_results(results4, filtered_df)

results = {
    "Variable": [
        "ESG_SCORE",
        "ENV_SCORE",
        "SOC_SCORE",
        "GOV_SCORE",
        "LnAsset",
        "Leverage",
        "Asset Turnover",
        "R-squared (Within)",
        "R-squared (Between)",
        "R-squared (Overall)",
        "F-statistic",
        "F-statistic (robust)",
        "F-test for Poolability",
        "P-value (Poolability F-test)",
        "Number of Entities",
        "Number of Time Periods",
    ],
    "Model 1": model1_results,
    "Model 2": model2_results,
    "Model 3": model3_results,
    "Model 4": model4_results,
}

df = pd.DataFrame(results)
df = df.fillna("-")
print(df)
df.to_csv("comparison_models_with_entities_time.csv", index=False)

try:
    from IPython.display import display
    display(
        df.style.format(
            {
                "Model 1": lambda x: f"{x:.4f}" if isinstance(x, (int,
float)) else x,
                "Model 2": lambda x: f"{x:.4f}" if isinstance(x, (int,
float)) else x,
                "Model 3": lambda x: f"{x:.4f}" if isinstance(x, (int,
float)) else x,
                "Model 4": lambda x: f"{x:.4f}" if isinstance(x, (int,
float)) else x,
            }
```

```
        ).set_caption("Ebit Margin as Dependent Variable")
    )
except ImportError:
    pass

df_roa = filtered_df.copy()

df_roa = df_roa.rename(columns={
    'ROA': 'Y',
    'ESG_SCORE': 'D',
    'Leverage': 'X1',
    'Asset Turnover': 'X2',
    'LnAsset': 'X3'
})

def demean_panel(series, group):
    return series - series.groupby(group).transform('mean')

def run_dml_roa_esg(ml_l_model, ml_m_model, ml_l_name):
    X = pd.DataFrame({
        'X1': demean_panel(df_roa['X1'], df_roa['RIC']),
        'X2': demean_panel(df_roa['X2'], df_roa['RIC']),
        'X3': demean_panel(df_roa['X3'], df_roa['RIC'])
    })
    y = demean_panel(df_roa['Y'], df_roa['RIC'])
    d = demean_panel(df_roa['D'], df_roa['RIC'])
    data = DoubleMLData.from_arrays(X, y, d)
    dml = DoubleMLPLR(data, ml_l=ml_l_model, ml_m=ml_m_model, n_folds=5)
    dml.fit()
    theta = dml.coef[0]
    se = dml.se[0]
    t_stat = theta / se
    p_val = 2 * (1 - norm.cdf(abs(t_stat)))
    ci_low = theta - 1.96 * se
    ci_high = theta + 1.96 * se

    return {
        'Variable': 'ESG_SCORE',
        'ML_l': ml_l_name,
        'ML_m': 'Lasso',
        'Theta': theta,
        'Std. Error': se,
        't-stat': t_stat,
        'P-value': p_val,
        'CI Lower': ci_low,
        'CI Upper': ci_high
    }

lasso = LassoCV(cv=5, random_state=42)
rf = RandomForestRegressor(max_depth=4, random_state=42)
xgb   =   XGBRegressor(max_depth=4,   learning_rate=0.1,   n_estimators=100,
random_state=42)

results_model5 = [
    run_dml_roa_esg(xgb, lasso, 'XGBoost'),
    run_dml_roa_esg(rf, lasso, 'Random Forest')
]

results_df_model5 = pd.DataFrame(results_model5)
n_obs = len(df_roa)
n_ric = df_roa['RIC'].nunique()
n_year = df_roa['YEAR'].nunique()
```

```
mean_years_per_ric = df_roa.groupby('RIC')['YEAR'].nunique().mean()
mean_ric_per_year = df_roa.groupby('YEAR')['RIC'].nunique().mean()

meta_info = {
    'N. observation': n_obs,
    'Entities': n_ric,
    'Time Periods': n_year,
    'Avg Obs Year per Entity': round(mean_years_per_ric, 2),
    'Avg Obs Entities per Year': round(mean_ric_per_year, 2)
}

print("\nMODEL 5 RESULTS - ROA \n")
print(tabulate(results_df_model5,    headers='keys',    tablefmt='github',
floatfmt=".4f"))

print("\nSAMPLE INFORMATION\n")
for k, v in meta_info.items():
    print(f"{k}: {v}")
df_roa2 = filtered_df.copy()
df_roa2 = df_roa2.rename(columns={
    'ROA': 'Y',
    'ENV_SCORE': 'D1',
    'SOC_SCORE': 'D2',
    'GOV_SCORE': 'D3',
    'Leverage': 'X1',
    'Asset Turnover': 'X2',
    'LnAsset': 'X3'
})

def demean_panel(series, group):
    return series - series.groupby(group).transform('mean')
def run_dml_custom_roa2(treatment_var, ml_l_model, ml_m_model, ml_l_name):
    X = pd.DataFrame({
        'X1': demean_panel(df_roa2['X1'], df_roa2['RIC']),
        'X2': demean_panel(df_roa2['X2'], df_roa2['RIC']),
        'X3': demean_panel(df_roa2['X3'], df_roa2['RIC'])
    })
    y = demean_panel(df_roa2['Y'], df_roa2['RIC'])
    d = demean_panel(df_roa2[treatment_var], df_roa2['RIC'])
    data = DoubleMLData.from_arrays(X, y, d)
    dml = DoubleMLPLR(data, ml_l=ml_l_model, ml_m=ml_m_model, n_folds=5)
    dml.fit()
    theta = dml.coef[0]
    se = dml.se[0]
    t_stat = theta / se
    p_val = 2 * (1 - norm.cdf(abs(t_stat)))
    ci_low = theta - 1.96 * se
    ci_high = theta + 1.96 * se
    return {
        'Variable': treatment_var,
        'ML_l': ml_l_name,
        'ML_m': 'Lasso',
        'Theta': theta,
        'Std. Error': se,
        't-stat': t_stat,
        'P-value': p_val,
        'CI Lower': ci_low,
        'CI Upper': ci_high
    }

lasso = LassoCV(cv=5, random_state=42)
rf = RandomForestRegressor(max_depth=4, random_state=42)
```

```
xgb = XGBRegressor(max_depth=4, learning_rate=0.1, n_estimators=100,
random_state=42)
results_model6 = []
for D in ['D1', 'D2', 'D3']:
    results_model6.append(run_dml_custom_roa2(D, xgb, lasso, 'XGBoost'))
    results_model6.append(run_dml_custom_roa2(D, rf, lasso, 'Random
Forest'))
results_df_model6 = pd.DataFrame(results_model6)
label_map = {
    'D1': 'ENV_SCORE',
    'D2': 'SOC_SCORE',
    'D3': 'GOV_SCORE'
}
results_df_model6['Variable'] =
results_df_model6['Variable'].replace(label_map)
n_obs = len(df_roa)
n_ric = df_roa['RIC'].nunique()
n_year = df_roa['YEAR'].nunique()
mean_years_per_ric = df_roa.groupby('RIC')['YEAR'].nunique().mean()
mean_ric_per_year = df_roa.groupby('YEAR')['RIC'].nunique().mean()

meta_info = {
    'N. observation': n_obs,
    'Entities': n_ric,
    'Time Periods': n_year,
    'Avg Obs Year per Entity': round(mean_years_per_ric, 2),
    'Avg Obs Entities per Year': round(mean_ric_per_year, 2)
}

print("\n MODEL 6 RESULTS - ROA \n")
print(tabulate(results_df_model6, headers='keys', tablefmt='github',
floatfmt=".4f"))

print("\nSAMPLE INFORMATION\n")
for k, v in meta_info.items():
    print(f"{k}: {v}")
df_ebit = filtered_df.copy()

df_ebit = df_ebit.rename(columns={
    'EBIT_MARGIN': 'Y',
    'ESG_SCORE': 'D',
    'Leverage': 'X1',
    'Asset Turnover': 'X2',
    'LnAsset': 'X3'
})

def demean_panel(series, group):
    return series - series.groupby(group).transform('mean')

def run_dml_ebit_esg(ml_l_model, ml_m_model, ml_l_name, ml_m_name):
    # De-meaned variables for panel structure (by RIC)
    X = pd.DataFrame({
        'X1': demean_panel(df_ebit['X1'], df_ebit['RIC']),
        'X2': demean_panel(df_ebit['X2'], df_ebit['RIC']),
        'X3': demean_panel(df_ebit['X3'], df_ebit['RIC'])
    })
    y = demean_panel(df_ebit['Y'], df_ebit['RIC'])
    d = demean_panel(df_ebit['D'], df_ebit['RIC'])

    data = DoubleMLData.from_arrays(X, y, d)
    dml = DoubleMLPLR(data, ml_l=ml_l_model, ml_m=ml_m_model, n_folds=5)
    dml.fit()
```

```
    theta = dml.coef[0]
    se = dml.se[0]
    t_stat = theta / se
    p_val = 2 * (1 - norm.cdf(abs(t_stat)))
    ci_low = theta - 1.96 * se
    ci_high = theta + 1.96 * se
    return {
        'Variable': 'ESG_SCORE',
        'ML_l': ml_l_name,
        'ML_m': ml_m_name,
        'Theta': theta,
        'Std. Error': se,
        't-stat': t_stat,
        'P-value': p_val,
        'CI Lower': ci_low,
        'CI Upper': ci_high
    }

lasso = LassoCV(cv=5, random_state=42)
rf = RandomForestRegressor(max_depth=4, random_state=42)
xgb   =   XGBRegressor(max_depth=4,   learning_rate=0.1,   n_estimators=100,
random_state=42)
results_model7 = [
    run_dml_ebit_esg(lasso, lasso, 'Lasso', 'Lasso'),
    run_dml_ebit_esg(lasso, lasso, 'Lasso', 'Lasso')
]
results_df_model7 = pd.DataFrame(results_model7)
n_obs = len(df_ebit)
n_ric = df_ebit['RIC'].nunique()
n_year = df_ebit['YEAR'].nunique()
mean_years_per_ric = df_ebit.groupby('RIC')['YEAR'].nunique().mean()
mean_ric_per_year = df_ebit.groupby('YEAR')['RIC'].nunique().mean()
meta_info = {
    'N. observation': n_obs,
    'Entities': n_ric,
    'Time Periods': n_year,
    'Avg Obs Year per Entity': round(mean_years_per_ric, 2),
    'Avg Obs Entities per Year': round(mean_ric_per_year, 2)
}

print("\n MODEL 7 RESULTS - Ebit Marigin\n")
print(tabulate(results_df_model7,    headers='keys',    tablefmt='github',
floatfmt=".4f"))
print("\nSAMPLE INFORMATION\n")
for k, v in meta_info.items():
    print(f"{k}: {v}")
df_ebit2 = filtered_df.copy()
df_ebit2 = df_ebit2.rename(columns={
    'EBIT_MARGIN': 'Y',
    'ENV_SCORE': 'D1',
    'SOC_SCORE': 'D2',
    'GOV_SCORE': 'D3',
    'Leverage': 'X1',
    'Asset Turnover': 'X2',
    'LnAsset': 'X3'
})

def demean_panel(series, group):
    return series - series.groupby(group).transform('mean')
def run_dml_custom_ebit2(treatment_var, ml_l_model, ml_m_model, ml_l_name):
    X = pd.DataFrame({
        'X1': demean_panel(df_ebit2['X1'], df_ebit2['RIC']),
```

```
        'X2': demean_panel(df_ebit2['X2'], df_ebit2['RIC']),
        'X3': demean_panel(df_ebit2['X3'], df_ebit2['RIC'])
    })
    y = demean_panel(df_ebit2['Y'], df_ebit2['RIC'])
    d = demean_panel(df_ebit2[treatment_var], df_ebit2['RIC'])

    data = DoubleMLData.from_arrays(X, y, d)

    dml = DoubleMLPLR(data, ml_l=ml_l_model, ml_m=ml_m_model, n_folds=5)
    dml.fit()
    theta = dml.coef[0]
    se = dml.se[0]
    t_stat = theta / se
    p_val = 2 * (1 - norm.cdf(abs(t_stat)))
    ci_low = theta - 1.96 * se
    ci_high = theta + 1.96 * se

    return {
        'Variable': treatment_var,
        'ML_l': ml_l_name,
        'ML_m': 'Lasso',
        'Theta': theta,
        'Std. Error': se,
        't-stat': t_stat,
        'P-value': p_val,
        'CI Lower': ci_low,
        'CI Upper': ci_high
    }
lasso = LassoCV(cv=5, random_state=42)
rf = RandomForestRegressor(max_depth=4, random_state=42)
xgb = XGBRegressor(max_depth=4, learning_rate=0.1, n_estimators=100,
random_state=42)

results_model8 = []
for D in ['D1', 'D2', 'D3']:
    results_model8.append(run_dml_custom_ebit2(D, lasso, lasso, 'Lasso'))
    results_model8.append(run_dml_custom_ebit2(D, xgb, lasso, 'XGBoost'))
    results_model8.append(run_dml_custom_ebit2(D, rf, lasso, 'Random
Forest'))
results_df_model8 = pd.DataFrame(results_model8)
label_map = {
    'D1': 'ENV_SCORE',
    'D2': 'SOC_SCORE',
    'D3': 'GOV_SCORE'
}
results_df_model8['Variable'] =
results_df_model8['Variable'].replace(label_map)
n_obs = len(df_ebit2)
n_ric = df_ebit2['RIC'].nunique()
n_year = df_ebit2['YEAR'].nunique()
mean_years_per_ric = df_ebit2.groupby('RIC')['YEAR'].nunique().mean()
mean_ric_per_year = df_ebit2.groupby('YEAR')['RIC'].nunique().mean()
meta_info = {
    'N. observation': n_obs,
    'Entities': n_ric,
    'Time Periods': n_year,
    'Avg Obs Year per Entity': round(mean_years_per_ric, 2),
    'Avg Obs Entities per Year': round(mean_ric_per_year, 2)
}
print("\n MODEL 8 RESULTS - Ebit Margin \n")
print(tabulate(results_df_model8, headers='keys', tablefmt='github',
floatfmt=".4f"))
```

```
print("\nSAMPLE INFORMATION\n")
for k, v in meta_info.items():
    print(f"{k}: {v}")
results_df_model5['Model'] = 'MODEL 5'
results_df_model6['Model'] = 'MODEL 6'
results_df_model7['Model'] = 'MODEL 7'
results_df_model8['Model'] = 'MODEL 8'
summary_df = pd.concat([
    results_df_model5,
    results_df_model6,
    results_df_model7,
    results_df_model8
], ignore_index=True)
from tabulate import tabulate

print("\nRISULTATI RIASSUNTIVI DI TUTTI I MODELLI\n")
print(tabulate(summary_df,        headers='keys',        tablefmt='github',
floatfmt=".4f"))
```

References:

- Almeyda, R.; Darmansya, A. (2019). The influence of environmental, social, and governance (ESG) disclosure on firm financial performance. IPTEK J. Proc. Ser. 5, 278–290.
- Atan, R., Alam, M. M., Said, J., & Zamri, M. (2019). The Impacts of environmental, social, and governance factors on firm performance: Panel study of Malaysian companies.
- Athey, S., & Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. Journal of Economic Perspectives, 31(2), 3–32. Hsiao, C. (2014). Analysis of Panel Data (3rd ed.). Cambridge University Press.
- Aydoğmuş, M., Gülay, G., & Ergun, K. (2022). Impact of ESG performance on firm value and profitability. Borsa Istanbul Review, 22, S119-S127.
- Bach, R., et al. (2024). Machine learning in empirical economics: Applications and perspectives. Review of Economics and Statistics (forthcoming).
- Baltagi, B. H. (2005). Econometric Analysis of Panel Data. Wiley.
- Baltagi, Badi H. (2021). Econometric Analysis of Panel Data, 6th Edition, Wiley.
- Barnea, A., & Rubin, A. (2010). Corporate Social Responsibility as a Conflict Between Shareholders. Journal of Business Ethics, 97(1), 71–86.
- Beja-Battais, P. (2023). Overview of AdaBoost: Reconciling its views to better understand its dynamics. arXiv preprint arXiv:2310.18323.
- Belloni, Chernozhukov, and Hansen (2014) "Inference on Treatment Effects after Selection among High-Dimensional Controls." The Review of Economic Studies, 81(2), 608–650.
- Bhaskaran, R. K., Ting, I. W. K., Sukumaran, S. K., & Sumod, S. D. (2020).
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
- Brammer, S., Brooks, C., & Pavelin, S. (2006). Corporate social performance and stock returns: U.K. Evidence from disaggregate measures. Financial Management, 35, 97–116
- Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
- Capizzi, V., Gioia, E., Giudici, G., & Tenca, F. (2021). THE DIVERGENCE OF ESG RATINGS: AN ANALYSIS OF ITALIAN LISTED COMPANIES. Journal of Financial Management, Markets and Institutions, 9(2), Article 2150006. https://doi.org/10.1142/S2282717X21500067

- Carroll, A. B. (1999). "Corporate social responsibility: Evolution of a definitional construct." Business & Society, 38(3), 268-295.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- Chernozhukov, V., et al. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1–C68.
- Ching, H.Y.; Gerab, F. (2017). Sustainability reports in Brazil through the lens of signaling, legitimacy and stakeholder theories. Soc. Responsib. J.
- Clarke, D., Paris, N., & Villena-Roldán, B. (2023). (Frisch-Waugh-Lovell)': On the Estimation of Regression Models by Row. arXiv preprint arXiv:2311.15829.
- Clarke, D., & Polselli, V. (2024). Causal inference with double machine learning in panel data settings. Journal of Econometric Methods (forthcoming).
- Connelly B.L., Ketchen D.J., Slater S.F., (2011). Toward a "Theoretical textbook" for sustainability research in marketing. J.Acard Mark. Sci.
- Cristianini, N. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press.
- Damodaran, A. (2014). Applied corporate finance. John Wiley & Sons.
- De Lucia, C., Pazienza, P., & Bartlett, M. (2020). Does good ESG lead to better financial performances by firms? Machine learning and logistic regression models of public enterprises in Europe. Sustainability, 12(13), 5317.
- Demers, E.; Hendrikse, J.; Joos, P.; Lev, B. (2021). ESG did not immunize stocks during the COVID-19 crisis, but investments in intangible assets did. J. Bus. Financ. Acc. 48, 433–462.
- Dhamani, N. (2024). Introduction to Generative AI. Simon and Schuster.
- Drukker, D. M. (2003). "Testing for Serial Correlation in Linear Panel-Data Models." The Stata Journal, 3(2), 168–177.
- Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). "The impact of corporate sustainability on organizational processes and performance." Management Science.
- Environmental, social and governance initiatives and wealth creation forfirms: An empirical examination. Managerial and Decision Economics, 41(5), 710–729.
- European Commission. (2021, December 21). Questions and answers on the Carbon Border Adjustment Mechanism (CBAM). European Commission. Retrieved from

https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_21_3661/QANDA_21_3661_EN.pdf

- European Union. (2024). *ESG Ratings Regulation - Enhancing transparency and oversight*. Retrieved from K&L Gates: https://www.klgates.com
- Fawagreh, K., Gaber, M. M., & Elyan, E. (2014). Random forests: from early developments to recent advancements. Systems Science & Control Engineering: An Open Access Journal, 2(1), 602-609.
- Freeman, R. E. (1984). Strategic Management: A Stakeholder Approach. Boston: Pitman.
- 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.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.
- Gentile M., Cura F., & Teti E. (2023). Divergent ESG ratings. Journal of Sustainable Finance & Investment.
- Giannopoulos, G., Fagernes, R. V. K., Elmarzouky, M., & Hossain, K. A. B. M. A. (2022). The ESG disclosure and the financial performance of Norwegian listed firms. Journal of Risk and Financial Management, 15, 237.
- Gillian, S. L., Hartzell, J. C., & Starks, L. T. (2009). "Institutional Investors and Shareholder Activism: Evidence from CalPERS." Journal of Financial Economics.
- Gillian, S. L., Koch, A., & Starks, L. T. (2021). "The Effect of ESG Performance on Financial Performance: Evidence from a Comprehensive Sample." Journal of Financial Economics, 142(3), 766-790
- Gogoi, A., & Kao, F. J. (2024). and Marvin Minsky (1927–2016). Biomedical Imaging: Advances in Artificial Intelligence and Machine Learning, 131.Chapter 2: The AI Winter and Rise of Statistical Learning (1970s – 1980s)
- Greene, W. H. (2008). The econometric approach to efficiency analysis. The measurement of productive efficiency and productivity growth, 1(1), 92-250.
- Greene, W. H. (2012). Econometric Analysis (7th ed.). Pearson.
- Greene, W. H. (2018). Econometric Analysis (8th ed.). Pearson.
- Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (5th ed.). McGraw-Hill.

- Han, J. J., Kim, H. J., & Yu, J. (2016). Empirical study on relationship between corporate social responsibility and financial performance in Korea. Asian Journal of Sustainability and Social Responsibility, 1, 61–76.
- Harrington, P. (2012). Machine learning in action. Manning Publications.
- Hart, S. L. (1995). "A Natural-Resource-Based View of the Firm." Academy of Management Review, 20(4), 986-1014.
- Hausman, J. A. (1978). "Specification Tests in Econometrics." Econometrica, 46(6), 1251-1271.
- Hebbar, P. A., Kumar, M. M., & Mathur, A. (2022). Theory, concepts, and applications of artificial neural networks. In Applied Soft Computing (pp. 153-176). Apple Academic Press.
- Higgins, R. C. (2016). Analysis for financial management. McGraw-Hill.
- Huber, R.; Hirsch, B. (2017). Behavioral effects of sustainability-oriented incentive systems. Bus. Strategy Environ. 26, 163–181.
- Jensen, M. C., & Meckling, W. H. (1976). "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." Journal of Financial Economics, 3(4), 305-360.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., ... & Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas (pp. 87–90). IOS Press.
- Knaus, M. C. (2022). Double machine learning and matching for treatment effect estimation. Journal of Business & Economic Statistics, 40(1), 74–86.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
- Krüger, P. (2015). Corporate goodness and shareholder wealth. Journal of Financial Economics, 115(2), 304–329.
- Landi, G., & Sciarelli, M. (2019). Towards a more ethical market: The impact of ESG rating on corporate financial performance. Social Responsibility Journal, 15(1), 11–27.
- Latham & Watkins. (2024). *Key provisions in the new EU ESG Ratings Regulation*. Retrieved from https://www.lw.com
- Li, F., & Polychronopoulos, A. (2020). What a difference an ESG ratings provider makes. Research affiliates, 24, 1-15.

- Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. Journal of the American Statistical Association, 58(304), 993–1010.
- Marsat, S., & Williams, B. (2011). CSR and market valuation: International evidence, International Conference of the French Finance Association (AFFI). May 11-13, 2011.
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (pp. 51–56).
- Merchant K, Van der Stede Wim and Zoni Laura (2014). Sistemi di controllo di gestione. Pearson Italia.
- Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.
- Mooneeapen, O.; Abhayawansa, S.; Khan, N.M. (2022). The influence of the country governance environment on corporate environmental, social and governance (ESG) performance. Sustain. Acc. Manag. Policy J. 13, 953–985.
- Müller, A., & Guido, S. (2016). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media.
- Murphy, K. P. (2012). Machine learning: A probabilistic perspective. MIT Press.
- Myers, S. C. (1984) \. The Capital Structure Puzzle. The Journal of Finance, 39(3), 575–592.
- Naeem, M., Ullah, H., Shahid, J., & Kakakhel, S. J. (2022). The impact of ESG practices on firm performance: Evidence from emerging countries. Indian Journal of Economics and Business, 20, 1.
- Nollet, J., Filis, G., & Mitrokostas, E. (2016). Corporate social responsibility and financial performance: A non-linear and disaggregated approach. Economic Modelling, 52, 400–407.
- Park, E., Kim, K. J., & Kwon, S. J. (2017). Corporate social responsibility as a determinant of consumer loyalty: An examination of ethical standard, satisfaction, and trust. Journal of business research, 76, 8-13.
- Park, S.R.; Jang, J.Y. (2021). The Impact of ESG Management on Investment Decision: Institutional Investors' Perceptions of Country-Specific ESG Criteria. Int. J. Financial Stud. 9, 48.
- Pearl, J. (2009). Causality: Models, Reasoning and Inference (2nd ed.). Cambridge University Press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

- Penman, S.H. (2001). Financial Statement Analysis and Security Valuation. McGraw-Hill Education.
- Pérez, F., Granger, B. E., & Hunter, J. D. (2007). Python: An Ecosystem for Scientific Computing. Computing in Science & Engineering, 13(2), 13–21.
- Porter, M. E., & Kramer, M. R. (2018). Creating shared value: How to reinvent capitalism—And unleash a wave of innovation and growth. In Managing sustainable business: An executive education case and textbook (pp. 323-346). Dordrecht: Springer Netherlands.
- Refinitiv. (2023). An overview of Global Industry Classification Standard (GICS). Tratto da Global Industry Classification Standard (GICS): https://www.refinitiv.com/en/financial-data/reference-data/classifications/business-and- industry-classifications/global-industry-classification-standard-system.
- Robinson, P. M. (1988). Root-N-consistent semiparametric regression. Econometrica, 56(4), 931–954.
- Rosenblatt, F. (2021). The Perceptron: A Probabilistic Model for Information Storage and Organization (1958).
- Rule, A., Tabard, A., & Hollan, J. D. (2018). Exploration and explanation in computational notebooks. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1–12).
- Saygili, E., Arslan, S., & Birkan, A. O. (2022). ESG practices and corporate financial performance: Evidence from Borsa Istanbul. Borsa Istanbul Review, 22(3), 525-533.
- Sarkar, D., Bali, R., & Sharma, T. (2018). Practical machine learning with Python: A problem-solver's guide to building real-world intelligent systems. Academic press.
- Schmalensee, R. (1989). Inter-industry studies of structure and performance. Handbook of Industrial Organization.
- Shan, X., Song, Y., & Song, P. (2024). How ESG performance impacts corporate financial performance: a DuPont analysis
- Shen, Y. (2023). ESG and firm performance: A literature review. BCP Business & Management, 46, 283-288.
- Shirley, P. (2015). Interactive Data Visualization for the Web. O'Reilly Media.
- Sikacz, H., & Wołczek, P. (2018). ESG analysis of companies included in the RESPECT Index based on Thomson Reuters EIKON database. Prace Naukowe Uniwersytetu Ekonomicznego We Wrocławiu, (520), 115-127.

- Silano, S. (2016). Cosa significa ESG. Retrieved from morningstar.it: https://www.morningstar.it/it/news/148200/cosa-significa-esg.aspx
- Stoxx Ltd. (2023). Broad Indices: Stoxx Europe 600 index. Tratto da https://www.stoxx.com/document/Bookmarks/CurrentFactsheets/SXXGR.pdf
- Subasi, A. (2020). Practical machine learning for data analysis using Python. Academic Press.
- Thomson Reuters, 2017a, Thomson Reuter ESG Score. Data fact sheet, https://financial.thomsonreuters.
- Thomson Reuters, 2017b, Thomson Reuters ESG Scores. Methodology, https://financial.thomsonreuters.
- Thomson Reuters, 2017c, Thomson Reuters ESG data and solutions, https://financial.thomsonreuters.com/content/dam/openweb/documents/pdf/financial/esg-research-brochure.pdf (20.08.2017).
- Thomson Reuters, 2017d, Infostream, Q217, https://financial.thomsonreuters.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267–288.
- Tiffen, R. (2025). Two Recursive Frisch-Waugh-Lovell Algorithms and Applications in Representing Bias with Multiple Omitted Variables.
- Toosi, A., Bottino, A. G., Saboury, B., Siegel, E., & Rahmim, A. (2021). A brief history of AI: how to prevent another winter (a critical review). PET clinics, 16(4), 449-469.
- Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley.
- Turing, A. M. (2021). Computing machinery and intelligence (1950).
- UN Global Compact & The World Bank. (2005). "Who Cares Wins: Connecting Financial Markets to the Principles of Sustainable Development."
- Varian, H. R. (2014). Big Data: New Tricks for Econometrics. Journal of Economic Perspectives, 28(2), 3–28.
- Velte, P. (2017). Does ESG performance have an impact on financial performance? Evidence from Germany.
- Wang, W. (2024). Principles of Machine Learning: The Three Perspectives. Springer Nature.
- Wang, Z., Liao, K., & Zhang, Y. (2022). Does ESG screening enhance or destroy stock portfolio value? Evidence from China. Emerging Markets Finance and Trade, 58(10), 2927-2941.

- Whelan, T., Atz, U., Holt, T. V., & Clark, C. (2021). ESG and financial performance: Uncovering the relationship by aggregating evidence from 1,000 plus studies published between 2015-2020. NYU Stern Center for Sustainable Business, Rockefeller Asset Management.
- Wong, T., & Barahona, M. (2023). Deep incremental learning models for financial temporal tabular datasets with distribution shifts. arXiv preprint arXiv:2303.07925.
- Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. MIT Press.
- Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.
- Xu, J., Liu, F., & Shang, Y. (2021). R&D investment, ESG performance and green innovation performance: evidence from China. Green Innovation performance, 50(3), 737-756.
- Zaras, A., Passalis, N., & Tefas, A. (2022). Neural networks and backpropagation. In Deep Learning for Robot Perception and Cognition (pp. 17-34). Academic Press.
- https://equileap.com/equileap-reports/

- https://equileap.com/equileap-reports/

- https://finance.ec.europa.eu/news/esg-rating-activities-2024-04-10_en

- https://finance.ec.europa.eu/news/esg-rating-activities-2024-04-10_en

- https://plana.earth/academy/eu-esg-regulations

- https://plana.earth/academy/eu-esg-regulations

- https://www.bloomberg.com

- https://www.bloomberg.com

- https://www.cdp.net/en/companies/companies-scores/a-list-europe-2022

- https://www.cdp.net/en/companies/companies-scores/a-list-europe-2022

- https://www.dlapiper.com/en/insights/publications/2024/05/european-parliament-adopts-regulation-on-esg-ratings

- https://www.dlapiper.com/en/insights/publications/2024/05/european-parliament-adopts-regulation-on-esg-ratings

- https://www.hrw.org/news/2024/05/24/questions-and-answers-new-eu-law-corporate-value-chains)

- https://www.hrw.org/news/2024/05/24/questions-and-answers-new-eu-law-corporate-value-chains)

- https://www.msci.com/research-and-insights/2022-esg-trends-to-watch

- https://www.msci.com/research-and-insights/2022-esg-trends-to-watch
- https://www.skadden.com/insights/publications/2024/07/eu-adopts-legislation
- https://www.skadden.com/insights/publications/2024/07/eu-adopts-legislation
- https://www.spglobal.com/en/annual-reports/2023
- https://www.spglobal.com/en/annual-reports/2023
- https://www.sustainalytics.com/corporate-esg-survey-report
- https://www.sustainalytics.com/corporate-esg-survey-report
- https://www.sustainalytics.com/esg-research/resource/investors-esg-blog/the-state-of-esg-risk-across-industries--three-key-takeaways-from-our-annual-industry-reports?utm_source=chatgpt.com
- https://www.sustainalytics.com/esg-research/resource/investors-esg-blog/the-state-of-esg-risk-across-industries--three-key-takeaways-from-our-annual-industry-reports?utm_source=chatgpt.com.