UNIVERSITÀ
DI PAVIA

Not all that glitters is green

-

an Extension employing Instrumental Variables and Propensity Score Matching

Master Thesis

Digital Management

Prof. Dr. Paola Cerchiello

June 20, 2024

Freya Reiss

Schreinergasse 12

72415 Grosselfingen

541927

International Business and Entrepreneurship

Digital Management

European Management

Contents

List of Figures

1 Abstract

This master thesis delves into the phenomenon of Greenwashing within Eurostoxx600 companies, employing advanced econometric techniques such as Instrumental Variables (IV) and Propensity Score Matching (PSM) to enhance model robustness and derive a truly causal understanding of the drivers of Greenwashing. The study investigates the divergence between internal self-reported and external third-party ESG (Environmental, Social, and Governance) ratings, uncovering significant discrepancies indicative of potential Greenwashing practices. Key findings underscore the impact of corporate governance structures, including CSR Committees and gender diversity, on sustainability perceptions and Greenwashing risks. This research offers critical insights for policymakers and stakeholders committed to promoting genuine corporate sustainability.

Abstract (Italian)

Questa tesi di laurea magistrale esamina il fenomeno del greenwashing all'interno delle aziende dell'Eurostoxx600, utilizzando tecniche econometriche avanzate come le Variabili Strumentali (IV) e il Propensity Score Matching (PSM) per migliorare la robustezza del modello e derivare una comprensione veramente causale dei driver del greenwashing. Lo studio esplora la divergenza tra le valutazioni ESG (Ambientali, Sociali e di Governance) autoriportate internamente e quelle di terze parti esterne, rivelando discrepanze significative che suggeriscono potenziali pratiche di greenwashing. I risultati principali mettono in evidenza l'impatto delle strutture di governance aziendale, inclusi i comitati CSR e la diversitá di genere, sulle percezioni della sostenibilitá e sui rischi di greenwashing. La ricerca offre importanti intuizioni per i decisori politici e gli stakeholder impegnati a promuovere una vera sostenibilitá aziendale.

2 Introduction

In today's increasingly eco-conscious market, the phenomenon of Greenwashing has emerged as a pervasive and misleading practice that can undermine genuine efforts towards sustainability [Delmas and Burbano, 2011]. Greenwashing, a deceptive marketing tactic where companies overstate or fabricate the environmental benefits of their products or practices, not only misleads consumers but also hampers the progress toward real environmental sustainability [Lyon and Montgomery, 2015]. As the demand for sustainable goods and services continues to grow, distinguishing between authentic green initiatives and greenwashed claims becomes crucial for consumers, investors, and regulators alike [Parguel et al., 2011]. The importance of transparency and authenticity in corporate sustainability efforts cannot be overstated, as they form the foundation upon which trust and progress in environmental stewardship are built [Parguel et al., 2011], [Balluchi et al., 2020]. The paper 'Not all that glitters is green: Empirical evidence on the drivers of Greenwashing from Eurostoxx600' by Costanza Bosone, Paola Cerchiello, and Yana Kostiuk provides a profound exploration into the underlying drivers and implications of Greenwashing within the context of the Eurostoxx600 companies. This study stands out for its innovative approach to addressing the dichotomy between internal and external perceptions of sustainability, thus shedding light on the complexities of Greenwashing practices in the corporate world. The research highlights the ambiguity and misleading nature of conventional Environmental, Social, and Governance (ESG) ratings. The authors advocate for a departure from standardized 'black-box' metrics towards more comprehensive and transparent measures [Pope and Wæraas, 2016]. By integrating internally disclosed and externally generated data, the study unveils the disparity between a company's self-reported sustainability efforts and the external perception of these efforts. Through a rigorous regression analysis, the paper reveals the incentives and deterrents that companies face in the area of Greenwashing. The paper emphasizes the critical role of Corporate Social Responsibility (CSR) Committees and gender diversity within corporations in influencing sustainability practices and perceptions. A unique contribution of the paper is its dissection of the variance between internal and external ESG scores, which serves as a proxy for the likelihood of a company engaging in Greenwashing practices. This approach not only provides insights into the effectiveness of voluntary CSR strategies in mitigating Greenwashing but also underscores the necessity for tangible sustainability initiatives and enhanced gender diversity to foster positive external and internal evaluations of a company's commitment to sustainability. The findings from Bosone, Cerchiello, and Kostiuk's study from the year 2024 make a compelling case for the need to refine and rethink the metrics and methodologies used to assess corporate sustainability. By illuminating the drivers of Greenwashing and the discrepancy between perceived and actual sustainability practices, the paper contributes significantly to the ongoing dialogue on sustainable finance and corporate social responsibility [Dura and Ghicajanu, 2012]. This research not only enriches our understanding of Greenwashing but also paves the way for future investigations aimed at fostering genuine environmental stewardship within the corporate sector. The rise of Greenwashing reflects a broader trend within the global marketplace where environmental sustainability has become a key competitive edge. As companies vie for the attention of increasingly environmentally conscious consumers, the temptation to embellish or overstate green credentials has grown. This trend not only poses a challenge to consumer trust but also complicates the landscape for investors and regulators striving to support genuine sustainability initiatives [Kanter, 2009]. Against this backdrop, the detailed investigation conducted by Bosone, Cerchiello, and Kostiuk provides essential insights into the nuanced dynamics of Greenwashing, offering a critical toolset for distinguishing between substantive and symbolic sustainability efforts [Mattis, 2008]. Utilizing an innovative methodology that leverages both internal and external ESG scores, the authors embark on a meticulous analysis to uncover the layers of Greenwashing within the corporate sector. By analyzing a dataset comprising over 5000 observations from the Eurostoxx600 companies, the research unravels the complex interplay between corporate profitability, CSR practices, and external perceptions of sustainability. The study's findings reveal that the presence of CSR Committees, often hailed as beacons of corporate sustainability, does not necessarily correlate with a reduction in Greenwashing practices [Dura and Ghicajanu, 2012]. Instead, the study suggests that these Committees may inadvertently contribute to the phenomenon of Greenwashing, underscoring the importance of substantiating CSR initiatives with tangible outcomes. The implications of this research extend far beyond the academic realm, offering valuable lessons for corporations, policymakers, and consumers alike. The study highlights the critical need for corporations to integrate sustainability and social initiatives into their core business strategies, moving beyond mere symbolic actions. Policymakers and regulators may find the research's insights into the limitations of current ESG metrics and the potential for Greenwashing Instrumental in designing more effective regulatory frameworks. Meanwhile, consumers equipped with a deeper understanding of Greenwashing practices can make more informed choices, supporting companies that demonstrate a true commitment to environmental responsibility. The paper by Bosone, Cerchiello, and Kostiuk not only enriches the discourse on Greenwashing but also acts as a clarion call for increased transparency, accountability, and authenticity in corporate sustainability efforts. As the global community continues to grapple with the pressing challenges of environmental degradation and climate change, fostering an environment where true green initiatives are celebrated and Greenwashing is rigorously challenged becomes imperative. The path to sustainable development requires collective action, and research such as this has a key role to play in guiding the way forward and ensuring that substance prevails over symbolism in the quest for sustainability. The regression model utilized in the study effectively captures the complex interplay between a company's self-reported sustainability efforts (internal scores) and the external perception of these efforts (external scores). By analyzing the Delta between these scores as a proxy for Greenwashing risk, the authors uncover insightful correlations that shed light on the drivers of deceptive environmental practices. This approach not only highlights the significance of transparency in ESG reporting, it helpes us understand how crucial it is in todays world to implement working mechanism to tackle the problems of sustainable and social governance.

2.1 Instrumental variables and Propensity Score Matching

While the initial model provides a solid foundation for understanding the dynamics of Greenwashing, the introduction of Instrumental Variables (IV) and Propensity Score Matching (PSM) techniques can further enhance its robustness and explanatory power [Angrist and Pischke, 2009, Rosenbaum and Rubin, 1983]. The inclusion of IV methods addresses potential endogeneity issues, ensuring that the relationships identified are not confounded by omitted variable bias or reverse causality. This is particularly relevant in the context of Greenwashing, where the causality between corporate practices and sustainability perceptions can be complex and bidirectional.

Instrumental Variables are used to isolate the exogenous variation in the explanatory variables, thereby providing more reliable estimates of causal relationships [Angrist and Krueger, 2001]. In the context of this study, an IV approach could be employed to disentangle the influence of external pressures (e.g., regulatory changes, consumer demand) on Greenwashing practices, ensuring that the observed relationships are not merely reflective of underlying market trends or company-specific factors not accounted for in the original model.

PSM, on the other hand, allows for the comparison of treatment and control groups on a common set of covariates, reducing the bias from confounding variables. By matching companies with similar characteristics but differing in their engagement with CSR practices or sustainability activities, PSM can offer deeper insights into the causal impact of such practices on Greenwashing risk. This technique is particularly useful for evaluating the effectiveness of CSR Committees and other governance mechanisms in mitigating the propensity for deceptive environmental claims [Furlow, 2010, Balluchi et al., 2020, Gatti et al., 2019].

In conclusion, extending the original regression model through the application of IV and PSM can significantly enrich the analysis, providing a more nuanced and causally informed understanding of the factors driving Greenwashing. These methodological enhancements not only bolster the rigor of the study but also offer valuable insights for practitioners, policymakers, and researchers aiming to promote authentic corporate sustainability practices. Through such advanced analytical techniques, the study stands to contribute even more profoundly to the critical discourse on sustainable finance and corporate responsibility, paving the way for more effective strategies to combat Greenwashing in the corporate world [Delmas and Burbano, 2011, Du, 2015, Lyon and Montgomery, 2015].

2.2 Theoretical Background

Given the intricate nature of sustainable finance, corporate governance, and the phenomenon of Greenwashing, this discourse aims to unpack the nuanced landscape these elements cohabit, drawing upon a rich tapestry of academic insights. The subsequent exposition synthesizes key findings from the literature, elucidating the multifaceted interplay between Environmental, Social, and Governance (ESG) considerations in corporate management and the pervasive issue of Greenwashing.

At the heart of sustainable finance lies the imperative to integrate ESG considerations into the corporate fabric, a movement driven by the escalating demand for companies and projects prioritizing sustainability [Liang and Renneboog, 2020]. This integration is not merely environmental stewardship but extends to ensuring social welfare and promoting governance equality. Notably, Liang and Renneboog [2020] underscore the strategic significance of embedding ESG factors into financial decision-making and investors' portfolio choices. However, this noble pursuit is often mired by Greenwashing, a deceptive practice where corporations disseminate misleading information about their ESG performance. Balluchi et al. [2020] highlight how such practices aim to skew stakeholders' perceptions favorably, despite a potential lack of substantial contributions to environmental or social welfare [Parguel et al., 2011].

A deeper exploration into the realm of Greenwashing reveals a complex array of drivers and consequences. Delmas and Burbano [2011] and Furlow [2010] pioneer this exploration, delineating how both external market pressures and internal organizational dynamics coalesce to foster an environment ripe for Greenwashing. Companies face immense pressure from consumers and investors to present themselves as environmentally conscious entities. This external pressure, coupled with internal factors such as corporate culture and communication strategies, often leads companies towards embellishing their sustainability efforts. The ramifications of Greenwashing extend beyond ethical concerns, as Furlow [2010] and Du [2015] assert, noting its potential to detrimentally impact financial performance and erode consumer trust.

Corporate governance emerges as a critical lens through which the intricacies of Greenwashing and ESG practices can be examined [Cherry and Sneirson, 2010]. The composition and actions of corporate boards, particularly concerning gender diversity and the establishment of CSR Committees, are pivotal in steering company-wide sustainability agendas. Azmat and Rentschler [2017] illuminate the positive influence of female leadership on corporate environmental investments and ethical climate perceptions, suggesting a nuanced relationship between governance structures and sustainability outcomes. Nevertheless, the role of CSR practices in genuinely advancing sustainability efforts remains contentious. While CSR initiatives are often lauded for their potential to contribute to societal and environmental well-being, critics like Cherry and Sneirson [2010] and Gennari and Salvioni [2019] argue that such measures can sometimes serve as facades, masking superficial engagement with sustainability.

The discourse surrounding sustainable finance and Greenwashing inevitably confronts the challenge of standardisation, or rather, the lack thereof [Imbens, 2004]. The heterogeneity in definitions and interpretations of ESG indicators complicates the assessment of corporate sustainability practices. Mattis [2008], Escrig-Olmedo et al. [2019] and Berg et al. [2022] advocate for a unified methodology and more comprehensive measures to navigate this complex terrain, highlighting a pressing need for clarity and consistency in evaluating corporate ESG performance.

Amidst these theoretical challenges, Propensity Score Matching and Instrumental Variables stand out as potent analytical tools capable of shedding light on the nuanced dynamics at play. PSM, as elucidated by Rosenbaum and Rubin [1983], offers a rigorous approach to controlling for selection bias, enabling more accurate estimation of the effects of Greenwashing by matching firms based on observable characteristics [Hirano and Imbens, 2004]. This methodological rigor paves the way for a clearer understanding of the impact of deceptive sustainability practices on various outcomes, including financial performance and stakeholder trust.

Similarly, the IV approach, championed by Angrist and Pischke [2009], addresses endogeneity concerns inherent in the analysis of Greenwashing and corporate governance. By leveraging Instruments that are correlated with potentially endogenous explanatory variables but not with the outcome variable, IV facilitates the identification of causal relationships [Hirano and Imbens, 2004]. This is particularly pertinent in unraveling the effects of governance practices or regulatory changes on sustainability reporting and performance, offering a pathway to discern the genuine impact of corporate actions on sustainability outcomes.

The synthesis of insights from PSM and IV methodologies underscores the complexity of accurately assessing the efficacy of CSR initiatives and governance reforms in promoting authentic sustainability practices. By moving beyond mere correlations to understanding causal relationships, these analytical tools enrich the discourse on sustainable finance, Greenwashing, and corporate governance. They highlight the need for a sophisticated evaluation framework that can navigate the nuanced interdependencies between corporate actions, regulatory environments, and stakeholder perceptions.

2.3 Data Description

To evaluate our theory, the article's authors; 'Not all that glitters is green' compiled a dataset comprising over 5,000 data points. They selected companies from the Eurostoxx600. Recognizing the limitations and general opaqueness of traditional ESG rating methodologies, they suggest moving away from the conventional, opaque scores offered by financial entities. Instead, an innovative and holistic measurement approach was introduced, that thoroughly analyzes both data disclosed internally by companies and data generated externally by stakeholders. This approach was made possible through the partnership with FinScience, a Milan-based fintech firm founded in 2017 by ex-Google senior managers and data experts, granting them access to their comprehensive sustainability scores [Bosone et al., 2024]. These scores reflect a dual perspective, evaluating a company's internally disclosed data (internal ESG score) and data generated by external stakeholders (external ESG score), on a scale from 0 to 100.

The data utilized is categorized into internal and external types, depending on the company willingness to reveal this kind of sensitive data. The internal ESG score evaluates a company's performance based on its self-reported and publicly disclosed data. In contrast, the external ESG score measures perceived company performance in ESG aspects, based on data from external stakeholders. Furthermore, FinScience's ESG Scores offer an assessment of corporate sustainability by evaluating company performance against the 17 UN Sustainable Development Goals (SDGs) as outlined in the UN 2030 Agenda Nations [2022]. These goals, along with 169 specific targets, provide a broad framework for governments to steer their policies and investments to meet current and future societal and environmental challenges [Fin, 2022].

The internal scores adopt a comprehensive approach, incorporating evaluations from Sustainability and CSR reports, corporate websites, and sustainability memberships or affiliations. For a membership to be deemed credible, a company is required to meet rigorous criteria concerning environmental performance or commitment. This includes having at least two reputable memberships focused on environmental issues. Additionally, these scores consider certifications, requiring at least two related to environmental standards, such as ISO 14001 and ISO 50001, to further affirm a company's dedication to environmental norms.

External data is gathered from esteemed sources, offering insights into both positive and negative sentiments regarding sustainability-related news about the company. This information is compiled from specialized websites, NGOs, vertical websites, and mainstream news outlets. The collected metrics also encompass controversies and reviews, essential for identifying cases where companies are penalized or fined for environmental breaches or their involvement in significantly polluting activities. Social media platforms are utilized as proxies for various digital assets, including news, forums, blogs, and reviews, facilitating a dynamic stream of information about a company's daily operations. This method ensures the capture of real-time insights from diverse digital channels, enabling a thorough understanding of a company's environmental impact and reputation.

Given that internal sources already include standard ESG metrics from renowned providers such as Refinitiv, S&P, and Bloomberg, these metrics are deliberately omitted from the estimation process to prevent redundancy. This strategy enhances the assessment's comprehensiveness and subtlety regarding sustainability factors for the evaluated companies.

By combining internal and external scores, an indicator of Greenwashing risk is formulated. Specifically, the difference between the internal and external scores, henceforth denoted as Delta Δ , serves as a measure of the potential of a company to disseminate misleading information about its sustainability efforts. A higher Δ indicates an increased risk, signifying a divergence between a company's external perception and its internal stakeholders' perspective. This calculation is performed only for companies with both internal and external scores available, resulting in a dataset of 467 companies from the Eurostoxx600 index.

$$\Delta = internal\_score - external\_score \tag{1}$$

In the data set also traditional measures for financial performance are accounted for, such as Revenue per Share (RPSH), ROA, ROE, Market Capitalization and Credit

Rating. All companies are actors who are responsible for broader stakeholders' interests. Hereby the authors of the original paper not just consider the fair governance mechanism but also the positive impact of female representation in CSR Committees, boards and in the general Company [Tirole, 2010, Ayuso et al., 2014, Aguilera et al., 2008], therefore including the variables Board Size, percentage of females on the board, percentage of female employees and the general existence of a CSR Committee. Because the effectivenes of CSR Committees is at the center of the analysis the CSR reports are deeply investigated, focusing on:

- Quality of sustainability reports, meaning if there is a third party involved controlling and verifying the initial Reports
- The presence of sustainability or CSR areas on the official company website. This point tries to assess the level of communication the company seeks to communicate its efforts

The authors control for several parameters, first of all for the company size, by establishing a four-class system which ranks the firms from small to large. The sectors of the firms which are controlled in the regression are: communication services, consumer discretionary, consumer staples, energy, financial, healthcare, industrial, materials, and utilities following the Global Industry Classification Standards (GICS), see Table 23 for further references.

As a last step the whole dataset is then standardised. Now the Author's started with the implementation of their baseline Model for the three Scores:

$$\begin{aligned}\Delta = \alpha_1 &+ \beta_1 RPSH + \beta_2 ROA + \beta_3 ROE + \beta_4 Market\_Cap + \beta_5 Credit \\ &+ \beta_6 CSR\_COMM + \beta_7 Board\_size + \beta_8 Women\_board + \beta_9 Women\_emp \\ &+ \beta_{10} FESize + \beta_{11} FESector + \epsilon_i\end{aligned} \tag{2}$$

$$\begin{aligned} Internal = \alpha_1 + \beta_1 RPSH + \beta_2 ROA + \beta_3 ROE + \beta_4 Market\_Cap + \beta_5 Credit \\ + \beta_6 CSR\_COMM + \beta_7 Board\_size + \beta_8 Women\_board + \beta_9 Women\_emp \\ + \beta_{10} FESize + \beta_{11} FESector + \epsilon_i \end{aligned} \tag{3}$$

$$\begin{aligned} External = \alpha_1 + \beta_1 RPSH + \beta_2 ROA + \beta_3 ROE + \beta_4 Market\_Cap + \beta_5 Credit \\ + \beta_6 CSR\_COMM + \beta_7 Board\_size + \beta_8 Women\_board + \beta_9 Women\_emp \\ + \beta_{10} FESize + \beta_{11} FESector + \epsilon_i \end{aligned} \tag{4}$$

In the regression analysis conducted by the authors, traditional financial indicators enter the equation as regressors from β_1 to β_5. These indicators, including Revenue per Share ($RPSH$), Return on Assets (ROA), Return on Equity (ROE), Market Capitalization ($Market\_Cap$), and Credit Rating ($Credit$), are considered. Moving forward to β_6 to β_9, governance-level fairness indicators, namely: CSR Committee (CSR_COMM), Board Size ($Board\_size$), Women on board($Female\_board$), and Women employees ($Women\_emp$), are incorporated. Each specification encompasses all companies indexed by i from 1 to N in the dataset, see in the Appendix Table 21 for an overview of the Variables and the companies . The error term is denoted as ϵ_i.

To ensure the reliability and robustness of the findings, the authors adopt a rigorous methodology. Initially, standard errors are clustered at the sector level, aiming to address potential correlations or heteroscedasticity within sectors. This approach allows for consideration of any sector-specific patterns that could influence the accuracy of the estimates, thus ensuring a more precise and robust evaluation of the relationships under scrutiny. Additionally, to account for heterogeneity within the dataset, fixed effects are introduced at both the size level (FESize) and sector-level (FEsector). By incorporating these fixed effects, the authors account for unobservable factors that may systematically vary within the size of companies or be unique to each economic sector [Bosone et al., 2024]. This meticulous handling of heterogeneity strengthens the robustness of the findings.

In addition to equation 2, which operates on (Δ), the authors extend the analysis to two alternative specifications by introducing an internal score (equation 3) and an external score (equation 4). Each model undergoes standard perturbation by incorporating fixed effects at both the size-level and sector-level.

Delta Score

The research examines the causes of Greenwashing in companies, utilizing equation 2 (Δ) and presenting the results in Table 1.

Table 1: Regression Results - Delta Score

	Model A	Model B	Model C
Dependent Var.:	Delta	Delta	Delta
Constant	-0.5476** (0.1617)		
CSR Committee	0.6704*** (0.1139)	0.7002*** (0.1144)	0.5762** (0.1283)
RPSH	-0.0566*** (0.0112)	-0.0491** (0.0141)	-0.0513** (0.0123)
ROA	0.0278 (0.0605)	-0.0061 (0.0562)	0.0068 (0.0441)
ROE	0.0183. (0.0087)	0.0233* (0.0086)	0.0182* (0.0078)
Board Size	0.0874. (0.0442)	0.1193. (0.0532)	0.1240. (0.0574)
Female Board	-0.0537 (0.0408)	-0.0430 (0.0405)	-0.0243 (0.0402)
Women Employment	-0.2287* (0.0720)	-0.2319* (0.0749)	-0.2090*** (0.0352)
Market Cap	0.0592 (0.0510)	0.0814 (0.0469)	0.0208 (0.0423)
Credit Rating	-0.0073 (0.0070)	-0.0041 (0.0072)	0.0003 (0.0074)
Fixed-Effects:	——————-	——————	——————-
Size	No	Yes	Yes

(continued)

	Model A	Model B	Model C
GICS_cluster	No	No	Yes
______________	________________	________________	________________
S.E.: Clustered	by: GICS_cluster	by: GICS_cluster	by: GICS_cluster
Observations	467	467	467
R2	0.11391	0.12394	0.20632
Within R2	–	0.12142	0.07925

It begins with a standard OLS model (column 1), adds size-level fixed effects (column 2), and then sector-level fixed effects (column 3). The researchers identify a negative correlation between revenues per share and Δ, indicating that companies may be more inclined to greenwash as profits decline [Bowen and Aragon-Correa, 2014] and [Du, 2015]. This finding aligns with previous research linking Greenwashing to market and investor responses. Despite mixed evidence, return on equity (ROE) shows a significant positive correlation, suggesting a nuanced relationship between financial performance and sustainability efforts ([Alareeni and Hamdan, 2020] and [Gregory, 2021]). Market Capitalization and Credit Rating did not show significant correlations with Greenwashing, suggesting limited variation in credit ratings among Eurostoxx600 companies ([Kim and Li, 2021] and [Jang et al., 2020]).

The study also finds a significant negative correlation between the percentage of female employees and the risk of Greenwashing, supporting the idea that gender diversity can enhance corporate sustainability (Azmat and Rentschler [2017] ;Jiang and Akbar [2018]). The size of the board correlates positively with the risk of Greenwashing, implying larger boards may face inefficiencies [Raheja, 2005]. CSR Committees, while intended to support sustainability, show a positive correlation with Greenwashing, indicating their presence alone does not guarantee genuine sustainability efforts (Aggarwal and Kadyan [2014]; Bazillier and Vauday [2009]; Sterbenk et al. [2022]).

Overall, the researchers highlight a complex interaction between financial distress, gender diversity, and external pressures in driving Greenwashing practices.

Internal vs. External Score

The study delves into the perceptions of sustainability from both internal and external stakeholders, employing equations for internal (equation 3) and external scores (equation 4) and factoring in sector and size-level effects. Internal scores are derived from companies' self-reported sustainability activities, including quality checks, affiliations, and involvement in sustainability-related controversies, serving as a proxy for companies' self-assessment of sustainability performance. The regression results are displayed in Table 2. Notably, a negative correlation was found between revenues per share and internal scores, indicating that companies may enhance their self-reported sustainability evaluations in times of financial downturns. Return on equity (ROE) displayed a positive correlation with internal sustainability scores, suggesting its influence on companies' self assessment, although its effectiveness as a sole performance metric is debated (Aguilera et al. [2015]; Arditti [1967]; Gregory [2021]).

Table 2: Regression Results - Internal Score

	Model A	Model B	Model C
Dependent Var.:	Internal	Internal	Internal
Constant	-0.8197** (0.1727)		
CSR Committee	0.9476*** (0.1542)	0.9180*** (0.1502)	0.8451*** (0.1577)
RPSH	-0.0368* (0.0114)	-0.0466* (0.0141)	-0.0469** (0.0114)
ROA	-0.0050 (0.0844)	0.0404 (0.0942)	0.0331 (0.0911)
ROE	0.0488*** (0.0036)	0.0451*** (0.0045)	0.0440*** (0.0060)
Board Size	0.2231*** (0.0377)	0.1811** (0.0387)	0.1887** (0.0438)

(continued)

	Model A	Model B	Model C
Female Board	0.0118 (0.0545)	-0.0022 (0.0516)	0.0133 (0.0518)
Women Employment	-0.1089. (0.0513)	-0.1116. (0.0542)	-0.0597 (0.0324)
Market Cap	0.1028 (0.0606)	0.0623 (0.0569)	0.0414 (0.0508)
Credit Rating	-0.0050 (0.0068)	-0.0094 (0.0064)	-0.0068 (0.0067)
Fixed-Effects:	————————	————————	————————
Size	No	Yes	Yes
GICS_cluster	No	No	Yes
____________	______________	______________	______________
S.E.: Clustered	by: GICS_cluster	by: GICS_cluster	by: GICS_cluster
Observations	467	467	467
R2	0.18366	0.19987	0.24035
Within R2	–	0.13430	0.11071

The research also highlighted a positive correlation between Board Size and internal sustainability perceptions, indicating that larger boards might foster a better internal view of sustainability efforts, though this relationship's implications on agency costs remain contested (Raheja [2005]; Kathy Rao et al. [2012]). The percentage of women on the board did not significantly impact the internal scores, contrasting with the positive correlation found with the presence of CSR Committees, which suggests their perceived importance in sustainability performance despite caution from the literature regarding their actual effectiveness in CSR engagement (Bazillier and Vauday [2009]).

External scores, based on data from stakeholders like NGOs and news outlets, showed diminished significance of traditional financial indicators (Revenues per Share and ROE) in external perceptions of sustainability. However, a strong financial performance was still viewed positively. The research discovered that external stakeholders place higher value on tangible outcomes, such as gender equality in employment, over formal achievements like CSR Committees' establishment. This finding underscores a divergence in priorities between internal and external perceptions, with external stakeholders focusing more on observable indicators of sustainability efforts.

Table 3: Regression Results - External Score

	Model A	Model B	Model C
Dependent Var.:	External	External	External
Constant	-0.2039 (0.1165)		
CSR Committee	0.1801* (0.0751)	0.1010 (0.0611)	0.1923* (0.0629)
RPSH	0.0371* (0.0140)	0.0142 (0.0123)	0.0170 (0.0125)
ROA	-0.0473 (0.0482)	0.0572 (0.0527)	0.0302 (0.0605)
ROE	0.0330* (0.0136)	0.0211 (0.0116)	0.0273* (0.0094)
Board Size	0.1453* (0.0532)	0.0480 (0.0506)	0.0504 (0.0536)
Female Board	0.0902 (0.0555)	0.0575 (0.0533)	0.0493 (0.0582)
Women Employment	0.1979* (0.0765)	0.1994* (0.0652)	0.2304** (0.0610)
Market Cap	0.0392 (0.0570)	-0.0423 (0.0249)	0.0198 (0.0254)
Credit Rating	0.0047 (0.0113)	-0.0053 (0.0081)	-0.0086 (0.0072)
Fixed-Effects:			
Size	No	Yes	Yes
GICS_cluster	No	No	Yes
S.E.: Clustered	by: GICS_cluster	by: GICS_cluster	by: GICS_cluster
Observations	467	467	467
R2	0.09895	0.18485	0.24248
Within R2	–	0.06087	0.05762

The study concludes that while internal corporate governance actors emphasize structural features like financial performance and CSR Committees, external stakeholders prioritize visible commitments to gender equality and sustainability practices. This divergence highlights the evolving expectations and scrutiny companies face in their sustainability endeavors, aligning with Delmas and Burbano [2011] on the broader trend where observable and measurable indicators are given precedence by external stakeholders in their assessment of a company's sustainability performance.

3 Instrumental variables

3.1 General Framework

The Instrumental Variable approach is a widely utilized methodology in econometrics and statistics for addressing issues of endogeneity in regression models. The significance of the IV approach extends beyond mere technical correction; it embodies a deeper philosophical commitment to uncovering true causal relationships in the social sciences, economics, health research, and beyond. By providing a methodological bridge between observational data and causal inference, the IV approach enables researchers to explore otherwise inaccessible questions, especially when controlled experiments are infeasible or unethical. This comprehensive analysis aims to elucidate the theoretical foundations, practical applications, challenges, and broader implications of the IV approach, underscoring its critical role in empirical research [Angrist and Pischke, 2009]. Endogeneity, where explanatory variables are correlated with the error term, poses a significant challenge in causal inference, leading to biased and inconsistent estimates in ordinary least squares (OLS) regressions. IV and 2SLS methods offer a way to circumvent these issues, enabling researchers to uncover causal effects even in complex settings [Angrist and Krueger, 2001]. Endogeneity arises in three main forms: omitted variable bias, measurement error, and simultaneity. Omitted variable bias occurs when a model fails to include one or more relevant variables that influence both the independent and dependent variables. Measurement errors in explanatory variables leads to attenuation bias, where the estimated coefficients are biased towards zero. Simultaneity arises in models where the causal direction between variables is not clear-cut, typically seen in supply and demand models. Each of these issues invalidates the key OLS assumption that explanatory variables are exogenous, resulting in biased OLS estimates [Angrist and Pischke, 2009].

The IV approach allows for more robust estimates of causal relationships by using Instruments that are correlated with the endogenous explanatory variables but uncorrelated with the error term. The key property that makes an Instrument valid is its ability to affect the dependent variable only through its correlation with the endogenous explanatory variables. To formally describe the IV approach, consider a basic linear model:

$$y = \beta_0 + \beta_1 x + \epsilon, \tag{5}$$

where y is the dependent variable, x is the endogenous explanatory variable, β_0 and β_1 are parameters to be estimated, and ϵ is the error term. The endogeneity of x arises when it is correlated with ϵ, i.e., $\text{Cov}(x, \epsilon) \neq 0$, which leads to biased and inconsistent estimates of β_1 when using ordinary least squares (OLS). The IV approach introduces an Instrument z that satisfies two key conditions:

- Relevance: The Instrument z is correlated with the endogenous variable x, i.e., $\text{Cov}(z, x) \neq 0$.

- Exogeneity: The Instrument z is uncorrelated with the error term ϵ, i.e., $\text{Cov}(z, \epsilon) = 0$.

Given these conditions, the IV estimator of β can be expressed as:

$$\hat{\beta}_{IV} = \frac{Cov(z, y)}{Cov(z, x)}, \tag{6}$$

which provides a consistent estimate of β even if x_i is endogenous [Angrist and Pischke, 2009].

The practical application of these conditions requires careful consideration and rigorous testing. The relevance condition can be assessed through statistical measures such as the F-statistic in the first stage of a two-stage least squares (2SLS) regression, where a higher value indicates a stronger Instrument. However, satisfying the exogeneity condition is more challenging, as it hinges on substantive theoretical arguments and empirical strategies to argue that the Instrument does not share common causes with the outcome variable beyond its association with the endogenous predictors. Given these conditions, the IV estimator for β_1 can be obtained through two-stage least squares (2SLS). The first stage involves regressing the endogenous variable x on the Instrument z (and possibly other exogenous variables):

- First Stage: Regress the endogenous variable x on the Instrument z (and possibly other exogenous variables) to obtain the predicted values $\hat{x}$:

$$x = \pi_0 + \pi_1 z + \nu, \tag{7}$$

  where π_0 and π_1 are parameters, and ν is the error term. For now, the variable z is assumed to be the only reason why the error term ϵ and x are correlated.

- Second Stage: The predicted values $\hat{x}$ from the first stage regression are then used in the second stage as explanatory variables, where y is regressed on the fitted values $\hat{x}$:

$$y = \alpha_0 + \alpha_1 \hat{x} + \eta, \tag{8}$$

  where α_0 and α_1 are parameters, and η is the error term. The IV estimate of β_1 is given by the estimate of α_1, which is consistent under the conditions of Instrument validity.

The use of IVs can significantly robustify models against the biases introduced by endogeneity, allowing for a more reliable estimation of causal effects. However, the choice of Instruments is crucial, as invalid Instruments can lead to even more biased estimates. Therefore, careful consideration and testing of the relevance and exogeneity conditions are essential in the application of the IV approach [Angrist and Pischke, 2009]. Applying the IV approach in empirical research is challenging, chiefly among them the identification of suitable Instruments. The quest for Instruments that genuinely satisfy both relevance and exogeneity is a substantive endeavor that often involves creative thinking and deep theoretical understanding. Researchers may look to natural experiments, policy changes, or unique variations in the data as potential Instruments. Yet, the validity of these Instruments must be meticulously justified and empirically tested. One of the main challenges in IV analysis is the risk of weak Instruments, where the correlation between the Instrument and the endogenous explanatory variable is too low to provide a reliable estimate. Weak Instruments can lead to biased estimates and inflated standard errors, compromising the credibility of the causal inference [Angrist and Pischke, 2009]. Diagnostic tests, such as the before mentioned F-statistic in the first-stage regression, are essential tools for assessing Instrument strength and guarding against the weak Instrument problem. Furthermore, the interpretation of IV estimates requires careful consideration [Angrist and Pischke, 2009]. IV estimates represent local average treatment effects (LATE), applicable only to a specific subset of the population - namely, the compliers for whom the Instrument effectively induces variation in the exposure. This limitation underscores the importance of understanding the context and mechanism through which the Instrument operates, as it may limit the generalizability of the findings.

The IV approach is particularly useful in addressing the endogeneity issue, where there's a two-way causation or an omitted variable bias between independent variables and the outcome. In the context of Greenwashing, endogeneity might arise due to reverse causality (e.g., companies with poor real ESG performance might be more inclined to greenwash) or omitted variables (e.g., unobserved corporate culture or values influencing both ESG efforts and Greenwashing simultaneously).

The use of IV in the Greenwashing context can help identify causal effects by leveraging variables that influence the predictor of interest (e.g., presence of CSR Committees) but are not directly related to the outcome variable (e.g., propensity to Greenwash). An Instrumental variable could be environmental or regulatory changes or shocks in public awareness about sustainability issues, which might affect companies' tendencies to engage in Greenwashing but are not directly related to their actual ESG performance. In this application of the Instrumental variable model, initially the variable which is suspected to introduce endogenity is defined, which, in this context, is the indicator variable for the presence of CSR Committees. Consequently, a variable that could serve as an Instrument for this endogenous variable is seeked. Subsequently, three potential Instruments will be explored.

Since the CSR_COMM variable is binary, the standard two-stage least squares (2SLS) regression model is not appropriate for analyzing its effects. Instead, a two-stage residual inclusion (2SRI) approach got employed, which is specifically designed for handling endogeneity in models where the outcome of the first stage or the endogenous regressor is binary. The 2SRI method, also known as the control function approach, involves two main stages:

1. First Stage: In the first stage, the endogenous binary variable (CSR_COMM) is modeled using a binary choice model, typically a probit or logistic regression. Here the endogenous binary variable (CSR_COMM) is modeled using a logistic regression. This model predicts the probability of CSR_COMM being 1, based on the Instrumental variable ($mean\_trend$, CSR_Rating, or SDR_Rating) and other relevant covariates. The key output from this stage are the estimates of the endogenous variable and the residuals from the regression, which capture the portion of CSR_COMM that is unexplained by the Instrumental variable and covariates:

$$\text{logit}(\Pr(CSR\_COMM = 1)) = \pi_0 + \pi_1 z + \nu, \tag{9}$$

   where π_0 and π_1 are parameters, and ν is the error term. The predicted probability from this logistic model is used to obtain the residuals, which are included in the second stage.

2. Second Stage: The main outcome variable (e.g., Δ for Greenwashing) is regressed on the predicted values of CSR_COMM (from the first stage) and the residuals from the first stage to control for endogeneity:

$$y = \alpha_0 + \alpha_1 CSR\_\hat{C}OMM + \gamma\hat{\nu} + \epsilon, \tag{10}$$

where α_0 and α_1 are parameters, $CSR\_\hat{C}OMM$ is the predicted value from the first stage, ϵ is the error term, and $\hat{\nu}$ is the residual from the first stage.

Importantly, the residuals from the first stage are also included as an additional regressor in this model. The inclusion of these residuals helps to control for any remaining endogeneity in CSR_COMM, ensuring that the estimated effect of CSR_COMM on the outcome is unbiased.

In the 2SRI framework, the interpretation of the parameters differs slightly from the 2SLS approach: First, α_1, as described in the model is the Effect of $CSR\_\hat{C}OMM$. The coefficient α_1 represents the causal effect of the endogenous binary variable CSR_COMM on the outcome variable y (Delta Score, Internal Score or External Score), adjusted for the endogeneity. It captures the impact of CSR_COMM on y after accounting for the Instrument's influence.

Secondly the coefficient on the residuals ($\hat{\nu}$) from the first stage regression, often denoted as γ, indicates the remaining endogeneity in CSR_COMM that was not explained by the Instrument. A significant coefficient for $\hat{\nu}$ suggests that there is still some endogeneity in CSR_COMM that gets account for with the residuals. The Role of Residuals in the Second Stage can be better explained if we look at the case of significant $\hat{\nu}$. If the residuals from the first stage are significant in the second stage regression, it indicates that there is remaining endogeneity in the explanatory variable. The significance of these residuals suggests that the error term from the first stage regression is correlated with the error term in the second stage regression. This confirms that the original explanatory variable was endogenous. The significance of the residuals also indirectly validates the Instruments used. If the residuals are significant, it implies the Instruments are relevant in explaining the endogenous variation. Including these residuals helps to correct for the bias caused by endogeneity. It ensures that the second stage regression accounts for the part of the variation in the endogenous variable that is correlated with the error term, thereby providing more reliable and consistent estimates. The inclusion of the residuals helps to ensure that the estimate of α_1 is unbiased by controlling for this remaining endogeneity. The inclusion of significant residuals helps correct the bias that would otherwise affect the estimates of the coefficients of the endogenous variable. Significant residuals do not undermine the validity of the model; rather, they indicate that the endogeneity problem is being acknowledged and corrected.

Thus, the 2SRI approach ensures that the estimated effect of the endogenous binary variable on the outcome is consistent and unbiased, making it suitable for models with binary endogenous variables.

3.2 Instrumental variable: Mean Sea Level trend per Country

The mean sea level trend analysis is obtained by averaging several measurement metrics from various locations around the world. The British Oceanographic Data Center provides the data for the Instrument. This factor is calculated on a per-country basis, providing an average increase or decrease in sea level for each country. Countries without coastlines are assigned a mean sea level trend of zero, as there is no change in water level in these areas. Utilizing the mean sea level trend per country as an Instrument for the presence of CSR Committees in firms is theoretically and empirically justifiable due to several reasons:

The effectiveness of the IV method is largely based on the choice of a proper external Instrument. The argument is that the rising sea level is a plausible Instrument that meets the relevance and exogeneity requirements.

Relevance

The Instrument must be correlated with the endogenous explanatory variable, in this case, CSR Committee (CSR_COMM). The rationale behind the relevance of $mean\_trend$

	Sea level	CSR Rating	SDR Rating
CSR Committee	0.04	0.01	4.16

Table 4: Covariance Matrix of CSR Committees and Instruments

is that environmental changes, such as sea level rise, may influence corporate social responsibility policies, particularly in sectors directly affected by environmental factors. Companies in countries experiencing significant environmental changes might be more compelled to form CSR Committees to address these issues actively.

The covariance between the Instrument *mean_trend* and the explanatory variable of interest CSR_COMM is 0.04, indicating a positive but weak relationship.

The covariance between the Instrument of the sea level and the dependent variable Delta is small and negative, as for the other dependent variables, see Table 5. Countries experiencing significant sea level rises may witness more direct impacts of climate change, increasing both public and corporate awareness. This heightened awareness likely accelerates the formation of CSR Committees focused on addressing these environmental challenges.

	Sea level	CSR Rating	SDR Rating
Delta	-0.08	-0.11	-0.03
Internal	-0.12	0.02	0.20
External	0.09	0.19	0.09

Table 5: Covariance Matrix of Dependent Variables and Instruments

Governments in countries facing drastic environmental changes, such as rising sea levels, might impose stricter environmental regulations. Companies may form CSR Committees as a proactive measure to ensure compliance and to spearhead initiatives that align with new legal frameworks. In regions severely affected by environmental issues like sea level rise, there is often increased pressure from stakeholders-including customers, investors, and local communities-for companies to demonstrate environmental responsibility. CSR Committees can be a strategic response to manage and meet these expectations, facilitating more effective stakeholder engagement and communication. Companies facing potential operational and strategic risks due to the implications of sea level rise might use CSR Committees to better position themselves against long-term environmental risks. These committees help develop strategies that address immediate concerns and prepare the company for future challenges.

Exogeneity

The Instrument must not be correlated with the error term in the regression model, implying that it affects the dependent variable (Δ) only through its effect on the endogenous explanatory variable (CSR_COMM). The mean sea level trend is likely exogenous in this context, as it is determined by long-term global climate patterns rather than corporate governance practices or company-specific factors that might directly influence Greenwashing practices.

Mean sea level trends are primarily influenced by climatic and geographical factors that are external to individual companies and their operational decisions. These trends can indirectly impact the economic environment of a country by affecting sectors like insurance, real estate, and agriculture, which in turn might influence corporate strategies regarding sustainability and CSR policies. However, the mean sea level trend itself is not directly caused by the company's internal policies or characteristics, ensuring that it remains exogenous to the error term of the Greenwashing likelihood regression equation.

Financially, companies in countries with significant mean sea level rise might face increased pressure from stakeholders to adopt CSR practices as a defensive strategy against potential climate-related risks. This stakeholder pressure could lead to the establishment of CSR Committees. However, the direct link between mean sea level trends and specific Greenwashing activities (captured by the delta score) remains indirect. Therefore, the mean sea level trend is likely exogenous, as its influence on Greenwashing is mediated through broader economic and policy responses rather than direct corporate actions.

From a social perspective, heightened awareness of climate change in countries experiencing significant sea level rise can lead to stronger societal expectations for corporate responsibility. This societal pressure can result in more companies forming CSR Committees. Nonetheless, societal awareness driven by sea level trends does not have a direct causal relationship with the likelihood of Greenwashing, ensuring that the Instrument's effect on the error term of the Greenwashing model is minimal.

Governance structures can also be influenced by environmental factors like mean sea level trends. In countries facing significant environmental risks, regulatory bodies and governance frameworks may enforce stricter environmental and CSR standards. Companies in these countries may establish CSR Committees to comply with these regulations and improve their governance practices. But on the other Hand, the governance channel operates through regulatory pressures rather than direct corporate behavior related to Greenwashing. This ensures that while the governance channel may influence the presence of CSR Committees, it does not directly impact the error term associated with Greenwashing likelihood.

Empirical validation of the exogeneity assumption can be complex. According to Angrist and Krueger [2001], an Instrumental variable should not be related to the outcome except through its effect on the endogenous regressor. Studies have shown that natural experiments or external environmental factors, like climate trends, are often valid Instruments because they influence economic activities without being directly affected by the decisions of the entities being studied [Angrist and Krueger, 2001].

In the first stage of the Instrumental variable model, the aim was to validate the relevance of the Instrument and identify optimal covariates for predicting the binary endogenous variable CSR_COMM. The analysis began by checking for multicollinearity among the independent variables, which could potentially influence the reliability of the regression coefficients. The Variance Inflation Factor (VIF) was calculated for each variable, and except for $Market\_cap$ and $EBIT$, which both exhibited high VIF values suggesting significant multicollinearity, other variables were deemed acceptable. Three distinct model specifications were explored to ascertain the best set of predictors and validate the Instrumental variable:

- Basic Model (Model A): This model considered *mean_trend* alone as a predictor for CSR_COMM. The significance of *mean_trend* in this model (p-Value: 0.0443) with an AIC of 282.79 suggests that sea level trend can be a relevant predictor of CSR Committee presence, supporting its use as an Instrument.

Table 6: Model A: 1. Stage, Sea level

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.17	0.17	12.66	0.00
mean_trend	0.17	0.08	2.01	0.04

- Optimized Covariate Selection (Model B): Various combinations of covariates were tested to find an optimal model where *mean_trend* remains as a significant predictor with a low AIC. The best model from this exercise included the covaraites $RPSH$, ROE and $GICS\_cluster$, alongside *mean_trend*, which showed significant predictive power for CSR_COMM (p-Value: 0.0462) with an AIC of 282. Similar to Model A, this suggests a model configuration where the Instrument retains relevance and the model remains robust (see Table 7).

- Extended Model (Model C): This model included *mean_trend* alongside a broader set of financial and governance variables, which were deemed to be of importance in the first fixed effects regressions as displayed in Table 2, $RPSH$, ROE, $Women\_emp$, $Board\_size$, and $Credit$. Here, the *mean_trend* variable is not significant anymore, indicating that its explanatory power might be overshadowed or confounded by other included variables. The only variable of strong significance is the positive parameter for the size of the Board. The AIC improved to 276.39, suggesting a better model fit compared to the basic model, although at the cost of the Instrument's significance (see Table 8).

Table 7: Model B: 1. Stage, Sea level

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.92	0.46	4.21	0.00
mean_trend	0.18	0.09	2.00	0.05
RPSH	0.68	1.33	0.51	0.61
ROE	-0.02	0.19	-0.12	0.91
GICS_cluster2	0.22	0.64	0.35	0.73
GICS_cluster3	1.71	1.13	1.52	0.13
GICS_cluster4	0.78	1.13	0.69	0.49
GICS_cluster5	-0.17	0.54	-0.31	0.76
GICS_cluster6	0.96	0.85	1.13	0.26
GICS_cluster7	-0.06	0.54	-0.12	0.91
GICS_cluster8	1.09	0.85	1.29	0.20
GICS_cluster9	1.14	1.11	1.02	0.31

The analysis underscores the sensitivity of Instrumental variable effectiveness to model specifications. *mean_trend*'s varying significance across models highlights the importance of careful covariate selection to maintain the Instrument's relevance and exogeneity. The results from the optimized covariate selection model indicate that including specific variables such as $RPSH$, ROE or $GICS\_cluster$, which might directly influence CSR practices, helps in isolating the effect of *mean_trend* on CSR_COMM without introducing bias from multicollinearity.

The predictive probabilities generated from these models $(prob\_A, prob\_B, prob\_C)$ represent the likelihood of CSR Committee presence based on different model specifications. These probabilities are crucial as they serve as fitted values in the second stage of the IV approach, where the actual effect of CSR_COMM on the outcome variable Δ will be assessed.

In conclusion, the first stage of the IV model has demonstrated that while *mean_trend* can be a relevant Instrument for CSR_COMM, its effectiveness is heavily dependent

Table 8: Model C: 1. Stage, Sea level

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.19	0.32	6.90	0.00
mean_trend	0.10	0.09	1.09	0.28
RPSH	0.57	1.19	0.48	0.63
ROE	0.15	0.59	0.26	0.79
Women_emp	0.03	0.17	0.16	0.87
Board_size	0.74	0.22	3.36	0.00
Credit	0.03	0.03	1.15	0.25

on the accompanying set of covariates. This stage is critical for ensuring that the Instrument is not only statistically significant but also appropriately isolated from other influences that could distort the intended causal inference in the second stage of the IV analysis.

3.2.1 Results

In the second stage of the Instrumental variable regression, the three models were considered to analyze the effect of CSR Committee presence, represented by the variable $prob\_A$, $prob\_B$, and $prob\_C$, on the variable Δ, which reflects an aspect of corporate behavior potentially influenced by CSR activities. Diagnostic plots for all three models are provided in the Appendix Figure 1, Figure 2 and Figure 3, including Q-Q plots, residuals versus fitted values plots, and residuals plots.

Each model provides insights into the relationship under different sets of covariates as displayed in Table 24 The Basic IV Model (Model A) tests the effect of $prob\_A$ on Δ without additional covariates. The coefficient for $prob\_A$ is negative (-1.709) but not statistically significant (p-Value: 0.304), indicating no strong evidence that the presence of CSR Committees influences Δ under this model setup. The model's low R^2 value (0.0430) and adjusted R^2 value (0.0389) suggest that $prob\_A$ explains very little of the variability in Δ. The Wald test confirms the overall significance of the model predictors

Table 9: IV Sea level - Delta Score

	Dependent variable: Delta		
	(A)	(B)	(C)
Constant	1.555	-0.480	0.860
	(1.511)	(1.247)	(1.424)
Probability A	-1.709		
	(1.660)		
Probability B		-0.104	
		(1.402)	
Probability C			-0.886
			(1.604)
RPSH		-0.071	-0.038
		(0.044)	(0.047)
ROE		0.018	0.026
		(0.044)	(0.045)
Women Employment			-0.229***
			(0.045)
Board Size			0.167
			(0.090)
Credit			-0.006
			(0.009)
GICS Cluster 2		0.326	
		(0.183)	
GICS Cluster 3		0.472*	
		(0.233)	
GICS Cluster 4		1.061***	
		(0.276)	
GICS Cluster 5		0.000	
		(0.162)	
GICS Cluster 6		0.701**	
		(0.216)	
GICS Cluster 7		0.441**	
		(0.161)	
GICS Cluster 8		1.089***	
		(0.214)	
GICS Cluster 9		0.197	
		(0.246)	
Residual A	0.705***		
	(0.159)		
Residual B		0.603***	
		(0.153)	
Residual C			0.678***
			(0.158)
R^2	0.0430	0.1575	0.1090
Adjusted R^2	0.0389	0.1353	0.0954
Pseudo R^2	0.0129	0.0513	0.0710
Wald test p-Value	1.625e-05***	<2.2e-16***	<2.2e-16***
Breusch-Pagan p-Value	0.3627	0.1544	0.3861

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

(p-Value = 1.625e-05). The Pseudo R^2 value is 0.0129, indicating a very small proportion of the variance in Δ explained by the model. The Breusch-Pagan test p-Value of 0.3627 suggests no significant heteroscedasticity. The first stage residual parameter is of high significance and positive this means that the endogenous variable CSR_COMM had an upward bias in its effect on the outcome variable Δ before correcting for endogeneity with the *mean_trend* Instrument. After including the residuals, the corrected effect of the endogenous variable is adjusted downwards. We can observe this effect for all three Models.

In the Residuals vs. Fitted plot for Model A, the residuals appear to be randomly scattered around zero, indicating that the model does not suffer from major specification errors. There are no clear patterns, suggesting that the model adequately captures the linear relationship. The Q-Q Plot shows that the residuals roughly follow the theoretical quantiles, indicating that they are approximately normally distributed. However, there are slight deviations at the tails, suggesting the presence of outliers. The spread of residuals is fairly consistent across the range of fitted values, indicating no apparent problems with heteroscedasticity, this is in line with the findings of the Breusch-Pagan test see Figure 1 in the Appendix.

The IV Model with Selected Covariates (Model B) includes *prob_B* along with covariates such as $RPSH$, $GICS\_cluster$, and ROE. The *prob_B* coefficient is positive (0.104) but still statistically insignificant (p-Value: 0.941). In Addition, certain levels of the $GICS\_cluster$ variable are significant, suggesting sector-specific effects on Δ. Notably, GICS Cluster 3 (estimate = 0.472, p-Value = 0.043), GICS Cluster 4 (estimate = 1.061, p-Value = 0.0001), GICS Cluster 6 (estimate = 0.701, p-Value = 0.0012), GICS Cluster 7 (estimate = 0.441, p-Value = 0.0065), and GICS Cluster 8 (estimate = 1.089, p-Value = 5.15e-07) are significant. The model has a better fit, with an adjusted R^2 of 0.1353 and R^2 of 0.1575. The Wald test confirms the overall significance of the model predictors (p-Value < 2.2e-16). The Pseudo R^2 value is 0.0513, indicating an improved proportion of variance explained compared to Model A. The Breusch-Pagan test p-Value of 0.1544 suggests no significant heteroscedasticity. As before the highly significant residual shows us that the inclusion of the residuals and the usage of an Instrument accounts for a lot of endogenity in the Model.

For Model B, the Residuals vs. Fitted plot shows the residuals are randomly scattered, suggesting a well-specified model without obvious specification errors. There are no discernible patterns, supporting the adequacy of the linear model. This randomness is positive and indicates that the model's specifications are appropriate. The residuals in the Q-Q Plot follow the theoretical quantiles closely, with minor deviations at the extremes, indicating approximate normality and the presence of some outliers (Figure 2).

The Comprehensive IV Model (Model C) tests the influence of $prob\_C$ along with a full set of theoretically relevant covariates, including $RPSH$, ROE, $Women\_emp$, $Board\_size$, and $Credit$. The coefficient for $prob\_C$ is negative but again shows no significant effect on Δ (p-Value: 0.581). Interestingly, the $Women\_emp$ coefficient is negative (-0.229) and highly significant (p-Value = 4.80e-07), suggesting that the employment of women has a noticeable negative correlation with Δ. The Board Size has a positive and slightly significant effect on the proxy for Greenwashing (estimate = 0.167, p-Value = 0.065). The model demonstrates a moderate fit, with an adjusted R^2 of 0.0954 and R^2 of 0.1090. The Wald test confirms the overall significance of the model predictors (p-Value < 2.2e-16). The Pseudo R^2 value is 0.0710, indicating that a higher proportion of variance is explained by this model compared to Models A and B. The Breusch-Pagan test p-Value of 0.3861 suggests no significant heteroscedasticity.

The Model C Residuals Plot shows no discernible trends or patterns. The residuals are well scattered around the zero line, indicating that the model handles the data variability well. The residuals displayed in the Q-Q Plot generally follow the theoretical quantiles but show deviations at the tails, suggesting some outliers and slight departures from normality. The Residuals vs. Fitted plot shows a random distribution around zero, supporting the assumption of a well-specified model. There is no clear pattern, indicating that the model appropriately captures the relationship between the variables.

The scatter is uniformly distributed around the horizontal line, reaffirming a lack of heteroscedasticity. The Breusch-Pagan Test p-Value of 0.3861 suggests no significant heteroscedasticity, indicating stable variance of the residuals (Figure 3).

The Likelihood-Ratio test results indicate that Model C provides a significantly better fit compared to Model A, with a p-Value of 0.0058. However, there are no significant differences between Models A and B (p-Value: 0.3704) and between Models B and C (p-Value: 0.3517). The overall results indicate that the presence of CSR Committees, as modeled by the Instrumental variables, does not significantly impact Δ across the tested models. Across all models, the Instrumented CSR Committee variables exhibited a negative parameter, indicating that the presence of CSR Committees has a negative effect on the risk of Greenwashing, even though not significant. This analysis underscores the importance of model specification and the choice of covariates in understanding the dynamics of CSR impact on corporate behaviors.

All three models seem well-specified without obvious specification errors, as indicated by the random distribution of residuals and consistent spread across fitted values. This suggests that the models accurately capture the underlying relationships without omitted variable bias or incorrect functional form. The randomness in the residuals across all models and the general alignment with theoretical quantiles in the Q-Q plots suggest that the Instruments chosen for each model are likely valid. There is no evidence from these plots that the Instruments are weak or invalid, as the residuals do not show patterns indicative of endogeneity issues. Each model exhibits slight issues with outliers as shown in the Q-Q plots, which could affect the robustness of the regression coefficients. These plots indicate that the models are generally well-fitted, with each successive model (from A to C) potentially providing a better fit by capturing more variability and addressing the range of data more effectively.

Throughout all three Models the first stage residuals are significant. The significance of these residuals confirms the presence of endogeneity in the original explanatory variables (CSR_COMM). This means that without accounting for this endogeneity (i.e., if we hadn't used the 2SRI method), the estimated coefficients for these explanatory variables would have been biased and inconsistent. By including the residuals in the second stage, the 2SRI approach helps to correct for this bias, leading to more reliable and consistent estimates.

The significance of the included residuals in the second stage of the 2SRI analysis in all three models highlights the importance of addressing endogeneity in this analysis. This result validates the use of the 2SRI method to obtain unbiased and consistent estimates of the impact of the explanatory variables on the dependent variable.

3.2.2 Internal and External Score

The External and Internal Scores have been analyzed using the same methodology as the Delta Score IV Regression. For each score, three different models are examined, but similar to the one concerning the Delta-Score. In the first stage, Model A included only the Instrument. Model B included covariates selected based on an algorithm, while Model C incorporated the same covariates as in the Delta Score analysis. These three models were run for both the External and Internal Scores.

Internal Score

In the first stage of the analysis as seen in the Appendix Table 27, the basic Model A yields a surprising result with a significant Instrument and an AIC of 282.79. In Model B, the Instrument remains significant, but the AIC increases. In Model C, the Instrument (*mean_trend*) is not significant, but, the number of people on the board of a firm is positive and significant, and the AIC decreases to 276.39.

In the second stage, Model A shows a marginally significant variable for the Instrumented variable *prob_A* (estimate = -2.834, p-Value = 0.077). The residuals are highly significant (estimate = 1.143, p-Value < 4.19e-13), indicating some remaining endogeneity. The adjusted R^2 is very low (0.1095), and the R^2 is 0.1133. The Wald test indicates the overall significance of the predictors (p-Value = 1.897e-11). The Breusch-Pagan test

Table 10: IV Sea level - Internal Score

	Dependent variable: Internal		
	(A)	(B)	(C)
Constant	2.580	0.689	0.796
	(1.455)	(1.249)	(1.368)
Probability A	-2.834		
	(1.598)		
Probability B		-1.092	
		(1.404)	
Probability C			-0.842
			(1.542)
RPSH		-0.053	-0.021
		(0.044)	(0.045)
ROE		0.034	0.058
		(0.044)	(0.043)
Women Employment			-0.091*
			(0.043)
Board Size			0.337***
			(0.087)
Credit			-0.003
			(0.009)
GICS Cluster 2		-0.004	
		(0.183)	
GICS Cluster 3		0.443	
		(0.233)	
GICS Cluster 4		0.687*	
		(0.276)	
GICS Cluster 5		0.167	
		(0.163)	
GICS Cluster 6		0.499*	
		(0.216)	
GICS Cluster 7		0.288	
		(0.162)	
GICS Cluster 8		0.856***	
		(0.214)	
GICS Cluster 9		0.402	
		(0.247)	
Residual A	1.143***		
	(0.153)		
Residual B		1.102***	
		(0.153)	
Residual C			1.001***
			(0.152)
R^2	0.1133	0.1547	0.1765
Adjusted R^2	0.1095	0.1323	0.1640
Pseudo R^2	0.0129	0.0513	0.0710
Wald test p-Value	1.897e-11***	<2.2e-16***	<2.2e-16***
Breusch-Pagan p-Value	0.0092***	0.0258**	0.0543

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

shows evidence of heteroscedasticity in the Model (p-Value = 0.009164).

Model B does not have a significant Instrumented variable ($prob\_B$, estimate = -1.092, p-Value = 0.437), but some of the GICS clusters are significant, including GICS Cluster 4 (estimate = 0.687, p-Value = 0.013), GICS Cluster 6 (estimate = 0.499, p-Value = 0.021), and GICS Cluster 8 (estimate = 0.856, p-Value = 7.47e-05). The adjusted R^2 (0.1323) and R^2 (0.1547) improve significantly compared to Model A. The residuals in Model B are significant (estimate = 1.102, p-Value = 2.56e-12), indicating as in Model A remaining endogeneity. The Wald test confirms the overall model significance (p-Value < 2.2e-16). The Breusch-Pagan test also shows evidence of heteroscedasticity (p-Value = 0.02576).

Model C displys that the Instrumented variable ($prob\_C$, estimate = -0.842, p-Value = 0.585) is not significant, but the Board Size variable is highly significant (estimate = 0.337, p-Value = 0.00012) as well as the variable $Women\_emp$ (estimate = -0.091, p-Value = 0.035). The adjusted R^2 (0.164) and R^2 (0.1765) are higher than in Models A and B. The residuals in Model C are significant (estimate = 1.001, p-Value = 1.15e-10), suggesting some endogeneity persists throughout all three Models. The Wald test confirms the overall model significance (p-Value < 2.2e-16). The Breusch-Pagan test suggests again borderline heteroscedasticity (p-Value = 0.05426).

In all models, the significant residuals suggest that some endogeneity existed in the Model and that the use of an Instrument is crucial in obtaining unbiased Instruments. In the LR-test, just for the comparison of Model A to Model C we get a significant difference with $Pr(> Chisq)$ at 0.0058** also the Model with the highest $Pesudo - R^2$ is Model C. Given the other diagnostics, Model C appears to be the most robust. It has the highest adjusted R^2 and R^2 values, indicating better explanatory power. The significant variables for Board Size and Women's Employment further validate the model's effectiveness in explaining the dependent variable. The Instrumented Variable of the presence of CSR Committees is not significant anymore but we now observe highly significant and positive residuals indicating again that the variable of interest experienced a strong and positive bias in the model without IV.

External Score

Among the three models for the External Score in the first stage of the IV Regression as seen in the Appendix Table 30, Model C emerged as the best based on the AIC criterion, having the lowest AIC value (276.39), see Table 11. This model includes the same covariates as Model C with the Delta Score. Although the *mean_trend* variable is not significant, the *Board_size* variable is highly significant, contributing significantly to the model's overall explanatory power. In contrast, Model A, while simpler and featuring a significant *mean_trend* variable, has a higher AIC, indicating a poorer fit compared to Model C. Model B neither achieves a lower AIC nor includes significant additional covariates, making it the least preferred model among the three.

Since logistic regression in the first stage was used, it is again necessary to derive the appropriate value for the Instrumented variables and for the residuals. In Model A, the coefficient of the second stage for the Instrumented variable (*prob_A*) is not significant, and both the R^2 (0.0119) and the adjusted R^2 (0.0076) values are very low. The residuals in Model A, however, are significant (estimate = 0.368, p-Value = 0.023), indicating just a little remaining endogeneity that was addressed by the IV. The Wald test for Model A is not significant (p-Value = 0.1112), suggesting that the overall model predictors are not jointly significant. The Breusch-Pagan test (p-Value = 0.2331) indicates no significant heteroscedasticity. The Pseudo R^2 for Model A is 0.0129, indicating a very low explanatory power.

Model B reveals that some of the clusters are significant, but not the Instrumented variable (*prob_B*) (estimate = -1.488, p-Value = 0.300). The R^2 (0.1166) and adjusted R^2 (0.0933) values are higher than in Model A, accompanied by a highly significant Wald test (p-Value = 2.034e-09). Notably, several GICS clusters have significant effects, including GICS Cluster 2 (estimate = -0.476, p-Value = 0.012), GICS Cluster 4 (estimate = -0.696, p-Value = 0.014), GICS Cluster 6 (estimate = -0.419, p-Value = 0.058), GICS

Table 11: IV Sea level - External Score

	Dependent variable: External		
	(A)	(B)	(C)
Constant	0.912	1.541	-0.264
	(1.535)	(1.277)	(1.440)
Probability A	-1.002		
	(1.686)		
Probability B		-1.488	
		(1.435)	
Probability C			0.241
			(1.622)
RPSH		0.038	0.029
		(0.045)	(0.047)
ROE		0.016	0.033
		(0.045)	(0.045)
Women Employment			0.220***
			(0.045)
Board Size			0.169
			(0.091)
Credit			0.005
			(0.009)
GICS Cluster 2		-0.476*	
		(0.188)	
GICS Cluster 3		-0.142	
		(0.239)	
GICS Cluster 4		-0.696*	
		(0.283)	
GICS Cluster 5		0.202	
		(0.166)	
GICS Cluster 6		-0.419*	
		(0.221)	
GICS Cluster 7		-0.287*	
		(0165)	
GICS Cluster 8		-0.531*	
		(0.219)	
GICS Cluster 9		0.205	
		(0.252)	
Residual A	0.368*		
	(0.162)		
Residual B		0.467**	
		(0.157)	
Residual C			0.235
			(0.160)
R^2	0.0119	0.1166	0.0883
Adjusted R^2	0.0076	0.0933	0.0744
Pseudo R^2	0.0129	0.0513	0.0710
Wald Test p-Value	0.1112	2.03e-09***	4.83e-12***
Breusch-Pagan p-Value	0.2331	1.78e-05***	2.57e-05***

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

Cluster 7 (estimate = -0.287, p-Value = 0.083), and GICS Cluster 8 (estimate = -0.531, p-Value = 0.016). The residuals in Model B are also significant (estimate = 0.467, p-Value = 0.003), indicating remaining endogeneity. The Breusch-Pagan test (p-Value = 1.781e-05) indicates significant heteroscedasticity, it means that the assumption of homoscedasticity (constant variance of the errors) is violated. While the OLS estimates of the coefficients remain unbiased, they are no longer efficient. This means that the estimates do not have the minimum possible variance, making them less reliable. In addition, the standard errors of the OLS estimates are biased in the presence of heteroscedasticity. This bias leads to incorrect test statistics and confidence intervals, potentially causing incorrect inferences about the significance of the predictors. The Pseudo R^2 for Model B is 0.0513, showing improved explanatory power compared to Model A.

In Model C the variable $Women\_emp$ exhibits a positive and highly significant effect on the External Score (estimate = 0.220, p-Value = 1.75e-06). This finding aligns with the original paper, which also concluded that a higher number of female employees in the workforce is linked to a more positive external reception of the company. The R^2 (0.0883) and adjusted R^2 (0.0744) values are low, and the Wald test for Model C is highly significant (p-Value = 4.833e-12), indicating that the overall model is statistically robust. The residuals in Model C (estimate = 0.235, p-Value = 0.142) are not significant, suggesting that the Instrument is not as useful as with the other Scores. The Breusch-Pagan test (p-Value = 2.576e-05) indicates again significant heteroscedasticity, meaning we have non efficient estimates. The Pseudo R^2 for Model C is 0.0710, indicating the best explanatory power among the three models.

Next the LR-test got used to asses which Model is the best one. First, the test between Model A and Model B shows no significant difference (p-Value = 0.3704). Next, the test between Model A and Model C shows a significant difference (p-Value = 0.005805), indicating that Model C provides a significantly better fit than Model A. Lastly, Model B and Model C show also no significant difference (p-Value = 0.3517).

Given the diagnostics, Model C is the best model. It demonstrates significant coefficients for key variables, has the highest Pseudo R^2 value, and the Wald test indicates strong overall model significance. The Likelihood-Ratio test between Model A and Model C confirms that Model C is significantly better. Despite the significant heteroscedasticity indicated by the Breusch-Pagan test, Model C provides a robust and comprehensive explanation of the External Score.

3.3 Instrumental variable: Per Country Ratings

3.3.1 Application

Next we analysis two more Instruments, to check the robustness of our findings. The two Instruments are actually displaying a similar metric, the Sustainability development ratings per country just derived from different points of view. The first Instrument is the CSR-Rating per Country and the second one is the SDR Rating per country
The methodology for deriving Corporate Social Responsibility (CSR) or Sustainable Development Report (SDR) scores per country involves several key steps to ensure consistency and accuracy across various companies, industries, countries and legislative frameworks. The further References for this Methodology can be found at CSR [2008]. In the first Ranking, the researchers categorize CSR performance into twelve subcategories, which further roll up into four main categories. Additionally, they manage various special issue topics to address CSR issues that do not conform to the standard subcategory schema. The researchers then convert each data point from their sources into a numerical rating on a scale from 0 to 100, where 100 indicates a positive rating. This numeric transformation allows for the standardisation of diverse data forms, such as numerical scores, qualitative markers like + or −, or relative rankings like "Top 50". Next, they normalize these scores to account for source-specific biases. By analyzing variations among sources for the same company, they can identify and adjust for these biases, ensuring a more consistent rating across sources. In the aggregation phase, they weight each source according to their evaluation of its credibility and relevance. They then

combine data for each country to generate base ratings at the subcategory level, which are further aggregated to the category level. In Addition, they trim their dataset by excluding ratings for approximately 19,354 companies for which they lack sufficient information. Finally, they conduct additional research on each rated company to gather information about the industry it operates in and other relevant details. This extensive process allows them to generate both industry and country-specific averages for CSR performance. In the analysis, the averaged CSR scores at the country level are utilized to enhance the robustness and applicability of the findings. The CSR score serves as an effective Instrumental variable for the CSR_COMM variable based on two critical assumptions: relevance and exogeneity. The relevance assumption is satisfied because the CSR score is highly likely to be correlated with the CSR_COMM variable. This correlation arises because both metrics aim to measure aspects of a company's corporate social responsibility performance, with the CSR score providing a quantitative reflection of broader per country legaslative CSR activities, including community involvement and ethical conduct. The CSR rankings should not be correlated with the error term in the regression of CSR Committees on Delta, suggesting that they are influenced by national policies and cultural norms rather than by individual firm practices or unobserved firm-level factors that could directly impact practices such as Greenwashing. The legislative framework is generally shaped by historical, political, and socio-economic factors that are independent of individual firm decisions on CSR and Greenwashing. This independence helps in maintaining the integrity of the CSR score as an unbiased Instrument in the analysis. By meeting these assumptions, the CSR score provides a theoretically justified and robust tool for analyzing the extent of Greenwashing by leveraging its influence on the CSR communication practices of companies. This analysis can reveal discrepancies between public CSR commitments and actual environmental performance, thereby enhancing the understanding of Greenwashing dynamics within corporate practices.

In the SDR Ranking, Countries are ranked by their overall score. The overall score measures the total progress towards achieving all 17 SDGs [SDR, 2023]. The score can be interpreted as a percentage of SDG achievement. A score of 100 indicates that all SDGs have been achieved. The References for the SDR Ranking is SDR [2023]. The Sustainable Development Report (SDR) rating is derived using 97 global indicators, with an additional 27 indicators specifically for OECD countries due to better data availability. Data selection involves choosing indicators based on global relevance, statistical adequacy, timeliness, coverage (with data available for at least 80% of UN Member States with a population over one million), and the measurability of distance to targets. The data sources include a mix of official sources, such as FAO, ILO, OECD, and the World Bank, and unofficial sources, such as household surveys, civil society organizations, and peer-reviewed journals.

To handle missing data, the SDR includes countries that have data for at least 80% of the indicators or have been included in previous editions of the SDR with data for at least 75% of the indicators. Generally, missing data are not imputed to maintain accuracy, except in exceptional circumstances.

The construction of the SDG Index involves three main steps. First, performance thresholds are established by rescaling each indicator from 0 to 100, with 0 denoting the worst performance and 100 the best. The upper bounds are determined based on SDG targets, the "leave no one behind" principle, science-based targets, or the average of the top five performers, while the lower bound is set at the 2.5th percentile. Second, normalization is carried out by linearly transforming the variables to a scale of 0 to 100, ensuring comparability and ease of interpretation. Third, weighting and aggregation are performed, where each SDG is given equal weight to reflect a commitment to treat all goals equally. Scores for each goal are averaged across the 17 SDGs to obtain the final Index score. Sensitivity tests, including comparisons of arithmetic and geometric means and Monte Carlo simulations, ensure the robustness of the rankings.

This methodology ensures that the SDR rating provides a comprehensive and balanced assessment of countries' progress towards achieving the Sustainable Development Goals

(SDGs).

The first stage of the Instrumental variable regression aims to verify the relevance of the two Ratings as Instruments for the CSR_COMM variable, which seeks to predict the Delta score, a proxy for Greenwashing. Here's a summary and interpretation of the results from the models tested:

In this analysis, the first stages of two different Instrumental variables regressions get compared, both of which use country ratings to indicate how well a country is performing in terms of sustainability and CSR. The first stage results can be fund in the Appendix Table 25 and Table 26

- Model A (Basic Model): This logistic regression model tested the direct impact of CSR and SDR Rating on CSR_COMM. The coefficient for CSR_Rating is not significant (Estimate = 0.05291, Std. Error = 0.10344, p-Value = 0.609). This suggests that CSR_Rating may not be a strong Instrument for CSR_COMM, as it does not significantly explain the variation in CSR_COMM. The AIC value is 286.16, indicating the model fit. The coefficient for SDR_Rating is not significant as well (Estimate = 0.06991, Std. Error = 0.08336, p-Value = 0.402). The AIC value here is 280.87, which is worse then in the Model with the CSR Ranking Instrument.

- Model B (Optimized Variables): This model is based on an algorithm to find the most suitable covariates to make the Instrument as significant as possible. The algorithm didn't manage to find any combination to make the Instruments significant. The model for the CSR-Rating Instrument includes the covariates ($RPSH$, $female\_board$, $Women\_emp$).The coefficient for CSR-Rating remains insignificant (Estimate = 0.06226, Std. Error = 0.09845, p-Value = 0.527). Also, $female\_board$ shows a significant positive effect (Estimate = 0.42033, Std. Error = 0.16530, p-Value = 0.011). The AIC value improves to 284.99, suggesting a better model fit than Model A.

  The model for the SDR-Rating Instrument includes the covariates $RPSH$, ROA,

$Total\_assets$, ROE, $Women\_emp$, $GICS\_cluster$ based on the algorithm to find the best model. The coefficient for SDR_Rating shows a borderline significance (Estimate = 0.15389, Std. Error= 0.09199, p-Value= 0.0944). The amount of Total assets and ROA show significant effects. The AIC value is 271.47, indicating an improved model fit.

- Model C (Theoretically Informed Model): This model includes theoretically relevant variables ($RPSH, ROE, Women\_emp, Board\_sizeandCredit$). The coefficient for CSR Ratings is still not significant (Estimate= 0.01059, Std. Error= 0.11021, p-Value= 0.923). The Board Size is significant and positively associated with CSR Committees. The AIC value improves further to 277.51. The coefficient for the SDR Rating remains insignificant (Estimate = 0.06651, Std. Error= 0.07957, p-Value= 0.403). The Board Size is again significant and positively associated with the presence of CSR Committees. The AIC value is 271.99, indicating the best model fit so far.

The first-stage regressions show that neither CSR_Rating nor SDR_Rating are strong Instruments for predicting CSR_COMM, as their coefficients are generally not significant across different models. The best model fits (lowest AIC values) are achieved when additional theoretically relevant covariates are included, but even then, the Instruments themselves (CSR_Rating and SDR_Rating) do not show significant explanatory power. For CSR_Rating, the inclusion of theoretically chosen covariates (Model C) slightly improves the model fit, with $Board\_size$ showing a significant positive effect. For SDR_Rating, the best model (Model B) shows borderline significance for SDR_Rating and significant effects for $Total\_assets$ and ROA.

Overall, the analysis indicates that while the models can be improved by including relevant covariates, the Instruments (CSR_Rating and SDR_Rating) themselves may not be sufficiently strong or valid for predicting CSR_COMM in the first-stage regressions. This suggests that these Instruments may not be effective in explaining the variance in Delta, as indicated by the generally low significance levels and AIC values.

3.3.2 Results Delta Score

See the Table 25 for the results with the CSR-Rating as Instrumental variable and Table 26 for the results of the regression with the SDR-Rating as Instrument.

Delta Score and CSR Rating

In Model A, the coefficient for $prob\_A$ is -11.438 with a p-Value of 0.0873, which is not statistically significant. The coefficient for $resid\_A$ is 0.695 with a p-Value of 1.43×10^{-5}, indicating a significant relationship. The R^2 value is 0.0454, and the adjusted R^2 is 0.0413, suggesting that the model explains a small proportion of the variability in Δ. The Pseudo R^2 is 0.0010, indicating a very low proportion of variance explained. The Wald test p-Value is 6.889×10^{-6}, confirming the overall significance of the model predictors. The Breusch-Pagan test p-Value is 0.5538, indicating no significant heteroscedasticity. In the Residuals vs. Fitted plot, the residuals are randomly scattered around zero, indicating that the model does not suffer from major specification errors. The Q-Q Plot shows that the residuals roughly follow the theoretical quantiles, indicating approximate normality, though there are slight deviations at the tails suggesting the presence of outliers.

In Model B, the coefficient for $prob\_B$ is 2.410 with a p-Value of 0.373, indicating no significant relationship. But, the variable Women Employment has a significant coefficient of -0.217 with a p-Value of 2.95×10^{-6}, and the residuals from the first stage (Residual B) have a significant coefficient of 0.711 with a p-Value of 6.59×10^{-6}. These results indicate that Women Employment has a strong negative effect and the residuals from the first stage have a strong positive effect on the dependent variable. The R^2 value is 0.0996, and the adjusted R^2 is 0.0898, indicating a better fit than Model A. The Pseudo R^2 is 0.0263. The Wald test p-Value is $< 2.2 \times 10^{-16}$, confirming the overall significance of the model predictors.

The Residuals vs. Fitted plot shows randomly scattered residuals, suggesting a well-specified model. The Q-Q Plot shows residuals closely following the theoretical quantiles,

Table 12: IV CSR Rating - Delta Score

	Dependent variable: Delta		
	(A)	(B)	(C)
Constant	10.410	-2.193	-0.375
	(6.076)	(2.461)	(1.713)
Probability A	-11.438		
	(6.677)		
Probability B		2.410	
		(2.703)	
Probability C			0.509
			(1.931)
RPSH		-0.074	-0.050
		(0.051)	(0.048)
ROE			0.0211
			(0.0448)
Female Board		-0.105	
		(0.102)	
Women Employment		-0.217***	-0.235***
		(0.046)	(0.045)
Board Size			0.099
			(0.105)
Credit			-0.010
			(0.010)
Residual A	0.695***		
	(0.158)		
Residual B		0.711***	
		(0.156)	
Residual C			0.660***
			(0.158)
R^2	0.0454	0.0996	0.1072
Adjusted R^2	0.0413	0.0898	0.0936
Pseudo R^2	0.0010	0.0263	0.0670
Wald test p-Value	6.889e-06***	<2.2e-16***	<2.2e-16***
Breusch-Pagan p-Value	0.5538	0.1152	0.2502

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

with minor deviations at the extremes, indicating approximate normality and the presence of some outliers. However, there is a slight pattern in the residuals, which, combined with the Breusch-Pagan test p-Value of 0.0406, indicates some evidence of heteroscedasticity in the model.

In Model C, the coefficient for $prob\_C$ is 0.509 with a p-Value of 0.792, indicating no significant relationship. The $Women\_emp$ coefficient is significant (estimate = -0.235, p-Value $= 2.86 \times 10^{-7}$), indicating a noticeable negative correlation with Δ. The $resid\_C$ coefficient is also significant (estimate = 0.660, p-Value $= 3.52 \times 10^{-5}$). The R^2 value is 0.1072, and the adjusted R^2 is 0.0936. The Pseudo R^2 is 0.0669. The Wald test p-Value is $< 2.2 \times 10^{-16}$, confirming the overall significance of the model predictors. The Breusch-Pagan test p-Value is 0.2502, indicating no significant heteroscedasticity. The Residuals vs. Fitted plot shows no discernible trends or patterns, with residuals well-scattered around the zero line, indicating a well-specified model. The Q-Q Plot shows residuals generally following the theoretical quantiles but with deviations at the tails, suggesting some outliers and slight departures from normality.

The First Stage Insignificance of the The CSR-Rating, which serves as the Instrument, did not significantly predict the endogenous variable. This could be indicative of a weak Instrument, meaning that the CSR-Rating does not have a strong direct influence on the endogenous variable. Despite the weak Instrument, the inclusion of the first stage residuals in the second stage regression corrects for the endogeneity. The residuals represent the portion of the endogenous variable that is unexplained by the Instrument but correlated with the outcome's error term, thus reducing bias. The primary role of the first stage residuals is to address the bias caused by endogeneity. The significant coefficients suggest that including these residuals helps to obtain unbiased and consistent estimates of the effect of the endogenous variable on the outcome. The results highlight that while the CSR-Rating might not be a strong Instrument, its inclusion through the residuals still plays a crucial role in mitigating endogeneity. This underscores the importance of examining not just the first stage predictive power of an Instrument but

also its overall contribution to addressing endogeneity in the model.
Each model exhibits slight issues with outliers as shown in the Q-Q plots, which could affect the robustness of the regression coefficients. These plots indicate that the models are generally well-fitted, with each successive model (from A to C) potentially providing a better fit by capturing more variability and addressing the range of data more effectively.
The Likelihood-Ratio test between Model A and Model B shows a p-Value of 0.06676, indicating that adding the covariates in Model B provides a marginally better fit than The LR test for the comparison of Model A and Model C shows a p-Value of 0.0022, indicating that Model C provides a significantly better fit compared to Model A. Finally the Likelihood-Ratio test between Model B and Model C shows a p-Value of 0.0032, indicating that Model C provides a significantly better fit compared to Model B. These results suggest that the addition of covariates in Models B and C significantly improves the fit of the models, with Model C providing the best fit overall.

Delta Score and SDR Rating

In Model A, the coefficient for *prob_A* is 4.008 with a p-Value of 0.331, indicating no significant relationship. The coefficient for $resid\_A$ is 0.668 with a p-Value of 3.93×10^{-5}, indicating the addressed remaining endogenity in the Model. The R^2 value is 0.0379, and the adjusted R^2 is 0.0338, suggesting that the model explains a small proportion of the variability in Δ. The Pseudo R^2 is 0.0026, indicating a very low proportion of variance explained. The Wald test p-Value is 0.0002, confirming the overall significance of the model predictors. The Breusch-Pagan test p-Value is 0.4961, indicating no significant heteroscedasticity.
In the Residuals vs. Fitted plot, the residuals are randomly scattered around zero, indicating that the model does not suffer from major specification errors. The Q-Q Plot shows that the residuals roughly follow the theoretical quantiles, indicating approximate normality, though there are slight deviations at the tails suggesting the presence of

Table 13: IV SDR Rating - Delta Score

	Dependent variable: Delta		
	(A)	(B)	(C)
Constant	-3.653	-1.674*	-0.824
	(3.756)	(0.769)	(1.492)
Probability A	4.008		
	(4.118)		
Probability B		1.423·	
		(0.830)	
Probability C			1.016
			(1.672)
RPSH		-0.061	-0.055
		(0.043)	(0.047)
Total Assets		0.104·	
		(0.059)	
ROA		0.048	
		(0.058)	
ROE		0.007	0.019
		(0.043)	(0.045)
Women Employment		-0.231***	-0.237***
		(0.056)	(0.045)
Board Size			0.074
			(0.093)
Credit			-0.012
			(0.009)
GICS Cluster 2		0.472*	
		(0.183)	
GICS Cluster 3		0.473*	
		(0.203)	
GICS Cluster 4		0.905***	
		(0.264)	
GICS Cluster 5		0.117	
		(0.192)	
GICS Cluster 6		0.809***	
		(0.205)	
GICS Cluster 7		0.372*	
		(0.162)	
GICS Cluster 8		0.841***	
		(0.198)	
GICS Cluster 9		-0.054	
		(0.226)	
Residual A	0.668***		
	(0.161)		
Residual B		0.537***	
		(0.155)	
Residual C			0.644***
			(0.160)
R^2	0.0379	0.1986	0.1052
Adjusted R^2	0.0338	0.1718	0.0915
Pseudo R^2	0.0026	0.1301	0.0706
Wald test p-Value	0.0002***	<2.2e-16***	<2.2e-16***
Breusch-Pagan p-Value	0.4961	0.147	0.2625

Note: (standard error); ·p<0.1; *p<0.05; **p<0.01; ***p<0.001

outliers.

In Model B, $prob\_B$ has a positive and marginally significant relationship with Δ (p-Value of 0.087). Several GICS clusters also have significant coefficients: GICS Cluster 2 (estimate = 0.472, p-Value = 0.010), GICS Cluster 3 (estimate = 0.473, p-Value = 0.020), GICS Cluster 4 (estimate = 0.905, p-Value = 0.000663), GICS Cluster 6 (estimate = 0.809, p-Value = 9.03×10^{-5}), GICS Cluster 7 (estimate = 0.372, p-Value = 0.022), and GICS Cluster 8 (estimate = 0.841, p-Value = 2.54×10^{-5}). The variable for the percentage of Women employed in the Firm also has a significant coefficient of -0.231 with a p-Value of 4.13×10^{-5}, indicating a strong negative effect. Additionally, the first-stage residuals (Residual B) have a significant positive coefficient of 0.537 with a p-Value of 5.64×10^{-4}. The R^2 value is 0.1986, and the adjusted R^2 is 0.1718, indicating a better fit than Model A. The Pseudo R^2 is 0.1301. The Wald test p-Value is $< 2.2 \times 10^{-16}$, confirming the overall significance of the model predictors. The Breusch-Pagan test p-Value is 0.147, indicating no significant heteroscedasticity in the model.

In the Residuals vs. Fitted plot, the randomly scattered residuals suggest a well-specified model. The Q-Q Plot shows residuals closely following the theoretical quantiles, with minor deviations at the extremes, indicating approximate normality and the presence of some outliers.

In the last Model C, the coefficient for $prob\_C$ is not significant. But the $Women\_emp$ coefficient is significant (estimate = -0.237, p-Value = 2.69×10^{-7}), indicating a noticeable negative correlation with Δ. The $resid\_C$ coefficient is also significant (estimate = 0.644, p-Value = 6.67×10^{-5}). The R^2 value is 0.1052, and the adjusted R^2 is 0.0915. The Pseudo R^2 is 0.0706. The Wald test p-Value is $< 2.2 \times 10^{-16}$, confirming the overall significance of the model predictors. The Breusch-Pagan test p-Value is 0.2625, indicating no significant heteroscedasticity in Model C. The Residuals vs. Fitted plot shows no discernible trends or patterns, with residuals well-scattered around the zero line, indicating a well-specified model. The Q-Q Plot shows residuals generally following the theoretical quantiles but with deviations at the tails, suggesting some outliers and slight departures from normality.

As a last test the Likelihood-Ratio Test between the Models has been used. The comparison of Model A and Model B shows a p-Value of 0.0007, indicating that Model B provides a significantly better fit than Model A. The Likelihood-Ratio Test between Model A and Model C shows a p-Value of 0.0020, indicating that Model C provides a significantly better fit than Model A. The Likelihood-Ratio Test between Model B and Model C shows a p-Value of 0.0355, indicating that Model C provides a significantly better fit than Model B.

The three models appear to be well-specified with no clear signs of specification errors. The residuals in each model exhibit randomness, and the Q-Q plots demonstrate a good fit with theoretical quantiles, suggesting that the Instruments used are likely appropriate. These visual checks do not reveal any patterns that would indicate issues of endogeneity, implying that the Instruments are neither weak nor invalid. It is crucial to highlight that the first-stage Instrument was not significant, whereas the first-stage residuals were highly significant in the second stage for all models. This finding implies that the Instrument is effective in addressing some endogeneity but does not fully resolve it. Consequently, there may still be some endogeneity present in the models, possibly due to the Instrument not fully capturing the endogenous variable or the existence of omitted variables that are correlated with both the endogenous variable and the outcome variable.

3.3.3 Results Internal Score

As before, now the Internal Score is tested with the new Instruments to determine if they provide additional insights into the behavior of these scores. The three different models are composed as usual: Model A includes only the Instrument variable, Model B selects covariates using an algorithm to achieve the most significant results, and Model C includes the same covariates as in the Delta Score analysis.

Internal Score and CSR Rating IV

In the first stage as seen in Table 28, the basic Model A shows no significant effect of the Instrument on the presence of CSR Committees. Model B, which includes the Instrument along with $RPSH$, $female\_board$, and $Women\_emp$, indicates that only the amount of females in a company's board is significant. However, the AIC improves only slightly compared to Model A. Model C, which still includes the same covariates as in all previous Models, exhibits the lowest AIC of 277.51, also has one highly significant variable, which is the Board Size.

In the second stage, Model A continues to perform poorly, with the Instrument variable remaining insignificant and a very low R^2. Specifically, the model shows an adjusted R^2 of 0.0976, indicating that the predictors explain very little of the variance in the dependent variable. The Pseudo R^2 is 0.0009. The Wald test (p-Value = 4.761e-10) and the Breusch-Pagan test (p-Value = 0.03789) indicate some model significance and heteroscedasticity. In Addition, the significant residuals (p-Value = 1.91e-12) suggest that some endogeneity persists, which the Instrument alone cannot address. This interpretation confirms that while the Instruments help to account for endogeneity, some unobserved factors influencing the dependent variable remain.

In Model B, the coefficient for *prob_B* is 7.847 with a p-Value of 0.0033, indicating a significant positive relationship. The variable of female Employment has a marginally significant coefficient of -0.080 with a p-Value of 0.077. The variables of RPSH and the percentage of females on the board are both significant at the 5% significance level, both exhibit negative coefficients. The residuals from the first stage (Residual B) have a significant coefficient of 1.078 with a p-Value of 7.92×10^{-12}. These results indicate that *prob_B* has a strong positive effect, Women Employment has a marginal negative effect, and the residuals from the first stage have a strong positive effect on the dependent variable. The R^2 value is 0.1270, and the adjusted R^2 is 0.1175, indicating a better fit than Model A. The Pseudo R^2 is 0.0263. The Wald test p-Value is $< 2.2 \times 10^{-16}$, confirming the overall significance of the model predictors.

Table 14: IV CSR-Rating - Internal Score

	Dependent variable: Internal		
	(A)	(B)	(C)
Constant	-0.913	-7.141**	-3.289*
	(5.895)	(2.423)	(1.644)
Probability A	1.003		
	(6.478)		
Probability B		7.847**	
		(2.662)	
Probability C			3.768*
			(1.853)
RPSH		-0.108*	-0.063
		(0.050)	(0.046)
ROE			0.0413
			(0.0429)
Female Board		-0.201*	
		(0.101)	
Women Employment		-0.080·	-0.111*
		(0.045)	(0.043)
Board Size			0.111
			(0.100)
Credit			-0.016
			(0.009)
Residual A	1.113***		
	(0.154)		
Residual B		1.078***	
		(0.154)	
Residual C			0.950***
			(0.152)
R^2	0.1015	0.1270	0.1781
Adjusted R^2	0.0976	0.1175	0.1656
Pseudo R^2	0.0010	0.0263	0.0670
Wald test p-Value	4.76e-10***	< 2.2e-16***	< 2.2e-16***
Breusch-Pagan test p-Value	0.0379**	0.0406**	0.1938

Note: *p<0.1; **p<0.05; ***p<0.01

Model C exhibits a significant Instrumented variable ($prob\_C$, estimate = 3.768, p-Value = 0.0425) and a significant variable for $Women\_emp$ (estimate = -0.111, p-Value = 0.0104). The adjusted R^2 is 0.1656, suggesting a higher explanatory power compared to the other models. The Pseudo R^2 is 0.0670. The Wald test (p-Value < 2.2e-16) suggests that overall, the model predictors are significant, but individually some are not. The Breusch-Pagan test (p-Value = 0.1938) shows no significant heteroscedasticity. However, the significant residuals (p-Value = 8.71e-10) indicate that some endogeneity persists. From an economic perspective, the significance of the $Women\_emp$ variable in Model C aligns with the theory that diversity and inclusivity in the workforce can enhance a company's internal CSR scores. The significance of GICS Cluster 8 in Model B suggests that certain industry sectors have unique characteristics that influence internal CSR scores, possibly due to higher regulatory scrutiny or stakeholder pressure.

The difference of Model A and B, as explored by the Likelihood-Ratio test is not significant. The difference between Model A and Model C is in the LR-test significant meaning adding the additional covariates improves the Model performance. Next Model B and Model C are compared here we also find a significant Improvement. Meaning adding the covariates of Model C results in more explanatory Power then the covariates of Model B. This Result is coherent with the output for the Pseudo R^2, Model C has the highest value 0.0670.

In summary, Model A remains ineffective with negligible explanatory power. Model B shows improvement with significant GICS clusters, indicating some explanatory power. Model C, with its higher adjusted R^2 and significant variables for $prob\_C$ and $Women\_emp$, suggests the highest explanatory power among the models. This suggests that the inclusion of both female employment and specific GICS clusters provides a more comprehensive explanation of the dependent variable. But nevertheless the values for the different R^2 are very low with a max of just 17% of the Variance beeing explained by the Model

All three models exhibit significant and positive coefficients for the first stage residuals in the second stage of the regression. This finding indicates that, even though the Instrument (CSR-Rating) did not have a significant effect on the endogenous variable in the first stage, it successfully accounts for a substantial amount of endogeneity in the model when the residuals are included in the second stage. This result suggests that the Instrument may not be very strong in terms of its direct predictive power on the endogenous variable (weak Instrument), but the inclusion of the residuals helps to correct for endogeneity bias in the estimator. In other words, the residuals from the first stage capture the part of the endogenous variable that is correlated with the error term in the outcome equation, which is essential for addressing endogeneity.

Internal Score and SDR Rating IV

In the first stage as seen in Table 29, Model A exhibits no significant variables. Model B includes several covariates besides the Instrument variable, such as $RPSH$, $Total\_assets$, ROA, ROE, $Women\_emp$, and $GICS\_cluster$. The parameters for the Instrument, the amount of total assets, and ROA are significant. Additionally, the AIC value improves significantly compared to Model A, indicating a better model fit. In Model C, as usual, only the Board Size is positive and highly significant. This suggests that Board Size generally has a substantial impact on the presence of CSR Committees. In the second stage, Model A demonstrates that even with just one covariate, the Instrumented CSR Committee variable has a positive but not significant effect on the Internal Score (estimate = 7.704). Due to the inclusion of only one covariate, the R^2 (0.0992) and adjusted R^2 (0.0953) values are very low. The residuals are highly significant (estimate = 1.060, p-Value = 2.33e-11), indicating some remaining endogeneity.

In Model B, the Instrumented variable of CSR Committees ($prob\_B$) remains highly significant and positive (estimate = 3.582, p-Value = 2.10e-05). Two GICS clusters also show significant coefficients: GICS Cluster 7 (estimate = 0.338, p-Value = 0.0387) and GICS Cluster 8 (estimate = 0.492, p-Value = 0.0135). Additionally, the residuals from the first stage (Residual B) are significant (estimate = 0.887, p-Value = 1.99e-08),

Table 15: IV SDR Rating - Internal Score

	Dependent variable: Internal		
	(A)	(B)	(C)
Constant	-7.020	-3.472***	-3.529*
	(3.612)	(0.772)	(1.424)
Probability A	7.704		
	(3.960)		
Probability B		3.582***	
		(0.833)	
Probability C			4.033*
			(1.596)
RPSH		-0.059	-0.065
		(0.043)	(0.045)
Total Assets		0.047	
		(0.059)	
ROA		0.063	
		(0.058)	
ROE		0.034	0.041
		(0.044)	(0.043)
Women Employment		-0.074	-0.114**
		(0.056)	(0.043)
Board Size			0.098
			(0.089)
Credit			-0.017
			(0.009)
GICS Cluster 2		-0.025	
		(0.183)	
GICS Cluster 3		0.063	
		(0.204)	
GICS Cluster 4		0.444·	
		(0.265)	
GICS Cluster 5		0.272	
		(0.193)	
GICS Cluster 6		0.240	
		(0.206)	
GICS Cluster 7		0.338*	
		(0.163)	
GICS Cluster 8		0.492*	
		(0.198)	
GICS Cluster 9		0.019	
		(0.227)	
Residual A	1.060***		
	(0.155)		
Residual B		0.887***	
		(0.155)	
Residual C			0.897***
			(0.153)
R^2	0.0992	0.1823	0.1746
Adjusted R^2	0.0953	0.1550	0.1620
Pseudo R^2	0.0026	0.1301	0.0706
Wald test p-Value	3.301e-10***	<2.2e-16***	<2.2e-16***
Breusch-Pagan test p-Value	0.0270**	0.0082**	0.2156

Note: ·p<0.1; *p<0.05; **p<0.01; ***p<0.001

indicating a strong positive effect. The R^2 improves significantly to 0.1823, suggesting better explanatory power, with an adjusted R^2 of 0.1550. The Breusch-Pagan test p-Value is 0.0082, indicating some evidence of heteroscedasticity in the model.

In Model C, the Instrumented CSR Committee variable remains positive and slightly significant as in Model B (estimate = 4.033, p-Value = 0.012). Additionally, the number of females employed in a firm remains significant and negative (estimate = -0.114, p-Value = 0.009). The residuals in Model C are significant (estimate = 0.897, p-Value = 8.4e-09), suggesting persistent endogeneity.

From an economic perspective, the presence of a CSR-Committee is viewed positively by internal stakeholders because it signifies a company's commitment to sustainable and socially responsible practices. This aligns with stakeholder theory, which posits that firms should consider the interests of all stakeholders, not just shareholders [Ong et al., 2023].

The puzzling negative relationship between female employment and the internal score in Model C might indicate underlying biases or structural issues within the company that need to be addressed. This result contradicts the general expectation that higher female representation should correlate with better CSR practices, reflecting a potential area for further investigation. Model B stands out as the best model based on its diagnostic parameters and significant variables. It highlights the importance of considering a comprehensive set of covariates to understand the internal dynamics influencing CSR practices. The strong significance of the CSR Committee variable across models underscores its critical role in shaping internal perceptions of sustainability and social responsibility.

The Wald test results indicate significant differences between the models, confirming that including additional covariates improves the model fit. The Pseudo R^2 values suggest that Model B has the highest explanatory power, followed by Model C and Model A. The Breusch-Pagan test results indicate heteroskedasticity in Models A and B but not in Model C.

The Likelihood-Ratio test values provide further insight into the relative fit of the models. Comparing Model A and Model B, the test shows a significant improvement in model fit when additional covariates are included. Similarly, comparing Model A and Model C, the inclusion of further covariates also significantly improves the model fit. Comparing Model B and Model C, the test indicates that the different sets of covariates lead to significantly different model fits. These results suggest that a comprehensive set of covariates is necessary to capture the factors influencing CSR practices effectively.

3.3.4 Results External Score

External Score and CSR Rating IV

The Basic Model A exhibits no significant effects in the first stage, the AIC is at $286,16$. Model B includes the covarites $RPSH$, ROE and $GICS\_cluster$ together with the first stage Instrument. Just the amount of female board member exhibits a significant effect on the presences of CSR Committees. Model C with the covariates $RPSH$, ROE, $Women\_emp$, $Board\_size$ and $Credit$ just shows a significant effect for the Board Size and the lower AIC of the three Models of $277,51$. All first stage Results can be found in the Appendix in Table 31.

Model A shows significant coefficients for the Instrumented variable $prob\_A$ (estimate = 17.7082, p-Value = 0.00899). But, the R^2 value is very low (0.02439), indicating that only 2.44% of the variance in the dependent variable "External" is explained by this model. Although the model is statistically significant overall (Wald test p-Value = 0.009063), its explanatory power is limited. The residuals are significant (estimate = 0.3457, p-Value = 0.03140), indicating some remaining endogeneity. The Breusch-Pagan test for heteroscedasticity is not significant (p-Value = 0.4638), suggesting no issues with variance stability in Model A.

Model B improves upon Model A by including additional covariates. In Model B, the Instrumented variable of CSR Committees ($prob\_B$) remains significant and positive (estimate = 6.066, p-Value = 0.0272). The variable Women Employment has a highly

Table 16: IV CSR-Rating - External Score

	Dependent variable: External		
	(A)	(B)	(C)
Constant	-16.116**	-5.521*	-3.472*
	(6.143)	(2.492)	(1.724)
Probability A	17.708**		
	(6.750)		
Probability B		6.066*	
		(2.738)	
Probability C			3.861*
			(1.944)
RPSH		-0.024	-0.004
		(0.051)	(0.048)
ROE			0.0194
			(0.0451)
Female Board		-0.096	
		(0.104)	
Women Employment		0.217***	0.204***
		(0.046)	(0.045)
Board Size			-0.008
			(0.105)
Credit			-0.005
			(0.010)
Residual A	0.346*		
	(0.160)		
Residual B		0.281·	
		(0.158)	
Residual C			0.197
			(0.159)
R^2	0.0244	0.0765	0.0952
Adjusted R^2	0.0202	0.0665	0.0815
Pseudo R^2	0.0010	0.0263	0.0670
Wald Test p-Value	0.0091**	5.359e-08***	1.494e-12***
Breusch-Pagan Test p-Value	0.4638	8.07e-05***	9.862e-05***

Note: (standard error); ·p<0.1; *p<0.05; **p<0.01; ***p<0.001
**p<0.05

significant positive effect (estimate = 0.217, p-Value = 4.21e-06). Additionally, the residuals from the first stage (Residual B) are marginally significant (estimate = 0.281, p-Value = 0.0756), indicating some remaining endogeneity. The R^2 improves to 0.0765, suggesting moderate explanatory power, with an adjusted R^2 of 0.0665. The Breusch-Pagan test p-Value is 8.07e-05, indicating significant heteroscedasticity in the model

Model C incorporates theoretically relevant covariates and demonstrates significant effects and superior diagnostic metrics compared to the other models. The variable $Women\_emp$ exhibits a positive and highly significant effect on the External Score (estimate = 0.2039, p-Value = 9.17e-06). The coefficient for $prob\_C$ is also significant (estimate = 3.8615, p-Value = 0.0476). The R^2 value (0.0952) and the adjusted R^2 value (0.0815) are notable, and the Wald test for Model C is highly significant (p-Value = 1.494e-12), indicating that the overall model is statistically robust. The residuals in Model C (estimate = 0.1968, p-Value = 0.2167) are not significant, suggesting that most endogeneity has been addressed by the Instrument and covariates. The Breusch-Pagan test for Model C also indicates significant heteroscedasticity (p-Value = 9.862e-05).

In Model A and Model B, the significant residuals suggest that some endogeneity persists, which the Instrument alone cannot address. In contrast, the non-significant residuals in Model C suggest that the Instrument and covariates effectively control for endogeneity, leading to unbiased estimates of the effect of the endogenous variable ($prob\_C$) on the outcome. This indicates that Model C is the most robust among the three models evaluated.

The Likelihood-Ratio (LR) tests further compare the models. The comparison between Model A and Model B shows that adding $RPSH$, ROE, and $GICS\_cluster$ does not significantly improve the model fit (p-Value = 0.06676), although it is close to the significance threshold. Next, the comparison between Model A and Model C shows that adding $RPSH$, ROE, $Women\_emp$, $Board\_size$, and $Credit$ significantly improves the model fit (p-Value = 0.0022). Lastly, between Model B and Model C the LR-test shows a significant improvement when adding the additional covariates (p-Value = 0.003209). The Pseudo R^2 also results in the highest numbers for Model C. Given these diagnostics, Model C emerges as the best model due to its higher Pseudo R^2 value, significant Likelihood-Ratio tests, and the effective control for endogeneity, as indicated by the non-significant residuals.

External Score and SDR Rating IV

In Model A, the first stage revealed no significant effects for the independent variables (Table 32). This indicates that the Instrument used in Model A does not effectively explain the variability in the dependent variable, suggesting a poor model fit. Model B includes the variables $RPSH$, $Total\_assets$, ROA, ROE, $Women\_emp$, and $GICS\_cluster$. The analysis of Model B shows positive and significant coefficients for $Total\_assets$, indicating that the amount of total assets has a positive and significant impact on the dependent variable. Conversely, the coefficients for ROA are negative and significant, suggesting that return on assets has a negative effect. Additionally, some of the $GICS\_clusters$ exhibit significant effects, highlighting their relevance in the model. Model C uses the same covariates as in all previous IV estimations of Model C. The results indicate a positive and highly significant effect for the size of the board. This finding is consistent with previous analyses, where the Board Size was a crucial factor influencing the dependent variable. In summary, the new Instrument SDR Rating was tested across three different models. Model A showed no significant effects, indicating an ineffective Instrument. Model B demonstrated significant positive effects for $Total\_assets$ and negative effects for ROA and some $GICS\_clusters$. Model C, consistent with previous analyses, revealed a significant positive effect for Board Size. Based on these findings, Model B and Model C provide valuable insights into the relationship between the Instrument and the possible endogenous variable, with Model C continuing to highlight the importance of Board Size. The first stage regression results can be found in the Appendix Table 32.

Table 17: IV SDR Rating - External Score

	Dependent variable: External		
	(A)	(B)	(C)
Constant	-3.238	-1.786*	-3.110*
	(3.806)	(0.782)	(1.498)
Probability A	3.554		
	(4.172)		
Probability B		2.283**	
		(0.844)	
Probability C			3.444*
			(1.679)
RPSH		0.016	0.000
		(0.043)	(0.047)
Total Assets		-0.093	
		(0.060)	
ROA		0.007	
		(0.059)	
ROE		0.032	0.022
		(0.044)	(0.045)
Women Employment		0.245***	0.204***
		(0.057)	(0.046)
Board Size			0.012
			(0.094)
Credit			-0.003
			(0.009)
GICS Cluster 2		-0.713***	
		(0.186)	
GICS Cluster 3		-0.606**	
		(0.206)	
GICS Cluster 4		-0.766**	
		(0.268)	
GICS Cluster 5		0.159	
		(0.195)	
GICS Cluster 6		-0.889***	
		(0.208)	
GICS Cluster 7		-0.127	
		(0.165)	
GICS Cluster 8		-0.615**	
		(0.201)	
GICS Cluster 9		0.102	
		(0.230)	
Residual A	0.321*		
	(0.163)		
Residual B		0.300·	
		(0.157)	
Residual C			0.156
			(0.161)
R^2	0.0099	0.1694	0.0958
Adjusted R^2	0.0056	0.1417	0.0820
Pseudo R^2	0.0026	0.1301	0.0706
Wald test p-Value	0.161	1.367e-14***	6.827e-13***
Breusch-Pagan test p-Value	0.2911	5.796e-08***	4.1e-05***

Note: (standard error); ·p<0.1; *p<0.05; **p<0.01; ***p<0.001

In the second stage of our analysis, Model A still fails to achieve any significant results. With an R^2 of just 0.0099, the explanatory power of this model is highly questionable, indicating that it is ineffective in explaining the variance in the dependent variable. The diagnostic tests for Model A show a very low Pseudo R^2 (0.0026) and the Wald test (p-Value = 0.161) confirms the overall insignificance of the predictors. The Breusch-Pagan test for heteroscedasticity is not significant (p-Value = 0.2911), suggesting no issues with variance stability in Model A. The residuals are significant (estimate = 0.321, p-Value = 0.0496), indicating some remaining endogeneity.

Model B, on the other hand, exhibits positive and highly significant coefficients for the Instrumented variable *prob_B* (estimate = 2.283, p-Value = 0.007). Additionally, several GICS clusters also show significant effects, including GICS Cluster 2 (estimate = -0.713, p-Value = 0.00014), GICS Cluster 3 (estimate = -0.606, p-Value = 0.00346), GICS Cluster 4 (estimate = -0.766, p-Value = 0.00451), GICS Cluster 6 (estimate = -0.889, p-Value = 2.37e-05), and GICS Cluster 8 (estimate = -0.615, p-Value = 0.00232). The variable Women Employment is also significant (estimate = 0.245, p-Value = 1.93e-05). The R^2 value for Model B improves considerably to 0.1694 with an adjusted R^2 of 0.1417, suggesting a better fit and greater explanatory power. This indicates that the inclusion of these covariates significantly enhances the model's ability to explain the dependent variable. The diagnostic tests for Model B show a higher Pseudo R^2 (0.1301) and the Wald test confirms the overall model significance (p-Value = 1.367e-14). The Breusch-Pagan test indicates significant heteroscedasticity (p-Value = 5.796e-08), suggesting potential issues with variance stability. The residuals in Model B are marginally significant (estimate = 0.300, p-Value = 0.05681), indicating the presence of endogeneity.

The last Model C demonstrates significant coefficients for the Instrumented variable *prob_C* (estimate = 3.444, p-Value = 0.041) and the number of women employed in the company (estimate = 0.204, p-Value = 1.02e-05). Consistent with previous findings, this underscores the importance of female employment in influencing the dependent variable. Model C demonstrates an improvement in R^2 to 0.0958 with an adjusted R^2 of 0.0820, which, while not surpassing the improvement observed in Model B, still indicates a significant explanatory power. The diagnostic tests for Model C show a moderate Pseudo R^2 (0.0706) and the Wald test indicates overall model significance (p-Value = 6.827e-13). The Breusch-Pagan test indicates significant heteroscedasticity (p-Value = 4.1e-05). The residuals in Model C (estimate = 0.156, p-Value = 0.333) are not significant, suggesting that most endogeneity has been addressed by the Instrument and covariates.

The first-stage residuals in Model A and Model B are significant, indicating some remaining endogeneity. In contrast, Model C has non-significant first-stage residuals, which means the covariates and especially the Instrumented variable in Model C effectively control for endogeneity, leading to unbiased estimates of the effect of the endogenous variable ($prob\_C$) in the second stage.

The Likelihood-Ratio (LR) tests further compare the three different models: Model A and Model B show in the comparison that adding $RPSH$, ROE, and GICS clusters significantly improves the model fit (p-Value = 0.0007344). The LR-test for Model A and Model C reveals that adding $RPSH$, ROE, $Women\_emp$, $Board\_size$, and $Credit$ significantly improves the model fit (p-Value = 0.002022). Lastly the LR-test for the Model B and Model C, indicates that Model B with additional GICS clusters and other covariates provides a better fit (p-Value = 0.03552). In summary, Model A remains ineffective with negligible explanatory power. Model B and Model C both demonstrate significant positive effects for female employment, with Model B also highlighting significant effects for several GICS clusters. The R^2 improvement in Model B is the most pronounced, indicating it as the most effective model in this stage of the analysis. This suggests that the inclusion of specific GICS clusters provides a more comprehensive explanation of the dependent variable.

It is logical that SDR and CSR ratings from external institutions are highly relevant to the external perception of a company and therefore stronger Instruments. If a

country receives a favorable rating in a sustainability scoring system, then this can significantly influence companies, customer and all other stakeholders. The importance of SDR (Sustainable Development Report) and CSR (Corporate Social Responsibility) cannot be overstated. These ratings serve as benchmarks for a country's commitment to sustainable practices, influencing its perception among the public, investors, and other stakeholders. High ratings in these systems signal to companies that the country is dedicated to sustainable and ethical practices. Companies today are increasingly aware of environmental and social issues and are proactively establishing tools and committees to improve their ESG (Environmental, Social, and Governance) scores.
In countries with high CSR or SDR ratings, it is expected that companies have a greater incentive to address their sustainability concerns more diligently. A high sustainability rating often indicates a robust regulatory framework within the country, compelling companies to adhere to stringent ESG standards. Consequently, these companies are more motivated to adopt sustainable practices.
Therefore, an external observer might reasonably assume that if a company operates in a country with high SDG ratings, the company itself likely upholds high SDG standards. This alignment with the country's sustainability practices would, in turn, contribute to a high external ESG score for the company.
This more close connection between the CSR and SDR Rating and the External Score is probably the reason for the more significant results of the Instrumented variable.
In the Appendix Regression Tables for all Stages of the 2SRI, for all IVs and for all Scores are provided Table 24 until Table 32

4 Propensity Score Matching

4.1 General Framework

The method of Propensity Score Matching is a statistical technique used to estimate the effect of a treatment, policy, or intervention by accounting for the covariates that predict receiving the treatment. PSM addresses the issue of selection bias and omitted variable bias (OVB) in observational studies, where treatment assignment is not random and may be correlated with the outcome of interest. The propensity score is defined as the probability of receiving the treatment given observed covariates [Angrist and Pischke, 2009, Wooldridge, 2010]. The propensity score theorem, according to Rosenbaum and Rubin (1983), extends this idea to estimation strategies that rely on matching instead of regression, where the causal variable of interest is a treatment dummy. To formalize the PSM approach, consider a binary treatment variable T, where $T = 1$ for individuals receiving the treatment and $T = 0$ for individuals in the control group. The goal is to estimate the average treatment effect (ATE) or the average treatment effect on the treated (ATT). The propensity score $p(X_i)$ is the conditional probability of receiving the treatment given covariates X_i:

$$p(X_i) \equiv E[D_i \mid X_i] \perp\!\!\!\perp P[D_i = 1 \mid X_i]. \tag{11}$$

Formally, the following theorem is used: Suppose the conditional independence assumption holds such that the potential outcome is independent of the assignment of treatment given a set of covariates $Y_{0i}, Y_1 \perp\!\!\!\perp D_i|x_i$. Then the potential outcomes are independent of the treatment assignment given the propensity score which is, as previously defined, just a scalar function of all the covariates $Y_{0i}, Y_1 \perp\!\!\!\perp D_i|p(X_i)$ [Hirano and Imbens, 2004]. The propensity score theorem says that you only need to control for covariates that affect the probability of being treated. In addition, the result hint that there is only a need to control for the probability of treatment itself. In practice, the propensity score theorem is usually used for estimation in four steps:

- As an initial step the propensity score $p(X_i)$ is estimated, with a parametric model like logit or probit.
- Then the treatment effect is estimated, this can be done by either matching on the estimated score from the previous step or use a weighting scheme [see Imbens, 2004, for an overview].

- Assessing Balance: It's crucial to assess the balance of covariates in the matched sample to ensure that the matching process has successfully created comparable treatment and control groups.
- Estimating Treatment Effects: Once a balanced matched sample is obtained, the average treatment effect on the treated (ATT) or the overall average treatment effect (ATE) can be estimated.

As mentioned the estimation of propensity scores involves modeling the treatment assignment using a logistic regression or other appropriate models based on the covariates X_i:

$$P(D_i = 1|X_i) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\ldots+\beta_k X_k)}}, \tag{12}$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are the coefficients estimated from the logistic regression model. After estimating the propensity scores, individuals in the treatment and control groups are matched based on their scores. Several matching techniques can be used, including nearest-neighbor matching, caliper matching, and stratification matching [Rosenbaum and Rubin, 1983, Rubin, 1973]. The most common approach is nearest-neighbor matching, where each treated unit is matched with one or more control units with the closest propensity score. Once the matching is completed, the treatment effect is estimated by comparing the outcomes of matched treated and control units. For the ATT, the difference in outcomes between treated units and their matched controls is calculated [Rubin, 1973, Stuart, 2010]:

$$ATT = \frac{1}{N_t} \sum_{i:T_i=1} (Y_i - Y_{m(i)}), \tag{13}$$

where N_t is the number of treated units, Y_i is the outcome for the treated unit i, and $Y_{m(i)}$ is the outcome for the control unit matched to i [Gu and Rosenbaum, 1993]. Propensity Score Matching helps to create a balanced comparison group by simulating a randomized experimental design, thereby mitigating the effects of confounding variables. It is crucial, however, to ensure a good match between treated and control units and to assess the balance of covariates after matching to validate the effectiveness of the matching process [Hansen, 2004]. PSM can significantly enhance the robustness of causal inferences in several ways:

- Reduction of Confounding: By matching units with similar propensity scores, PSM reduces confounding by observed covariates, leading to more accurate estimates of the treatment effect [Rosenbaum, 1991].

- Improvement of Study Design: PSM can improve the design of observational studies by creating a quasi-experimental setup, making them more akin to randomized controlled trials [Sekhon, 2008, Diamond and Sekhon, 2013].

- Flexibility in Matching: PSM offers flexibility in choosing the matching algorithm and criteria, allowing researchers to tailor the matching process to the specific context of their study [Stuart and Green, 2008].

- Enhancement of External Validity: By focusing on comparable units within the overlap region of the propensity score distribution, PSM can enhance the external validity of causal inferences.

Despite these advantages, PSM has limitations, including reliance on the assumption of no unobserved confounders (selection on observables) and the potential for poor matches if the propensity score model is incorrectly specified or if there is limited common support. Furthermore, the choice of matching algorithm and the handling of unmatched units can affect the estimates of treatment effects [Stuart and Green, 2008]. In conclusion, PSM is a powerful tool for enhancing the robustness of causal inferences in observational studies. It enables researchers to account for confounding by observed covariates, thereby improving the credibility of their findings. As a result, careful implementation of PSM, including the selection of a suitable propensity score model, matching algorithm, and thorough assessment of balance and common support, is essential for obtaining reliable results. [see Angrist and Pischke, 2009, Chapter 3.3.2]

4.2 Application

PSM is a technique used to control for selection bias by matching firms with CSR Committees to similar firms without such Committees, based on a range of observable characteristics [Imbens, 2004, Rosenbaum and Rubin, 1983]. This method attempts to mimic a randomized controlled trial, thereby allowing for a clearer interpretation of the effects of Greenwashing on perceived CSR performance. In applying PSM to the study of Greenwashing, the presence of CSR Committees is defined as the treatment variable [Angrist and Pischke, 2009]. To initiate the propensity score matching process, it was essential first to identify variables that influence not only the various Greenwashing scores but also the presence of CSR Committees. Much of the groundwork for identifying key factors affecting the different scores was laid in prior analyses. Consequently, this phase focused primarily on assessing the impact of different covariates on CSR Committees [Hirano and Imbens, 2004]. To pinpoint the most critical covariates, several regression approaches were employed. The analysis began with linear regression, both with and without clustered standard errors, to establish a baseline understanding [Angrist and Pischke, 2009]. Subsequently, models incorporating clustering by industry and by country were explored to capture any sector-specific or geographic variations in the data. Finally, stepwise regression was utilized to refine the selection of relevant variables further, ensuring a robust set of covariates for the matching process. This comprehensive approach ensured that the variables selected for the propensity score model were those most pertinent to both the scores and the presence of CSR Committees. The main variables were the Revenue per Share, Total assets and the percentage of women on the Board, which are now used for the matching [Azmat and Rentschler, 2017]. Companies that have established CSR Committees would then be matched with companies that have not, based on their propensity scores. In this study different matching algorithms have been considered [Du, 2015]

- Nearest Neighbor Matching is the most basic form of PSM, where each treated unit is matched with one or more control units that have the closest propensity scores.

This method is prized for its simplicity and flexibility, allowing for straightforward implementation. However, it may result in poor matches if close neighbors are not available, potentially introducing bias into the analysis [Rubin, 1973, Stuart, 2010].

- Optimal Matching improves on this by minimizing the total sum of distances between matched pairs across all pairs, using combinatorial optimization techniques. This method is effective in enhancing match quality and is efficient in larger datasets. Despite these advantages, the computational intensity and complexity of the algorithm can be prohibitive, particularly for less technically inclined researchers [Gu and Rosenbaum, 1993, Hansen, 2004].
- Genetic Matching uses genetic algorithms to iteratively adjust the weights of covariates in the propensity score model to achieve optimal balance. This method dynamically adapts to the data, potentially leading to superior covariate balance. But, the computational demands are significant, and there is a risk of overfitting the model to the sample data [Sekhon, 2008, Diamond and Sekhon, 2013].
- Full Matching categorizes units into subsets that each contain at least one treated and one control unit, maximizing the use of available data. While this method offers flexibility in group sizes and ensures broad utilization of data, it can complicate subsequent analyses due to variable group sizes and potential for incomplete matching in heterogeneous datasets [Rosenbaum, 1991].
- Subclassification stratifies the data into several strata based on the propensity scores and then compares outcomes within these strata. This reduces reliance on the exact specification of the propensity score model and is relatively easy to implement. Nonetheless, strata may have residual imbalances, and the method can lead to data fragmentation, reducing the effective sample size within each stratum [Cochran, 1968, Rosenbaum and Rubin, 1983].
- Coarsened Exact Matching (CEM) temporarily coarsens variables into broader

categories, matches exactly on these categories, and then refines the analyses within these matched groups [Blackwell et al., 2009]. CEM typically achieves excellent balance on coarsened variables and is robust to model specifications. The main drawback is the loss of detailed information due to coarsening, which can obscure nuanced differences between units.

Each of these methods offers unique benefits and poses specific challenges. Now the question is which matching algorithm to choose? Nearest neighbor and optimal matching are suitable for datasets where close matches are expected to be available, while genetic and full matching are better suited for more complex datasets with multiple covariates influencing treatment assignment. After matching, a balancing table got used, which assesses the distribution of covariates between treated and control groups post-matching. This step is critical as it helps verify whether the matching process has successfully minimized pre-existing differences between the groups, an essential condition for unbiased estimation of the treatment effect [Furlow, 2010]. This also enables Method Comparison, by applying and comparing multiple matching methods, researchers can identify which method provides the best balance for their specific dataset and research questions, optimizing the quality of their analyses. It also demonstrates the Robustness of the analysis, by using multiple methods and assessing the balance systematically helps demonstrate the robustness of the study findings, providing stronger evidence for stakeholders and decision-makers.

4.3 Results

The analysis commenced with an algorithm designed to evaluate various matching models to identify the most suitable one. In this algorithm, each previously described matching algorithm was tested systematically. The key criterion for selection was the improvement in the balance of differences between the control group and the treatment group. This improvement guided the decision on which model to choose. The primary objective was to create two homogeneous samples where the main difference stemmed from one group of firms having CSR Committees while the other group did not. By ensuring that the samples were otherwise similar in terms of key characteristics, the aime is to isolate the effect of CSR Committees on the outcomes of interest. The algorithm iteratively checked each matching process, refining the selection of firms in both groups to enhance comparability. Several iterations of testing and refinement were conducted to assess the effectiveness of each matching model. Various metrics were employed to evaluate the balance between the groups, such as standardised mean differences and variance ratios. The algorithm continued to test the model until the optimal balance Matching algorithm was found. Ultimately, the optimal matching model was determined by its ability to minimize differences between the control and treatment groups across all covariates except for the presence of CSR Committees. This rigorous approach ensured that the resulting matched samples provided a robust basis for comparing firms with and without CSR Committees, thereby isolating the impact of CSR Committee presence on the observed outcomes. The chosen 'Optimal Matching' model was deemed the best at achieving the desired homogeneity between the two groups, thus allowing for a more accurate assessment of the influence of CSR Committees. Afterward finding the best matching algorithm the application of optimal pair matching began, to account for the differences between treated and control units, using the covariates $RPSH$, $Total\_assets$, and $female\_board$. The matching procedure employed a propensity score estimated through logistic regression. However, it should be noted that there were fewer control units than treated units, resulting in some treated units not finding a match. Initially, the dataset consisted of 467 observations, which were reduced to 84 matched observations post-matching.

The optimal matching method was used to perform 1:1 optimal pair matching based on the propensity score, which was estimated using logistic regression [Gu and Rosenbaum, 1993]. The covariates used for matching were $RPSH$, $Total\_assets$, and $Female\_board$. Because there were fewer control units than treated units, not all treated units were matched [Rosenbaum, 1991, Stuart, 2010]. The summary shows that out of the original

467 observations, only 84 were matched (42 control and 42 treated), targeting the Average Treatment effect on the Treated (ATT). In the balance table both unadjusted (Diff.Un) and adjusted (Diff.Adj) differences for the covariates and the propensity score distance are presented. After matching, the differences between the treatment and control groups were significantly reduced [Stuart and Green, 2008]: These adjusted differences indicate a substantial improvement in the balance between the control and treatment groups, suggesting that the matching procedure was effective in creating comparable groups. The final sample sizes indicate that all 42 control units were matched, while 383 treated units remained unmatched, reflecting the initial warning about the imbalance in the number of control and treated units [Rubin, 2001].

Covariate	**Type**	**Difference Unadjusted**	**Difference Adjusted**
distance	Distance	0.6740	0.0131
RPSH	Continuous	0.0608	0.0208
Total_assets	Continuous	0.3047	0.0374
Female_board	Continuous	0.4065	−0.3118

Balance Measures - Delta Score - Before and After Matching

The results indicated significant improvements in balance, with the adjusted differences for all covariates being close to zero, suggesting that the matching process effectively reduced covariate imbalance between the treated and control groups.

Density plots were created to visualize the distributions of CSR_COMM,$RPSH$, $Total\_assets$, and $female\_board$ before and after matching, see Figure 36 until Figure 43 in the Appendix. The density plots for $RPSH$ and $Total\_assets$ demonstrated that the distributions were similar pre- and post-matching, indicating a successful balancing of these covariates. Just the density plot for $female\_board$ showed some residual imbalance, we also observe for this variable the lowest levels of improvement. Subsequent to matching, a linear regression model was fitted to the matched data to evaluate the effect of the treatment. The initial model, which included only the treatment variable, revealed a significant treatment effect. This model was then extended to incorporate additional covariates such as $Board\_size$, $Women\_emp$, $female\_board$, $RPSH$, ROA, ROE, $Total\_assets$, and $GICS\_cluster$. The expanded model continued to show a significant treatment effect along with significant contributions from some of the added covariates.

Further analysis involved fitting fixed-effects models to control for unobserved heterogeneity. These models included fixed effects for Size and $GICS\_cluster$. The fixed-effects model with Size showed a significant positive effect of the CSR Committee on Delta, and a significant negative effect of $Women\_emp$. The inclusion of both Size and $GICS\_cluster$ in the fixed-effects model confirmed the robustness of the CSR Committee's positive effect on Delta.

Additionally, clustered models were estimated to account for different clustering structures, including subclass, $GICS\_cluster$, and Country. The clustered models consistently indicated a significant treatment effect, underscoring the robustness of the findings across various clustering specifications.

A comparison table summarizing the results from the three fixed-effects models highlighted the consistent positive effect of the CSR Committee on Delta across all specifications. The final analysis confirmed that the CSR Committee had a significant positive impact on Delta, while $Women\_emp$ had a significant negative impact. These findings were robust across different model specifications and clustering methods.

Table 18: Model PSM - Delta Score - Fixed Effects

	Model A	Model B	Model C
Dependent Var.:	Delta	Delta	Delta
Constant	-0.3371 (0.2572)		
CSR Committee	0.6321** (0.1830)	0.6398** (0.1902)	0.5192* (0.2210)
RPSH	0.5410 (0.3951)	0.4260 (0.5718)	0.7196 (0.5506)

(continued)

	Model A	Model B	Model C
ROA	-0.0498 (0.1020)	-0.0631 (0.1161)	-0.0480 (0.1110)
ROE	-0.0408 (0.4329)	-0.1044 (0.4718)	-0.1040 (0.4978)
Board Size	0.2657. (0.1378)	0.2797* (0.1260)	0.2457. (0.1351)
Female Board	-0.0070 (0.1033)	-0.0249 (0.1114)	0.0070 (0.1134)
Women Employment	-0.2391** (0.0830)	-0.2366** (0.0849)	-0.1787 (0.1182)
EBIT	7.647 (5.197)	7.036 (5.415)	7.043 (6.107)
Market Cap	-8.289 (5.589)	-7.657 (5.822)	-7.661 (6.536)
Credit	-0.0397. (0.0226)	-0.0395 (0.0271)	-0.0376 (0.0278)
Fixed-Effects:	———————	———————	———————-
Size	No	Yes	Yes
GICS_cluster	No	No	Yes
————————	————————	————————	————————
S.E.: Clustered	by: subclass	by: subclass	by: subclass
Observations	84	84	84
R2	0.29499	0.30183	0.36736
Within R2	–	0.27655	0.19390

When these findings are compared with those of the original paper, similarities for some variables but notable differences for others can be observed.

First, let us focus on the differences. Revenue per shareholder is now slightly positive and significant across all three models, implying that as revenue increases, firms are more inclined to engage in Greenwashing. In the original paper, these findings were reversed, showing a negative effect. This discrepancy could be attributed to the tendency of well-performing companies to maintain or increase their revenue by focusing on short-term solutions, which may lead to Greenwashing practices. Economically, this could be because higher revenues might enable firms to invest in image-enhancing strategies that do not necessarily align with genuine sustainability efforts, opting instead for superficial or deceptive practices to appear environmentally friendly.

The number of females on the board now also shows a positive, though very small, effect. This slight increase could be economically justified by the growing trend of gender diversity in corporate governance [Mu and Lee, 2023]. It has to be considered that the small size effect suggests that simply having more women on the board does not significantly deter Greenwashing, perhaps because the underlying corporate culture and policies are more influential than board composition alone.

On the other hand, a persistent, slightly positive, and significant relationship between the presence of CSR Committees and Greenwashing can be found. This indicates that companies with CSR Committees face a higher risk of engaging in Greenwashing activities compared to those without such Committees. This finding is rather surprising because CSR Committees are expected to have a positive impact on the company's sustainability measures and not incentivize Greenwashing. Economically, this could be explained by the potential misuse of CSR Committees as marketing tools rather than genuine oversight bodies. Companies might establish these Committees to signal their commitment to sustainability without implementing substantial changes, leading to increased Greenwashing activities.

A similar finding to the original paper is the number of women employed within a firm. Firms with more female employees are less likely to engage in Greenwashing practices. This result is consistent with the findings of the original paper [Bosone et al., 2024] and also aligns with previous studies by Azmat and Rentschler [2017] and Jiang and Akbar [2018]. From an economic perspective, this could be due to diverse workforces fostering more ethical and socially responsible decision-making processes, thus reducing the likelihood of deceptive practices. Both ROA and ROE remain positive and not significant in our analysis, similar to the original paper's findings. Economically, this suggests that profitability metrics such as ROA and ROE do not have a strong direct influence on Greenwashing behaviors. This might be because these financial performance indicators are more reflective of overall business efficiency and profitability, rather than specific ethical or sustainability-related practices.

In conclusion, the optimal pair matching effectively balanced the covariates between the treated and control groups, particularly for $RPSH$ and $Total\_assets$. The subsequent regression analysis consistently demonstrated the significant positive impact of the CSR Committee on Delta, alongside a significant negative impact of $Women\_emp$. These results were corroborated through various model specifications and clustering methods, providing robust evidence of the treatment effect. The density plots affirmed the balance achieved for $RPSH$ and $Total\_assets$, though some imbalance persisted for $female\_board$, warranting cautious interpretation in related analyses.

Internal and External Score

For both the internal and external scores, it is first assessed which variables influence both the outcome variable and the variable of interest to identify influential variables for matching. After evaluating several different matching algorithms, similar to the Delta model, both scores revealed optimal results using the optimal matching algorithm [Gu and Rosenbaum, 1993, Hansen, 2004]. This table summarizes the balance measures before and after matching for the covariates used in the analysis. The unadjusted differences (Diff.Un) represent the initial imbalance between the treatment and control groups. The adjusted differences (Diff.Adj) indicate the remaining imbalance after the matching process.

Covariate	**Type**	**Difference Unadjusted**	**Difference Adjusted**
distance	Distance	0.6740	0.0131
RPSH	Contin.	0.0608	0.0208
Total_assets	Contin.	0.3047	0.0374
Female_board	Contin.	0.4065	−0.3118

Balance Measures - for the Internal and External Score - Before and After Matching

The difference decreased significantly from 0.6740 to 0.0131, indicating a substantial improvement in the balance of propensity scores between the groups [Stuart and Green, 2008]. For the *Female_board* variable the difference changed from 0.4065 to -0.3118, indicating a significant adjustment, although the negative adjusted difference suggests a reversal in the initial imbalance direction [Rosenbaum, 1991]. Overall, the matching procedure effectively reduced the differences between the treatment and control groups for all covariates, resulting in more comparable groups [Rubin, 1973]. The Propensity Score Matching Result for External Score are displayed in Table 19. Regarding the dependent variable of the external score, the most significant changes can be observed. No variable appears to be significant, indicating that after propensity score matching, none of the variables continue to impact the external perception of the company by customers. This suggests that other factors must influence a company's external score [Rubin, 2001].

Table 19: PSM - External Score - Fixed Effects

	Model A	Model B	Model C
Dependent Var.:	External	External	External
Constant	-0.4694* (0.1999)		
CSR Committee	0.2434 (0.2161)	0.2055 (0.2047)	0.3505 (0.2232)
RPSH	0.5328 (0.4242)	0.1704 (0.5100)	-0.3770 (0.5195)
ROA	0.0095 (0.0801)	0.0815 (0.0782)	0.0761 (0.0917)
ROE	0.0987 (0.5679)	0.3091 (0.4965)	0.0726 (0.4875)
Board Size	0.0600 (0.1078)	-0.0649 (0.1240)	-0.0295 (0.1232)
Female Board	0.0403 (0.1152)	0.0438 (0.1034)	0.0257 (0.1181)
Women Employment	0.1645 (0.1068)	0.1484 (0.1151)	0.1612 (0.1291)
Credit	0.0313 (0.0218)	0.0193 (0.0236)	0.0136 (0.0239)

(continued)

	Model A	Model B	Model C
Fixed-Effects:	———————	——————-	——————-
Size	No	Yes	Yes
GICS_cluster	No	No	Yes
______________	_______________	______________	______________
S.E.: Clustered	by: subclass	by: subclass	by: subclass
Observations	84	84	84
R2	0.08940	0.17612	0.28056
Within R2	–	0.06695	0.06281

For the internal score see Table 20, a strong significance of the CSR Committee variable across all three model specifications can be observed. Additionally, Board Size and the number of female employees within a company appear to be important determinants of the internal score. Beyond these variables, no other variable consistently exhibits a significant effect across all model specifications.

Table 20: PSM - Internal Score - Fixed Effects

	Model A	Model B	Model C
Dependent Var.:	Internal	Internal	Internal
Constant	-0.6096* (0.2317)		
CSR Committee	0.9514*** (0.2163)	0.9183*** (0.2220)	0.8935** (0.2586)
RPSH	1.084* (0.4298)	0.7673 (0.5512)	0.6480 (0.6222)
ROA	-0.0462 (0.1105)	-0.0123 (0.1251)	-0.0085 (0.1120)
ROE	-0.0568 (0.3798)	0.0216 (0.4121)	-0.1492 (0.3998)
Board Size	0.3246* (0.1320)	0.2434. (0.1227)	0.2265. (0.1247)

(continued)

	Model A	Model B	Model C
Female Board	0.0022 (0.1477)	-0.0089 (0.1515)	0.0196 (0.1352)
Women Employment	-0.1705. (0.0952)	-0.1810. (0.0984)	-0.1052 (0.1213)
Credit	-0.0202 (0.0203)	-0.0268 (0.0221)	-0.0299 (0.0221)
Fixed-Effects:	————	————	————
Size	No	Yes	Yes
GICS_cluster	No	No	Yes
————	————	————	————
S.E.: Clustered	by: subclass	by: subclass	by: subclass
Observations	84	84	84
R2	0.35990	0.37901	0.44570
Within R2	–	0.27629	0.23164

For both dependent variables, the distribution of the variables before and after matching has to be examined. Although the number of observations was reduced, the distribution of the variables remained comparable and relatively similar.

5 Conclusion

This section summarizes all findings, highlighting significant results across various variables and models.

Delta Score

In the fixed effects model, clustered for company size and GICS cluster, the significant variables impacting the Delta Score were RPSH (Revenue Per Share), which had a negative impact. This suggests that higher revenue per share is associated with a lower risk of Greenwashing activities, indicating that financially stable firms are less likely to engage in deceptive sustainability practices. The CSR Committee indicator variable had a positive impact on the risk of Greenwashing, implying that firms with CSR Committees might use them as facades for Greenwashing. The percentage of females employed exhibited a negative impact, suggesting that gender diversity within the workforce contributes to more genuine and effective sustainability practices, as also proposed by Mu and Lee [2023].

In the Instrumental variable regression with the sea level trend as an Instrument for the presence of CSR Committees, only the percentage of women employed was significant and negative. This reinforces the idea that gender diversity reduces the risk of Greenwashing. Similarly, in the regression with SDR country rankings as Instruments, the percentage of women in the company remained the only significant and negative variable. This consistency underscores the importance of gender diversity in reducing the risk of Greenwashing [Mu and Lee, 2023]. The significance of this variable under Instrumental variable regression highlights its robustness as a determinant of genuine sustainability practices. Additionally, in the Instrumental variable approach with industry sector-specific clusters, the GICS Clusters for Energy (4), Consumer Staples (2), Healthcare (6), and Materials (8) were all highly significant and positive across the different Instruments and models. This finding indicates that companies in these sectors have a higher probability of engaging in Greenwashing activities.

From an economic perspective, it makes sense that companies in these sectors are more prone to Greenwashing. The Energy sector, due to its high environmental impact, faces intense scrutiny and pressure to appear sustainable, leading to higher incentives for Greenwashing. The Consumer Staples sector often involves products that directly impact consumers' daily lives, making sustainability claims a powerful marketing tool that companies might exploit. In Healthcare, the focus on sustainability is increasingly becoming important for reputation and trust, but the complexity and regulatory environment might lead companies to superficially enhance their sustainability profile. Lastly, the Materials sector deals with raw materials and manufacturing processes that are typically scrutinized for their environmental impact, creating a higher likelihood of Greenwashing to improve public perception.

Across all three Instruments, some persistent but not statistically significant patterns were identified. RPSH was consistently negative, while ROE and Board Size were positive. A negative RPSH suggests that firms with higher revenue per shareholder are less likely to have high delta scores, focusing more on genuine financial performance. A positive ROE implies that more profitable firms might be more prone to Greenwashing. Similarly, a positive Board Size indicates that firms with larger boards might have higher delta scores due to increased complexity in decision-making and less effective oversight.

The propensity score matching model confirmed these findings, with CSR Committees and Board Size exhibiting positive parameters, suggesting a higher risk of Greenwashing. The presence of CSR Committees exhibited a positive parameter, indicating that firms with CSR Committees still show a higher risk of Greenwashing, supporting the earlier finding that these Committees might be used superficially [Balluchi et al., 2020]. The positive and significant Board Size, suggests that larger boards are associated with a higher risk of Greenwashing, potentially due to a more complex and time consuming decision making process [Gatti et al., 2019]. The percentage of females employed consistently had a negative parameter value, confirming that higher female employment reduces the risk of Greenwashing [Du, 2015]. Diverse perspectives and ethical considerations brought in by female employees can lead to more authentic and effective sustainability practices. These analyses highlight the economic implications for firms and stakeholders, emphasizing the importance of financial stability, genuine CSR efforts, gender diversity, and effective governance in promoting true sustainability and mitigating Greenwashing risks.

Internal Score

The ESG internal score assesses a company's performance using self-reported corporate data. In the fixed effects model, clustered for company size and GICS cluster, several significant variables impacted the Internal Score. RPSH (Revenue Per Share) had a positive impact, suggesting that companies with higher revenue per share tend to have a better internal perception of their sustainability. This indicates that financial performance may boost confidence in sustainability efforts [Du, 2015]. The CSR Committee indicator also showed a positive impact, implying that firms with CSR Committees perceive themselves as more sustainable. Board Size had a positive impact, suggesting that larger boards contribute to a more favorable internal sustainability perception, possibly due to diverse expertise and increased governance [Anyigbah et al., 2023]. ROE (Return on Equity) was positively associated with the Internal Score, indicating that more profitable firms perceive themselves as more sustainable, likely because profitability provides more resources for sustainability activities and practices. Using the sea level trend as an Instrument for the presence of CSR Committees, significant variables included Board Size and the percentage of females in the workforce. Board Size had a positive impact, highlighting the importance of governance structures in shaping sustainability perception [Anyigbah et al., 2023]. However, the percentage of females in the workforce had a consistently negative impact, which is counterintuitive. This could indicate underlying biases in measuring internal sustainability scores or challenges in integrating diverse perspectives into sustainability initiatives.

In the Instrumental variable regression with CSR and SDR country rankings as Instruments, significant variables included the presence of CSR Committees, RPSH, the percentage of females on the board, and the percentage of women employed. The presence of CSR Committees and the percentage of females on the board had positive impacts, suggesting that formal structures and gender diversity at the board level contribute positively to internal sustainability perceptions. Conversely, the negative impact of the percentage of women employed suggests potential internal biases or the need for more inclusive sustainability approaches.

Furthermore, the inclusion of industry sector-specific clusters in the Instrumental variable approach revealed that the GICS Cluster for the Materials sector (8) was consistently significant and positive across various Instruments and models. This suggests that firms within the Materials sector are more likely to give themselves a higher internal score. Economically, this is understandable as the Materials sector encompasses activities such as extraction, processing, and manufacturing of raw materials, which are inherently environmentally intensive. Companies in this sector often face intense regulatory scrutiny and public pressure to demonstrate sustainable practices. As a result, these companies may resort to give themselves a better rating to enhance their public image and comply with regulatory expectations without fully committing to substantial environmental practices. The high environmental stakes and regulatory demands in the Materials sector create a strong incentive for companies to project an image of sustainability, even if it involves exaggerating or falsifying their actual efforts. These findings emphasize the need for stricter regulations and greater transparency within the Materials sector to curb Greenwashing and ensure authentic sustainability initiatives.

The propensity score matching model confirmed the significance of CSR Committees, RPSH, and Board Size. The positive impact of CSR Committees and RPSH aligns with previous findings, indicating that financial performance and formal sustainability structures enhance internal perceptions. Board Size was positively significant, reinforcing the role of governance in shaping sustainability perceptions [Delmas and Burbano, 2011, Anyigbah et al., 2023].

Overall, these findings provide valuable insights into the factors influencing a company's internal perception of its sustainability. Financial performance, indicated by RPSH and ROE, plays a crucial role in boosting confidence in sustainability efforts [Du, 2015]. Economically, this suggests that financially successful firms are likely to invest more in SDG supporting initiatives and view themselves as leaders in this area. The consistent positive impact of CSR Committees highlights the importance of formal sustainability structures. Firms with dedicated CSR Committees likely have more organized and strategic approaches to sustainability, leading to higher internal scores. On the other hand, this also suggests that the presence of a CSR Committee can enhance a firm's image internally, regardless of the actual effectiveness of its sustainability practices [Lyon and Montgomery, 2015]. The role of Board Size as a positive factor indicates that larger boards, with their diverse expertise and increased governance capabilities, contribute to better sustainability perceptions. This underscores the importance of strong governance structures in driving sustainability agendas within firms [Anyigbah et al., 2023]. On the other hand, the negative impact of the percentage of women employed raises important questions about internal biases and the need for more inclusive sustainability strategies. While diversity is crucial, internal perceptions may not always align with the actual benefits of having a diverse workforce, pointing to potential areas for improving internal communication and recognition of diverse contributions [Gatti et al., 2019].The consistently negative but not significant impact across all Models of credit scores on internal sustainability perceptions suggests that financially prudent firms may have a more realistic or cautious view of their sustainability challenges. This could reflect a broader understanding of the complexities involved in achieving true sustainability, leading to more conservative internal assessments. The consistency of certain variables, such as CSR Committees and percentage of females employed, across different models enhances the reliability of these findings. The variations in significance of other variables, like the size of the board, RPSH or the percentage of females on the board, highlight the nuanced nature of internal sustainability perceptions and the importance of considering multiple factors and perspectives. Overall, these findings emphasize the multifaceted nature of sustainability perceptions within firms, influenced by financial performance, governance structures, formal sustainable and social governance initiatives, and potentially internal biases. Addressing these factors holistically can help firms enhance both their actual and perceived sustainability, contributing to more effective and genuine sustainability practices.

External Score

The External Score is derived from standards and external data sourced from reputable entities that provide insights into the sentiment surrounding news related to sustainability matters concerning the company. This data is collected from specialized websites, NGOs, vertical websites, and mainstream news sources, including metrics such as controversies and reviews to capture instances where companies face sanctions or fines due to environmental violations or their involvement in highly polluting activities [Bosone et al., 2024]. Social media also serves as a proxy for other digital assets, creating a dynamic information stream about a company's daily performance [Yan and Yang, 2024].

In the fixed effects model, clustered for company size and GICS cluster, several variables significantly impacted the External Score. RPSH (Revenue Per Share) had a positive impact, suggesting that companies with higher revenue per share are perceived more favorably in terms of sustainability. This indicates that financial performance may enhance a company's external reputation regarding sustainability [Du, 2015]. The presence of CSR Committees also had a positive impact, implying that firms with CSR Committees are viewed more positively in terms of sustainability by external entities. Board Size showed a positive impact, suggesting that larger boards contribute to a better external sustainability perception, possibly due to increased oversight and governance [Anyigbah et al., 2023]. ROE (Return on Equity) was positively associated with the External Score, indicating that more profitable firms are perceived as more sustainable. The percentage of females in the company also had a positive impact, suggesting that gender diversity within the workforce enhances a company's external sustainability reputation.

In the Instrumental variable regression with the sea level trend as an Instrument, the percentage of women employed had a positive impact, suggesting that gender diversity is perceived favorably in terms of sustainability [Lyon and Montgomery, 2015]. In the regression with CSR and SDR country rankings as Instruments, significant variables included CSR Committees and the percentage of females in the workforce, both positively impacting external sustainability perceptions. ROE consistently had a positive impact across all models, indicating that more profitable firms are perceived to have better sustainability practices, likely due to their resources and better management practices.

Moreover, the Instrumental variable approach with industry sector-specific clusters showed that the GICS Clusters for Consumer Discretionary (2), Energy (4), Healthcare (6), and Materials (8) sectors were highly significant and positive across different Instruments and models. This suggests that companies within these sectors tend to have lower externally derived ESG Scores.

This observation is sensible for several reasons: Companies in the Consumer Discretionary sector often face intense scrutiny regarding their supply chains, product safety, and overall environmental impact, resulting in lower external ESG scores. The Energy sector, being associated with high levels of environmental risk and carbon emissions, finds it challenging to achieve high external ESG scores, even with claimed sustainability initiatives. In the Healthcare sector, environmental and social impacts of pharmaceutical production, alongside ethical issues, can adversely affect their external ESG ratings. Similarly, the Materials sector, which involves significant environmental degradation and pollution from mining and chemical processing, tends to suffer from poor external ESG perceptions.

These findings highlight the importance of targeted regulations and improved transparency in these sectors to foster genuine sustainability practices and enhance their external ESG scores.

The propensity score matching model did not yield significant results, indicating that this method was less effective in identifying impactful variables for the External Score in this context.

Economically, these results provide insights into how external entities perceive a company's sustainability efforts. The consistent positive impact of CSR Committees highlights the importance of formal sustainability structures in enhancing a company's external reputation. The positive impact of gender diversity further strengthens the idea that diverse workforces are generally viewed favorably for sustainability.

Overall, these findings underscore the multifaceted nature of sustainability perceptions, influenced by financial performance, governance structures, formal ESG practices, and potential biases. Addressing these factors holistically can help firms enhance their sustainability reputation, leading to more effective and genuine sustainability practices.

5.1 Overall Discussion

The analysis of the Delta Score, Internal Score, and External Score reveals several overarching themes and insights into the factors that influence sustainability perceptions and performance. These results provide a comprehensive understanding of how different variables affect a company's risk of Greenwashing, its self-reported sustainability performance, and its external reputation. Our comprehensive analysis across multiple models, including fixed effects, Instrumental variables, and propensity score matching, reveals several consistent and significant patterns regarding the factors influencing sustainability perceptions within firms.

Firstly, revenue per shareholder (RPSH) consistently emerged as a significant variable, albeit with differing impacts depending on the context. In the context of the Delta Score, a negative RPSH suggests that financially stable firms are less likely to engage in Greenwashing, possibly due to their focus on genuine financial performance rather than superficial CSR activities. Conversely, for the Internal Score, higher RPSH is associated with better internal sustainability perceptions, indicating that financial performance boosts confidence in sustainability efforts. These contrasting impacts highlight the complexity of financial metrics and their varying interpretations based on the specific sustainability dimension being examined.

The role of CSR Committees was another critical factor. The presence of CSR Committees generally had a positive and significant impact on both internal and external sustainability scores, suggesting that formal sustainability structures enhance a firm's sustainability reputation. However, the Delta Score analysis paradoxically indicated that firms with CSR Committees might have a higher risk of Greenwashing, implying that these committees could sometimes serve as facades rather than drivers of genuine sustainability efforts. This finding emphasizes the need for stakeholders to critically evaluate the authenticity of a firm's CSR activities.

Gender diversity within the workforce consistently showed a significant impact across different models. A higher percentage of female employees correlated with a lower risk of Greenwashing, reinforcing the notion that gender diversity promotes more authentic and effective sustainability practices [Lyon and Montgomery, 2015, Mu and Lee, 2023]. In the context of the Internal Score, the percentage of females employed has a negative impacts, and in the External Score, the percentage of females on the board has again a positive impact. These mixed results highlight the complexity of gender diversity's role in sustainability perceptions. While gender diversity is generally beneficial, internal perceptions may be influenced by biases or varying expectations. Mu and Lee [2023] also highlight how gender diversity impacts perceptions and actual sustainability practices, noting the nuanced effects observed in different contexts. This aligns with the original findings where we observed that a higher number of female employees is associated with a lower likelihood of engaging in Greenwashing. Moreover, the negative impact of female representation on the board for internal scores underscores the value of diverse perspectives in enhancing sustainability efforts and governance.

Board Size also emerged as a significant factor, with larger boards generally associated with higher sustainability scores. This finding suggests that larger boards, with their diverse expertise and increased governance capabilities, contribute positively to sustainability perceptions. However, the association with higher Greenwashing risks in the Delta Score analysis indicates that the effectiveness of larger boards may depend on their ability to maintain effective oversight and avoid superficial CSR activities.

Lastly, return on equity (ROE) was positively linked to higher sustainability scores across different models, indicating that more profitable firms are perceived to have better sustainability practices. This relationship may stem from the resources and better management practices that profitability enables, allowing firms to invest more comprehensively in sustainability initiatives.

Overall, these findings highlight the multifaceted nature of sustainability perceptions, influenced by financial performance, governance structures, formal ESG practices, and potential biases. Addressing these factors holistically can help firms enhance their sustainability reputation, leading to more effective and genuine sustainability practices. As forecasted by Delmas and Burbano [2011], after years of uncontrolled Greenwashing, consumers tend to become more cynical about green claims, making it hard for companies to deceive them by the mere establishment of CSR Committees. Statistically, the use of various models, including fixed effects, Instrumental variables, and propensity score matching, provides robust insights into the factors influencing sustainability scores. The consistency of certain variables, such as financial performance, CSR Committees, and Board Size, across different models enhances the reliability of these findings. The variations in the significance of other variables, such as gender diversity and Financial parameters, highlight the nuanced nature of sustainability perceptions and the importance of considering multiple factors and perspectives. Financial performance, governance structures, and formal SDG or ESG initiatives are crucial in promoting genuine sustainability practices and enhancing both internal and external reputations. Addressing potential biases and ensuring the effective implementation of sustainability strategies can help firms mitigate the risk of Greenwashing and achieve more authentic and impactful sustainability outcomes.

References

The csrhub ratings methodology, 2008. URL `https://www.csrhub.com/csrhub-esg-ratings-methodology`. Accessed: 05-04-2024.

Finscience, 2022. URL `https://finscience.com`. Accessed: 23-10-2022.

The sustainable development report, 2023. URL `https://dashboards.sdgindex.org/rankings`. Accessed: 15-04-2024.

Priyanka Aggarwal and Aarti Kadyan. Greenwashing: The darker side of csr. *Indian Journal of Applied Research*, 4(3):61–66, 2014.

Ruth V Aguilera, Igor Filatotchev, Howard Gospel, and Gregory Jackson. An organizational approach to comparative corporate governance: Costs, contingencies, and complementarities. *Organization science*, 19(3):475–492, 2008.

RV Aguilera, K Desender, MK Bednar, and JH Lee. Connecting the dots: Bringing external corporate governance into the corporate governance puzzle. *Academy of Management Annals*, 9(1):483–573, 2015.

Bahaaeddin Ahmed Alareeni and Allam Hamdan. Esg impact on performance of us s&p 500-listed firms. *Corporate Governance: The International Journal of Business in Society*, 20(7):1409–1428, 2020.

Joshua D Angrist and Alan B Krueger. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*, 15(4):69–85, 2001.

Joshua D. Angrist and Joern-Steffen Pischke. *Mostly Harmless Econometrics*. Princeton University Press, Princeton, 2009. ISBN 9781400829828. doi: doi:10.1515/9781400829828. URL `https://doi.org/10.1515/9781400829828`.

Emmanuel Anyigbah, Yusheng Kong, Bless Kofi Edziah, Ahotovi Thomas Ahoto, and Wilhelmina Seyome Ahiaku. Board characteristics and corporate sustainability reporting: Evidence from chinese listed companies. *Sustainability*, 15(4):3553, 2023.

Fred D Arditti. Risk and the required return on equity. *The Journal of Finance*, 22(1): 19–36, 1967.

Silvia Ayuso, Miguel A Rodríguez, Roberto García-Castro, and Miguel A Ariño. Maximizing stakeholders' interests: An empirical analysis of the stakeholder approach to corporate governance. *Business & society*, 53(3):414–439, 2014.

Fara Azmat and Ruth Rentschler. Gender and ethnic diversity on boards and corporate responsibility: The case of the arts sector. *Journal of Business Ethics*, 141:317–336, 2017.

Federica Balluchi, Arianna Lazzini, and Riccardo Torelli. Csr and greenwashing: A matter of perception in the search of legitimacy. *Accounting, accountability and society: Trends and perspectives in reporting, management and governance for sustainability*, pages 151–166, 2020.

Rémi Bazillier and Julien Vauday. The greenwashing machine: Is csr more than communication. 2009.

Florian Berg, Julian F Koelbel, and Roberto Rigobon. Aggregate confusion: The divergence of esg ratings. *Review of Finance*, 26(6):1315–1344, 2022.

Matthew Blackwell, Stefano Iacus, Gary King, and Giuseppe Porro. cem: Coarsened exact matching in stata. *The Stata Journal*, 9(4):524–546, 2009.

Costanza Bosone, Paola Cerchiello, and Yana Kostiuk. Csr under attack: Empirical evidence from the eurostoxx600 on stakeholders' perception of greenwashing. *Available at SSRN 4840850*, 2024.

Frances Bowen and J Alberto Aragon-Correa. Greenwashing in corporate environmentalism research and practice: The importance of what we say and do. *Organization & Environment*, 27(2):107–112, 2014.

Miriam A Cherry and Judd F Sneirson. Beyond profit: Rethinking corporate social responsibility and greenwashing after the bp oil disaster. *Tul. L. Rev.*, 85:983, 2010.

William G Cochran. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, pages 295–313, 1968.

Magali A Delmas and Vanessa Cuerel Burbano. The drivers of greenwashing. *California management review*, 54(1):64–87, 2011.

Alexis Diamond and Jasjeet S Sekhon. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3):932–945, 2013.

Xingqiang Du. How the market values greenwashing? evidence from china. *Journal of Business Ethics*, 128:547, 2015.

Codruţa Dura and Mihaela Ghicajanu. Overview of the state of corporate social responsibility within multinational companies. *Annals of the University of PetrosÌ§ani. Economics*, 12(4):85–94, 2012.

Elena Escrig-Olmedo, María Ángeles Fernández-Izquierdo, Idoya Ferrero-Ferrero, Juana María Rivera-Lirio, and María Jesús Muñoz-Torres. Rating the raters: Evaluating how esg rating agencies integrate sustainability principles. *Sustainability*, 11(3): 915, 2019.

Nancy E Furlow. Greenwashing in the new millennium. *The Journal of Applied Business and Economics*, 10(6):22, 2010.

Lucia Gatti, Peter Seele, and Lars Rademacher. Grey zone in–greenwash out. a review of greenwashing research and implications for the voluntary-mandatory transition of csr. *International Journal of Corporate Social Responsibility*, 4(1):1–15, 2019.

Francesca Gennari and Daniela M Salvioni. Csr committees on boards: The impact of the external country level factors. *Journal of management and Governance*, 23(3): 759–785, 2019.

Richard Paul Gregory. Why roe, roa, and tobin's q in regressions aren't good measures of corporate financial performance for esg criteria. *Available at SSRN 3775789*, 2021.

Xing Sam Gu and Paul R Rosenbaum. Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2(4):405–420, 1993.

Ben B Hansen. Full matching in an observational study of coaching for the sat. *Journal of the American Statistical Association*, 99(467):609–618, 2004.

Keisuke Hirano and Guido W Imbens. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164:73–84, 2004.

Guido W Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, 86(1):4–29, 2004.

Ga-Young Jang, Hyoung-Goo Kang, Ju-Yeong Lee, and Kyounghun Bae. Esg scores and the credit market. *Sustainability*, 12(8), 2020.

Xinfeng Jiang and Ahsan Akbar. Does increased representation of female executives improve corporate environmental investment? evidence from china. *Sustainability*, 10 (12):4750, 2018.

J Kanter. Study: For consumers, green is greenwash. *The New York Times*, pages 4–8, 2009.

Kathyayini Kathy Rao, Carol A Tilt, and Laurence H Lester. Corporate governance and environmental reporting: an australian study. *Corporate Governance: The international journal of business in society*, 12(2):143–163, 2012.

Sang Kim and Zhichuan Li. Understanding the impact of esg practices in corporate finance. *Sustainability*, 13(7), 2021.

Hao Liang and Luc Renneboog. Corporate social responsibility and sustainable finance: A review of the literature. *European Corporate Governance Institute–Finance Working Paper*, (701), 2020.

Thomas P Lyon and A Wren Montgomery. The means and end of greenwash. *Organization & Environment*, 28(2):223–249, 2015.

M Mattis. Csr-washing is the new greenwashing. money watch, cbs news, 2008.

Honglei Mu and Youngchan Lee. Greenwashing in corporate social responsibility: A dual-faceted analysis of its impact on employee trust and identification. *Sustainability*, 15(22):15693, 2023.

United Nations. The 17 goals, 2022. URL `https://sdgs.un.org/goals`.

Tze San Ong, Jing Zhou, Boon Heng Teh, and Assunta Di Vaio. Equity ownership concentration's impact on corporate internal control: the moderating effects of female directors and board compensation. *Environment, Development and Sustainability*, pages 1–29, 2023.

Béatrice Parguel, Florence Benoît-Moreau, and Fabrice Larceneux. How sustainability ratings might deter 'greenwashing': A closer look at ethical corporate communication. *Journal of business ethics*, 102:15–28, 2011.

Shawn Pope and Arild Wæraas. Csr-washing is rare: A conceptual framework, literature review, and critique. *Journal of Business Ethics*, 137:173–193, 2016.

Charu G Raheja. Determinants of board size and composition: A theory of corporate boards. *Journal of financial and quantitative analysis*, 40(2):283–306, 2005.

Paul R Rosenbaum. A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):597–610, 1991.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Donald B Rubin. Matching to remove bias in observational studies. *Biometrics*, pages 159–183, 1973.

Donald B Rubin. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2:169–188, 2001.

Jasjeet S Sekhon. Multivariate and propensity score matching software with automated balance optimization: the matching package for r. *Journal of Statistical Software, Forthcoming*, 2008.

Yvette Sterbenk, Sara Champlin, Kasey Windels, and Summer Shelton. Is femvertising the new greenwashing? examining corporate commitment to gender equality. *Journal of Business Ethics*, 177(3):491–505, 2022.

Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1): 1, 2010.

Elizabeth A Stuart and Kerry M Green. Using full matching to estimate causal effects in nonexperimental studies: examining the relationship between adolescent marijuana use and adult outcomes. *Developmental psychology*, 44(2):395, 2008.

Jean Tirole. *The theory of corporate finance*. Princeton university press, 2010.

Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

Xiao Yan and Chengning Yang. The impacts of external sustainability: Institutional investors' sustainable identity, corporate environmental responsibility, and green innovation. *Sustainability*, 16(5):1961, 2024.

6 Appendix

All Tables, Graphs and Figures are derived by the Author

6.0.1 Descriptive Statistic

Table 23: Company Names and Their Sectors

Company Name	**Sector**
Ashmore Group PLC	Financials
Mercedes-Benz Group AG	Consumer Discretionary
William Hill Ltd	Consumer Discretionary
3i Group PLC	Financials
A2A SpA	Utilities
AAK AB (publ)	Consumer Staples
AB Skf	Industrials
Abb Ltd	Industrials
ABN Amro Bank NV	Financials
Acciona SA	Utilities
Accor SA	Consumer Discretionary
ACS Actividades de Construccion y Servicios SA	Industrials
Adecco Group AG	Industrials
Adidas AG	Consumer Discretionary
Admiral Group PLC	Financials
Adyen NV	Information Technology
Aegon NV	Financials
Aena SME SA	Industrials
Aeroports de Paris SA	Industrials
Airbus SE	Industrials
Aker BP ASA	Energy

Company Name	**Sector**
Akzo Nobel NV	Materials
Alcon AG	Health Care
Allianz SE	Financials
Alstom SA	Industrials
Amadeus IT Group SA	Information Technology
Amplifon SpA	Health Care
ams OSRAM AG	Information Technology
Amundi SA	Financials
Andritz AG	Industrials
Anglo American PLC	Materials
Anheuser Busch Inbev SA	Consumer Staples
AP Moeller - Maersk A/S	Industrials
ArcelorMittal SA	Materials
Arkema SA	Materials
ASM International NV	Information Technology
ASML Holding NV	Information Technology
ASR Nederland NV	Financials
Assa Abloy AB	Industrials
Assicurazioni Generali SpA	Financials
Associated British Foods PLC	Consumer Staples
Assura PLC	Real Estate
AstraZeneca PLC	Health Care
Atlas Copco AB	Industrials
Atos SE	Information Technology
Avast PLC	Information Technology
AVEVA Group PLC	Information Technology
Aviva PLC	Financials

Company Name	**Sector**
AXA SA	Financials
B&M European Value Retail SA	Consumer Discretionary
BAE Systems PLC	Industrials
Banco Bilbao Vizcaya Argentaria SA	Financials
Banco de Sabadell SA	Financials
Banco Santander SA	Financials
Bank of Ireland Group PLC	Financials
Bankinter SA	Financials
Barclays PLC	Financials
Barratt Developments P L C	Consumer Discretionary
Barry Callebaut AG	Consumer Staples
Basf Se	Materials
BAWAG Group AG	Financials
Bayer AG	Health Care
Bayerische Motoren Werke AG	Consumer Discretionary
Bechtle AG	Information Technology
Beiersdorf AG	Consumer Staples
Bellway PLC	Consumer Discretionary
Berkeley Group Holdings PLC	Consumer Discretionary
BHP Group Ltd	Materials
Biomerieux SA	Health Care
BNP Paribas SA	Financials
Boliden AB	Materials
Bollore SE	Communication Services
Bouygues SA	Industrials
BP PLC	Energy
Brenntag SE	Industrials

Company Name	**Sector**
British American Tobacco PLC	Consumer Staples
Britvic PLC	Consumer Staples
BT Group PLC	Communication Services
Bunzl plc	Industrials
Burberry Group PLC	Consumer Discretionary
Bureau Veritas SA	Industrials
Caixabank SA	Financials
Capgemini SE	Information Technology
Carl Zeiss Meditec AG	Health Care
Carlsberg A/S	Consumer Staples
Carnival PLC	Consumer Discretionary
Carrefour SA	Consumer Staples
CD Projekt SA	Communication Services
Cellnex Telecom SA	Communication Services
Centrica PLC	Utilities
Chocoladefabriken Lindt & Spruengli AG	Consumer Staples
Chr Hansen Holding A/S	Materials
Clariant AG	Materials
Close Brothers Group PLC	Financials
CNH Industrial NV	Industrials
CNP Assurances SA	Financials
Coca-Cola Co	Consumer Staples
Coloplast A/S	Health Care
Commerzbank AG	Financials
Compagnie de Saint Gobain SA	Industrials
Compagnie Financiere Richemont SA	Consumer Discretionary
Compagnie Generale des Etablissements Michelin SCA	Consumer Discretionary

Company Name	**Sector**
Compass Group PLC	Consumer Discretionary
Continental AG	Consumer Discretionary
ConvaTec Group PLC	Health Care
Corbion NV	Materials
Countryside Partnerships PLC	Consumer Discretionary
Covestro AG	Materials
Credit Agricole SA	Financials
Credit Suisse Group AG	Financials
CRH PLC	Materials
Croda International PLC	Materials
Danone SA	Consumer Staples
Danske Bank A/S	Financials
Dassault Systemes SE	Information Technology
Davide Campari Milano NV	Consumer Staples
DCC PLC	Industrials
Dechra Pharmaceuticals PLC	Health Care
Demant A/S	Health Care
Deutsche Boerse AG	Financials
Deutsche Lufthansa AG	Industrials
Deutsche Post AG	Industrials
Deutsche Telekom AG	Communication Services
Deutsche Wohnen SE	Real Estate
Diageo PLC	Consumer Staples
DiaSorin SpA	Health Care
Direct Line Insurance Group PLC	Financials
Dometic Group AB (publ)	Consumer Discretionary
DS Smith PLC	Materials

Company Name	**Sector**
DSV A/S	Industrials
Dufry AG	Consumer Discretionary
E.ON SE	Utilities
Edenred SE	Information Technology
EDP Energias de Portugal SA	Utilities
EDP Renovaveis SA	Utilities
Eiffage SA	Industrials
Electricite de France SA	Utilities
Electrolux AB	Consumer Discretionary
Elekta AB (publ)	Health Care
Enagas SA	Utilities
Endesa SA	Utilities
Enel SpA	Utilities
Engie SA	Utilities
Eni SpA	Energy
Entain PLC	Consumer Discretionary
Epiroc AB	Industrials
Equinor ASA	Energy
Erste Group Bank AG	Financials
EssilorLuxottica SA	Consumer Discretionary
Essity AB (publ)	Consumer Staples
Etablissementen Franz Colruyt NV	Consumer Staples
Eurazeo SE	Financials
Eurofins Scientific SE	Health Care
Euronext NV	Financials
Evolution AB (publ)	Consumer Discretionary
Evonik Industries AG	Materials

Company Name	**Sector**
EVRAZ plc	Materials
Exor NV	Financials
Experian PLC	Industrials
Faurecia SE	Consumer Discretionary
Ferguson PLC	Industrials
Ferrari NV	Consumer Discretionary
Ferrovial SA	Industrials
Flutter Entertainment PLC	Consumer Discretionary
Fortum Oyj	Utilities
Fresenius Medical Care AG & Co KGaA	Health Care
Fresenius SE & Co KGaA	Health Care
Fuchs Petrolub SE	Materials
G4S Ltd	Financials
Galp Energia SGPS SA	Energy
Games Workshop Group PLC	Consumer Discretionary
GEA Group AG	Industrials
Gecina SA	Real Estate
Genmab A/S	Health Care
Genus PLC	Health Care
Georg Fischer AG	Industrials
Gerresheimer AG	Health Care
Getinge AB	Health Care
Givaudan SA	Materials
Glanbia PLC	Consumer Staples
GlaxoSmithKline PLC	Health Care
Glencore PLC	Materials
Grainger PLC	Real Estate

Company Name	**Sector**
Grifols SA	Health Care
H & M Hennes & Mauritz AB	Consumer Discretionary
Halma PLC	Information Technology
Hannover Rueck SE	Financials
Hargreaves Lansdown PLC	Financials
Hays PLC	Industrials
HeidelbergCement AG	Materials
Heineken Holding NV	Consumer Staples
Hellofresh SE	Consumer Staples
Henkel AG & Co KGaA	Consumer Staples
Hermes International SCA	Consumer Discretionary
Hexagon AB	Information Technology
Hikma Pharmaceuticals PLC	Health Care
Hiscox Ltd	Financials
HomeServe PLC	Industrials
Howden Joinery Group PLC	Industrials
HSBC Holdings PLC	Financials
Huhtamaki Oyj	Materials
Husqvarna AB	Industrials
Iberdrola SA	Utilities
ICA Gruppen AB	Consumer Staples
IG Group Holdings PLC	Financials
Iliad SA	Communication Services
IMCD NV	Industrials
IMI PLC	Industrials
Imperial Brands PLC	Consumer Staples
Inchcape PLC	Consumer Discretionary

Company Name	Sector
Industria de Diseno Textil SA	Consumer Discretionary
Infineon Technologies AG	Information Technology
Informa PLC	Communication Services
ING Groep NV	Financials
Inmobiliaria Colonial SOCIMI SA	Real Estate
InterContinental Hotels Group PLC	Consumer Discretionary
International Consolidated Airlines Group SA	Industrials
Interpump Group SpA	Industrials
Intertek Group PLC	Industrials
Intesa Sanpaolo SpA	Financials
Investor AB	Financials
Ipsen SA	Health Care
Iss A/S	Industrials
ITV PLC	Communication Services
J Sainsbury PLC	Consumer Staples
JD Sports Fashion PLC	Consumer Discretionary
JDE Peets NV	Consumer Staples
Jeronimo Martins SGPS SA	Consumer Staples
Johnson Matthey PLC	Materials
Just Eat Takeaway.com NV	Consumer Discretionary
Kaz Minerals Ltd	Materials
KBC Groep NV	Financials
Kering SA	Consumer Discretionary
Kerry Group PLC	Consumer Staples
Kesko Oyj	Consumer Staples
Kingfisher PLC	Consumer Discretionary
Kingspan Group PLC	Industrials

Company Name	**Sector**
Kinnevik AB	Financials
Klepierre SA	Real Estate
Knorr Bremse AG	Industrials
Kone Oyj	Industrials
Koninklijke Ahold Delhaize NV	Consumer Staples
Koninklijke DSM NV	Materials
Koninklijke Philips NV	Health Care
Koninklijke Vopak NV	Energy
L E Lundbergforetagen AB (publ)	Financials
L'Air Liquide Societe Anonyme pour l'Etude et l'Exploitation des Procedes Georges Claude SA	Materials
L'Oreal SA	Consumer Staples
Land Securities Group PLC	Real Estate
Lanxess AG	Materials
LEG Immobilien SE	Real Estate
Legal & General Group PLC	Financials
Legrand SA	Industrials
Leonardo SpA	Industrials
Linde PLC	Materials
Lloyds Banking Group PLC	Financials
Logitech International SA	Information Technology
London Stock Exchange Group PLC	Financials
Lonza Group AG	Health Care
LVMH Moet Hennessy Louis Vuitton SE	Consumer Discretionary
M&G PLC	Financials
Marks and Spencer Group PLC	Consumer Staples
Mediobanca Banca di Credito Finanziario SpA	Financials

Company Name	**Sector**
Meggitt PLC	Industrials
Melrose Industries PLC	Industrials
Merck KGaA	Health Care
Metso Outotec Corp	Industrials
Moncler SpA	Consumer Discretionary
Mondi PLC	Materials
Mowi ASA	Consumer Staples
MTU Aero Engines AG	Industrials
Muenchener Rueckversicherungs Gesellschaft in Muenchen AG	Financials
National Grid PLC	Utilities
Natixis SA	Financials
Naturgy Energy Group SA	Utilities
Natwest Group PLC	Financials
Nemetschek SE	Information Technology
Neste Oyj	Energy
Nestle SA	Consumer Staples
Next PLC	Consumer Discretionary
NN Group NV	Financials
Nokia Oyj	Information Technology
Nokian Tyres plc	Consumer Discretionary
Nordea Bank Abp	Financials
Norsk Hydro ASA	Materials
Novartis AG	Health Care
Novozymes A/S	Materials
Ocado Group PLC	Consumer Staples
OMV AG	Energy
Orange SA	Communication Services

Company Name	**Sector**
Orion Oyj	Health Care
Orkla ASA	Consumer Staples
Orsted A/S	Utilities
Partners Group Holding AG	Financials
Pearson PLC	Communication Services
Pernod Ricard SA	Consumer Staples
Persimmon PLC	Consumer Discretionary
Phoenix Group Holdings PLC	Financials
Polski Koncern Naftowy Orlen SA	Energy
Poste Italiane SpA	Financials
Prosiebensat 1 Media SE	Communication Services
Prudential PLC	Financials
Prysmian SpA	Industrials
Publicis Groupe SA	Communication Services
Puma SE	Consumer Discretionary
Qiagen NV	Health Care
Raiffeisen Bank International AG	Financials
Randstad NV	Industrials
Rational AG	Industrials
Reckitt Benckiser Group PLC	Consumer Staples
Red Electrica Corporacion SA	Utilities
Relx PLC	Industrials
Remy Cointreau SA	Consumer Staples
Renault SA	Consumer Discretionary
Rentokil Initial PLC	Industrials
Repsol SA	Energy
Rexel SA	Industrials

Company Name	**Sector**
Rheinmetall AG	Industrials
Rightmove PLC	Communication Services
Rio Tinto PLC	Materials
Roche Holding AG	Health Care
Rockwool International A/S	Industrials
Rolls-Royce Holdings PLC	Industrials
Rotork PLC	Industrials
Royal Mail PLC	Industrials
RSA Insurance Group Ltd	Financials
Rwe AG	Utilities
Ryanair Holdings PLC	Industrials
Safran SA	Industrials
Sage Group PLC	Information Technology
Sandvik AB	Industrials
Sanofi SA	Health Care
SAP SE	Information Technology
SBM Offshore NV	Energy
Schibsted ASA	Communication Services
Schindler Holding AG	Industrials
Schneider Electric SE	Industrials
Schroders PLC	Financials
Scor SE	Financials
Securitas AB	Industrials
SEGRO PLC	Real Estate
SES SA	Communication Services
Severn Trent PLC	Utilities
SGS SA	Industrials

Company Name	**Sector**
Shell PLC	Energy
Siemens AG	Industrials
Siemens Energy AG	Industrials
Siemens Gamesa Renewable Energy SA	Industrials
Siemens Healthineers AG	Health Care
SIG Combibloc Group AG	Materials
Signature Aviation Ltd	Industrials
Signify NV	Industrials
Sika AG	Materials
Skandinaviska Enskilda Banken AB	Financials
Skanska AB	Industrials
Smith & Nephew PLC	Health Care
Smiths Group PLC	Industrials
Smurfit Kappa Group PLC	Materials
Snam SpA	Utilities
Societe Generale SA	Financials
Sodexo SA	Consumer Discretionary
Solvay SA	Materials
Sonova Holding AG	Health Care
Spie SA	Industrials
Spirax-Sarco Engineering PLC	Industrials
SSE PLC	Utilities
St James's Place PLC	Financials
Stadler Rail AG	Industrials
Standard Chartered PLC	Financials
Stellantis NV	Consumer Discretionary
STMicroelectronics NV	Information Technology

Company Name	**Sector**
Stora Enso Oyj	Materials
Storebrand ASA	Financials
Straumann Holding AG	Health Care
Suez SA	Utilities
Svenska Handelsbanken AB	Financials
Swatch Group AG	Consumer Discretionary
Swiss Life Holding AG	Financials
Swiss Re AG	Financials
Swisscom AG	Communication Services
Symrise AG	Materials
Tate & Lyle PLC	Consumer Staples
Taylor Wimpey PLC	Consumer Discretionary
TeamViewer AG	Information Technology
Tecan Group AG	Health Care
TechnipFMC PLC	Energy
Tele2 AB	Communication Services
Telecom Italia SpA	Communication Services
Telefonaktiebolaget LM Ericsson	Information Technology
Telefonica SA	Communication Services
Telenor ASA	Communication Services
Teleperformance SE	Industrials
Telia Company AB	Communication Services
Temenos AG	Information Technology
Tenaris SA	Energy
Terna Rete Elettrica Nazionale SpA	Utilities
Tesco PLC	Consumer Staples
Thales SA	Industrials

Company Name	**Sector**
THG PLC	Consumer Discretionary
thyssenkrupp AG	Materials
Tomra Systems ASA	Industrials
TotalEnergies SE	Energy
Trainline PLC	Consumer Discretionary
Travis Perkins PLC	Industrials
Trelleborg AB	Industrials
Tryg A/S	Financials
TUI AG	Consumer Discretionary
Ubisoft Entertainment SA	Communication Services
UBS Group AG	Financials
Ucb SA	Health Care
Umicore SA	Materials
UniCredit SpA	Financials
Unilever PLC	Consumer Staples
Uniper SE	Utilities
United Internet AG	Communication Services
United Utilities Group PLC	Utilities
UPM-Kymmene Oyj	Materials
Valmet Oyj	Industrials
Veolia Environnement SA	Utilities
Verbund AG	Utilities
Victrex PLC	Materials
Vifor Pharma AG	Health Care
Vinci SA	Industrials
Virgin Money UK PLC	Financials
Viscofan SA	Consumer Staples

Company Name	**Sector**
Vivendi SE	Communication Services
Vodafone Group PLC	Communication Services
voestalpine AG	Materials
Volkswagen AG	Consumer Discretionary
Volvo AB	Industrials
Vonovia SE	Real Estate
Wartsila Oyj Abp	Industrials
Wendel SE	Financials
WH Smith PLC	Consumer Discretionary
Whitbread PLC	Consumer Discretionary
Wienerberger AG	Materials
Wizz Air Holdings PLC	Industrials
WM Morrison Supermarkets Ltd	Consumer Staples
Wolters Kluwer NV	Industrials
WPP PLC	Communication Services
Yara International ASA	Materials
Zalando SE	Consumer Discretionary
Zurich Insurance Group AG	Financials
Novo Nordisk A/S	Health Care
Alten SA	Information Technology
Pandora A/S	Consumer Discretionary
Sartorius AG	Health Care
Sartorius Stedim Biotech SA	Health Care
Swedbank AB	Financials
Swedish Match AB	Consumer Staples
Unibail-Rodamco-Westfield SE	Real Estate
Valeo SE	Consumer Discretionary

Company Name	**Sector**
VAT Group AG	Industrials
Vestas Wind Systems A/S	Industrials
Alfa Laval AB	Industrials
Belimo Holding AG	Industrials
British Land Company PLC	Real Estate
Deutsche Bank AG	Financials
EQT AB	Financials
Geberit AG	Industrials
Getlink SE	Industrials
Julius Baer Gruppe AG	Financials
Koninklijke KPN NV	Communication Services
Kuehne und Nagel International AG	Industrials
Lundin Energy AB	Energy
Lafarge SA	Materials
Abrdn PLC	Financials

Type of variable	**Variable**	**Description**	**Source**
Dependent variable	Delta	Difference in absolute Terms between internal and external scores.	Authors' calculations
	Internal score	Measures the company's ESG performance based on corporate self-reported and disclosed data.	FinScience
	External score	Measures the company's ESG 'perceived' performance based on alternative external stakeholder-generated data.	FinScience
Explanatory variables	Revenue per Share ($RPSH$)	Total Revenue for the fiscal year divided by Diluted Weighted Average Shares Outstanding	Refinitiv
	ROA	Return on Assets	Refinitiv
	ROE	Return on Equity	Refinitiv
	Market Capitalization ($Market\_cap$)	Market value of the requested issue share type	Refinitiv
	Credit rating ($Credit$)	Agency-equivalent credit rating	Refinitiv
	CSR Committee (CSR_COMM)	The presence of a CSR committee or team in the company	Refinitiv
	Board Size ($Board\_size$)	The total number of board members at the end of the fiscal year	Refinitiv
	Women on board($female\_board$)	Percentage of females on the board	Refinitiv
	Women employees ($Women\_emp$)	Percentage of women employees	Refinitiv
Instrumental Variable	Sea level trend ($mean\_trend$)	average change of Sea level per country	Author' calculation
	SDR-Rating (SDR_Rating)	Sustainable development report Rating per Country	sdgindex.org
	CSR-Rating (CSR_Rating)	Ratings of Corporate Social Responsibility (CSR) performance for a broad range of countries	CSRHub

Table 21: Description of Variables

No.	GICS Sector	Number of Firms
1	Communication Services & Information Technology	52 (11.1%)
2	Consumer Discretionary	55 (11.8%)
3	Consumer Staples	43 (9.2%)
4	Energy	16 (3.4%)
5	Financials & Real Estate	93 (20.0%)
6	Healthcare	41 (8.8%)
7	Industrials	95 (20.3%)
8	Materials	45 (9.6%)
9	Utilities	27 (5.8%)
Total	**467**	

Table 22: Distribution of companies across sectors. Source: Refinitiv (2022) and authors' elaboration.

6.0.2 Instrumental Variable

Delta Sore: First Stage and Diagnostics

Table 24: First Stage - Delta Score - Mean Sea Level Trend

	Second Stage variable: Delta		
	(A)	(B)	(C)
Constant	2.170***	1.919***	2.195***
	(0.171)	(0.456)	(0.318)
mean_trend	0.171*	0.176*	0.099
	(0.085)	(0.088)	(0.091)
RPSH		0.678	0.570
		(1.332)	(1.192)
ROE		-0.023	0.153
		(0.195)	(0.587)
GICS_cluster2		0.223	
		(0.644)	
GICS_cluster3		1.715	
		(1.130)	
GICS_cluster4		0.783	
		(1.128)	
GICS_cluster5		-0.166	
		(0.537)	
GICS_cluster6		0.960	
		(0.848)	
GICS_cluster7		-0.064	
		(0.541)	
GICS_cluster8		1.094	
		(0.850)	
GICS_cluster9		1.136	
		(1.111)	
Women_emp			0.027
			(0.169)
Board_size			0.738***
			(0.220)
Credit			0.034
			(0.030)
AIC	282.79	291.95	276.39

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

Figure 1: Diagnostic plots for Delta Model A - Sea level

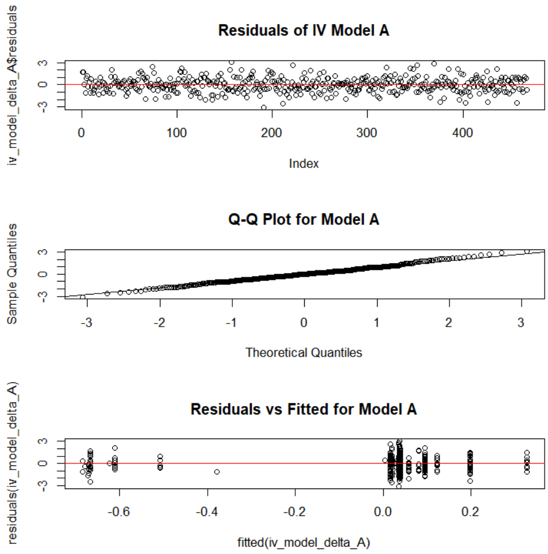


Figure 2: Diagnostic plots for Delta Model B - Sea level

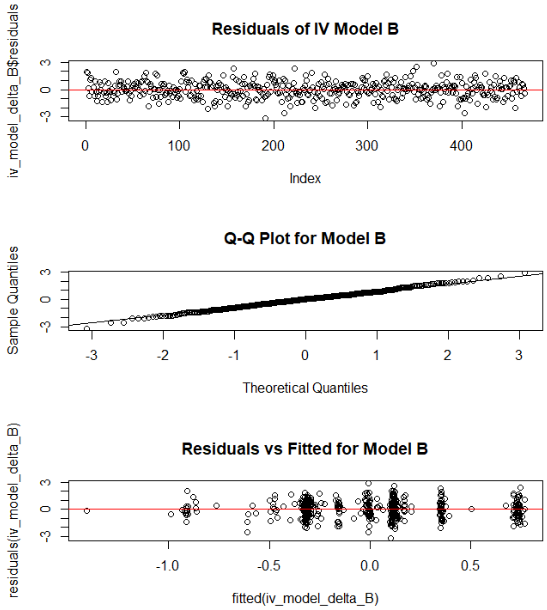


Figure 3: Diagnostic plots for Delta Model C - Sea level

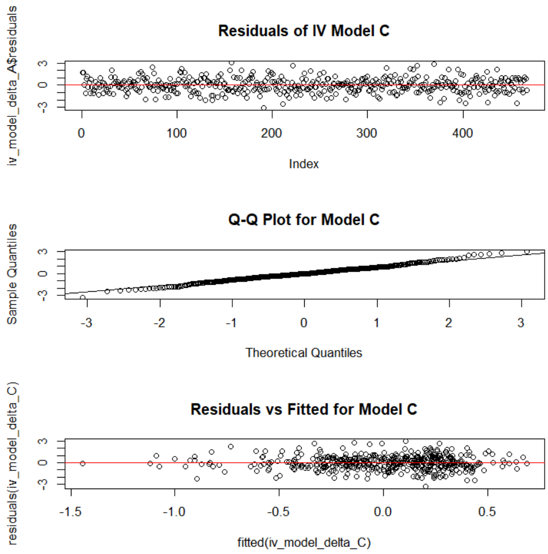


Table 25: First Stage - Delta Score - CSR Ratings

	Second Stage variable: Delta		
	(A)	(B)	(C)
Constant	-0.540	-0.930	1.720
	(5.577)	(5.302)	(5.919)
CSR_Rating	0.053	0.062	0.011
	(0.103)	(0.098)	(0.110)
RPSH		1.006	0.602
		(1.379)	(1.223)
Female_board		0.420*	
		(0.165)	
Women_emp		-0.039	0.053
		(0.167)	(0.168)
ROE			0.169
			(0.668)
Board_size			0.773***
			(0.216)
Credit			0.034
			(0.030)
AIC	286.16	284.99	277.51

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

Figure 4: Diagnostic plots for Delta Model A - CSR-Rating

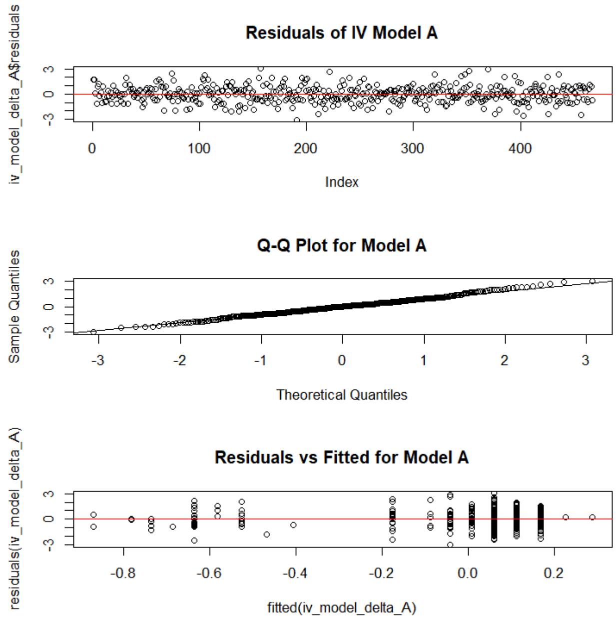


Figure 5: Diagnostic plots for Delta Model B - CSR-Rating

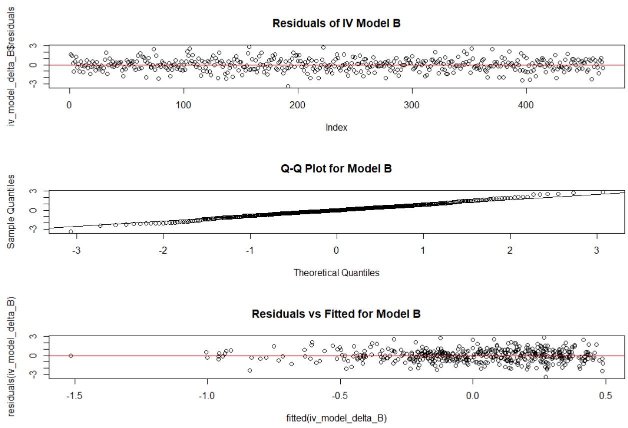


Figure 6: Diagnostic plots for Delta Model C - CSR-Rating

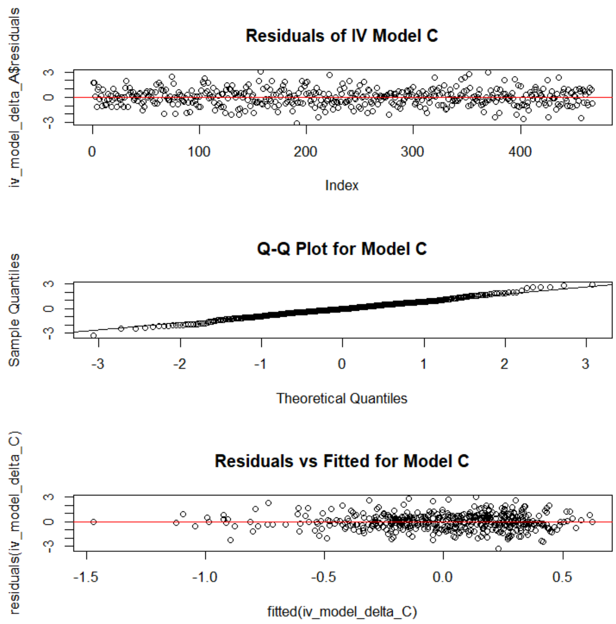


Table 26: First Stage - Delta Score - SDR Rating

	Second Stage variable: Delta		
	(A)	(B)	(C)
Constant	-3.387	-9.408	-3.098
	(6.819)	(7.474)	(6.517)
SDR_Rating	0.070	0.154$^{\cdot}$	0.067
	(0.083)	(0.092)	(0.080)
RPSH		0.215	0.605
		(1.162)	(1.243)
Total_assets		2.974*	
		(1.400)	
ROA		-0.394*	
		(0.182)	
ROE		0.978	0.083
		(0.872)	(0.457)
Women_emp		0.110	0.076
		(0.219)	(0.171)
GICS_cluster2		-0.035	
		(0.704)	
GICS_cluster3		1.500	
		(1.130)	
GICS_cluster4		0.319	
		(1.165)	
GICS_cluster5		-1.254*	
		(0.622)	
GICS_cluster6		0.815	
		(0.888)	
GICS_cluster7		-0.297	
		(0.587)	
GICS_cluster8		0.925	
		(0.882)	
GICS_cluster9		0.858	
		(1.148)	
Board_size			0.790***
			(0.217)
Credit			0.031
			(0.030)
AIC	280.87	271.47	271.99

Note: (standard error); $^{\cdot}$p<0.1; *p<0.05; **p<0.01; ***p<0.001

Figure 7: Diagnostic plots for Delta Model A - SDR-Rating

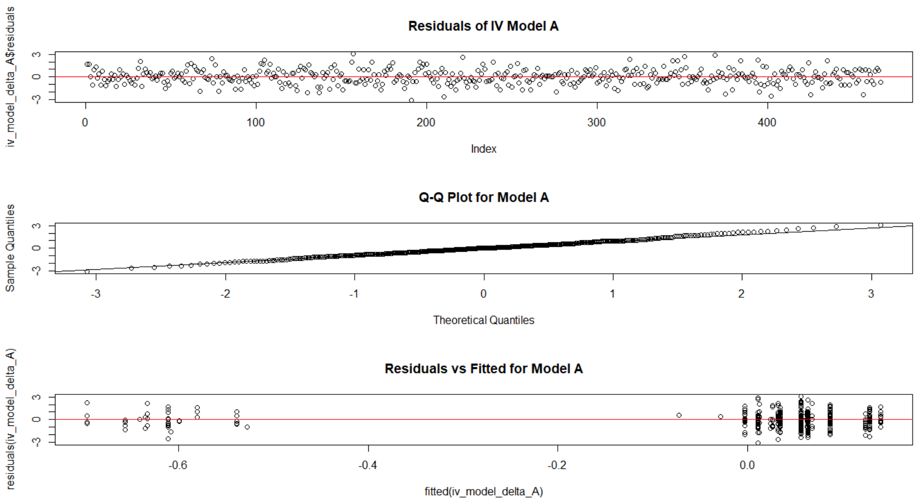


Figure 8: Diagnostic plots for Delta Model B - SDR-Rating

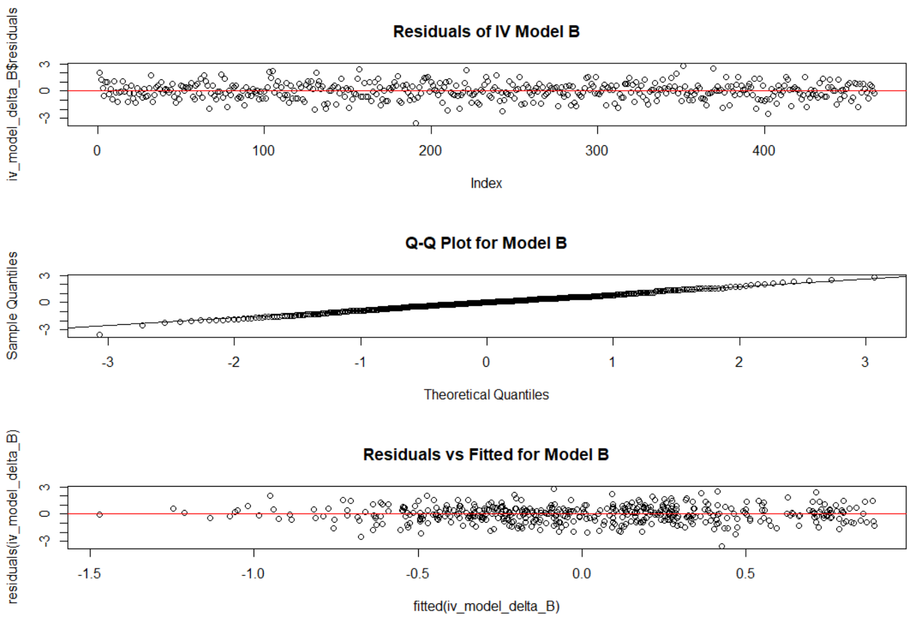
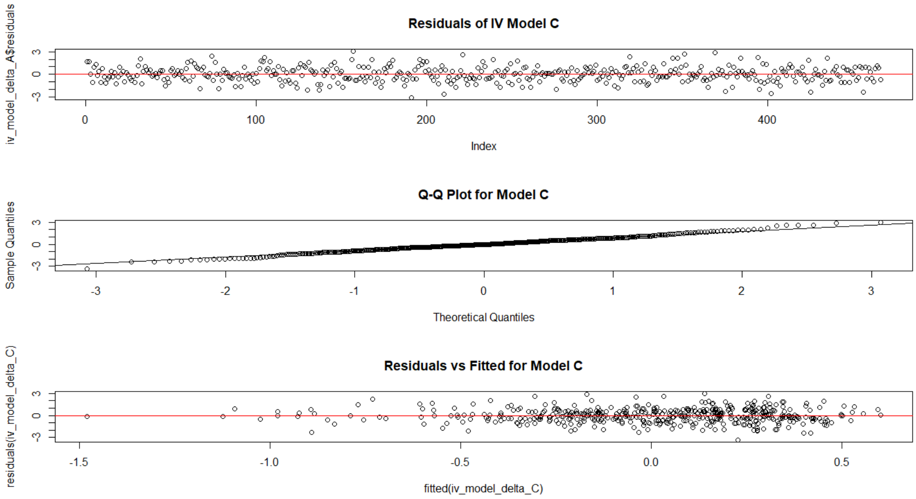


Figure 9: Diagnostic plots for Delta Model C - SDR-Rating



Internal Score: First Stage and Diagnostics

Table 27: First Stage - Internal Score - Mean Sea Level Trend

	Second Stage variable: Internal		
	(A)	(B)	(C)
Constant	2.170***	1.919***	2.195***
	(0.171)	(0.456)	(0.318)
mean_trend	0.171*	0.176*	0.099
	(0.085)	(0.088)	(0.091)
RPSH		0.678	0.570
		(1.332)	(1.192)
ROE		-0.023	0.153
		(0.195)	(0.587)
GICS_cluster2		0.223	
		(0.644)	
GICS_cluster3		1.715	
		(1.130)	
GICS_cluster4		0.783	
		(1.128)	
GICS_cluster5		-0.166	
		(0.537)	
GICS_cluster6		0.960	
		(0.848)	
GICS_cluster7		-0.064	
		(0.541)	
GICS_cluster8		1.094	
		(0.850)	
GICS_cluster9		1.136	
		(1.111)	
Women_emp			0.027
			(0.169)
Board_size			0.738***
			(0.220)
Credit			0.034
			(0.030)
AIC	282.79	291.95	276.39

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

Figure 10: Diagnostic plots for Internal Model A - Sea level

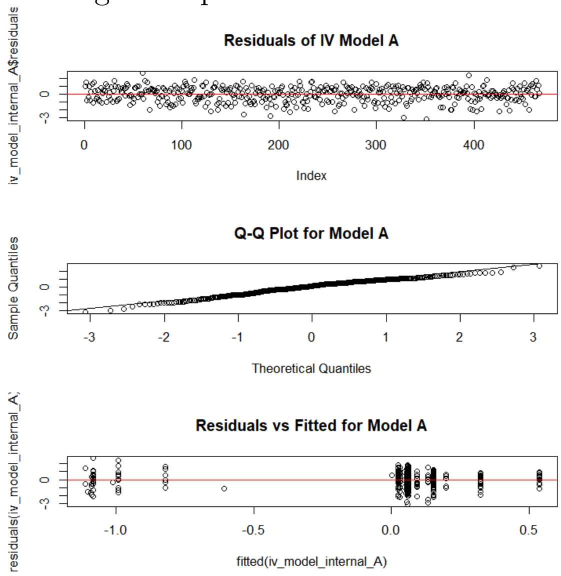


Figure 11: Diagnostic plots for Internal Model B - Sea level

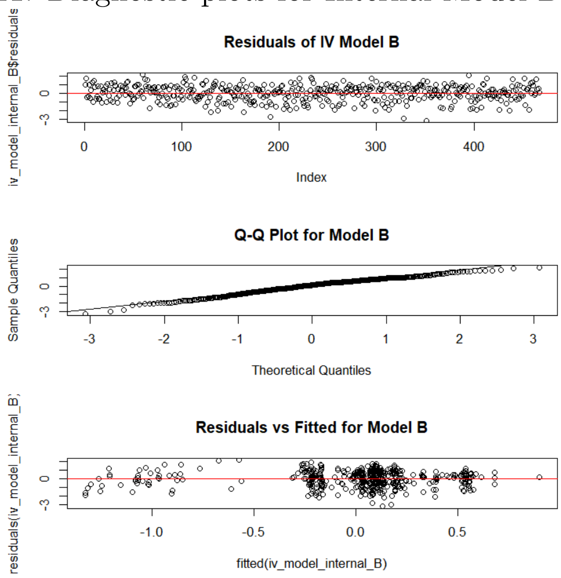


Figure 12: Diagnostic plots for Internal Model C - Sea level

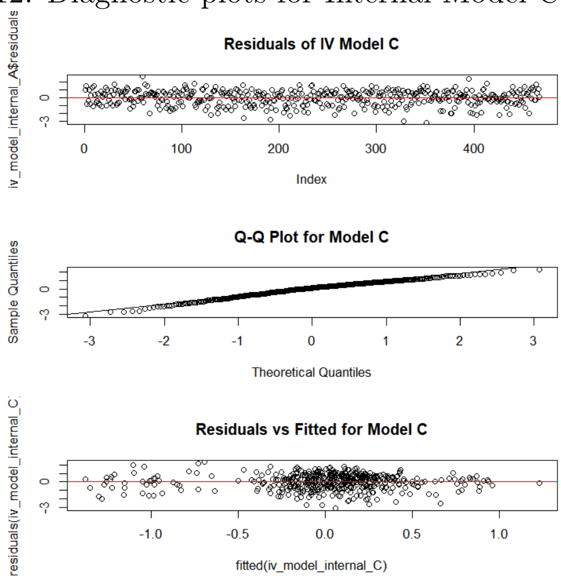


Table 28: First Stage - Internal Score - CSR Rating

	Second Stage variable: Internal		
	(A)	(B)	(C)
Constant	-0.540	-0.930	1.720
	(5.577)	(5.302)	(5.919)
CSR_Rating	0.053	0.062	0.011
	(0.103)	(0.098)	(0.110)
RPSH		1.006	0.602
		(1.379)	(1.223)
Female_board		0.420*	
		(0.165)	
Women_emp		-0.039	0.053
		(0.167)	(0.168)
ROE			0.169
			(0.668)
Board_size			0.773***
			(0.216)
Credit			0.034
			(0.030)
AIC	286.16	284.99	277.51

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

Figure 13: Diagnostic plots for Internal Model A - CSR-Rating

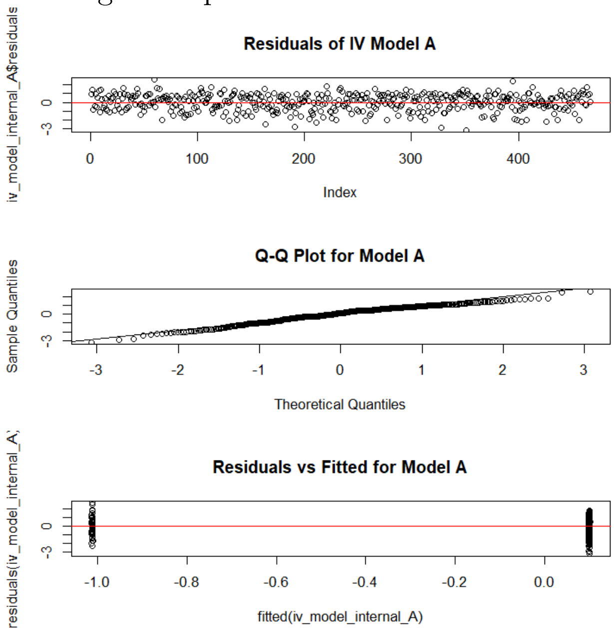


Figure 14: Diagnostic plots for Internal Model B - CSR-Rating

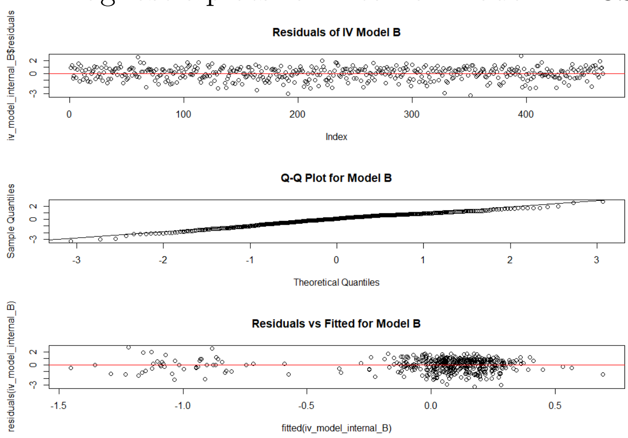


Figure 15: Diagnostic plots for Internal Model C - CSR-Rating

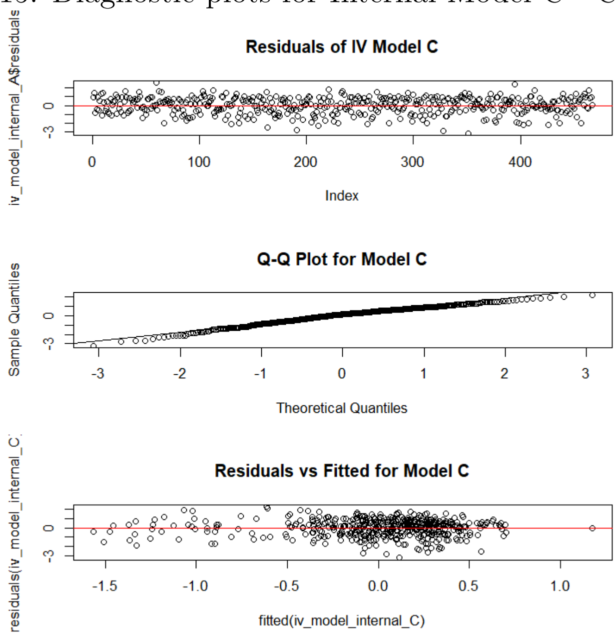


Table 29: First Stage - Internal Score - SDR Rating

	Second Stage variable: Internal		
	(A)	(B)	(C)
Constant	-3.387	-9.408	-3.098
	(6.819)	(7.474)	(6.517)
SDR_Rating	0.070	0.154·	0.067
	(0.083)	(0.092)	(0.080)
RPSH		0.215	0.605
		(1.162)	(1.243)
Total_assets		2.974*	
		(1.400)	
ROA		-0.394*	
		(0.182)	
ROE		0.978	0.083
		(0.872)	(0.457)
Women_emp		0.110	0.076
		(0.219)	(0.171)
GICS_cluster2		-0.035	
		(0.704)	
GICS_cluster3		1.500	
		(1.130)	
GICS_cluster4		0.319	
		(1.165)	
GICS_cluster5		-1.254*	
		(0.622)	
GICS_cluster6		0.815	
		(0.888)	
GICS_cluster7		-0.297	
		(0.587)	
GICS_cluster8		0.925	
		(0.882)	
GICS_cluster9		0.858	
		(1.148)	
Board_size			0.790***
			(0.217)
Credit			0.031
			(0.030)
AIC	280.87	271.47	271.99

Note: (standard error); ·p<0.1; *p<0.05; **p<0.01; ***p<0.001

Figure 16: Diagnostic plots for Internal Model A - SDR-Rating

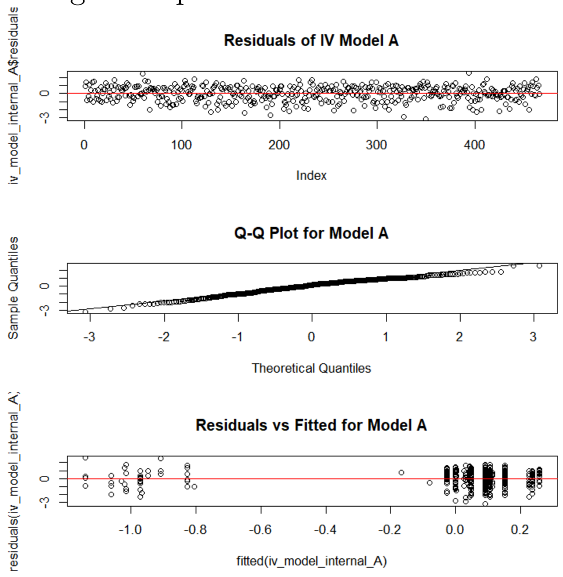


Figure 17: Diagnostic plots for Internal Model B - SDR-Rating

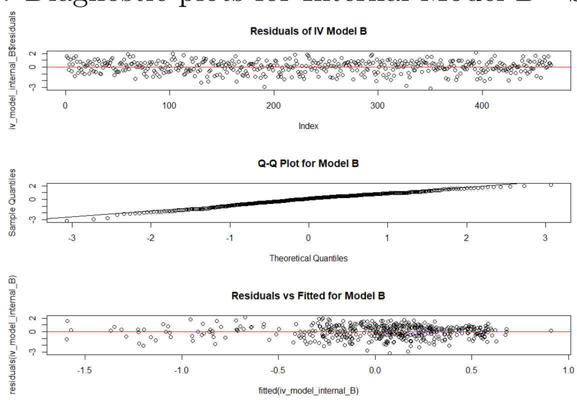
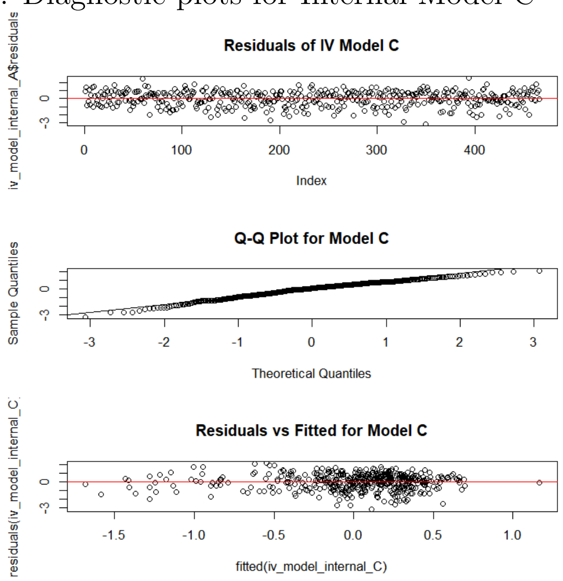


Figure 18: Diagnostic plots for Internal Model C - SDR-Rating



External Score: First Stage and Diagnostics

Table 30: First Stage - External Score - Mean Sea Level Trend

	Second Stage variable: External		
	(A)	(B)	(C)
Constant	2.170***	1.919***	2.195***
	(0.171)	(0.456)	(0.318)
mean_trend	0.171*	0.176*	0.099
	(0.085)	(0.088)	(0.091)
RPSH		0.678	0.570
		(1.332)	(1.192)
ROE		-0.023	0.153
		(0.195)	(0.587)
GICS_cluster2		0.223	
		(0.644)	
GICS_cluster3		1.715	
		(1.130)	
GICS_cluster4		0.783	
		(1.128)	
GICS_cluster5		-0.166	
		(0.537)	
GICS_cluster6		0.960	
		(0.848)	
GICS_cluster7		-0.064	
		(0.541)	
GICS_cluster8		1.094	
		(0.850)	
GICS_cluster9		1.136	
		(1.111)	
Women_emp			0.027
			(0.169)
Board_size			0.738***
			(0.220)
Credit			0.034
			(0.030)
AIC	282.79	291.95	276.39

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

Figure 19: Diagnostic plots for External Model A - Sea level

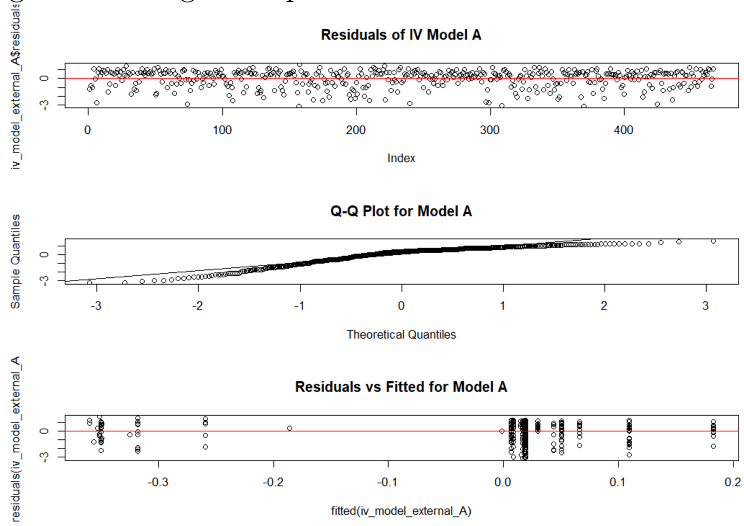


Figure 20: Diagnostic plots for External Model B - Sea level

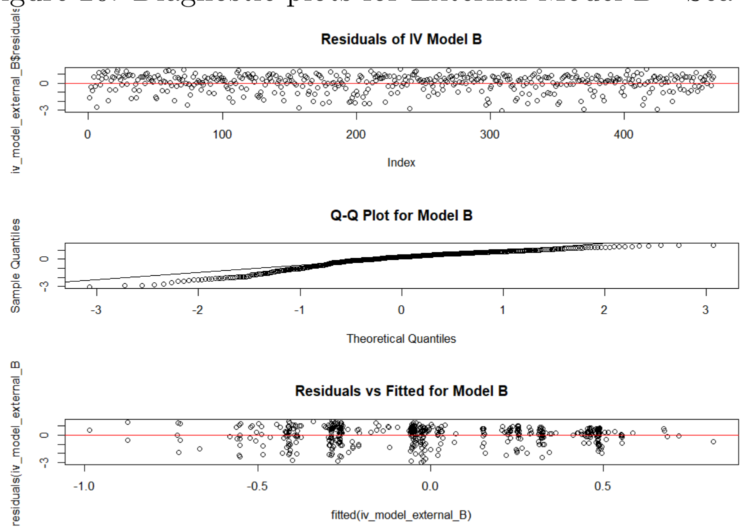


Figure 21: Diagnostic plots for External Model C - Sea level

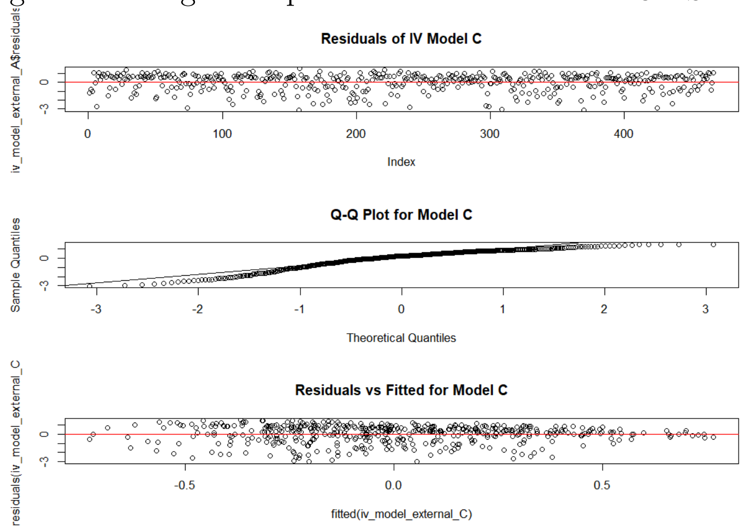


Table 31: First Stage - External Score - CSR Rating

	Second Stage variable: External		
	(A)	(B)	(C)
Constant	-0.540	-0.930	1.720
	(5.577)	(5.302)	(5.919)
CSR_Rating	0.053	0.062	0.011
	(0.103)	(0.098)	(0.110)
RPSH		1.006	0.602
		(1.379)	(1.223)
Female_board		0.420*	
		(0.165)	
Women_emp		-0.039	0.053
		(0.167)	(0.168)
ROE			0.169
			(0.668)
Board_size			0.773***
			(0.216)
Credit			0.034
			(0.030)
AIC	286.16	284.99	277.51

Note: (standard error); *p<0.1; **p<0.05; ***p<0.01

Figure 22: Diagnostic plots for External Model A - CSR-Rating

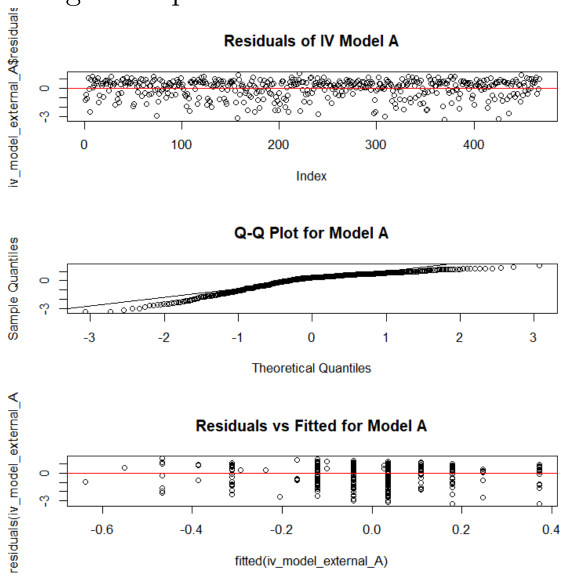


Figure 23: Diagnostic plots for External Model B - CSR-Rating

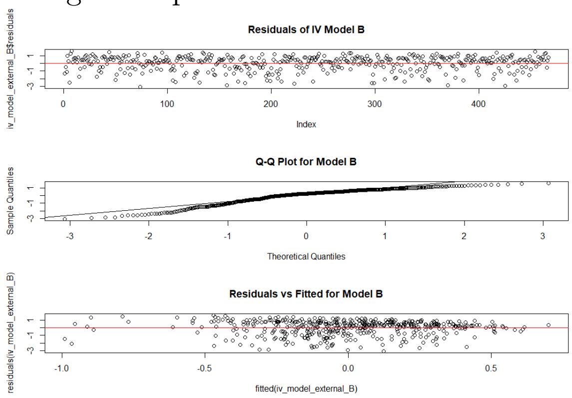


Figure 24: Diagnostic plots for External Model C - CSR-Rating

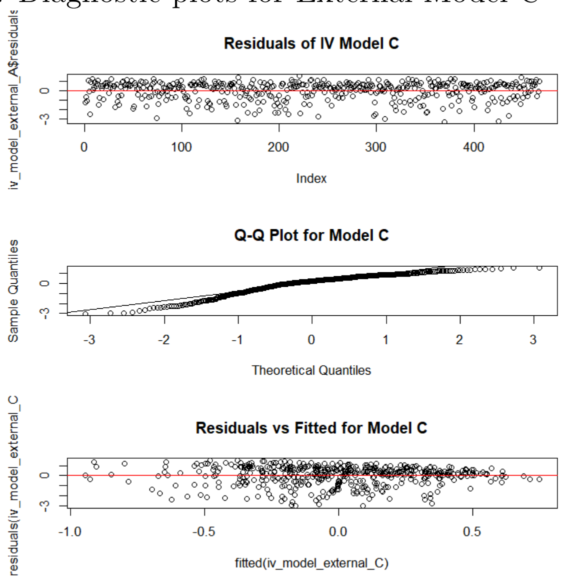


Table 32: First Stage - External Score - SDR Rating

	Second Stage variable: External		
	(A)	(B)	(C)
Constant	-3.387	-9.408	-3.098
	(6.819)	(7.474)	(6.517)
SDR_Rating	0.070	0.154$^{\cdot}$	0.067
	(0.083)	(0.092)	(0.080)
RPSH		0.215	0.605
		(1.162)	(1.243)
Total_assets		2.974*	
		(1.400)	
ROA		-0.394*	
		(0.182)	
ROE		0.978	0.083
		(0.872)	(0.457)
Women_emp		0.110	0.076
		(0.219)	(0.171)
GICS_cluster2		-0.035	
		(0.704)	
GICS_cluster3		1.500	
		(1.130)	
GICS_cluster4		0.319	
		(1.165)	
GICS_cluster5		-1.254*	
		(0.622)	
GICS_cluster6		0.815	
		(0.888)	
GICS_cluster7		-0.297	
		(0.587)	
GICS_cluster8		0.925	
		(0.882)	
GICS_cluster9		0.858	
		(1.148)	
Board_size			0.790***
			(0.217)
Credit			0.031
			(0.030)
AIC	280.87	271.47	271.99

Note: (standard error); $^{\cdot}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Figure 25: Diagnostic plots for External Model A - SDR-Rating

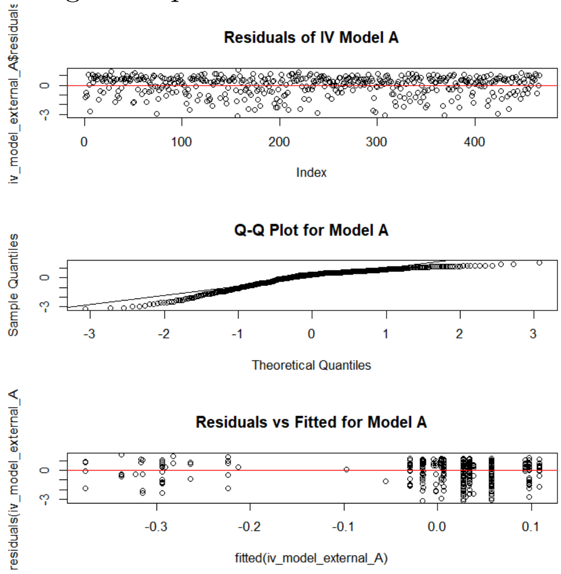


Figure 26: Diagnostic plots for External Model B - SDR-Rating

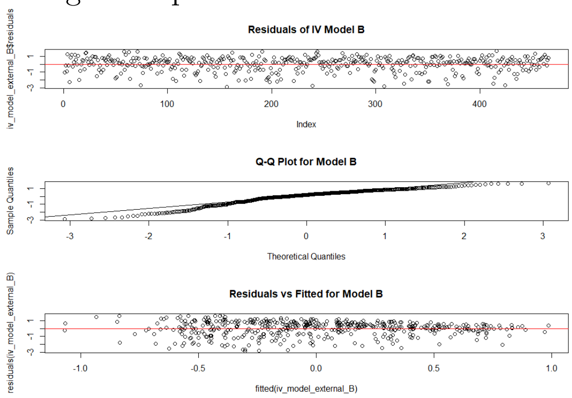
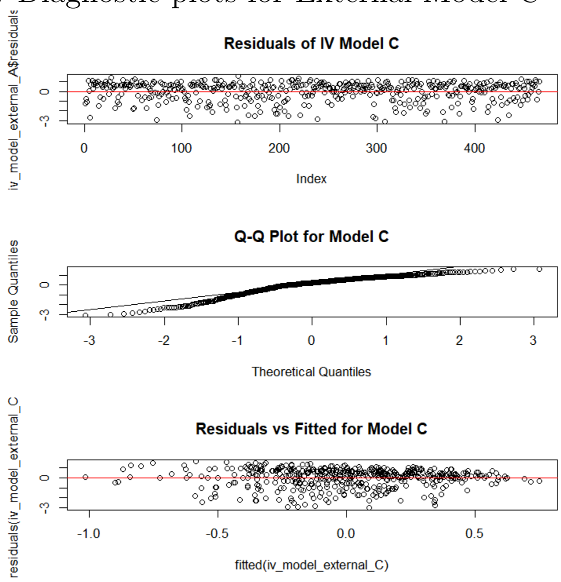


Figure 27: Diagnostic plots for External Model C - SDR-Rating



6.1 PSM

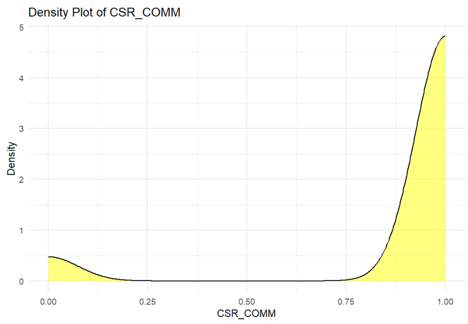


Figure 28: Density CSR-COMM

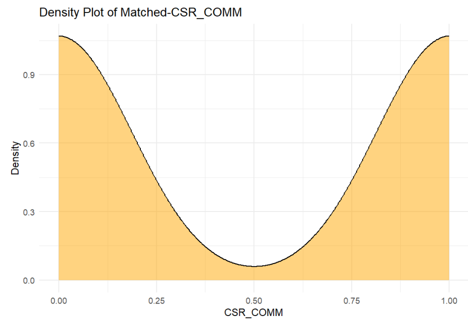


Figure 29: Density Matched CSR-COMM

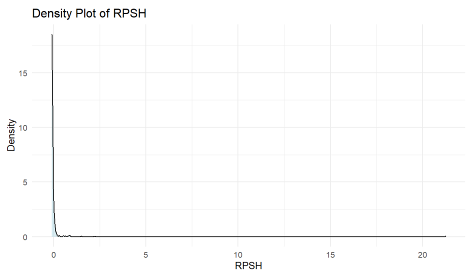


Figure 30: Density RPSH

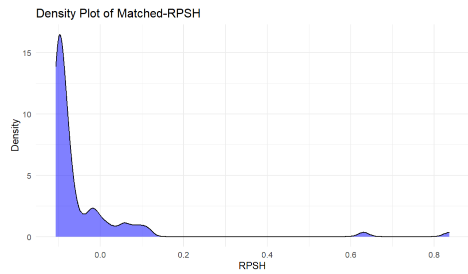


Figure 31: Density Matched RPSH

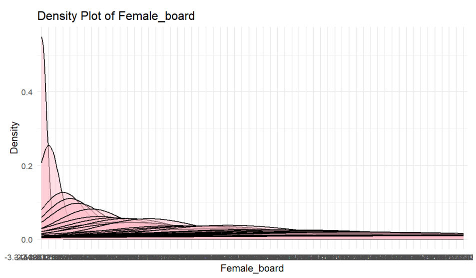


Figure 32: Density Female Board

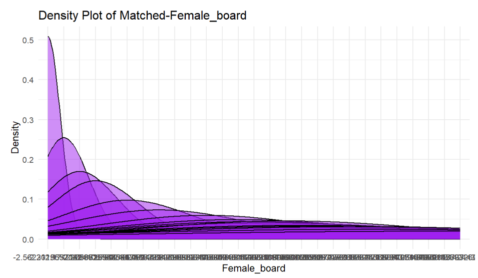


Figure 33: Density Matched Female Board

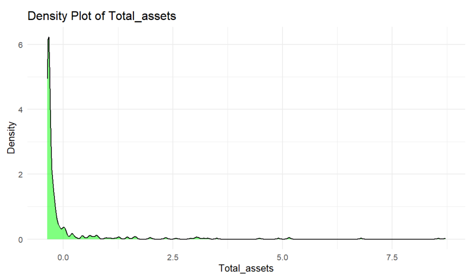


Figure 34: Density Total Assets

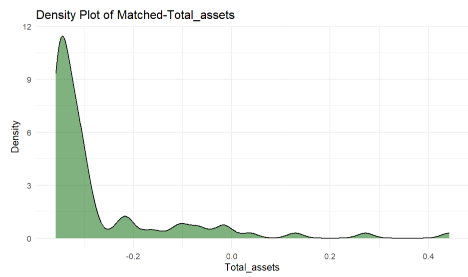


Figure 35: Density Matched Total Assets

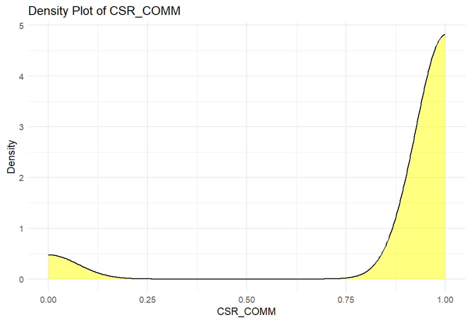


Figure 36: Density CSR-COMM

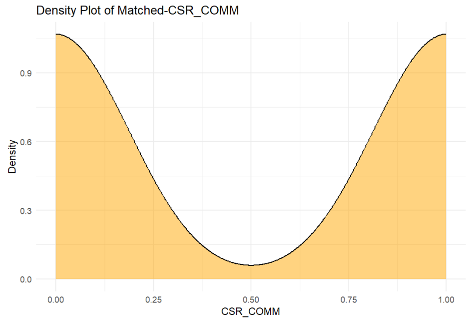


Figure 37: Density Matched CSR-COMM

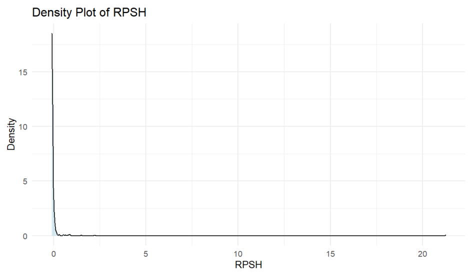


Figure 38: Density RPSH

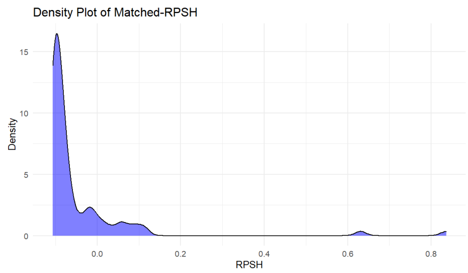


Figure 39: Density Matched RPSH

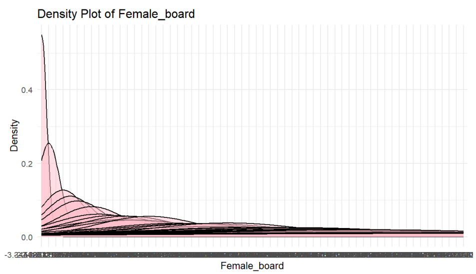


Figure 40: Density Female Board

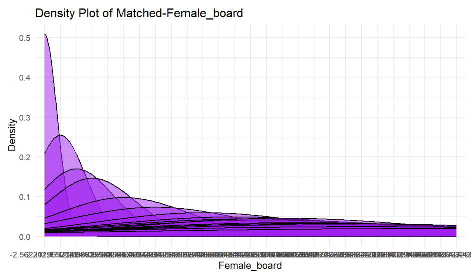


Figure 41: Density Matched Female Board

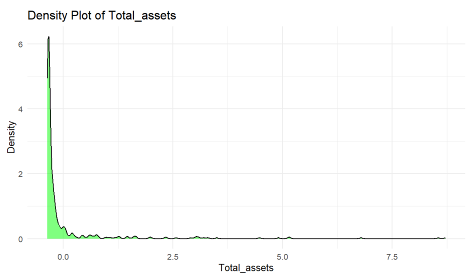


Figure 42: Density Total Assets

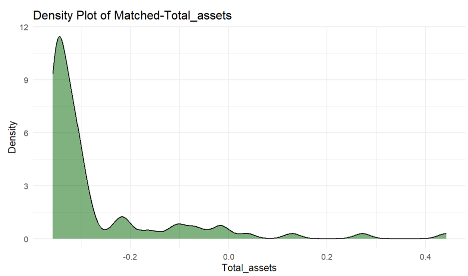


Figure 43: Density Matched Total Assets

7 Code

```
rm(list=ls())

library_packages <- c("quickmatch", "Rglpk", "optmatch", "Matching", "rgenoud",
                      "cobalt", "MatchIt", "data.table", "miceadds", "dplyr",
                      "fixest", "readxl", "gtsummary", "broom", "countrycode",
                      "AER", "car", "purrr", "xtable", "knitr", "kableExtra", "fixest", "ggplot2")
lapply(library_packages, library, character.only = TRUE)

# Data----

ESG_delta_rating <- read_excel("Studium/P_SS24/Thesis/R-Script+Data von Paola/ESG_delta_rating.xlsx")

data<-ESG_delta_rating
data_numerici<-data[,-c(1:2, 11:13)]

#standardize data
clever_scale = function(input_df, exclude_var = c()){
  # input_df is a data.frame
  # if variable is constant or included in 'exclude_var', no transformation

  output_matrix = c()

```

```r
  for (col in colnames(input_df)){
    val = input_df %>% pull(col)
    if (col %in% exclude_var){
      std_val = val
    } else if (uniqueN(val) == 1){
      std_val = scale(val, center = T, scale = F) %>% as.numeric
    }else {
      std_val = scale(val, center = T, scale = T) %>% as.numeric
    }
    output_matrix = output_matrix %>%
      cbind(std_val)
  }
  output_matrix = output_matrix %>%
    as.data.frame() %>%
    setNames(colnames(input_df)) %>%
    `rownames<-`(rownames(input_df)) %>%
    as.matrix()

  return(output_matrix)
}

data_numerici$Credit<-as.factor(data_numerici$Credit)
data_scaled<-clever_scale(data_numerici, exclude_var = c("CSR_COMM", "GICS_cluster", "Credit"))
data_standard<-as.data.frame(data_scaled)
data_standard$GICS_cluster<-as.factor(data_standard$GICS_cluster)
data_standard$Size <- cut(data_standard$Total_assets,
```

```
                              breaks = quantile(data_standard$Total_assets, probs = 0:4/4, na.rm = TRUE),
                              labels = 1:4,
                              include.lowest = TRUE)


data_standard$Country<- data$Country
summary(data_standard)


# standard regression------

fit_delta<- lm(Delta ~ ., data=data_standard)
summary(fit_delta)

#fixed effects

mod_delta <- feols(Delta ~ CSR_COMM + RPSH + ROA + ROE + Board_size + Female_board + Women_emp + Market_cap + Credit,
                      data = data_standard)
mod_delta_s <- feols(Delta ~ CSR_COMM + RPSH + ROA + ROE + Board_size + Female_board + Women_emp + Market_cap + Credit | Size,
                       data = data_standard)
mod_delta_si <- feols(Delta ~ CSR_COMM + RPSH + ROA + ROE + Board_size + Female_board + Women_emp + Market_cap + Credit | Size + GICS_cluster,
                        data = data_standard)

```

```
# Summarizing models with clustering by GICS_cluster
summary(mod_delta, cluster = "GICS_cluster")
summary(mod_delta_s, cluster = "GICS_cluster")
summary(mod_delta_si, cluster = "GICS_cluster")


comparison_table1 <- etable(list(mod_delta, mod_delta_s, mod_delta_si),
                            cluster = "GICS_cluster",
                            dict = c(CSR_COMM = "CSR Committee",
                                     RPSH = "RPSH",
                                     ROA = "ROA",
                                     ROE = "ROE",
                                     Board_size = "Board Size",
                                     Female_board = "Female Board",
                                     Women_emp = "Women Employment",
                                     Market_cap = "Market Cap",
                                     Credit = "Credit Rating"))

# Print the table
print(comparison_table1)

#cluster GICS
cluster <- lm.cluster(Delta ~ ., data = data_standard, cluster = c("GICS_cluster"))
#cluster Country
```

```
clusterc <- lm.cluster(Delta ~ Board_size + Women_emp + `CSR_COMM` + Total_assets+ Female_board+ RPSH+ ROA+ ROE+ GICS_cluster, data = data_standard, cluster = c("Country"))
summary(cluster)
summary(clusterc)

##### PSM -----

delta_data <- data_standard
# Define the treatment variable.
delta_data$treatment <- delta_data$CSR_COMM

# Matching and balancing------

#linear regression on CSR to find the variables that influence the most CSR_comm, we already now the variable which influence the delta variables, now we need to find variables that influence bothe CSR and Delta for a good matching
fit_csr <- glm(CSR_COMM ~ RPSH+ Total_assets+ ROA+ ROE+ Board_size+ Female_board+ Women_emp+ EBIT+ Market_cap+ Credit+ GICS_cluster,family = binomial(link = "logit"), data=data_standard)
summary(fit_csr)
# with clustered standard errors
summary(fit_csr, cluster=c("GICS_cluster"))
#stepwise
stepwise_model <- step(fit_csr, direction = "both")
summary(stepwise_model)

#with clustering at the industry level
```

```
fit_csr1 <- glm.cluster(CSR_COMM ~ RPSH+ Total_assets+ ROA+ ROE
    + Board_size+ Female_board+ Women_emp+ EBIT+ Market_cap+
    Credit,family = binomial(link = "logit"), data=data_standard,
     cluster=c("GICS_cluster"))
summary(fit_csr1)
#with clustering at country level
fit_csr2 <- glm.cluster(CSR_COMM ~ RPSH+ Total_assets+ ROA+ ROE
    + Board_size+ Female_board+ Women_emp+ EBIT+ Market_cap+
    Credit, family = binomial(link = "logit"), data=data_standard
    , cluster=c("Country"))
summary(fit_csr2)

matching_methods <- c("nearest", "optimal", "genetic", "full",
    "subclass", "cem", "cardinality","quick")
match_outputs <- list()

# Loop through each matching method
for (method in matching_methods) {
  # Select Covariates: Choose covariates that influence both
      the treatment and the outcome.
  #These should be variables that could affect a company's
      likelihood of engaging in greenwashing as well as their
      sustainability practices.
  # Perform matching using the current method
  m.out <- matchit(treatment ~ RPSH + Total_assets+ Female_
      board+ Market_cap,
                    data = delta_data, method = method)

  # Store the matchit output in the list
  match_outputs[[method]] <- m.out
```

```
  # Print the balance for the current method
  cat("Balance for method:", method, "\n")
  print(bal.tab(m.out, un = TRUE))
}

# Matching with the 'optimal' algorithm-------

#the best methode was optimal

m.out_delta <- matchit(treatment ~ RPSH + Total_assets+ Female_
    board, data = delta_data, method = "optimal")
m.out_delta

bal.tab(m.out_delta, un = TRUE)
print(bal.tab(m.out_delta, un=TRUE))

# Extract matched data
matched_delta <- match.data(m.out_delta)

#look at distribution before and after matching
#RPSH
par(mfrow=c(2,4))

ggplot(delta_data, aes(x = CSR_COMM)) +
  geom_density(fill = "yellow", alpha = 0.5) +
  labs(title = "Density Plot of CSR_COMM", x = "CSR_COMM", y =
     "Density") +
  theme_minimal()
ggplot(matched_delta, aes(x = CSR_COMM)) +
```

```r
  geom_density(fill = "orange", alpha = 0.5) +
  labs(title = "Density Plot of Matched-CSR_COMM", x = "CSR_COMM", y = "Density") +
  theme_minimal()
#RPSH
par(mfrow=c(2,1))
ggplot(delta_data, aes(x = RPSH)) +
  geom_density(fill = "lightblue", alpha = 0.5) +
  labs(title = "Density Plot of RPSH", x = "RPSH", y = "Density") +
  theme_minimal()
ggplot(matched_delta, aes(x = RPSH)) +
  geom_density(fill = "blue", alpha = 0.5) +
  labs(title = "Density Plot of Matched-RPSH", x = "RPSH", y = "Density") +
  theme_minimal()

#total assets
ggplot(delta_data, aes(x = Total_assets)) +
  geom_density(fill = "green", alpha = 0.5) +
  labs(title = "Density Plot of Total_assets", x = "Total_assets", y = "Density") +
  theme_minimal()
ggplot(matched_delta, aes(x = Total_assets)) +
  geom_density(fill = "darkgreen", alpha = 0.5) +
  labs(title = "Density Plot of Matched-Total_assets", x = "Total_assets", y = "Density") +
  theme_minimal()

# Create density plot for Female_board
```

```r
ggplot(delta_data, aes(x = as.factor(Female_board))) +
  geom_density(fill = "pink", alpha = 0.5) +
  labs(title = "Density Plot of Female_board", x = "Female_board", y = "Density") +
  theme_minimal()
ggplot(matched_delta, aes(x = as.factor(Female_board))) +
  geom_density(fill = "purple", alpha = 0.5) +
  labs(title = "Density Plot of Matched-Female_board", x = "Female_board", y = "Density") +
  theme_minimal()

# Regression with PSM -----

####OLS
fit_delta_psm<- lm(Delta ~ treatment, data=matched_delta)
summary(fit_delta_psm)

#now fit the data again with more variables then just the treatment variabl
fit_delta_psm1<- lm(Delta ~ treatment + Board_size + Women_emp +Female_board+RPSH+ROA+ROE + Total_assets+GICS_cluster, data=matched_delta)
summary(fit_delta_psm1)

##### Fixed Effects models

# Base
mod_delta_psm = feols(Delta ~ CSR_COMM + RPSH + ROA + ROE + Board_size + Female_board + Women_emp+ EBIT+  Market_cap+ Credit, data = matched_delta)
```

```r
summary(mod_delta_psm, cluster=c("subclass"))
# fe on size
mod_delta_psm_s = feols(Delta ~ CSR_COMM + RPSH + ROA + ROE + Board_size + Female_board + Women_emp+ EBIT+  Market_cap+ Credit | Size, data = matched_delta)
summary(mod_delta_psm_s, cluster=c("subclass"))
#fe on industry
mod_delta_psm_si <- feols(Delta ~ CSR_COMM + RPSH + ROA + ROE +  Board_size + Female_board + Women_emp+ EBIT+  Market_cap+ Credit | Size+ GICS_cluster, data = matched_delta)
summary(mod_delta_psm_si, cluster=c("subclass"))

comparison_table_psm <- etable(list(mod_delta_psm, mod_delta_psm_s, mod_delta_psm_si),
                               cluster = "GICS_cluster",
                               dict = c(CSR_COMM = "CSR Committee",
                                        RPSH = "RPSH",
                                        ROA = "ROA",
                                        ROE = "ROE",
                                        Board_size = "Board Size",
                                        Female_board = "Female Board",
                                        Women_emp = "Women Employment",
                                        Market_cap = "Market Cap",
                                        Credit = "Credit Rating"))
```

```

# Print the table
print(comparison_table_psm)

##### Clustering

#cluster subclasses
clusters_psm <- lm.cluster(Delta~ treatment+ RPSH+ Total_assets
    + ROA+ ROE+ Board_size+ Female_board+ Women_emp+ Country+
    EBIT+  Market_cap+ Credit+ GICS_cluster, data = matched_delta
    , cluster = c("subclass"))
#cluster GICS
cluster_psm <- lm.cluster(Delta ~ treatment+ RPSH+ Total_assets
    + ROA+ ROE+ Board_size+ Female_board+ Women_emp+ Country+
    EBIT+  Market_cap+ Credit, data = matched_delta, cluster = c(
    "GICS_cluster"))
#cluster Country
clusterc_psm <- lm.cluster(Delta ~ treatment+ Board_size +
    Women_emp + Total_assets+ Female_board+ RPSH+ ROA+ ROE+GICS_
    cluster, data = matched_delta, cluster = c("Country"))

summary(cluster_psm)
summary(clusterc_psm)
summary(clusters_psm)


#Summary----
comparison_table <- etable(list(mod_delta_psm, mod_delta_psm_s,
     mod_delta_psm_si),
                            cluster = "subclass",
```

```
                          dict = c(CSR_COMM = "CSR Committee",
                                   RPSH = "RPSH",
                                   ROA = "ROA",
                                   ROE = "ROE",
                                   Board_size = "Board Size",
                                   Female_board = "Female Board",
                                   Women_emp = "Women Employment",
                                   Market_cap = "Market Cap",
                                   Rating = "Rating"))


##### Instrumental variable Sea level -----
#Merging data
delta_data <- data_standard

Sea_level_trends <- read.csv("Studium/P_SS24/Thesis/R-Script+Data von Paola/Sea_level_trends.csv",header=TRUE, sep = ';')
#View(Sea_level_trends)
# 1. Convert Country Names to ISO Codes: If a lookup table or library is available, we can convert the Country names in the ESG dataset to their corresponding ISO codes.
delta_data$iso3 <- countrycode(delta_data$Country, "country.name", "iso3c")

mean_sea_level <- Sea_level_trends %>%
  group_by(iso3) %>%
  summarise(mean_trend = mean(Trend, na.rm = TRUE))  # na.rm = TRUE removes NA values from the calculation
```

```
# 2. Merge Datasets: Once both datasets use the same country
    identification system, they can be merged based on this
    common field.
sea_data <- delta_data %>%
  left_join(mean_sea_level, by = "iso3")
#there are three countries with no match CHE: Switzerland, AUT:
     Austria, LUX: Luxembourg, simply because they have no access
     to a sea, set thier mean_sea_level to zero
sea_data <- sea_data %>%
  mutate(mean_trend = ifelse(is.na(mean_trend), 0, mean_trend))
csd<-cov(sea_data$mean_trend, sea_data$Delta)

# First stage: relevance of the instrument -----

######################### the endogenous regressor CSR_COMM is
     binary

# check for multicolinearity by anaylsing the VIF
#If VIF is high (commonly a VIF value greater than 10 is
    considered high), it suggests significant multicollinearity
    between the independent variables.
model <- glm(CSR_COMM ~ RPSH+ Total_assets+ ROA+ ROE+ Board_
    size+ Female_board+ Women_emp+ GICS_cluster+ Market_cap+ EBIT
    + Credit, family = binomial(link = "logit"), data = sea_data)
vif_values <- vif(model)
print(vif_values)


```

```r
#######In the following I tried different IV models to find the one with the best output

#A Basic model
model_A <- glm(CSR_COMM ~ mean_trend,family = binomial, data = sea_data)
summary(model_A)

#B Model based on optimization
first_stage <- glm(CSR_COMM ~ mean_trend+ RPSH+ Total_assets+ ROA+ ROE+ Board_size+ Female_board+ Women_emp+ GICS_cluster+ Credit, family = binomial, data = sea_data)
summary(first_stage)

#Find the combintaion of covaraites with a significant mean_trend parameter and the lowest AIC
covariates <- c( "RPSH","Total_assets", "ROA", "ROE", "Board_size", "Female_board", "Women_emp","GICS_cluster", "Credit" )

#find best suitable covarites
run_regression <- function(vars) {
  formula <- as.formula(paste("CSR_COMM ~ mean_trend +", paste(vars, collapse = " + ")))
  model <- glm(formula,family = "binomial", data = sea_data)
  summ <- summary(model)
  list(AIC= AIC(model),p_value = coef(summ)["mean_trend", "Pr(>|z|)"], vars = vars)
}
results <- list()
for(i in 1:length(covariates)) {
```

```r
  combinations <- combn(covariates, i, simplify = FALSE)
  results[[i]] <- map(combinations, run_regression)
}
flattened_results <- do.call(c, results)
# Convert each element into a tibble
results_df <- bind_rows(lapply(flattened_results, function(x) {
  tibble(
    AIC = x$AIC,
    p_value = x$p_value,
    vars = paste(x$vars, collapse = ", ")
  )
}))
best_model_df <- results_df %>%
  filter(p_value < 0.05)
best_model <- results_df %>%
  filter(p_value < 0.05)%>%
  arrange(desc(AIC))%>%
  slice(1)  # Selects the top row after arranging
# Print the best model details
print(best_model)

model_B <- glm(CSR_COMM ~ mean_trend+ RPSH+ ROE+ GICS_cluster,
   family = "binomial", data = sea_data)
summary(model_B, cluster=c("GICS_cluster"))

#C include variables that make theoretical sense
model_C <- glm(CSR_COMM ~ mean_trend+ RPSH + ROE + Women_emp +
   Board_size + Credit, family = "binomial", data = sea_data)
summary(model_C,cluster=c("Country"))

```

```
sea_data$prob_A <- predict(model_A, type = "response")
sea_data$prob_B <- predict(model_B, type = "response")
sea_data$prob_C <- predict(model_C, type = "response")

sea_data$resid_A <- residuals(model_A, type = "response")
sea_data$resid_B <- residuals(model_B, type = "response")
sea_data$resid_C <- residuals(model_C, type = "response")

# Second Stage and Testing----

#_____________________________________________________

# Model A
iv_model_delta_A <- lm(Delta ~ prob_A + resid_A, data = sea_
    data)

# Model B
iv_model_delta_B <- lm(Delta ~ prob_B + RPSH + ROE + GICS_
    cluster + resid_B, data = sea_data)

# Model C
iv_model_delta_C <- lm(Delta ~ prob_C + RPSH + ROE + Women_emp
    + Board_size + Credit + resid_C, data = sea_data)

#_____________________________________________________


#3.1 Wald test and Breusch-Pagan test
summary(iv_model_delta_A, diagnostics= TRUE)
waldtest(iv_model_delta_A, vcov = NeweyWest(iv_model_delta_A))
```

```
bptest(iv_model_delta_A, ~ fitted(iv_model_delta_A), data = sea
    _data)

summary(iv_model_delta_B, diagnostics= TRUE)
waldtest(iv_model_delta_B, vcov = NeweyWest(iv_model_delta_B))
bptest(iv_model_delta_B, ~ fitted(iv_model_delta_B), data = sea
    _data)

summary(iv_model_delta_C, diagnostics= TRUE)
waldtest(iv_model_delta_C, vcov = NeweyWest(iv_model_delta_C))
bptest(iv_model_delta_C, ~ fitted(iv_model_delta_C), data = sea
    _data)

#3.2 Likelyhoodratio test

lr_test_AB <- lrtest(model_A, model_B)
lr_test_AC <- lrtest(model_A, model_C)
lr_test_BC <- lrtest(model_B, model_C)

#3.3 Calculate Pseudo R-squared for each model
pseudo _r2_A <- with(summary(model_A), 1 - deviance / null.
    deviance)
pseudo_r2_B <- with(summary(model_B), 1 - deviance / null.
    deviance)
pseudo_r2_C <- with(summary(model_C), 1 - deviance / null.
    deviance)

# Display Pseudo R-squared values
pseudo_r2_values <- data.frame(Model = c("Model A", "Model B",
    "Model C"), Pseudo_R2 = c(pseudo_r2_A, pseudo_r2_B, pseudo_r2
```

```
    _C))

# Display results of LR tests and Pseudo R-squared
list(
  LR_test_AB = lr_test_AB,
  LR_test_AC = lr_test_AC,
  LR_test_BC = lr_test_BC,
  Pseudo_R2_Values = pseudo_r2_values
)

#3.4 Check for Residual Normality and Homoscedasticity
par(mfrow=c(3,1))

residuals_plot_A <- plot(iv_model_delta_A$residuals, main = "
    Residuals of IV Model A")
abline(h = 0, col = "red")
# Normality of residuals
qqnorm(residuals(iv_model_delta_A), main = "Q-Q Plot for Model
    A")
qqline(residuals(iv_model_delta_A))
# Homoscedasticity check
plot(fitted(iv_model_delta_A), residuals(iv_model_delta_A),
    main = "Residuals vs Fitted for Model A")
abline(h = 0, col = "red")

residuals_plot_B <- plot(iv_model_delta_B$residuals, main = "
    Residuals of IV Model B")
abline(h = 0, col = "red")
qqnorm(residuals(iv_model_delta_B), main = "Q-Q Plot for Model
    B")
```

```r
qqline(residuals(iv_model_delta_B))
# Homoscedasticity check
plot(fitted(iv_model_delta_B), residuals(iv_model_delta_B), main = "Residuals vs Fitted for Model B")
abline(h = 0, col = "red")

residuals_plot_C <- plot(iv_model_delta_A$residuals, main = "Residuals of IV Model C")
abline(h = 0, col = "red")
qqnorm(residuals(iv_model_delta_C), main = "Q-Q Plot for Model C")
qqline(residuals(iv_model_delta_C))
# Homoscedasticity check
plot(fitted(iv_model_delta_C), residuals(iv_model_delta_C), main = "Residuals vs Fitted for Model C")
abline(h = 0, col = "red")

#3.5 robust standard errors
summary(iv_model_delta_A, vcov = NeweyWest)
summary(iv_model_delta_B, vcov = NeweyWest)
summary(iv_model_delta_C, vcov = NeweyWest)

##### Instrumental variable CSR Rating-----------
# merging data
delta_data <- data_standard
CSRHub_Rating <- read_excel("Studium/P_SS24/Thesis/R-Script+Data von Paola/CSRHub_Rating.xlsx")
CSRHub_Rating <- subset(CSRHub_Rating, select = -Number_of_Companies)

```

```
delta_data <- delta_data %>%
  mutate(Country = ifelse(Country == "Ireland; Republic of", "
      Ireland", Country))
CSR_data <- delta_data %>%
  left_join(CSRHub_Rating, by = "Country")

CSR_data$CSR_Rating <- as.numeric(CSR_data$CSR_Rating)
summary(CSR_data)
summary(CSR_data$CSR_Rating)

ccsrd<-cov(CSR_data$CSR_Rating, CSR_data$Delta)
# First stage: Relevance of the instrument -----

#A Basic model
model_A <- glm(CSR_COMM ~ CSR_Rating,family = "binomial", data
    = CSR_data)
summary(model_A)
exp(coef(model_A))

#B find best suitable covarites
run_regression <- function(vars) {
  formula <- as.formula(paste("CSR_COMM ~ CSR_Rating +", paste(
      vars, collapse = " + ")))
  model <- glm(formula,family = "binomial", data = CSR_data)
  summ <- summary(model)
  list(AIC= AIC(model),p_value = coef(summ)["CSR_Rating", "Pr
      (>|z|)"], vars = vars)
}
results <- list()
for(i in 1:length(covariates)) {
```

```
  combinations <- combn(covariates, i, simplify = FALSE)
  results[[i]] <- map(combinations, run_regression)
}
flattened_results <- do.call(c, results)
# Convert each element into a tibble
results_df <- bind_rows(lapply(flattened_results, function(x) {
  tibble(
    AIC = x$AIC,
    p_value = x$p_value,
    vars = paste(x$vars, collapse = ", ")
  )
}))
best_model_df <- results_df %>%
  arrange(p_value, desc(AIC))
best_model <- results_df %>%
  arrange(p_value,desc(AIC))%>%
  slice(1)  # Selects the top row after arranging
# Print the best model details
print(best_model)

model_B <- glm(CSR_COMM ~ CSR_Rating+ RPSH + Female_board +
    Women_emp, family = "binomial", data = CSR_data)
summary(model_B,cluster=c("Country"))

#C include variables that make theoretical sense
model_C <- glm(CSR_COMM ~ CSR_Rating+ RPSH + ROE + Women_emp +
    Board_size + Credit, family = "binomial", data = CSR_data)
summary(model_C,cluster=c("Country"))
exp(coef(model_C))
```

```
CSR_data$prob_A <- predict(model_A, type = "response")
CSR_data$prob_B <- predict(model_B, type = "response")
CSR_data$prob_C <- predict(model_C, type = "response")

CSR_data$resid_A <- residuals(model_A, type = "response")
CSR_data$resid_B <- residuals(model_B, type = "response")
CSR_data$resid_C <- residuals(model_C, type = "response")

# Second Stage and Testing----
#_____________________________________________________

# Model A
iv_model_delta_A <- lm(Delta ~ prob_A + resid_A, data = CSR_
    data)

# Model B
iv_model_delta_B <- lm(Delta ~ prob_B + RPSH + Female_board +
    Women_emp + resid_B, data = CSR_data)

# Model C
iv_model_delta_C <- lm(Delta ~ prob_C + RPSH + ROE + Women_emp
    + Board_size + Credit + resid_C, data = CSR_data)

#_____________________________________________________


#3.1 Wald Test and Breusch-Pagan Test
summary(iv_model_delta_A, diagnostics= TRUE)
waldtest(iv_model_delta_A, vcov = NeweyWest(iv_model_delta_A))
```

```
bptest(iv_model_delta_A, ~ fitted(iv_model_delta_A), data = CSR_data)

summary(iv_model_delta_B, diagnostics= TRUE)
waldtest(iv_model_delta_B, vcov = NeweyWest(iv_model_delta_B))
bptest(iv_model_delta_B, ~ fitted(iv_model_delta_B), data = CSR_data)

summary(iv_model_delta_C, diagnostics= TRUE)
waldtest(iv_model_delta_C, vcov = NeweyWest(iv_model_delta_C))
bptest(iv_model_delta_C, ~ fitted(iv_model_delta_C), data = CSR_data)

#3.2 Likelyhoodratio test
lr_test_AB <- lrtest(model_A, model_B)
lr_test_AC <- lrtest(model_A, model_C)
lr_test_BC <- lrtest(model_B, model_C)

#3.3 Calculate Pseudo R-squared for each model
pseudo_r2_A <- with(summary(model_A), 1 - deviance / null.deviance)
pseudo_r2_B <- with(summary(model_B), 1 - deviance / null.deviance)
pseudo_r2_C <- with(summary(model_C), 1 - deviance / null.deviance)

pseudo_r2_B
# Display Pseudo R-squared values
pseudo_r2_values <- data.frame(Model = c("Model A", "Model B", "Model C"), Pseudo_R2 = c(pseudo_r2_A, pseudo_r2_B, pseudo_r2
```

```
_C))

# Display results of LR tests and Pseudo R-squared
list(
  LR_test_AB = lr_test_AB,
  LR_test_AC = lr_test_AC,
  LR_test_BC = lr_test_BC,
  Pseudo_R2_Values = pseudo_r2_values
)

#3.4 check residuals
#Check for Residual Normality and Homoscedasticity

par(mfrow=c(3,1))
residuals_plot_A <- plot(iv_model_delta_A$residuals, main = "
   Residuals of IV Model A")
abline(h = 0, col = "red")
# Normality of residuals
qqnorm(residuals(iv_model_delta_A), main = "Q-Q Plot for Model
   A")
qqline(residuals(iv_model_delta_A))
# Homoscedasticity check
plot(fitted(iv_model_delta_A), residuals(iv_model_delta_A),
   main = "Residuals vs Fitted for Model A")
abline(h = 0, col = "red")

residuals_plot_B <- plot(iv_model_delta_B$residuals, main = "
   Residuals of IV Model B")
abline(h = 0, col = "red")
```

```r
qqnorm(residuals(iv_model_delta_B), main = "Q-Q Plot for Model B")
qqline(residuals(iv_model_delta_B))
# Homoscedasticity check
plot(fitted(iv_model_delta_B), residuals(iv_model_delta_B),
    main = "Residuals vs Fitted for Model B")
abline(h = 0, col = "red")

residuals_plot_C <- plot(iv_model_delta_A$residuals, main = "Residuals of IV Model C")
abline(h = 0, col = "red")
qqnorm(residuals(iv_model_delta_C), main = "Q-Q Plot for Model C")
qqline(residuals(iv_model_delta_C))
# Homoscedasticity check
plot(fitted(iv_model_delta_C), residuals(iv_model_delta_C),
    main = "Residuals vs Fitted for Model C")
abline(h = 0, col = "red")

#3.5 robust standard errors
summary(iv_model_delta_A, vcov = NeweyWest)
summary(iv_model_delta_B, vcov = NeweyWest)
summary(iv_model_delta_C, vcov = NeweyWest)

##### Instrumental variable SDR Ranking-----------
# merging data (https://dashboards.sdgindex.org/rankings)
delta_data <- data_standard
SDR_Rating <- read_excel("Studium/P_SS24/Thesis/R-Script+Data von Paola/SDR_Rankings.xlsx")

```

```r
delta_data <- delta_data %>%
  mutate(Country = ifelse(Country == "Ireland; Republic of", "
     Ireland", Country))

combined_SDR_data <- delta_data %>%
  left_join(SDR_Rating, by = "Country")

combined_SDR_data$SDR_Rating <- as.numeric(combined_SDR_data$
   Score)
summary(combined_SDR_data)
#exclude the one missing value
SDR_data <- combined_SDR_data[combined_SDR_data$Country!="
   Bermuda",]
summary(SDR_data$SDR_Rating)
csdrd<-cov(SDR_data$SDR_Rating, SDR_data$Delta)

# First stage: Relevance of the instrument -----

#In the following I tried different IV models to find the one
   with the best output

#A Basic model
model_A <- glm(CSR_COMM ~ SDR_Rating,family = "binomial", data
   = SDR_data)
summary(model_A)
exp(coef(model_A))

#D include covaiates based on algorithm
run_regression <- function(vars) {
```

```
  formula <- as.formula(paste("CSR_COMM ~ SDR_Rating +", paste(
     vars, collapse = " + ")))
  model <- glm(formula,family = "binomial", data = SDR_data)
  summ <- summary(model)
  list(AIC= AIC(model),p_value = coef(summ)["SDR_Rating", "Pr
     (>|z|)"], vars = vars)
}
results <- list()
for(i in 1:length(covariates)) {
  combinations <- combn(covariates, i, simplify = FALSE)
  results[[i]] <- map(combinations, run_regression)
}
flattened_results <- do.call(c, results)
# Convert each element into a tibble
results_df <- bind_rows(lapply(flattened_results, function(x) {
  tibble(
    AIC = x$AIC,
    p_value = x$p_value,
    vars = paste(x$vars, collapse = ", ")
  )
}))
best_model_df <- results_df %>%
  arrange(p_value, desc(AIC))
best_model <- results_df %>%
  arrange(p_value, desc(AIC))%>%
  slice(1)  # Selects the top row after arranging
# Print the best model details
print(best_model)

```

```
model_B <- glm(CSR_COMM ~ SDR_Rating+ RPSH+ Total_assets+ ROA+
    ROE+ Women_emp+ GICS_cluster, family = "binomial", data = SDR
    _data)
summary(model_B,cluster=c("Country"))
exp(coef(model_B))


#C include variables that make theoretical sense
model_C <- glm(CSR_COMM ~ SDR_Rating+ RPSH + ROE + Women_emp +
    Board_size + Credit, family = "binomial", data = SDR_data)
summary(model_C,cluster=c("Country"))
exp(coef(model_C))

SDR_data$prob_A <- predict(model_A, type = "response")
SDR_data$prob_B <- predict(model_B, type = "response")
SDR_data$prob_C <- predict(model_C, type = "response")

SDR_data$resid_A <- residuals(model_A, type = "response")
SDR_data$resid_B <- residuals(model_B, type = "response")
SDR_data$resid_C <- residuals(model_C, type = "response")

# Second Stage and Testing----
#______________________________________________________

# Model A
iv_model_delta_A <- lm(Delta ~ prob_A + resid_A, data = SDR_
    data)

# Model B
```

```r
iv_model_delta_B <- lm(Delta ~ prob_B + RPSH + Total_assets+ ROA+ ROE+ Women_emp+ GICS_cluster + resid_B, data = SDR_data)

# Model C
iv_model_delta_C <- lm(Delta ~ prob_C + RPSH + ROE + Women_emp + Board_size + Credit + resid_C, data = SDR_data)

#_____________________________________________________

#3.1 Wald test and Breusch-Pagan test
summary(iv_model_delta_A, diagnostics= TRUE)
waldtest(iv_model_delta_A, vcov = NeweyWest(iv_model_delta_A))
bptest(iv_model_delta_A, ~ fitted(iv_model_delta_A), data = SDR_data)

summary(iv_model_delta_B, diagnostics= TRUE)
waldtest(iv_model_delta_B, vcov = NeweyWest(iv_model_delta_B))
bptest(iv_model_delta_B, ~ fitted(iv_model_delta_B), data = SDR_data)

summary(iv_model_delta_C, diagnostics= TRUE)
waldtest(iv_model_delta_C, vcov = NeweyWest(iv_model_delta_C))
bptest(iv_model_delta_C, ~ fitted(iv_model_delta_C), data = SDR_data)

#3.2 Likelyhoodratio test

lr_test_AB <- lrtest(model_A, model_B)
lr_test_AC <- lrtest(model_A, model_C)
lr_test_BC <- lrtest(model_B, model_C)
```

```
#3.3 Calculate Pseudo R-squared for each model
pseudo_r2_A <- with(summary(model_A), 1 - deviance / null.
    deviance)
pseudo_r2_B <- with(summary(model_B), 1 - deviance / null.
    deviance)
pseudo_r2_C <- with(summary(model_C), 1 - deviance / null.
    deviance)

pseudo_r2_B
# Display Pseudo R-squared values
pseudo_r2_values <- data.frame(Model = c("Model A", "Model B",
    "Model C"), Pseudo_R2 = c(pseudo_r2_A, pseudo_r2_B, pseudo_r2
    _C))

# Display results of LR tests and Pseudo R-squared
list(
  LR_test_AB = lr_test_AB,
  LR_test_AC = lr_test_AC,
  LR_test_BC = lr_test_BC,
  Pseudo_R2_Values = pseudo_r2_values
)

#3.4 check residuals
#Check for Residual Normality and Homoscedasticity

par(mfrow=c(3,1))

residuals_plot_A <- plot(iv_model_delta_A$residuals, main = "
    Residuals of IV Model A")
```

```
abline(h = 0, col = "red")
# Normality of residuals
qqnorm(residuals(iv_model_delta_A), main = "Q-Q Plot for Model
    A")
qqline(residuals(iv_model_delta_A))
# Homoscedasticity check
plot(fitted(iv_model_delta_A), residuals(iv_model_delta_A),
    main = "Residuals vs Fitted for Model A")
abline(h = 0, col = "red")

residuals_plot_B <- plot(iv_model_delta_B$residuals, main = "
    Residuals of IV Model B")
abline(h = 0, col = "red")
qqnorm(residuals(iv_model_delta_B), main = "Q-Q Plot for Model
    B")
qqline(residuals(iv_model_delta_B))
# Homoscedasticity check
plot(fitted(iv_model_delta_B), residuals(iv_model_delta_B),
    main = "Residuals vs Fitted for Model B")
abline(h = 0, col = "red")

residuals_plot_C <- plot(iv_model_delta_A$residuals, main = "
    Residuals of IV Model C")
abline(h = 0, col = "red")
qqnorm(residuals(iv_model_delta_C), main = "Q-Q Plot for Model
    C")
qqline(residuals(iv_model_delta_C))
# Homoscedasticity check
plot(fitted(iv_model_delta_C), residuals(iv_model_delta_C),
    main = "Residuals vs Fitted for Model C")
```

```
abline(h = 0, col = "red")

#3.5 robust standard errors
summary(iv_model_delta_A, vcov = NeweyWest)
summary(iv_model_delta_B, vcov = NeweyWest)
summary(iv_model_delta_C, vcov = NeweyWest)
```