TESIS DE LA UNIVERSIDAD
DE ZARAGOZA

2024

482

Pablo Vilas Naval

Analysis of sustainability in financial markets from a quantitative approach

Director/es
Serrano Cinca, Carlos

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

ANALYSIS OF SUSTAINABILITY IN FINANCIAL MARKETS FROM A QUANTITATIVE APPROACH

Autor

Pablo Vilas Naval

Director/es

Serrano Cinca, Carlos

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Contabilidad y Finanzas

2022

PhD Dissertation

ANALYSIS OF SUSTAINABILITY IN FINANCIAL MARKETS FROM A QUANTITATIVE APPROACH

Pablo Vilas Naval

Supervisors:

Dr. Carlos Serrano Cinca
Dr. José Luis Sarto Marzal

-September 2022-

PhD Program in Accounting and Finance

Agradecimientos

La presente tesis doctoral no hubiera sido posible sin la comunidad de código abierto, y sin todos los profesionales que comparten altruistamente su conocimiento y mantienen gratuitamente paquetes como *pandas, statsmodels, o scikit-learn*. Tampoco hubiera sido posible sin la supervisión constante de Laura Andreu, la ayuda incondicional de José Luis Sarto y la motivación y autoconfianza que me brindó Carlos Serrano. Pero por encima de todo, está tesis no hubiera sido posible sin mi madre, que me orientó a comenzar esta etapa doctoral. Esta etapa en el Departamento de Finanzas y Contabilidad de la Universidad de Zaragoza me ha permitido, en poco tiempo, adquirir la base tecnológica necesaria para enfrentarme al futuro que nos aguarda.

Gracias a todo el mundo, familia, amigos y compañeros, que de un modo u otro me han apoyado durante estos tres años, consigno por último una mención de cariño especial a los compañeros de la sala de *personal investigador predoctoral en formación*.

Contents

List of Tables

List of Figures

Introduction

This doctoral thesis is motivated by the growing importance of nonfinancial factors in capital markets. Currently, in the United States, one dollar out of three under professional management follows some kind of sustainable investment strategy, up from one dollar out of eight in 2010 (USIF, 2010; USIF, 2020). In addition, half of the 28 trillion dollars of investments necessary to achieve the zero net carbon emissions by 2050 are not profitable (EUROSIF, 2021). In this context, capital markets can play a crucial role in financing the ecological transition. This doctoral thesis aims to contribute to this transition by analyzing the suitability of sustainability equity indices in guiding investment flows that follow environmental, social, and governance (ESG) criteria; the adequacy of ESG ratings to identify companies that are committed to environmental and social issues; the factors that push companies to improve their environmental and social indicators; and to what extent investors value the ESG ratings of mutual funds when making their investment decisions.

The first chapter of the doctoral thesis examines how the FTSE4Good sustainability equity indices apply the ESG criteria when listing or delisting companies from their benchmarks. The literature has focused on analyzing the financial performance of these indices (Cunha et al., 2020) or on studying whether financial markets reward the listing or delisting of companies in these indices (Hawn et al., 2018). However, little is known about how these indices apply the ESG criteria to their listing and delisting processes compared to conventional ones. In addition, one of the biggest concerns raised by sustainability products is the shadow of greenwashing (Berrone et al., 2017; Lyon & Montgomery, 2015). In other words, index providers can use the sustainability label as a marketing tool, and their sustainability index does not follow different criteria than those used by their conventional counterparts. The first chapter addresses this issue by using probit models and a cluster analysis. The results show that the most important factor followed by sustainability indices to include or exclude companies is market capitalization instead of ESG performance. Furthermore, the cluster analysis shows that the inclusion and exclusion criteria applied by some sustainability indices do not differ from the criteria applied by conventional indices—specifically in the exclusion process. This chapter posits that index providers should reduce the importance of the company size in their inclusion and exclusion processes to achieve a stronger differentiation from conventional indices.

Recent studies criticize the excessive influence of the company size on the achievement of a high ESG rating (Drempetic et al., 2020). Moreover, companies from developed countries, specifically European ones, usually get the best ESG scores (Demirbag et al., 2017; Liang & Renneboog, 2017). The second chapter analyzes all the companies with an ESG rating on the Refinitiv database to develop a methodology to obtain inclusive ESG ratings. These inclusive ratings would avoid that small companies or some regions may be excluded from the socially responsible investment flows. With these inclusive ratings, regulators can assess companies by their capability to fulfill environmental and social standards. Specifically, we propose a cross-sectional regression that captures the virtuous behavior of the company as its ESG excess relative to its size, country, and industry. This chapter contributes to the literature on the relation between corporate social responsibility (CSR) and financial performance from an inclusive perspective. The results show that companies that meet higher environmental and social standards than their counterparts have worse financial performance. The method we propose is useful for managers, investors, regulators, and researchers because it provides a comparable indicator among companies that captures their virtuous behavior.

The third chapter of the thesis examines the role of public attention toward companies on CSR. The legitimacy theory framework views CSR as a tool for increasing and securing company legitimacy (Baldini et al., 2018; Cormier & Magnan, 2015; Hörisch et al., 2015). Companies react to greater public attention, such as greater exposure to public scrutiny, by increasing their commitment to CSR. However, the empirical studies that address this relation are scarce and the proxies for measuring attention are not always suitable. This chapter fills the gap by proposing a new method that uses Google Trends to obtain a yearly ranking of public attention to the companies of the S&P 500. This ranking places companies according to the number of web searches. First, our findings show a positive relationship between public attention and CSR performance. Second, following a quasi-experimental approach using matching procedures, we prove a causal relation from public attention to CSR. Specifically, we show that companies improve their CSR performance after a "shock" from public attention. Therefore, we conclude that public scrutiny and signaling would be an effective strategy for pressuring companies to improve their environmental and social records. Beyond conventional approaches (Choi et al., 2020; Da et al., 2011), our new method of using Google Trends is useful for researchers and practitioners.

The last chapter of this thesis addresses the influence of nonfinancial factors on the investment decisions of mutual fund investors. Studies based on surveys and stated choice experiments show that investors are willing to forgo financial performance in order to invest in sustainable products (Gutsche & Ziegler, 2019; Riedl & Smeets, 2017). Similarly, investors allocate more money to high-rated funds than to low-rated funds (Ammann et al., 2019; Hartzmark & Sussman, 2019). Studies have shown a positive and significant influence of sustainability factors on investment decisions. However, the contributions of these factors to explaining investment decisions and the influence of nonfinancial factors on fund flows compared to the influence of financial ones has remained unexplored. We address this gap by means of regressions and machine learning techniques (neural networks, random forests, and gradient boosting decision trees). Our models predict fund flows with an accuracy of about 70%. The dominance analysis and permutation feature importance show that nonfinancial factors have a limited contribution to the model's goodness of fit. Similarly, the analysis of standardized regression coefficients and Shapley additive explanations show that the nonfinancial factors have a limited effect on the fund flows. Thus, in this chapter we conclude that investors consider ESG performance, but the factors that matter most are past growth, management fees, and past returns.

This doctoral thesis comprises four chapters and a conclusions section. The first chapter entitled *Cluster analysis to validate the sustainability label of stock indices: An analysis of the inclusion and exclusion processes in terms of size and ESG ratings* analyzes sustainability equity indices. The second chapter entitled *In search of inclusive ESG ratings* studies the link between CSR and financial performance from an inclusive perspective. The third chapter entitled *The Influence of public attention on corporate social responsibility* examines the role of visibility and public scrutiny towards the company on its CSR engagement. The fourth chapter entitled *The limited role of sustainability in mutual fund investor decisions: A machine learning approach* explores the importance of non-financial variables on the investment decisions. Finally, the last section shows the conclusions of this thesis.

References

Ammann, M., Bauer, C., Fischer, S., & Müller, P. (2019). The impact of the Morningstar Sustainability Rating on mutual fund flows. European Financial Management, 25(3), 520–553. https://doi.org/https://doi.org/10.1111/eufm.12181

Baldini, M., Maso, L. D., Liberatore, G., Mazzi, F., & Terzani, S. (2018). Role of Country- and Firm-Level Determinants in Environmental, Social, and Governance Disclosure. Journal of Business Ethics, 150(1), 79–98. https://doi.org/10.1007/s10551-016-3139-1

Berrone, P., Fosfuri, A., & Gelabert, L. (2017). Does greenwashing pay off? Understanding the relationship between environmental actions and environmental legitimacy. Journal of Business Ethics, 144(2), 363–379. https://doi.org/10.1007/s10551-015-2816-9

Choi, D., Gao, Z., & Jiang, W. (2020). Attention to Global Warming. The Review of Financial Studies, 33(3), 1112–1145. https://doi.org/10.1093/rfs/hhz086

Cormier, D., & Magnan, M. (2015). The Economic Relevance of Environmental Disclosure and its Impact on Corporate Legitimacy: An Empirical Investigation. Business Strategy and the Environment, 24(6), 431–450. https://doi.org/https://doi.org/10.1002/bse.1829

Cunha, F. A. F. de S., de Oliveira, E. M., Orsato, R. J., Klotzle, M. C., Cyrino Oliveira, F. L., & Caiado, R. G. G. (2020). Can sustainable investments outperform traditional benchmarks? Evidence from global stock markets. Business Strategy and the Environment, 29(2), 682–697. https://doi.org/https://doi.org/10.1002/bse.2397

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. The Journal of Finance, 66(5), 1461–1499. https://doi.org/https://doi.org/10.1111/j.1540-6261.2011.01679.x

Demirbag, M., Wood, G., Makhmadshoev, D., & Rymkevich, O. (2017). Varieties of CSR: Institutions and socially responsible behaviour. International Business Review, 26(6), 1064–1074. https://doi.org/https://doi.org/10.1016/j.ibusrev.2017.03.011

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. Journal of Business Ethics, 167(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1

EUROSIF, (2018). European SRI study 2018. Available at: http://www.eurosif.org/wp-content/uploads/2018/11/European-SRI-2018-Study.pdf

Gutsche, G., & Ziegler, A. (2019). Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. Journal of Banking & Finance, 102, 193–214. https://doi.org/https://doi.org/10.1016/j.jbankfin.2019.03.007

Hartzmark, S. M., & Sussman, A. B. (2019). Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. The Journal of Finance, 74(6), 2789–2837. https://doi.org/https://doi.org/10.1111/jofi.12841

Hawn, O., Chatterji, A. K., & Mitchell, W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). Strategic Management Journal, 39(4), 949–976. https://doi.org/https://doi.org/10.1002/smj.2752

Hörisch, J., Johnson, M. P., & Schaltegger, S. (2015). Implementation of sustainability management and company size: A knowledge-based view. Business Strategy and the Environment, 24(8), 765–779. https://doi.org/https://doi.org/10.1002/bse.1844

Liang, H. ao, & Renneboog, L. (2017). On the foundations of corporate social responsibility. The Journal of Finance, 72(2), 853–910. https://doi.org/https://doi.org/10.1111/jofi.12487

Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. Organization & Environment, 28(2), 223–249. https://doi.org/10.1177/1086026615575332

Riedl, A., & Smeets, P. (2017). Why Do Investors Hold Socially Responsible Mutual Funds? The Journal of Finance, 72(6), 2505–2550. https://doi.org/https://doi.org/10.1111/jofi.12547

Social Investment Forum Foundation, 2010. Report on Responsible Investing Trends in the United States 2010. http://www.socialinvest.org

US SIF Foundation, 2020. Report on US Sustainable and Impact Investing Trends 2020. https://www.ussif.org/

Chapter 1.

Cluster analysis to validate the sustainability label of stock indices: An analysis of the inclusion and exclusion processes in terms of size and ESG ratings

Synopsis

Sustainability stock indices play an important role in guiding socially responsible funds to their constituents. Thus, to find out whether the term sustainability is more than just a label, we analyze the inclusion and exclusion criteria applied by sustainability indices, and we compare them with those applied by conventional indices. We analyze the level of sustainability and size of the companies included in and excluded from five sustainability indices compared to a control group of 11 conventional indices. Our results show that the level of sustainability influences the inclusion process and, to a lesser extent, the exclusion process of the five FTSE4Good indices. However, we find similar results for several conventional indices. In addition, the size criterion dominates the sustainability criterion in the inclusion and exclusion processes of sustainability indices like in conventional indices. Further, we use different cluster algorithms to determine that the inclusion and exclusion processes of four of the five sustainability indices are different from those of the conventional indices. Our results validate the use of the "sustainability" label for four of five sustainability indices but also show that further differentiation between sustainability and conventional indices is needed.

This chapter has been published in the Journal of Cleaner Production (2022 with Laura Andreu & José Luis Sarto). DOI: https://doi.org/10.1016/j.jclepro.2021.129862

1. Introduction

In the last two decades, two major trends have existed in asset management: the increase in socially responsible (SR) investment and the increase in passive management. Renneboog et al. (2008) define SR investment as a process that integrates social, environmental, and ethical considerations into investment decision-making. In 2018, assets under management in SR investment had increased to $12 trillion in the United States (US SIF, 2018) and to €11 trillion in Europe (EUROSIF, 2018). Similar to SR investment, passive management which replicates market indices with extremely low fees for investors, has grown considerably (Sushko and Turner, 2018). Therefore, the analysis of sustainability indices is important both as benchmarks for active SR investment and for tracking passive SR investment.

Conventional and sustainability indices measure the evolution of the performance of a set of stocks in a geographical area (e.g., United States, Europe, World). While conventional indices include and exclude companies based exclusively on financial criteria (e.g., market capitalization), sustainability indices also consider environmental, social, and governance (ESG) criteria. Studies have often referred to a firm's consideration and response to issues beyond the financial, technical, and legal requirements as corporate social responsibility (CSR) (Montiel, 2008; Liang and Renneboog, 2017; Abbas, 2020). However, there is no single definition of CSR (Carroll, 1999; Garriga and Melé, 2004; Ashrafi et al., 2018). Studies have also referred to the degree to which a company responds to environmental and social aspects as corporate social/ sustainability performance (CSP) which is usually measured using ESG ratings (Zhao and Murrell, 2016; Awaysheh et al., 2020; Drempetic et al., 2020). Thus, sustainability indices contain companies that meet high ESG standards.

BlackRock, Vanguard, and State Street dominate passive investment (Fichtner et al., 2017), while S&P Dow Jones, Morgan Stanley Capital International (MSCI) and FTSE Russell are the three largest index providers with a market share of 70% in 2018 (Walker, 2019). In this research, we analyze five FTSE4Good sustainability stock indices and some FTSE Russell conventional stock indices. We chose FTSE indices because they are some of the most important ones, and other studies have widely used the FTSE4Good series (Belghitar et al., 2014; Montoya-Cruz et al., 2020; El Ouadghiri et al., 2021). Nevertheless, we also include common conventional indices such as S&P 500, EuroStoxx 50, and the Stoxx 600 for robustness purposes.

The fundamental work of the index providers is to create and maintain an index. Once this work is done, the supplier can produce a myriad of sub-indices based on a subset of the same companies (Tabor and Molinas, 2020). FTSE Russell uses this method. The FTSE4Good criteria is applied to create the FTSE Developed Index Series and the FTSE Emerging Index Series, which cover over 23 developed countries and 20 emerging countries.[1] The FTSE4Good criteria comprises companies whose overall ESG rating is 3.3 or higher for developed markets and 2.9 or higher for emerging markets. On the other hand, the FTSE excludes the companies whose overall ESG rating is lower than 2.9 for developed countries and lower than 2.4 for emerging markets. Like other sustainability indices, the FTSE4Good excludes companies that have had major controversies and those with particular business activities such as tobacco or weapons.[2]

Petry et al. (2019) underline that index providers steer capital with their indices because the inclusion of firms or countries in an index can result in large inflows while exclusions can cause large outflows. Therefore, academics and regulators should thoroughly analyze the decisions made by indices because of their influence on financial markets and SR flows. Especially on SR investors given that they are willing to sacrifice returns for investing in SR products (Borgers and Pownall, 2014; Gutsche and Ziegler, 2019).

However, most studies focus on comparing the financial performance of conventional and sustainability indices (Schröder, 2007; Cunha et al., 2020; Chiappini et al., 2021) or on analyzing whether the inclusion or exclusion from a sustainability index affects its financial performance (Oberndorfer et al., 2013; Kappou and Oikonomou 2016; Durand et al., 2019). Instead of analyzing the financial performance, and more closely related to our analysis, some studies analyze the factors that explain why a company enters into or exits from a sustainability index (Pineiro-Chousa et al., 2019; Arribas et al., 2021). However, these studies only analyze sustainability indices,

[1] See FTSE Russell Factsheet.

[2] We do not only analyze FTSE indices; hence, we have decided to use the ESG score from Refinitiv that replaces the existing ASSET4® Equal Weighted Ratings (Refinitiv, 2019). This database is prestigious as numerous studies have used it to measure the CSP of companies (Miras-Rodríguez et al., 2015; Nuber et al., 2020; Rajesh and Rajendran, 2020). In addition, the number of companies in this database is much higher than in other rating databases.

while we analyze both sustainability and conventional ones. It is important to analyze both types of indices because the conclusions may not be specific to the sustainability indices but to the stock indices in general. Therefore, our first objective is to analyze the inclusion and exclusion criteria of sustainability and conventional indices in terms of CSP and size; and to analyze which of the two criteria dominates the other.

SR investors must reconcile two somewhat dual criteria when selecting investments: financial and nonfinancial. Therefore, in this study, we focus on the influence of company size (financial criteria) and ESG ratings (nonfinancial criteria) to determine the inclusions in and exclusions from sustainability indices. Moreover, as we apply the same analysis to sustainability and conventional indices, we can observe whether these criteria differ between sustainability and conventional indices.

As opposed to other studies, we argue that the use of relativized variables (percentile rank) is preferable to absolute values because they provide homogeneous and comparable values between indices and dates. These values show the position in which the company enters or leaves the index with respect to those companies that the index could have included or excluded. Specifically, for each index and each month in which the event occurs (inclusion or exclusion), we analyze the position in terms of CSP and size of companies that are included with regard to those companies that are not or the position of the companies that are excluded with regard to the companies that remain in the index. Thus, we analyze how important the CSP and the size are for companies that are included in or excluded from sustainability indices and from our control group of conventional indices.

The analysis and comparison between conventional and sustainability indices gives a better understanding of what “sustainability” is in the index industry. Our results show that the companies that the sustainability indices exclude are worse in terms of size than in terms of CSP. In fact, probit models show that CSP does not influence the exclusion process of certain indices. In the inclusion process, the CSP has more influence, although it does not predominate over the size criterion. Thus, our results show that sustainability indices first follow a size criterion and secondarily a CSP criterion. Similar to Drempetic et al. (2020), who put ESG rating agencies “under review” because the positive relationship between the size and the CSP may not provide SR investors with the information they need, our results show that index providers should improve the method that they use to create sustainability indices to differentiate these indices from conventional ones.

Petry et al. (2019) argue that index providers may establish standards for what constitutes a sustainable investment. Moreover, investors widely use sustainability indices as an indicator of CSP (Kappou and Oikonomou, 2016; Gómez-Bezares et al., 2017; Forcadell and Aracil, 2017). For both reasons, the second objective of this research is to test whether the FTSE4Good indices are worthy of the “sustainability” label. Thus, we use a cluster analysis to test whether the inclusion and exclusion criteria of sustainability indices in terms of CSP and size are different from that of conventional ones. The goal of a cluster analysis is to discover the natural grouping(s) of a set of individuals (Jain, 2010). Our analysis shows that four of the five sustainability indices are different from the conventional indices. Therefore, we validate the “sustainability” labeling of four sustainability indices. The development of methods to validate the labeling used by index providers is necessary due to the growth of passive investment in general and passive SR investment in particular. To the best of our knowledge, our research that is based on unsupervised learning techniques is the first to show that the criteria applied by sustainability indices differ from those applied by conventional indices.

The paper is organized as follows: Section 2 reviews the literature and introduces the hypotheses, section 3 describes our sample design and the methodology, section 4 shows the empirical results, section 5 discuss the findings obtained and section 6 concludes.

2. Literature review and hypotheses

The literature agrees that there is a growing awareness of the global links among environmental problems, socioeconomic issues related to poverty and inequality, and concerns about a healthy future for humanity (Hopwood et al., 2005). The literature uses the assumption that sustainability stock indices are an appropriate indicator for corporate environmental and social activities (Chatterji and Mitchell, 2018; Arribas et al., 2021). Consolandi et al. (2009) conclude that SR investors interpret the inclusion of a firm in a sustainability index as a “certification” of a high degree of CSR, while they interpret the deletion from an index as a loss in CSR status. However, Ziegler (2012) shows that factors that are not directly connected to corporate, environmental, or social activities may influence the composition of sustainability indices.

Table 1.1 provides an overview of the studies that use the permanence and inclusion (exclusion) of companies from sustainability indices as a proxy for good (poor) CSP. Specifically, there are several short-term event studies that analyze whether the inclusion in or exclusion from sustainability indices influence corporate financial performance. As indicated by Oberndorfer et al. (2013), the reliability of event studies is that the timing of the event is exogenous and thus the company cannot influence the event. Event studies assume that stronger CSP criteria than those of conventional indices guide the inclusion or exclusion decisions of sustainability indices. Thus, we hypothesize that sustainability indices follow CSP criteria in their inclusion process (H1A) and in their exclusion process (H1B):

Hypothesis 1A: The level of CSP influences the inclusion in sustainability indices.

Hypothesis 1B: The level of CSP influences the exclusion from sustainability indices.

In order to confirm that the influence of CSP on inclusions and exclusions is specific to sustainability indices as opposed to conventional ones, we replicate the analyses with the conventional indices as a control group.

Drempetic et al. (2020) indicate that the method used by index providers to score companies gives an advantage to large firms. This idea is consistent with the studies that find a positive relation between size and CSP (Orlitzky, 2001; Udayasankar, 2008; Hörisch et al., 2015). Therefore, whether the size criterion dominates the CSP influence in the inclusion and exclusion processes of sustainability indices, these indices may not provide SR investors with the information they need to make the correct decisions based on their beliefs. Thus, we hypothesize that the influence of CSP dominates the influence of size in the inclusion and exclusion processes of sustainability indices.

Hypothesis 2: The CSP criteria dominates over the size criterion in the inclusions in and exclusions from sustainability indices.

Table 1.1 Review of studies that use inclusion, permanence or exclusion from sustainability indices as a proxy for CSP

This table shows a brief literature review of the permanence of/ inclusion in/ exclusion from sustainability indices. The first column shows the authorship of the study, the second column shows the indices analyzed, the third column gives a brief description of the study, and the fourth column shows the main findings.

Authors	Index	Description	Main Findings
McWilliams and Siegel (2000)	Domini 400 Social Index	Analysis of the relation between financial performance and the permanence in the Domini 400 Social Index	CSP has a neutral impact on financial performance
Curran and Moran (2007)	FTSE4Good UK	Event study to analyze the abnormal daily returns associated with inclusions and exclusions	The abnormal daily returns associated with the event are not significant
Becchetti et al. (2008)	Domini 400 Index	Analysis of the relation between inclusion and permanence in the domini social index and financial performance	Permanence in the Domini Index reduces returns on equity but not when large and R&D investing companies are excluded
Doh et al. (2010)	Calvert social Index	Analysis of the positive/ negative shareholder wealth effect associated with a firm's addition/ deletion to the index	The abnormal returns associated with deletions are weakly negative
Artiach et al. (2010)	Dow Jones Sustainability World Index	Analysis of the accounting determinants to be a member of the index	Sustainability firms are significantly larger but do not have greater free cash flows or lower leverage than other firms
Ziegler and Schröder (2010)	Dow Jones Sustainability World and Dow Jones Stoxx Sustainability	Determinants of the inclusion of European firms in the Dow Jones Sustainability World Index and the Dow Jones Stoxx Sustainability Index	The composition of the index is also influenced by factors that do not necessarily have to be directly related to the environmental or social activities of the companies
Cheung (2011)	Dow Jones Sustainability	Event study to analyze the stock return, risk, and liquidity associated with the event (exclusions and inclusions) in US companies	There is no strong evidence that the announcement will have a significant impact on stock returns and risk
Ziegler (2012)	Dow Jones Sustainability World Index	Analysis of the effects of inclusion in the Dow Jones Sustainability World Index on corporate financial performance	Weak or neutral effect of inclusion in the index on corporate financial performance
Oberndorfer et al. (2013)	Dow Jones STOXX Sustainability Index and the Dow Jones Sustainability World Index	Event study using three-factor Fama and French and a t-GARCH(1,1) to analyze inclusions of German firms in sustainability indices	Stock markets penalize the inclusion of a firm in sustainability stock indices
Kaspereit and Lopatta (2016)	Dow Jones Sustainability Index Europe	Analysis of the effects of permanence in the Dow Jones Sustainability World Index on corporate financial performance	Positive association between CSP and market value
Kappou and Oikonomou (2016)	MSCI KLD 400	Analysis of the financial effects of additions to and deletions from the social index MSCI KLD 400	Addition in the index does not lead to material changes in its market price, whereas deletions are accompanied by negative cumulative abnormal returns
Chatterji and Mitchell (2018)	Dow Jones Sustainability Index World	Event study of reactions to the addition, continuation, and deletion from the index	Investors appear to punish firms that are added to or continue on the index
Pineiro-Chousa et al. (2019)	S&P500 Environmental and Socially Responsible Index	Determinants of changes in the composition of SRI indices	There is no single financial performance indicator that explains the exclusion from or inclusion in a sustainability index

The results of studies that compare the performance of sustainability and conventional indices are not conclusive (Cunha et al., 2020). In addition, the results of some studies question the suitability of sustainability indices as a reference for SR investment. Cortez et al. (2009) and Leite and Cortez (2014) conclude that conventional benchmarks explain the returns of SR funds better than sustainable benchmarks. Joliet and Titova (2018) analyze the relation between SR funds' investment decisions and CSP. They find that new inclusions in the portfolios of passive management funds are not related to the CSP of a company but to an increase in its size. Ziegler and Schröder (2010) discuss the reliability of the Dow Jones Sustainability Index as an indicator of CSP. Therefore, it would be important to know whether the criteria applied by both types of indices are sufficiently different to state that sustainability indices deserve a label that distinguishes them from conventional indices.

As opposed to the majority of the research that is focused on whether inclusions in or exclusions from a given sustainability index affect financial performance and whether the risk adjusted returns of sustainability indices are different from conventional ones, we propose the application of a cluster analyses to find out whether the inclusion and exclusion processes of sustainability and conventional indices are different. Aldenderfer and Blashfield (1984) summarize the goals of cluster analysis in four major aspects: development of a classification; investigation of useful conceptual schemes for grouping entities; hypothesis generation through data exploration; hypothesis testing or the attempt to determine if types defined through other procedures are in fact present in a data set. The fourth goal perfectly suits our objective of knowing whether the label "sustainability" is present in our set of indices.

Hypothesis 3A: There are differences in the inclusion criteria of sustainability and conventional indices in terms of CSP and size.

Hypothesis 3B: There are differences in the exclusion criteria of sustainability and conventional indices in terms of CSP and size.

Hypothesis 3C: There are differences in the inclusion and exclusion criteria of sustainability and conventional indices in terms of CSP and size.

3. Data and methodology

3.1 Data

We analyze five FTSE4Good sustainability indices for different geographic areas: FTSE4Good Global, FTSE4Good Developed 100, FTSE4Good US, FTSE4Good US 100, and FTSE4Good Europe. We select these indices because they are diversified (they have a high number of constituents) and can be tracked by passive SR investments. We focus on Europe and US because they are important financial areas. However, we also include global indices to make our study more comprehensive. We also select 11 conventional indices for these geographic areas from FTSE and from different providers such as S&P and Stoxx. Table 1.2 provides more information about the analyzed indices such as the geographic area or the index supplier.

Table 1.2 Indices analyzed by geographic area and data supplier
This table lists the 16 indices analyzed in this study and their geographic areas, the index type, the data suppliers, and the Refinitiv ticker.

Name	Market	Type	Index Supplier	Ticker
FTSE4Good Global	Global	Sustainability	FTSE Group	LFT4GBGL
FTSE4Good Developed 100	Global	Sustainability	FTSE Group	LFT4G100
FTSE4Good US	United States	Sustainability	FTSE Group	LFT4GBUS
FTSE4Good US 100	United States	Sustainability	FTSE Group	LFT4U100
FTSE4Good Europe	Europe	Sustainability	FTSE Group	LFT4GBEU
FTSE Global	Global	Conventional	FTSE Group	LFAWRLD
FTSE Global 100	Global	Conventional	FTSE Group	LFTSEGL
FTSE US	United States	Conventional	FTSE Group	LWIUSAM
FTSE US All caps	United States	Conventional	FTSE Group	LFAUSAM
FTSE Eurofirst 100	Europe	Conventional	FTSE Group	LFTEFC1E
FTSE Eurotop 100	Europe	Conventional	FTSE Group	LFTEU100
S&P 500	United States	Conventional	Standard & Poor's	LS&PCOMP
S&P 100	United States	Conventional	Standard & Poor's	LS&P100I
S&P EURO	Europe	Conventional	Standard & Poor's	LSPEUROP
STOXX 50	Europe	Conventional	Stoxx	LDJSTO50
STOXX 600	Europe	Conventional	Stoxx	LDJSTOXX

We use the country of domicile to determine the location of a company. We group these countries into geographic areas when necessary to resemble the geographic areas of the indices. Table A1.9 of the Appendix shows the distribution of the companies across years (June 2007- June 2017) and across the geographic areas. Our unbalanced panel data comprise 555,816 monthly observations belonging to 7,378 companies. The number of companies analyzed has increased over time which reflects the growth in the ESG rating industry (Saadaoui and Soobaroyen, 2018; Escrig-Olmedo et al., 2019).

There is no consensus on the inclusion of the governance dimension in CSP because the governance pillar overlaps with corporate governance issues, which differ from the other stakeholder issues (Hong et al., 2012; Krüger, 2015; Liang and Renneboog, 2017). However, the FTSE4Good indices use ESG criteria and companies with exposure to significant controversies are not eligible. Hence, we argue that the best proxy for a company's CSP is the ESG ratings. Specifically, we use the ESG score and the ESG combined score provided by Refinitiv. The ESG score is an overall score whose value depends on the company's performance in three dimensions (environmental, social, and governance). The ESG combined score reduces the overall score as a result of controversies in which a company has been involved.

Once we define the CSP proxies, it is necessary to define the variables related to the company size. Several studies in the field use sales, the number of employees, or total assets of a company as a proxy for size (Gallego-Álvarez et al., 2014; Gómez-Bezares, et al., 2017; Minutolo et al., 2019) but these size measures do not suit our analysis because of the large differences among industries. Hence, we use the market value of a company to measure size because market value is the main criteria followed by conventional indices. Specifically, we use the market value in American dollars to homogenize the sample because our sustainability and conventional indices belong to different geographic areas with different currencies.

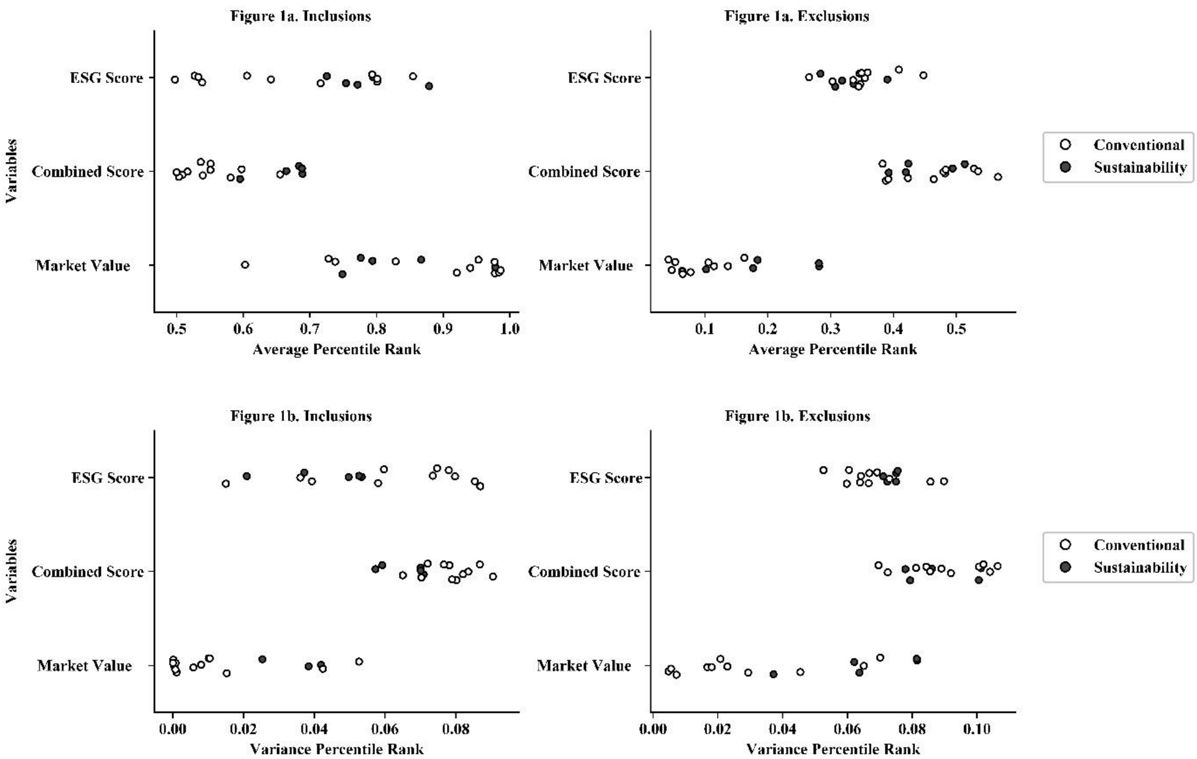
Table A1.10 of the Appendix shows the descriptive statistics of monthly observations for the ESG score, the ESG combined score, and the market value by geographic area; and Table A1.11 of the Appendix provides the descriptive statistics on the number of constituents of each index and the free float weight covered by our sample. The monthly composition of the indices was obtained from Refinitiv.

3.2 Percentile rank method

We define inclusions as those companies that did not belong to the index in the previous month but were added in the current month, and exclusions as those companies that belonged to the index in a given month and did not in the next month. In this study, we argue that what provides information is the ranking position, that is, the position at which the index adds or excludes the company with respect to those companies that the index could have included or excluded and not the absolute value of our variables. Hence, we measure the influence of the CSP and size criteria on the inclusion and exclusion processes of the indices by calculating the percentile rank of the ESG score, the ESG combined score, and the market value on a monthly basis.[3]

By using this method, we can analyze the position of inclusions and exclusions in different months. Increases or decreases in the CSP or size in the period of analysis would make the analysis based on the absolute value of different dates impossible. Additionally, the limited number of inclusions or exclusions for each review date of some indices prevents a monthly analysis. Moreover, the CSP and the size of companies differ among regions (Ferrel et al., 2016; Auer, 2018); therefore, the comparison of the absolute values of companies of different geographical areas would not be appropriate. Moreover, several SR products use a best-in-class approach, which is a method similar to the percentile rank.

Figure 1.1 Average and variance of the percentile rank of inclusions and exclusions
This figure shows the average and variance of the percentile rank for exclusions and inclusions in the three variables analysed (ESG score, ESG combined score, and market value) for conventional and sustainability indices.



3 We compute the percentile rank as the relative rank and not as the cumulative distribution. We obtain the rank of each company in its peer group and then we compute the relative rank of the company as (rank-1)/(# companies -1). Thus, the values range from zero to one.

In order to obtain the position in terms of the CSP and the size of the companies included in the index, we calculate the percentile rank, for each index and month, with the companies that belong to the same geographical area of the index and are not part of the index. Similarly, for the companies excluded from the index, we calculate the percentile rank, for each index and month, with the companies that belong to the index.

Figure 1.1A shows the average percentile rank (position) that the inclusions and exclusions have taken in the whole sample period in terms of CSP and size for each index. This average position measures how strong the indices consider the CSP and the size criteria. On the other hand, Figure 1.1B shows the variance in the percentile rank (volatility in the position) that the inclusions and exclusions display for the whole sample period in terms of the CSP and size for each index. The variance in the percentile rank shows how strongly the indices apply the CSP and the size criteria.

3.3 Methods for testing hypotheses 1 and 2

For hypotheses 1A and 1B, we use the T-test. We assume that an index follows the CSP criteria in the inclusion process whether the positions of inclusions are higher than the positions of companies that are not included. Similarly, we assume that an index follows the CSP criteria in the exclusion process whether the positions of exclusions are lower than the positions of maintenances.

For hypothesis 2, we use the T-test and the Bartlett's test of variance differences.[4] We assume that the CSP criteria dominate the size criterion in the inclusion process whether the CSP positions of inclusions are higher than the size positions. Similarly, we assume that the CSP criteria dominate size in the exclusion process whether the CSP positions of exclusions are lower than the size positions. We also test this hypothesis by comparing the variance of the percentile rank between CSP and size. We assume that indices apply the CSP criteria more firmly than the size criterion, whether the variance in the percentile rank of the CSP is lower than the variance in the percentile rank of size.

The position at which the index includes or excludes the company with respect to those companies that the index could have included or excluded is important. However, we also test whether there is a causal relationship between the position in

---

[4] We apply the T-test and the Bartlett's test using SciPy version 1.4.1 (Virtanen, et al., 2020). SciPy is an open-source scientific computing library for the Python programming language.

terms of CSP and size with the inclusion in or exclusion from sustainability indices. Thus, as an additional robustness check, we also test hypotheses 1 and 2 with the following two probit regressions for each index:

$$Inclusion_{it} = \beta_0 + \beta_1 ESG\ Score_{it} + \beta_2 ESG\ Combined_{it} + \beta_3 MV_{it} + \beta_4 ROA_{iy-1} + \beta_5 \frac{Total\ liabilities_{iy-1}}{Assets_{iy-1}} + \beta_6 \frac{Capital\ expenditures_{\ iy-1}}{Assets_{\ iy-1}} + \varepsilon_{it} \quad \textbf{(1)}$$

where $Inclusion_{it}$ is a dummy that equals one when company i is in its last month out of the index, that is, the index is going to add the company in the next month and zero otherwise. $ESG\ score_{it}$, $ESG\ Combined_{it}$, and MV_{it} are the percentile ranks of the ESG score, ESG combined score, and market value of company i in month t; ROA_{iy-1} is the return on assets (profitability); $\frac{Total\ liabilities_{iy-1}}{Assets_{\ iy-1}}$ is the total liabilities of the company divided by total assets (capital structure); and $\frac{Capital\ expenditures_{\ iy-1}}{Assets_{\ iy-1}}$ is the additions to fixed assets divided by total assets (capital intensity) of company i in the previous year (y-1).

Equation 2 is similar to equation 1 but here we examine the position of the companies excluded from an index.

$$Exclusion_{it} = \beta_0 + \beta_1 ESG\ Score_{it} + \beta_2 ESG\ Combined_{it} + \beta_3 MV_{it} + \beta_4 ROA_{iy-1} + \beta_5 \frac{Total\ liabilities_{iy-1}}{Assets_{iy-1}} + \beta_6 \frac{Capital\ expenditures_{\ iy-1}}{Assets_{\ iy-1}} + \varepsilon_{it} \quad \textbf{(2)}$$

where $Exclusion_{it}$ is a dummy that equals one when the company i is in its last month in the index, that is, the index is going to exclude the company in the next month, and zero otherwise. The remaining variables are defined as in Equation 1. As additional information, Table A1.12 of the Appendix shows the correlation matrix between the variables used in this research.

3.4 Method for testing hypothesis 3

To test hypotheses 3, we use a cluster analysis because the percentile rank method provides a standardization of the data (Milligan and Cooper, 1988; Bakoben et al., 2020). Several disciplines widely apply clustering techniques (see, Peters et al., 2013). In economics, it is used for applications such as the recognition of purchase patterns and the grouping of firms or analyzing stock trends (Xu and Wunsch, 2008). However, we have not found studies that use this method to validate or perform an index classification.

We assume that the differences between conventional and sustainability indices can be explained by the CSP and the size criteria. In order to carry out the cluster analysis, we assume that each index is characterized by the average of and the variance in the percentile rank of inclusions and exclusions in each of the variables analyzed: the ESG score, the ESG combined score, and the market value (see, Figure 1). Specifically, we analyze the average of and the variance in inclusions, the average of and the variance in exclusions, and both the average of and the variance in inclusions and exclusions.

Backer and Jain (1981) indicate that in cluster analyses, the elements are split into a number of more or less homogeneous subgroups on the basis of a subjectively chosen measure of similarity. However, according to Jain (2010), there is no single definition of similarity or cluster that consequently, has resulted in the publication of thousands of clustering algorithms. Therefore, for reducing any subjectivity and for increasing the robustness of our results, we run five clustering algorithms (k-means, agglomerative clustering, spectral clustering, mean shift, and affinity propagation). If the output of each algorithm is a cluster that is composed of the set of sustainability indices, then we can affirm that these indices are different from the rest and deserve the differentiating label of “4Good”.

The k-means, agglomerative clustering, and spectral clustering algorithms require the specification of the number of clusters (n) returned by the algorithm. As we do not know the ex-ante number of groups, we run these algorithms from n = 2 to 5. By contrast, the mean shift and affinity propagation methods do not require the specification of the number of clusters.

We have a high dimensionality problem that prevents the presentation of the groups returned by the algorithms in a two-dimensional plot. Hence, we apply the principal component analysis (PCA) to project our data on a lower dimensional space (two variables). PCA is widely employed to reduce the number of dimensions (see e.g., Jiang, et al., 2012; Ortas, et al., 2015). By using the PCA, we can plot the groups found in the cluster analysis.

4. Empirical results

4.1 Hypotheses 1

4.1.1 Are the inclusion processes of sustainability indices following a CSP criteria? (Hypothesis 1A)

In this subsection, we analyze the position (percentile rank) in terms of the CSP of companies included in sustainability indices. In order to confirm that the influence of CSP is specific to sustainability indices, we replicate the analysis in the control group of conventional indices. Table 1.3 shows the results of the T-test for the difference in means between the inclusions and the companies that the index does not include in terms of the ESG score and ESG combined score.

Table 1.3 Comparison of the companies included in the index versus the universe of companies that could be included

This table shows the results of the test of equal means of the ESG score and ESG combined score of inclusions and the universe of companies that could be included in the index. The first column shows the indices analyzed, the second and third columns show the number of companies that were included in the index (I) and the universe of companies that could have been included (U) in each index, respectively. The following columns show the average position (percentile rank) for I and U and the result of T-test (H_0: $\mu_I = \mu_U$) for each variable: ESG score, ESG combined score. Bartlett's test of equal variance was considered to calculate the T-test. The * and ** indicate statistical significance at the 5%, 1%, levels, respectively.

	#		ESG Score			Combined Score		
	I	U	I	U	Test	I	U	Test
FTSE4Good Global	586	101,804	0.754	0.498	26.7**	0.689	0.499	17.3**
FTSE4Good Developed 100	118	105,828	0.879	0.499	28.4**	0.595	0.5	4.3**
FTSE4Good US	183	19,979	0.772	0.497	16.5**	0.684	0.498	8.7**
FTSE4Good US 100	113	39,718	0.794	0.499	16.2**	0.665	0.499	7.2**
FTSE4Good Europe	244	20,897	0.725	0.497	15.4**	0.688	0.498	10.3**
FTSE Global	1,037	35,167	0.498	0.5	-0.3	0.516	0.499	1.9
FTSE Global 100	109	195,897	0.794	0.5	12.7**	0.551	0.5	1.9
FTSE US	207	14,361	0.539	0.499	1.9	0.536	0.499	1.8
FTSE US All caps	258	5,884	0.527	0.499	1.5	0.551	0.498	2.9**
FTSE Eurofirst 100	30	9,929	0.801	0.499	8.7**	0.509	0.5	0.2
FTSE Eurotop 100	63	15,935	0.801	0.499	12.0**	0.581	0.5	2.2*
S&P 500	208	69,906	0.642	0.499	7.1**	0.598	0.5	4.9**
S&P 100	42	22,514	0.716	0.5	4.9**	0.504	0.5	0.1
S&P EURO	52	31,868	0.606	0.5	2.6**	0.655	0.5	3.9**
STOXX 50	36	20,296	0.855	0.499	17.3**	0.5	0.5	0.0
STOXX 600	417	41,158	0.533	0.5	2.3*	0.539	0.5	2.8**

By focusing on the ESG score, we can conclude that the five sustainability indices follow an ESG criterion in their inclusion process. However, all conventional indices, except three, also include companies with high ESG scores. The FTSE4Good

Developed 100 reaches the best positions in terms of the ESG score, although the positions of inclusions in three conventional indices are higher than the other four sustainability indices. Regarding the ESG combined score, all the sustainability indices and five conventional indices show higher positions for inclusions than for companies that could be included in the indices.

Our results show that the sustainability indices consider the ESG score and the ESG combined score in their inclusion process. Therefore, we accept hypothesis 1A. However, some conventional indices also include companies with high ESG scores.

4.1.2 Are the exclusion processes of sustainability indices following CSP criteria? (Hypothesis 1B)

In this subsection, we analyze the positions in terms of the CSP of companies excluded from sustainability indices. In order to confirm that the influence of CSP is specific to sustainability indices, we replicate the analysis with the control group of conventional indices. Table 1.4 shows the results of the T-test for the difference in means between maintenances and exclusions in terms of the ESG score and the ESG combined score.

By focusing on the ESG score, we can conclude that the five sustainability indices consider the ESG score in their exclusion process. However, all conventional indices, except the STOXX 50, also follow an ESG criterion. Moreover, the exclusions from the FTSE Eurofirst 100 have the lowest positions in terms of the ESG score of all indices. By analyzing the ESG combined score, the results become heterogeneous for both groups of indices. Three sustainability indices and four conventional indices show lower positions for exclusions than for maintenances. Therefore, the controversies of the companies seemingly are not a very important factor in being excluded from a sustainability index. Moreover, exclusions from the FTSE US have the lowest position in terms of the ESG combined score.

Our results show that the sustainability indices consider the ESG score in their exclusion process but rarely consider the ESG combined score. Moreover, the exclusion process of conventional indices that are based on market value (size) achieves similar or even better CSP levels than sustainability indices that may indicate a relation between size and CSP. This positive correlation between size and CSP is noticeable in the empirical literature (Hasan, et al., 2018; Yen et al., 2019). Thus, we only accept hypothesis 1B for three sustainability indices.

Table 1.4 Comparison of the firms that remain in the index versus those excluded from the index
This table shows the results of the test of equal means of the ESG score and ESG combined score of exclusions and maintenances. The first column of the table shows the indices analyzed, the second and third columns show the number of companies that were excluded (E) and the number of companies that remained (M) in each index, respectively. The following columns show the average position (percentile rank) for E and M and the result of T-test (H_0: $\mu_E = \mu_M$) for the ESG score and ESG combined score. Bartlett's test of equal variance was considered to calculate the T-test. The * and ** indicate statistical significance at the 5%, 1%, levels, respectively.

	#		ESG Score			Combined Score		
	E	**M**	**E**	**M**	**Test**	**E**	**M**	**Test**
FTSE4Good Global	347	28,393	0.307	0.502	-12.5**	0.393	0.501	-7.0**
FTSE4Good Developed 100	111	1,975	0.39	0.506	-4.1**	0.494	0.5	-0.2
FTSE4Good US	102	3,621	0.318	0.505	-6.4**	0.424	0.502	-2.7**
FTSE4Good US 100	98	2,163	0.337	0.507	-5.7**	0.513	0.499	0.5
FTSE4Good Europe	124	8,780	0.284	0.503	-8.4**	0.42	0.501	-3.1**
FTSE Global	445	270,237	0.346	0.5	-14.2**	0.388	0.5	-9.0**
FTSE Global 100	106	3,970	0.409	0.502	-3.3**	0.482	0.5	-0.6
FTSE US	168	13,169	0.303	0.502	-10.5**	0.383	0.501	-5.3**
FTSE US All caps	75	32,198	0.359	0.5	-4.2**	0.392	0.5	-3.2**
FTSE Eurofirst 100	27	1,080	0.266	0.506	-4.2**	0.566	0.498	1.2
FTSE Eurotop 100	56	1,072	0.347	0.508	-4.0**	0.528	0.498	0.7
S&P 500	108	27,861	0.345	0.501	-5.6**	0.464	0.5	-1.3
S&P 100	36	1,577	0.355	0.503	-3.0**	0.534	0.499	0.7
S&P EURO	40	4,772	0.349	0.501	-3.3**	0.48	0.5	-0.4
STOXX 50	29	766	0.448	0.502	-1.0	0.483	0.5	-0.3
STOXX 600	331	31,294	0.336	0.502	-11.8**	0.423	0.501	-4.9**

4.2 Hypothesis 2

We use the average of and the variance in the percentile rank of inclusions and exclusions to test whether the CSP criteria dominates size in the inclusion and exclusion processes of the indices.

4.2.1 Mean test

We argue that CSP dominates size whether the position of inclusions (exclusions) is higher (lower) in terms of the CSP than in terms of size. Table 1.5 shows the average percentile rank of inclusions and exclusions in the three variables analyzed and the result of the T-test. The T-test compares the average percentile rank of the size variable against the average percentile rank of the two CSP variables (ESG score and ESG combined score).

The market value dominates the ESG score for all indices except for the FTSE4Good Global and the FTSE4Good US. In these two indices, the market value and the ESG score have a similar influence on their inclusion and exclusion processes. However, the market value (the size criterion) dominates in all indices when analyzing the ESG combined score. Therefore, we can conclude that the influence of size dominates the CSP in the conventional and sustainability indices.

Table 1.5 Comparison between the CSP and size criteria for inclusions and exclusions in terms of average

This table shows the results of the test of equal means between CSP criteria (ESG score and ESG combined score) and size criteria (market value) for inclusions and exclusions. The first column shows the indices analyzed, and the second row shows the average of the percentile rank of the variables analyzed for inclusions and exclusions: market value (MV), ESG score, and ESG combined score (Com. Score). The Test column shows the result of T-test of equal means (H_0: $\mu_S = \mu_{CSP}$) between size (S) and the two CSP variables. Bartlett's test of equal variance was considered to calculate the T-test. The * and ** indicate statistical significance at the 5%, 1%, levels, respectively.

	Inclusions					Exclusions				
	MV	ESG Score		Com. Score		MV	ESG Score		Com. Score	
	S	CSP	Test	CSP	Test	S	CSP	Test	CSP	Test
FTSE4Good Global	0.749	0.754	-0.4	0.689	4.3**	0.282	0.307	-1.2	0.393	-5.1**
FTSE4Good Developed 100	0.978	0.879	7.4**	0.595	17.3**	0.102	0.39	-9.1**	0.494	-11.1**
FTSE4Good US	0.794	0.772	1.0	0.684	4.5**	0.281	0.318	-0.9	0.424	-3.5**
FTSE4Good US 100	0.868	0.797	3.5**	0.667	8.1**	0.177	0.337	-4.2**	0.513	-8.2**
FTSE4Good Europe	0.776	0.725	2.8**	0.688	4.5**	0.184	0.284	-3.1**	0.42	-7.0**
FTSE Global	0.603	0.498	9.9**	0.516	8.0**	0.163	0.346	-11.3**	0.388	-12.9**
FTSE Global 100	0.977	0.794	7.9**	0.551	15.7**	0.064	0.409	-10.9**	0.482	-12.4**
FTSE US	0.941	0.539	18.5**	0.536	18.3**	0.047	0.303	-12.0**	0.383	-13.7**
FTSE US All caps	0.728	0.527	9.0**	0.551	7.7**	0.137	0.359	-5.3**	0.392	-5.7**
FTSE Eurofirst 100	0.953	0.801	4.3**	0.509	9.5**	0.064	0.266	-3.2**	0.566	-8.7**
FTSE Eurotop 100	0.978	0.801	7.0**	0.581	11.8**	0.065	0.347	-7.9**	0.528	-10.4**
S&P 500	0.922	0.642	13.0**	0.599	14.9**	0.115	0.345	-7.1**	0.464	-9.8**
S&P 100	0.985	0.715	6.9**	0.495	11.0**	0.042	0.355	-6.7**	0.534	-9.0**
S&P EURO	0.829	0.606	5.6**	0.655	4.4**	0.077	0.349	-6.0**	0.48	-7.1**
STOXX 50	0.986	0.855	6.4**	0.5	10.9**	0.053	0.448	-7.7**	0.483	-7.5**
STOXX 600	0.738	0.534	13.5**	0.541	13.1**	0.106	0.336	-13.7**	0.423	-17.0**

4.2.2 Variance test

We also test hypothesis 2 by analyzing the variance in the percentile rank. This variable measures how strongly indices apply the criteria to include (exclude) a company. A small variance in terms of size for exclusions indicates that the companies excluded from the index are always in a similar position in terms of size. We argue that the CSP criteria dominates the size criterion whether the variance in the percentile rank of inclusions and exclusions is lower in terms of the CSP than the variance in terms of size. Table 1.6 shows the variance in the percentile rank of inclusions and exclusions in the three variables analyzed as well as the result of Bartlett's test.

If we focus on the ESG score, we observe that in the inclusion process of all indices, except the FTSE4Good US, the size criterion dominates the CSP criteria. In the exclusion process of sustainability indices, the applications of CSP and size criteria are similar except for the FTSE4Good Developed 100. On the other hand, in conventional indices, the size criterion tends to dominate the CSP criteria. Therefore, analyzing the variance, we also conclude that all indices apply the size criterion more strongly than the CSP criteria in their inclusion and exclusion processes.

Table 1.6 Comparison between the CSP and size criteria for inclusions and exclusions in terms of variance

This table shows the results of the Bartlett's test of equal variances between CSP criteria (ESG score and ESG combined score) and size criteria (market value) for inclusions and exclusions. The first column shows the indices analyzed, and the second row shows the variance of the percentile rank of the variables analyzed for inclusions and exclusions: market value (MV), ESG score and ESG combined score (Com. Score). The Test column shows the result of Bartlett's test of equal variances (H_0: $\sigma^2_S = \sigma^2_{CSP}$) between size (S) and the two CSP variables. The * and ** indicate statistical significance at the 5%, 1%, levels, respectively.

	Inclusions					Exclusions				
	MV	ESG Score		Com. Score		MV	ESG Score		Com. Score	
	S	CSP	Test	CSP	Test	S	CSP	Test	CSP	Test
FTSE4Good Global	0.042	0.053	8.5**	0.07	38.1**	0.081	0.072	1.2	0.079	0.1
FTSE4Good Developed 100	0.0	0.021	319.6**	0.057	434.4**	0.037	0.074	13.2**	0.1	26.0**
FTSE4Good US	0.038	0.049	3.0	0.071	16.8**	0.081	0.074	0.2	0.085	0.1
FTSE4Good US 100	0.01	0.037	43.3**	0.059	76.8**	0.063	0.075	0.7	0.1	5.1*
FTSE4Good Europe	0.025	0.052	31.7**	0.07	60.2**	0.062	0.071	0.6	0.077	1.6
FTSE Global	0.042	0.073	77.0**	0.078	95.1**	0.065	0.052	5.0*	0.069	0.5
FTSE Global 100	0.0	0.058	466.7**	0.08	501.3**	0.021	0.085	48.5**	0.1	59.2**
FTSE US	0.01	0.087	196.8**	0.09	203.3**	0.017	0.059	63.1**	0.084	98.5**
FTSE US All caps	0.052	0.074	7.8**	0.083	13.5**	0.069	0.063	0.2	0.08	0.4
FTSE Eurofirst 100	0.001	0.035	61.1**	0.063	77.1**	0.017	0.086	14.9**	0.07	11.4**
FTSE Eurotop 100	0.001	0.039	155.8**	0.069	190.4**	0.005	0.066	75.5**	0.105	98.2**
S&P 500	0.008	0.086	237.8**	0.087	240.6**	0.045	0.066	3.8	0.091	12.9**
S&P 100	0.001	0.06	120.3**	0.079	131.1**	0.005	0.071	46.2**	0.099	56.3**
S&P EURO	0.006	0.077	68.0**	0.075	67.1**	0.022	0.059	8.7**	0.101	20.2**
STOXX 50	0.0	0.015	135.9**	0.07	189.7**	0.007	0.067	29.4**	0.086	35.2**
STOXX 600	0.015	0.079	253.4**	0.078	250.5**	0.029	0.064	49.3**	0.085	90.2**

In no index do the CSP criteria dominate the size criterion. Hence, we reject hypothesis 2. Therefore, size has more influence on the inclusion and exclusion processes than the CSP criteria. The primacy of size is also observed in Joliet and Titova (2018), who conclude that inclusions in the portfolios of passive management funds that replicate the composition of sustainability indices are related to increases in the sizes rather than the CSPs of companies.

4.3 Results of the robustness analyses for testing hypotheses 1 and 2

We also test the hypotheses 1 and 2 through regressions 1 and 2. Table 1.7 shows the results of the probit regression on the inclusions for each index. This table shows that the influence of the ESG score and the market value on the inclusion process is positive and statistically significant in all sustainability indices, while the influence of the ESG combined score is only positive and statistically significant for the FTSE4Good Europe. We can conclude that the expected influence of market value in the inclusion process is higher than the influence of the CSP criteria because they are measured with the same unit (percentile rank). In some sustainability indices, the influence of the return on assets of the company is positive and statistically significant. This finding is in line with those studies that show that well-performing companies are the ones that carry out more CSR activities (Waddock and Graves, 1997; Soytas et al., 2019). The total liabilities to assets are only statistically significant at 5% in one sustainability index. This result is in line with Ziegler and Schröder (2010) and Arribas et al., (2021) who find that the company's capital structure does not influence the inclusion process of sustainability indices. Regarding conventional indices, the market value positively influences the inclusion process. However, the relative value of the ESG score only has a positive influence on three indices.

Table 1.8 shows the results of the probit regression on the exclusions of each index. This table shows that a high position in terms of the ESG score reduces the possibilities of being excluded from the FTSE4Good Global, FTSE4Good US, and the FTSE4Good Europe. In terms of the ESG combined score, only this influence is statistically significant for the FTSE4Good Global and the FTSE4Good Europe. Moreover, the influence of size in the exclusion process is greater than the influence of CSP, like the inclusion process. In both groups of indices, the larger the company, the lower the probabilities of being excluded. In general, in conventional indices the ESG score and ESG combined score do not influence the exclusion process although the return on assets of the company seems to reduce the probability of exclusion from several conventional indices.

Table 1.7 Probit regression for analyzing the influence of CSP and size on the index inclusions

This table shows the results of equation 1 for each index of row 2 where the dependent variable is equal to 1 when the company is in its last month outside the index, that is, the company is going to be added to the index in the next month and 0 otherwise. Rows 3 to 9 show the coefficients, the standard errors in parentheses, and the significance of each variable of equation 1. Column 10 shows the number of observations, and column 11 the fit of the model. The * and ** indicate statistical significance at the 5% and 1% levels, respectively. (1)FTSE4Good Global, (2)FTSE4Good Developed 100, (3)FTSE4Good US, (4)FTSE4Good US 100, (5)FTSE4Good Europe, (6)FTSE Global, (7)FTSE Global 100, (8)FTSE US, (9)FTSE US All caps, (10)FTSE Eurofirst 100, (11)FTSE Eurotop 100, (12)S&P 500, (13)S&P 100, (14)S&P EURO, (15)STOXX 50, (16)STOXX 600.

	Sustainability Indices					**Conventional Indices**										
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
Intercept	-3.74**	-20.30**	-3.68**	-4.59**	-3.91**	-2.21**	-18.41**	-7.34**	-2.88**	-8.98**	-33.42**	-5.73**	-13.98**	-4.93**	-38.88**	-3.22**
	(0.081)	(1.848)	(0.158)	(0.257)	(0.158)	(0.048)	(1.716)	(0.405)	(0.130)	(1.405)	(4.344)	(0.254)	(1.873)	(0.320)	(6.427)	(0.092)
P. Rank ESG Score	0.85**	1.18**	0.83**	0.50*	0.32*	-0.85**	0.27	-0.79**	-2.61**	0.84*	0.91*	0.03	0.54	-1.32**	0.96*	-0.49**
	(0.099)	(0.273)	(0.197)	(0.234)	(0.160)	(0.169)	(0.202)	(0.266)	(0.533)	(0.394)	(0.374)	(0.171)	(0.359)	(0.312)	(0.488)	(0.132)
P. Rank Com. Score	0.12	-0.06	0.03	0.17	0.56**	0.86**	-0.22	0.35	2.63**	-0.52	0.24	0.11	-0.86**	1.27**	-0.31	0.48**
	(0.086)	(0.177)	(0.159)	(0.183)	(0.141)	(0.166)	(0.173)	(0.260)	(0.524)	(0.302)	(0.266)	(0.163)	(0.309)	(0.298)	(0.320)	(0.129)
P. Rank Market Value	0.88**	17.63**	1.12**	2.21**	1.61**	0.70**	16.88**	6.76**	1.87**	6.51**	32.19**	3.85**	12.83**	2.39**	36.72**	1.44**
	(0.075)	(1.893)	(0.165)	(0.297)	(0.147)	(0.060)	(1.792)	(0.444)	(0.142)	(1.409)	(4.398)	(0.271)	(1.923)	(0.334)	(6.468)	(0.093)
ROA	0.00	0.00	0.01*	-0.00	0.01**	0.00**	0.01	0.00	-0.00	0.01	-0.02	0.01**	0.00	-0.00	0.03	0.00
	(0.002)	(0.007)	(0.004)	(0.004)	(0.002)	(0.001)	(0.006)	(0.003)	(0.002)	(0.007)	(0.013)	(0.003)	(0.009)	(0.006)	(0.017)	(0.002)
Liabilities to Assets	0.18*	0.19	0.08	-0.20	0.15	-0.24**	-0.68**	-0.32	-0.13	0.62	-0.28	-0.33*	-0.64	0.49*	0.74	0.13
	(0.080)	(0.231)	(0.160)	(0.192)	(0.146)	(0.059)	(0.207)	(0.178)	(0.119)	(0.425)	(0.421)	(0.137)	(0.365)	(0.243)	(0.542)	(0.092)
Capital Expt.	-1.37**	-2.68*	-1.84*	-2.69**	-1.72*	0.75**	-4.84**	0.90*	0.02	-0.72	-1.77	0.30	-0.99	1.74*	-6.73*	-1.69**
	(0.376)	(1.136)	(0.761)	(0.986)	(0.752)	(0.163)	(1.114)	(0.460)	(0.466)	(2.255)	(2.298)	(0.395)	(1.524)	(0.834)	(2.946)	(0.452)
#	96,261	102,187	19,700	37,810	19,102	30,872	185,126	14,137	5,994	9,662	15,306	66,724	23,416	32,055	18,886	38,029
Pseudo R2	0.099	0.387	0.123	0.162	0.132	0.026	0.346	0.35	0.122	0.31	0.493	0.221	0.383	0.14	0.483	0.076

Table 1.8 Probit regression for analyzing the influence of CSP and size on the index exclusions

This table shows the results of equation 2 for each index of row 2 where the dependent variable is equal to 1 when the company is in its last month inside the index, that is, the company is going to be excluded from the index in the next month and 0 if the company remains. Rows 3 to 9 show the coefficients, the standard errors in parentheses, and the significance of each variable of equation 2. Column 10 shows the number of observations, and column 11 the fit of the model. The * and ** indicate statistical significance at the 5% and 1% levels, respectively. (1)FTSE4Good Global, (2)FTSE4Good Developed 100, (3)FTSE4Good US, (4)FTSE4Good US 100, (5)FTSE4Good Europe, (6)FTSE Global, (7)FTSE Global 100, (8)FTSE US, (9)FTSE US All caps, (10)FTSE Eurofirst 100, (11)FTSE Eurotop 100, (12)S&P 500, (13)S&P 100, (14)S&P EURO, (15)STOXX 50, (16)STOXX 600.

	Sustainability Indices					Conventional Indices										
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
Intercept	-1.69**	0.13	-1.24**	-0.91**	-1.27**	-2.38**	-0.29	-0.84**	-2.05**	-0.74	-0.25	-1.90**	0.66	-1.61**	-0.12	-1.36**
	(0.085)	(0.327)	(0.201)	(0.276)	(0.175)	(0.053)	(0.274)	(0.146)	(0.139)	(0.634)	(0.476)	(0.157)	(0.760)	(0.346)	(0.751)	(0.099)
P. Rank ESG Score	-0.45**	-0.20	-0.52*	-0.42	-0.43*	-0.07	0.16	0.20	0.09	-1.13*	-0.46	-0.21	-0.36	0.00	1.02	0.45**
	(0.117)	(0.224)	(0.228)	(0.227)	(0.194)	(0.146)	(0.234)	(0.289)	(0.373)	(0.439)	(0.360)	(0.236)	(0.412)	(0.342)	(0.540)	(0.147)
P. Rank Com. Score	-0.26*	-0.17	-0.16	-0.10	-0.49**	-0.16	-0.25	-0.28	-0.38	0.23	0.36	0.17	-0.20	0.07	-0.94*	-0.47**
	(0.100)	(0.211)	(0.186)	(0.208)	(0.174)	(0.131)	(0.210)	(0.231)	(0.324)	(0.363)	(0.294)	(0.191)	(0.363)	(0.299)	(0.459)	(0.127)
P. Rank Market Value	-0.79**	-3.62**	-1.05**	-2.14**	-1.45**	-1.62**	-5.12**	-7.15**	-2.02**	-4.11**	-6.13**	-2.56**	-8.75**	-3.11**	-9.79**	-3.24**
	(0.099)	(0.324)	(0.210)	(0.248)	(0.190)	(0.093)	(0.451)	(0.549)	(0.270)	(0.826)	(0.835)	(0.280)	(1.454)	(0.476)	(1.639)	(0.175)
ROA	-0.00	-0.03*	0.00	0.02*	-0.01	-0.01**	-0.02*	-0.02**	-0.00**	-0.03	-0.05*	-0.01**	-0.01	-0.02	-0.02	-0.01**
	(0.003)	(0.011)	(0.006)	(0.009)	(0.006)	(0.001)	(0.010)	(0.005)	(0.002)	(0.017)	(0.024)	(0.004)	(0.014)	(0.013)	(0.034)	(0.003)
Liabilities to Assets	-0.07	-0.59	-0.13	0.15	-0.06	0.03	-0.27	-0.22	-0.15	-0.22	0.04	-0.03	-1.23	0.27	-0.17	-0.07
	(0.110)	(0.346)	(0.242)	(0.308)	(0.209)	(0.075)	(0.309)	(0.205)	(0.191)	(0.693)	(0.512)	(0.217)	(0.803)	(0.381)	(0.763)	(0.127)
Capital Expt.	1.30*	1.95	1.03	-2.85	0.35	0.48*	-3.09	0.46	-0.30	7.11*	2.11	-0.46	-2.35	-2.87	11.51*	1.39**
	(0.569)	(1.642)	(1.397)	(1.942)	(1.189)	(0.220)	(1.669)	(0.700)	(0.769)	(3.464)	(2.411)	(0.833)	(2.813)	(2.404)	(4.920)	(0.501)
#	28,808	2,124	3,736	2,224	8,870	265,313	4,195	12,625	30,513	1,224	1,318	26,277	1,678	4,790	786	31,064
Pseudo R^2	0.07	0.313	0.091	0.189	0.126	0.126	0.347	0.379	0.15	0.345	0.407	0.187	0.42	0.238	0.494	0.221

The probit model shows that the ESG score influences the inclusion process of the five sustainability indices. However, the ESG score does not influence the exclusion process of the FTSE4Good Developed 100 and the FTSE4Good US 100. Hence, we again reject hypothesis 1B for these indices. Nevertheless, the probit model confirms our conclusions related to hypothesis 2, the size criterion dominates the CSP criteria in the inclusion and exclusion processes of sustainability indices. Our analysis also shows the importance of analyzing several indices because what is valid for one index may not be valid for others.

4.4 Hypotheses 3

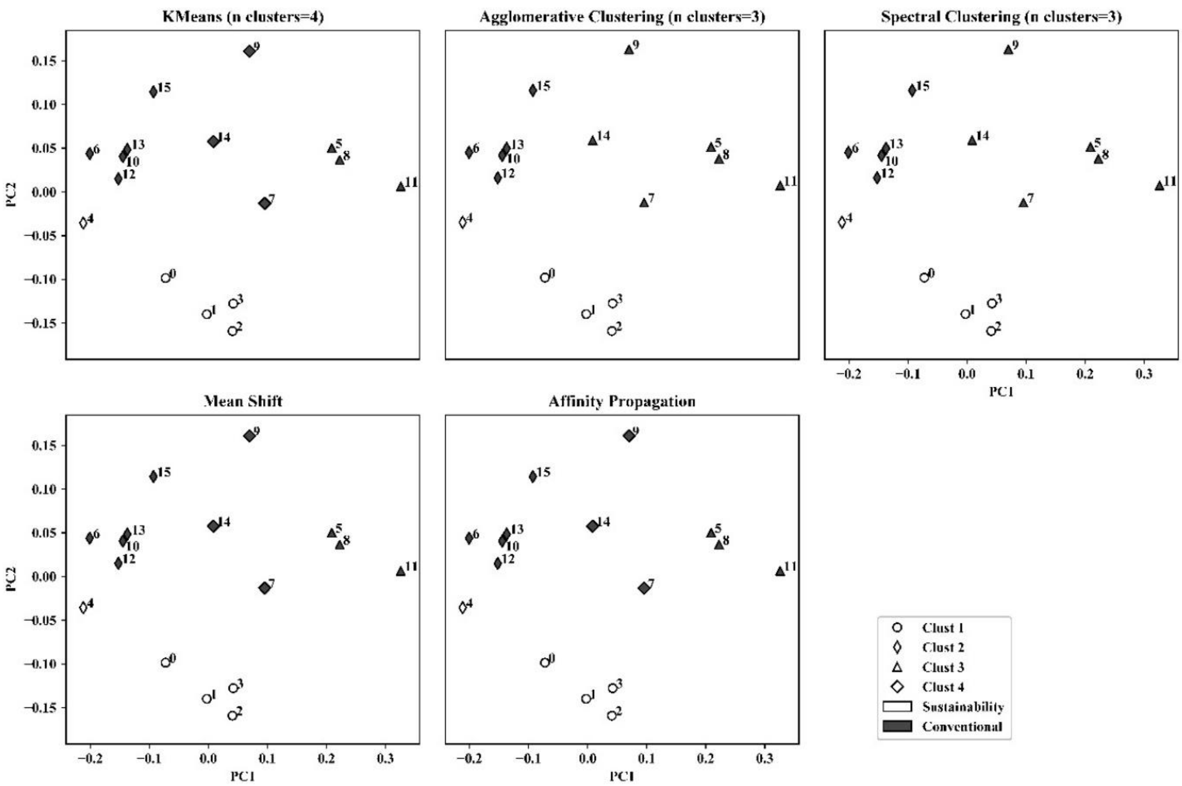
4.4.1 Are the inclusion processes of sustainability and conventional indices different in terms of CSP and size? (Hypothesis 3A)

While the evidence from hypotheses 1 and 2 shows that there are some differences between the criteria applied to conventional and sustainability indices, the differences are not easily observable as indicated by Figure 1. The results also show some differences between the criteria applied to the sustainability group. This heterogeneity within the sustainability group indicates that some sustainability indices are different from conventional ones, while other sustainability indices are similar to conventional ones. A Kruskal-Wallis, ANOVA, or similar tests are not able to capture these singularities. For that reason, in this subsection, we use a cluster analysis to find out whether there is a sustainability group, that is, whether there is a sustainable inclusion and exclusion process.

The variables used in the cluster analysis are the average of and the variance in the percentile rank of inclusions and exclusions for the ESG score, the ESG combined score, and the market value for the whole period analyzed. For the sake of clarity, we only plot the results of the algorithms that find a group comprised exclusively of sustainability indices. For k-means, agglomerative clustering, and spectral clustering, we plot the first n in detecting a sustainability group because up to n=5 there are no changes inside the sustainability group. As we have noted in the methodology section, we only use the PCA to reduce the dimensionality in order to be able to carry out the plots. Despite applying the PCA, the loss of information is minor because the explained variance ratio of the first principal component (PC1) in each figure is roughly 0.65, and the explained variance ratio of the second principal component (PC2) in each figure is roughly 0.2. Thus, 85% of the variance of our original variables is described by two components: PC1 and PC2.

To test whether there are differences in the inclusion criteria of sustainability and conventional indices, we apply the clustering techniques abovementioned, using the percentile rank information of those companies included in the indices. The results of each algorithm are shown in Table A1.13 of the Appendix and are plotted in Figure 1.2. This figure shows all the algorithms that detect the same sustainability group and provides evidence that the inclusion process of sustainability indices differs from that of conventional indices. Only the FTSE4Good Developed 100 is not in the sustainability group. Therefore, with regard to the inclusion processes, we accept hypothesis 3A for all sustainability indices except for the FTSE4Good Developed 100.

Figure 1.2 Best Clustering algorithms that find differences between the inclusion processes of conventional and sustainability indices
This figure shows those clustering algorithms that group some sustainability indices separately from conventional ones using the information in Table A1.13 of the Appendix. The explained variance ratios of the two principal component analysis are 0.70 (PC1) and 0.23(PC2). (0)FTSE4Good US 100, (1)FTSE4Good US, (2)FTSE4Good Global, (3)FTSE4Good Europe, (4)FTSE4Good Developed 100, (5)STOXX 600, (6)STOXX 50, (7)S&P EURO, (8)FTSE US All caps, (9)FTSE US, (10)FTSE Global 100, (11)FTSE Global, (12)FTSE Eurotop 100, (13)FTSE Eurofirst 100, (14)S&P 500 , (15)S&P 100.



4.4.2 Are the exclusion processes of sustainability and conventional indices different in terms of CSP and size? (Hypothesis 3B)

The results of each clustering technique that we applied to the exclusion processes are shown in Table A1.14 of the Appendix and are plotted in Figure 1.3. All algorithms confirm that the criteria of the FTSE4Good Global and the FTSE4Good US are different from those of the other indices. However, the other three sustainability indices appear clustered with the conventional indices. Only spectral clustering groups together four of five sustainability indices when *n*=4. In Figure 1.3, we plot the k-means (n=3), agglomerative clustering (n =3), spectral clustering (n=4), mean shift, and the affinity propagation. Therefore, we can conclude that in two of the five sustainability indices, there are substantial differences between their exclusion processes and those of conventional indices. Hence, we only accept hypothesis 3B for the FTSE4Good Global and the FTSE4Good US.

Figure 1.3 Best Clustering algorithms that find differences between the exclusion processes of conventional and sustainability indices

This figure shows those clustering algorithms that group some sustainability indices separately from conventional ones using the information in Table A1.14 of the Appendix. The explained variance ratios of the two principal component analysis are 0.667 (PC1) and 0.186(PC2). (0)FTSE4Good US 100, (1)FTSE4Good US, (2)FTSE4Good Global, (3)FTSE4Good Europe, (4)FTSE4Good Developed 100, (5)STOXX 600, (6)STOXX 50, (7)S&P EURO, (8)FTSE US All caps, (9)FTSE US, (10)FTSE Global 100, (11)FTSE Global, (12)FTSE Eurotop 100, (13)FTSE Eurofirst 100, (14)S&P 500 , (15)S&P 100.

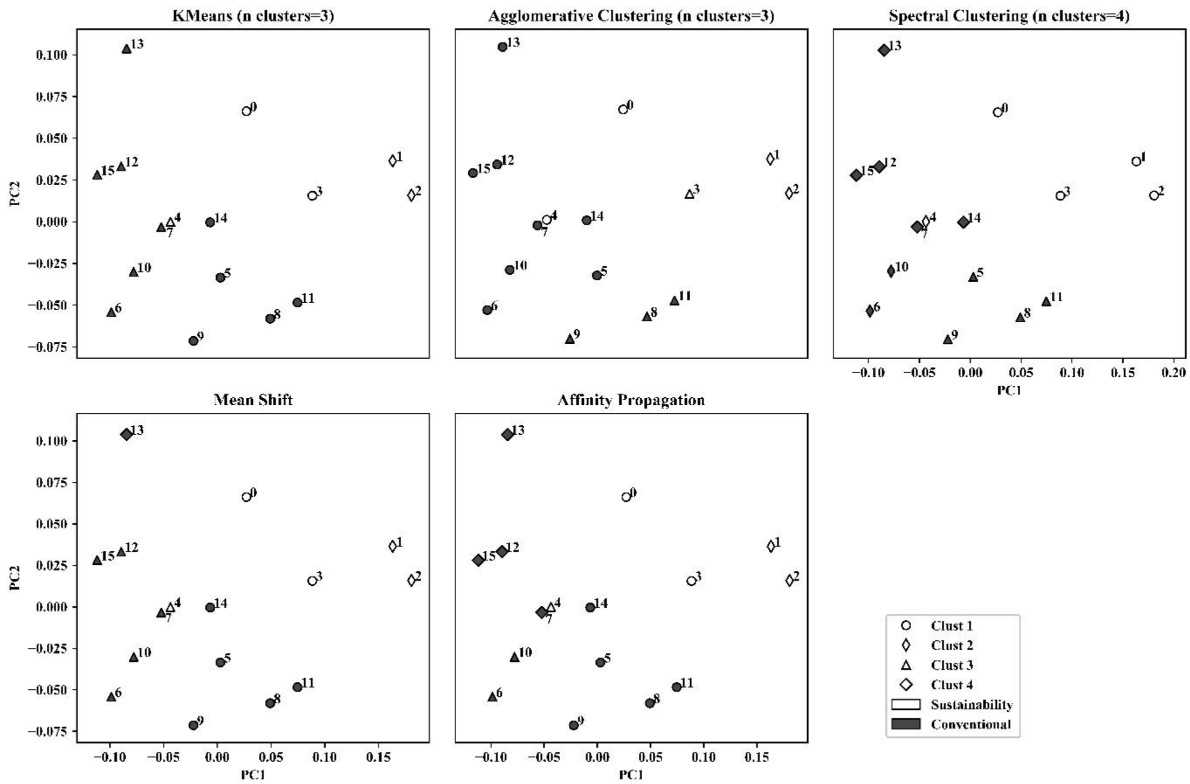
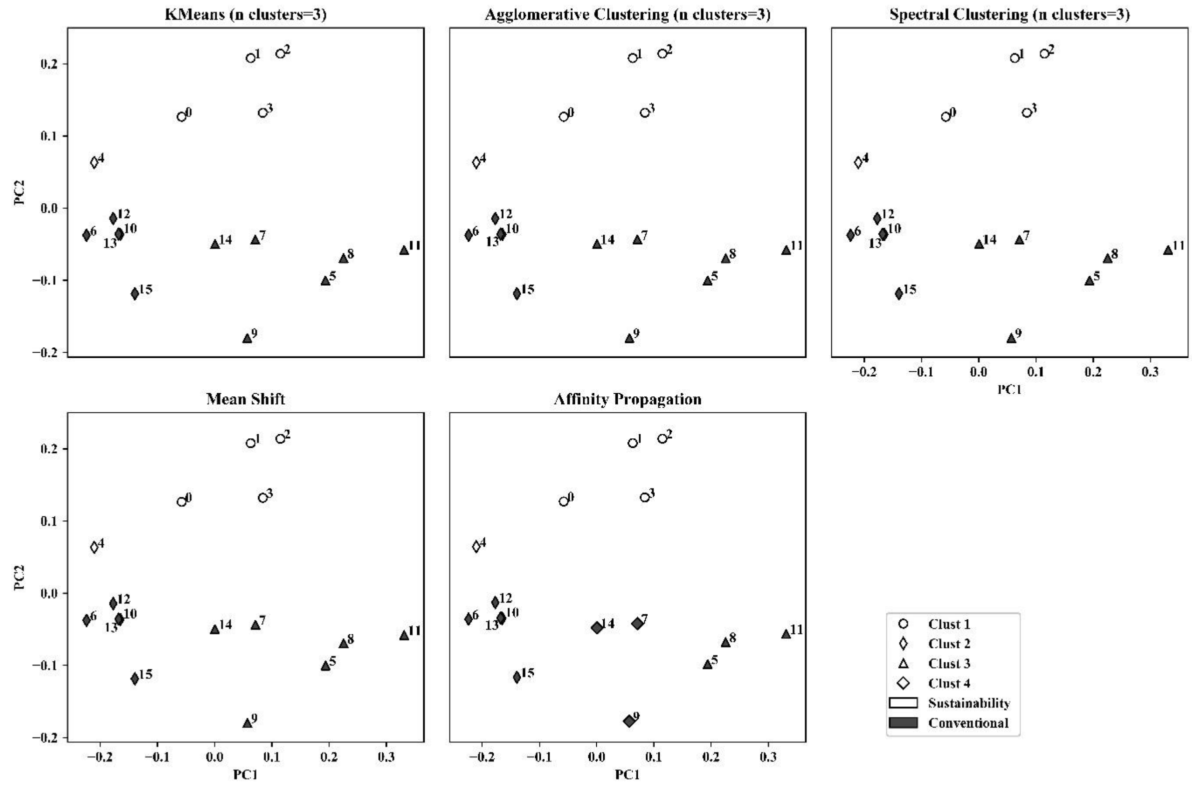


Figure 1.4 Best Clustering algorithms that find differences between the inclussion and exclusion processes of conventional and sustainability indices

This figure shows those clustering algorithms that group some sustainability indices separately from conventional ones using the information in Table A1.15 of the Appendix. The explained variance ratios of the two principal component analysis are 0.596 (PC1) and 0.274(PC2). (0)FTSE4Good US 100, (1)FTSE4Good US, (2)FTSE4Good Global, (3)FTSE4Good Europe, (4)FTSE4Good Developed 100, (5)STOXX 600, (6)STOXX 50, (7)S&P EURO, (8)FTSE US All caps, (9)FTSE US, (10)FTSE Global 100, (11)FTSE Global, (12)FTSE Eurotop 100, (13)FTSE Eurofirst 100, (14)S&P 500 , (15)S&P 100.



4.4.3 Are the inclusion and exclusion processes of sustainability and conventional indices different in terms of CSP and size? (Hypothesis 3C)

Now, we test whether there are differences in the inclusion and exclusion criteria of indices in terms of CSP and size considering both inclusion and exclusion processes. The results of each algorithm are shown in Table A1.15 of the Appendix and are plotted in Figure 1.4. The results show that the mean shift and affinity propagation techniques detect a cluster composed by four sustainability indices. Moreover, the techniques that require the definition of the number of clusters also return the same sustainability group when *n* is higher than two. The FTSE4Good Developed 100 does not appear with the other sustainability indices. However, Figure 1.4 shows that this index is situated between high capitalization indices with few constituents and sustainability indices. Although the FTSE4Good Developed 100 is a sustainability index, our results indicate that the criteria applied in the inclusion and exclusion processes are more similar to those of the FTSE Eurotop 100, FTSE Eurofirst 100, FTSE Global 100, S&P 100, and the STOXX 50 than to the other FTSE4Good indices. Thus, regarding the inclusion and exclusion processes together, we accept hypothesis 3C for all sustainability indices except the FTSE4Good Developed 100.

In short, we can conclude that the criteria applied by four of the five sustainability indices are different enough from the criteria applied by conventional indices to deserve the differentiating label of "4Good".

5. Discussion

In the sustainability index literature, studies that only analyze the inclusions in or exclusions from one or two indices are common. However, the analysis of more indices in order to obtain generalizable conclusions is also important. Moreover, on several occasions, studies do not address whether the results would be similar if they had applied the same analysis to a conventional index. In our research, we attempt to address these problems by analyzing several sustainability and conventional indices.

Our results show that in terms of CSP, the best (worst) companies are not always included (excluded) in sustainability indices. This finding is consistent with Ziegler and Schröder (2010) and Ziegler (2012) who argue that the composition of sustainability indices does not only rely on CSP. In fact, our study is the first to show that the main factor explaining the inclusion or exclusion from the index is the company’s size instead of its CSP. Moreover, the influence of the controversies seems to be minimal. This finding is consistent with Arribas et al. (2019) and Arribas et al. (2021) who find divergent effects on the influence of controversies on the composition of the Dow Jones Sustainability Index (DJSI) World.

Recently, Drempetic et al. (2020) question whether ESG ratings meet the expectations of SR investors due to their correlation with size. This observation also holds true for sustainability indices, given that our research shows that these indices are overly influenced by the company's market capitalization. Thus, we suggest that index providers should reduce the importance of the size criterion in their inclusion and exclusion processes to achieve a stronger differentiation from conventional indices.

According to Petry et al. (2019), the index industry exerts great power in deciding which companies or countries they include or exclude. Therefore, it is important to validate or develop alternative index classifications beyond the labels used by the index providers. To the best of our knowledge, this is the first study that uses a cluster analysis to compare the inclusion and exclusion processes of sustainability and conventional indices in order to validate the "sustainability" label of five FTSE4Good indices. Further research can apply this method to classify the huge number of indices that exist.

6. Conclusions

This work is motivated by the important role that sustainability indices play as a reference for SR investment and by their power in steering capital to their constituents. A growing body of literature uses sustainability indices as a proxy for high sustainability standards and for analyzing the relation between CSP and corporate financial performance. However, this literature does not examine the difference in the criteria applied by sustainability and conventional indices. We also provide original evidence of different selection criteria between both groups of indices by means of five clustering algorithms.

First, we observe a weak influence of company CSP on the exclusions from sustainability indices. In fact, some conventional indices exclude companies with lower CSP than sustainability indices. In addition, in sustainability indices, the size criterion prevails over the CSP criteria when determining which companies leave the index. Second, we observe that the CSP criteria are more relevant to define the inclusions of the companies in sustainability indices than the exclusions. Even though the size criterion still prevails over the CSP criteria, we conclude that sustainability indices are more "size" than "sustainability" indices, specifically when only exclusions are analyzed.

Finally, the cluster analysis shows that four of the five sustainability indices apply different criteria to the inclusion process than conventional indices. However, this is not observed in the exclusion process. When we jointly analyze inclusion and exclusion processes, we find evidence of different criteria applied by four of five sustainability indices as compared to conventional ones.

Our study has two main implications. First, although the cluster analysis shows differences between both groups, the criteria applied by the sustainability indices are excessively influenced by size. This influence should disappear to achieve a real differentiation from the conventional indices. Second, the clustering results indicate that the variables used in this paper are appropriate for classifying indices and can be applied to further research.

References

Abbas, J. (2020). Impact of total quality management on corporate green performance through the mediating role of corporate social responsibility. *Journal of Cleaner Production*, 242. https://doi.org/10.1016/j.jclepro.2019.118458

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster analysis. SAGE Publications, Inc. https://www.doi.org/10.4135/9781412983648

Arribas, I., Espinós-Vañó, M.D., García, F., & Morales-Bañuelos, P.B. (2019). The inclusion of socially irresponsible companies in sustainable stock indices. *Sustainability*, 11(7). https://doi.org/10.3390/su11072047

Arribas, I., Espinós-Vañó, M.D., García, F., & Riley, N. (2021). Do irresponsible corporate activities prevent membership in sustainable stock indices? The case of the dow jones sustainability index world. *Journal of Cleaner Production*, 298. https://doi.org/10.1016/j.jclepro.2021.126711

Artiach, T., Lee, D., Nelson, D., & Walker, J. (2010). The determinants of corporate sustainability performance. *Accounting & Finance*, 50(1), 31-51. https://doi.org/10.1111/j.1467-629X.2009.00315.x

Ashrafi, M., Adams, M., Walker, T.R., & Magnan, G. (2018). 'How corporate social responsibility can be integrated into corporate sustainability: A theoretical review of their relationships'. *International Journal of Sustainable Development & World Ecology*, 25(8), 672-682. https://doi.org/10.1080/13504509.2018.1471628

Auer, B.R. (2018). Green, greener, greenest: Identifying ecological trends and leading entities by means of environmental ratings. *International Review of Applied Economics*, 32(2), 139-162. https://doi.org/10.1080/02692171.2017.1332015

Awaysheh, A., Heron, R.A., Perry, T., & Wilson, J.I. (2020). On the relation between corporate social responsibility and financial performance. *Strategic Management Journal*, 41(6), 965-987. https://doi.org/10.1002/smj.3122

Backer, E., & Jain, A.K. (1981). A clustering performance measure based on fuzzy set decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(1), 66-75. https://doi.org/10.1109/TPAMI.1981.4767051

Bakoben, M., Bellotti, T., & Adams, N. (2020). Identification of credit risk based on cluster analysis of account behaviours. *Journal of the Operational Research Society,* 71(5), 775-783. https://doi.org/10.1080/01605682.2019.1582586

Becchetti, L., Di Giacomo, S., & Pinnacchio, D. (2008). Corporate social responsibility and corporate performance: Evidence from a panel of us listed companies. *Applied Economics,* 40(5), 541-567. https://doi.org/10.1080/00036840500428112

Belghitar, Y., Clark, E., & Deshmukh, N. (2014). Does it pay to be ethical? Evidence from the FTSE4Good. *Journal of Banking & Finance,* 47, 54-62. https://doi.org/10.1016/j.jbankfin.2014.06.027

Borgers, A.C., & Pownall, R.A. (2014). Attitudes towards socially and environmentally responsible investment. *Journal of Behavioral and Experimental Finance*, 1, 27-44. https://doi.org/10.1016/j.jbef.2014.01.005

Carroll, A.B. (1999). Corporate social responsibility: Evolution of a definitional construct. *Business & Society,* 38(3), 268-295. https://doi.org/10.1177/000765039903800303

Chatterji, A.K., & Mitchell, W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the dow jones sustainability index (DJSI). *Strategic Management Journal*, 39(4), 949-976. https://doi.org/10.1002/smj.2752

Cheung, A.W.K. (2011). Do stock investors value corporate sustainability? Evidence from an event study. *Journal of Business Ethics*, 99(2), 145-165. https://doi.org/10.1007/s10551-010-0646-3

Chiappini, H., Vento, G., & De Palma, L. (2021). The impact of Covid-19 lockdowns on sustainable indexes. *Sustainability*, 13(4). https://doi.org/10.3390/su13041846

Consolandi, C., Jaiswal-Dale, A., Poggiani, E., & Vercelli, A. (2009). Global standards and ethical stock indexes: The case of the dow jones sustainability stoxx index. *Journal of Business Ethics*, 87(1), 185-197. https://doi.org/10.1007/s10551-008-9793-1

Cortez, M.C., Silva, F., & Areal, N. (2009). The performance of European socially responsible funds. *Journal of Business Ethics*, 87(4), 573-588. https://doi.org/10.1007/s10551-008-9959-x

Cunha, F.A.F.D.S., de Oliveira, E.M., Orsato, R.J., Klotzle, M.C., Cyrino Oliveira, F.L., & Caiado, R.G.G. (2020). Can sustainable investments outperform traditional benchmarks? Evidence from global stock markets. *Business Strategy and the Environment*, 29(2), 682-697. https://doi.org/10.1002/bse.2397

Curran, M.M., & Moran, D. (2007). Impact of the FTSE4Good index on firm price: An event study. *Journal of Environmental Management*, 82(4), 529-537. https://doi.org/10.1016/j.jenvman.2006.02.010

Doh, J.P., Howton, S.D., Howton, S.W., & Siegel, D.S. (2010). Does the market respond to an endorsement of social responsibility? The role of institutions, information, and legitimacy. *Journal of Management*, 36(6), 1461-1485. https://doi.org/10.1177/0149206309337896

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. *Journal of Business Ethics*, 167(2), 333-360. https://doi.org/10.1007/s10551-019-04164-1

Durand, R., Paugam, L., & Stolowy, H. (2019). Do investors actually value sustainability indices? Replication, development, and new evidence on CSR visibility. *Strategic Management Journal*, 40(9), 1471-1490. https://doi.org/10.1002/smj.3035

El Ouadghiri, I., Guesmi, K., Peillex, J., & Ziegler, A. (2021). Public attention to environmental issues and stock market returns. *Ecological Economics*, 180. https://doi.org/10.1016/j.ecolecon.2020.106836

Escrig-Olmedo, E., Fernández-Izquierdo, Á.M., Ferrero-Ferrero, I., Rivera-Lirio, M.J., & Muñoz-Torres, J.M. (2019). Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles. *Sustainability*, 11(3), 1-16. https://doi.org/10.3390/su11030915

EUROSIF, (2018). European SRI study 2018. Available at: http://www.eurosif.org/wp-content/uploads/2018/11/European-SRI-2018-Study.pdf

Ferrell, A., Liang, H., & Renneboog, L. (2016). Socially responsible firms. *Journal of Financial Economics*, 122(3), 585-606. https://doi.org/10.1016/j.jfineco.2015.12.003

Fichtner, J., Heemskerk, E.M., & Garcia-Bernardo, J. (2017). Hidden power of the big three? Passive index funds, re-concentration of corporate ownership, and new financial risk. *Business and Politics*, 19(2), 298-326. https://doi.org/10.1017/bap.2017.6

Forcadell, F.J., & Aracil, E. (2017). European banks' reputation for corporate social responsibility. *Corporate Social Responsibility and Environmental Management*, 24(1), 14-24. https://doi.org/10.1002/csr.1402

Gallego-Álvarez, I., García-Sánchez, I.M., & da Silva Vieira, C. (2014). Climate change and financial performance in times of crisis. *Business Strategy and the Environment*, 23(6), 361-374. https://doi.org/10.1002/bse.1786

Garriga, E., & Melé, D. (2004). Corporate social responsibility theories: Mapping the territory. *Journal of Business Ethics*, 53(1), 51-71. https://doi.org/10.1023/B:BUSI.0000039399.90587.34

Gómez-Bezares, F., Przychodzen, W., & Przychodzen, J. (2017). Bridging the gap: How sustainable development can help companies create shareholder value and improve financial performance. *Business Ethics: A European Review*, 26(1), 1-17. https://doi.org/10.1111/beer.12135

Gutsche, G., & Ziegler, A. (2019). Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. *Journal of Banking & Finance*, 102, 193-214. https://doi.org/10.1016/j.jbankfin.2019.03.007

Hasan, I., Kobeissi, N., Liu, L., & Wang, H. (2018). Corporate social responsibility and firm financial performance: The mediating role of productivity. *Journal of Business Ethics*, 149(3), 671-688. https://doi.org/10.1007/s10551-016-3066-1

Hong, H., Kubik, J.D., & Scheinkman, J.A. (2012). Financial constraints on corporate goodness. *National Bureau of Economic Research Working Paper Series*, (6). https://doi.org/10.3386/w18476

Hopwood, B., Mellor, M., & O'Brien, G. (2005). Sustainable development: Mapping different approaches. *Sustainable Development*, 13(1), 38-52. https://doi.org/10.1002/sd.244

Hörisch, J., Johnson, M.P., & Schaltegger, S. (2015). Implementation of sustainability management and company size: A knowledge-based view. *Business Strategy and the Environment*, 24(8), 765-779. https://doi.org/10.1002/bse.1844

Jain, A.K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

Jiang, S., Ferreira, J., & González, M.C. (2012). Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, 25(3), 478-510. https://doi.org/10.1007/s10618-012-0264-z

Joliet, R., & Titova, Y. (2018). Equity SRI funds vacillate between ethics and money: An analysis of the funds’ stock holding decisions. *Journal of Banking & Finance*, 97, 70-86. https://doi.org/10.1016/j.jbankfin.2018.09.011

Kappou, K., & Oikonomou, I. (2016). Is there a gold social seal? The financial effects of additions to and deletions from social stock indices. *Journal of Business Ethics*, 133(3), 533-552. https://doi.org/10.1007/s10551-014-2409-z

Kaspereit, T., & Lopatta, K. (2016). The value relevance of SAM's corporate sustainability ranking and GRI sustainability reporting in the European stock markets. *Business Ethics: A European Review*, 25(1), 1-24. https://doi.org/10.1111/beer.12079

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics*, 115(2), 304-329. https://doi.org/10.1016/j.jfineco.2014.09.008

Leite, P., & Cortez, M.C. (2014). Style and performance of international socially responsible funds in Europe. *Research in International Business and Finance*, 30, 248-267. https://doi.org/10.1016/j.ribaf.2013.09.007

Liang, H., & Renneboog, L. (2017). On the foundations of corporate social responsibility. *The Journal of Finance*, 72(2), 853-910. https://doi.org/10.1111/jofi.12487

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, 21(5), 603-609. https://doi.org/10.1002/(SICI)1097-0266(200005)21:5<603::AID-SMJ101>3.CO;2-3

Milligan, G.W., & Cooper, M.C. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification*, 5(2), 181-204. https://doi.org/10.1007/BF01897163

Minutolo, M.C., Kristjanpoller, W.D., & Stakeley, J. (2019). Exploring environmental, social, and governance disclosure effects on the S&P 500 financial performance. *Business Strategy and the Environment*, 28(6), 1083-1095. https://doi.org/10.1002/bse.2303

Miras-Rodríguez, M.D.M., Carrasco-Gallego, A., & Escobar-Pérez, B. (2015). Has the CSR engagement of electrical companies had an effect on their performance? A closer look at the environment. *Business Strategy and the Environment*, 24(8), 819-835. https://doi.org/10.1002/bse.1848

Montiel, I. (2008). Corporate social responsibility and corporate sustainability: Separate pasts, common futures. *Organization & Environment*, 21(3), 245-269. https://doi.org/10.1177/1086026608321329

Montoya-Cruz, E., Ramos-Requena, J.P., Trinidad-Segovia, J.E., & Sánchez-Granero, M.Á. (2020). Exploring arbitrage strategies in corporate social responsibility companies. *Sustainability*, 12(16). https://doi.org/10.3390/su12166293

Nuber, C., Velte, P., & Hörisch, J. (2020). The curvilinear and time-lagging impact of sustainability performance on financial performance: Evidence from Germany. *Corporate Social Responsibility and Environmental Management*, 27(1), 232-243. https://doi.org/10.1002/csr.1795

Oberndorfer, U., Schmidt, P., Wagner, M., & Ziegler, A. (2013). Does the stock market value the inclusion in a sustainability stock index? An event study analysis for German firms. *Journal of Environmental Economics and Management*, 66(3), 497-509. https://doi.org/10.1016/j.jeem.2013.04.005

Orlitzky, M. (2001). Does firm size comfound the relationship between corporate social performance and firm financial performance? *Journal of Business Ethics*, 33(2), 167-180. https://doi.org/10.1023/A:1017516826427

Ortas, E., Álvarez, I., Jaussaud, J., & Garayar, A. (2015). The impact of institutional and social context on corporate environmental, social and governance performance of companies committed to voluntary corporate social responsibility initiatives. *Journal of Cleaner Production*, 108, 673-684. https://doi.org/10.1016/j.jclepro.2015.06.089

Peters, G., Crespo, F., Lingras, P., & Weber, R. (2013). Soft clustering – fuzzy and rough approaches and their extensions and derivatives. *International Journal of Approximate Reasoning*, 54(2), 307-322. https://doi.org/10.1016/j.ijar.2012.10.003

Petry, J., Fichtner, J., & Heemskerk, E. (2019). Steering capital: The growing private authority of index providers in the age of passive asset management. *Review of International Political Economy*, (2), 1-25. https://doi.org/10.1080/09692290.2019.1699147

Pineiro-Chousa, J., Romero-Castro, N., & Vizcaíno-González, M. (2019). Inclusions in and exclusions from the S&P 500 environmental and socially responsible index: A fuzzy-set qualitative comparative analysis. *Sustainability*, 11(4). https://doi.org/10.3390/su11041211

Rajesh, R., & Rajendran, C. (2020). Relating environmental, social, and governance scores and sustainability performances of firms: An empirical analysis. *Business Strategy and the Environment*, 29(3), 1247-1267. https://doi.org/10.1002/bse.2429

Refinitiv, (2019). Environmental, Social and Governance (ESG) Scores from Refinitiv.

Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, 32(9), 1723-1742. https://doi.org/10.1016/j.jbankfin.2007.12.039

Saadaoui, K., & Soobaroyen, T. (2018). An analysis of the methodologies adopted by CSR rating agencies. *Sustainability Accounting, Management and Policy Journal*, 9(1), 43-62. https://doi.org/10.1108/SAMPJ-06-2016-0031

Schröder, M. (2007). Is there a difference? The performance characteristics of SRI equity indices. *Journal of Business Finance & Accounting*, 34(2), 331-348. https://doi.org/10.1111/j.1468-5957.2006.00647.x

Soytas, M.A., Denizel, M., & Durak Usar, D. (2019). Addressing endogeneity in the causal relationship between sustainability and financial performance. *International Journal of Production Economics*, 210, 56-71. https://doi.org/10.1016/j.ijpe.2019.01.016

Sushko, V., & Turner, G. (2018). The implications of passive investing for securities markets. *BIS Quarterly Review*, March, 113-131. https://ssrn.com/abstract=3139242

Tabor, I., & Molinas, C. (2020, June). Active and Passive Investment: Their Coexistence in Portfolio Management. Instituto Español de Analistas Financieros.

Udayasankar, K. (2008). Corporate social responsibility and firm size. *Journal of Business Ethics*, 83(2), 167-175. https://doi.org/10.1007/s10551-007-9609-8

US SIF, (2018). Report on US sustainable, responsible and impact investing Trends 2018. Available at: https://www.ussif.org/files/Trends/Trends%202018%20executive%20summary%20FINAL.pdf

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., … and Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Waddock, S.A., & Graves, S.B. (1997). The corporate social performance-financial performance link. *Strategic Management Journal*, 18(4), 303-319. https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.CO;2-G

Walker, O. (2019, May 20). Index companies to feel the chill of fund managers' price war. Financial Times. https://www.ft.com/content/e886b2d2-e852-3071-85c1-c9a57113d8a5

Xu, R., & Wunsch, D. (2008). Clustering (Vol. 10). John Wiley & Sons.

Yen, M., Shiu, Y., & Wang, C. (2019). Socially responsible investment returns and news: Evidence from Asia. *Corporate Social Responsibility and Environmental Management*, 26(6), 1565-1578. https://doi.org/10.1002/csr.1833

Zhao, X., & Murrell, A.J. (2016). Revisiting the corporate social performance-financial performance link: A replication of Waddock and Graves. *Strategic Management Journal*, 37(11), 2378-2388. https://doi.org/10.1002/smj.2579

Ziegler, A. (2012). Is it beneficial to be included in a sustainability stock index? A panel data study for European firms. *Environmental and Resource Economics*, 52(3), 301-325. https://doi.org/10.1007/s10640-011-9529-z

Ziegler, A., & Schröder, M. (2010). What determines the inclusion in a sustainability stock index? A panel data analysis for European firms. *Ecological Economics*, 69(4), 848-856. https://doi.org/10.1016/j.ecolecon.2009.10.009

Appendix

Table A1.9 Number of firms analyzed each year by geographic area
This table shows the firms analyzed each year by geographic area. Column 1 reports countries and geographic areas while the following columns show the number of companies under analysis for each year. The last row shows the total number of companies under analysis for each year.

Geographical area	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Africa	8	20	50	67	149	153	149	147	142	140	136
America (ex Canada & United States)	33	67	95	152	162	163	168	174	186	276	269
Asia (ex Japan)	157	276	434	735	779	803	834	901	963	987	1,178
Canada	166	231	241	263	254	253	270	278	287	296	284
Europe	861	917	952	992	1,009	1,008	1,013	1,066	1,150	1,179	1,172
Japan	397	408	415	412	408	406	412	422	430	433	429
Oceania	97	183	267	306	320	351	366	412	432	446	436
United States	665	873	936	958	953	942	941	1,009	1,718	2,310	2,329
not available	20	22	23	23	21	20	18	17	18	25	25
Total	2,404	2,997	3,413	3,908	4,055	4,099	4,171	4,426	5,326	6,092	6,258

***Table A1.10* Descriptive statistics of ESG scores and market value by geographic area**

This table shows descriptive statistics by geographic area of the analyzed variables. Column 1 reports the different variables; column 2 reports different countries and geographic areas, and the following columns list the number of observations and some descriptive statistics including the average, standard deviation, minimum, quartiles and maximum value for each variable.

Variable	Geo. area	count	mean	std	min	25%	50%	75%	max
ESG Score	Africa	13,110	52.35	16.12	7.31	41.83	52.97	63.99	93.46
	America (ex Canada & United States)	20,721	48.05	17.50	7.55	34.57	48.24	61.29	93.39
	Asia (ex Japan)	95,212	46.51	18.17	8.06	32.07	45.38	60.79	94.96
	Canada	33,469	46.77	16.50	11.87	33.97	44.35	58.51	93.07
	Europe	133,931	56.67	16.89	7.68	44.23	57.56	69.64	95.79
	Japan	53,615	50.74	18.64	4.80	35.53	52.51	65.81	93.56
	Oceania	40,123	46.94	17.15	9.65	33.88	44.86	57.89	95.84
	United States	159,007	48.69	17.51	9.64	34.87	45.46	62.04	97.90
	not available	2,628	52.97	16.66	13.99	41.48	52.17	65.53	89.22
	All Sample	551,816	50.29	17.89	4.80	36.17	49.42	64.27	97.90
ESG Combined Score	Africa	13,110	48.75	16.05	7.31	37.19	48.31	60.59	89.81
	America (ex Canada & United States)	20,721	45.28	16.97	7.55	32.52	43.74	57.69	93.39
	Asia (ex Japan)	95,212	43.37	17.28	7.43	30.16	40.95	55.80	94.96
	Canada	33,469	43.63	14.92	11.87	32.70	41.24	52.77	89.54
	Europe	133,931	50.32	16.15	7.68	38.38	48.55	62.67	93.19
	Japan	53,615	47.56	17.60	4.80	33.58	46.42	61.64	92.73
	Oceania	40,123	43.57	15.55	9.65	32.09	41.76	52.69	95.84
	United States	159,007	42.37	14.80	9.64	31.90	39.76	50.21	95.59
	not available	2,628	47.69	15.85	13.99	36.20	46.36	58.97	89.22
	All Sample	551,816	45.43	16.36	4.80	33.27	43.11	56.74	95.84
Market Value($)	Africa	13,110	4,119	6,982	1.58	558	1,629	4,456	112,518
	America (ex Canada & United States)	20,721	7,299	12,271	3.35	1,679	3,674	7,980	181,504
	Asia (ex Japan)	95,212	9,517	20,867	0.20	2,092	4,382	9,285	686,341
	Canada	33,469	5,570	11,221	0.02	754	1,618	4,533	115,428
	Europe	133,931	10,903	21,996	1.85	1,542	3,697	9,634	355,280
	Japan	53,615	8,450	14,254	109.10	2,305	4,065	8,487	241,229
	Oceania	40,123	3,676	11,053	0.19	249	724	2,341	162,967
	United States	159,007	13,278	34,156	2.89	1,747	3,873	10,331	878,224
	not available	2,628	8,750	16,191	0.08	548	2,245	7,127	101,985
	All Sample	551,816	9,954	24,121	0.02	1,443	3,436	8,572	878,224

Table A1.11 Descriptive statistics on index compositions and their components

This table shows descriptive statistics for five different years for the examined benchmarks. Specifically, the table reports three statistics for each year: the monthly average number of constituents in the index (N), the monthly average number of constituents covered by our sample (NC) and the monthly average of the free float weight covered by our sample (WC).

	2008			2010			2012			2014			2016		
	N	**NC**	**W.C**	**N**	**NC**	**W.C**	**N**	**NC**	**W.C**	**N**	**NC**	**W.C**	**N**	**NC**	**W.C**
FTSE4Good Global	695	593	93%	658	584	94%	728	650	94%	757	680	93%	809	726	93%
FTSE4Good Developed 100	104	94	94%	103	95	94%	103	95	94%	104	95	93%	105	95	93%
FTSE4Good US	147	131	96%	131	124	98%	155	142	96%	167	151	93%	183	165	92%
FTSE4Good US 100	101	93	96%	102	97	98%	102	94	96%	103	94	93%	103	92	92%
FTSE4Good Europe	287	241	88%	273	238	89%	290	258	90%	297	268	91%	336	302	91%
FTSE Global	7,950	2,633	82%	7,354	3,376	87%	7,359	3,570	87%	7,483	3,727	87%	7,766	4,545	90%
FTSE Global 100	104	95	95%	95	88	87%	103	94	94%	104	95	94%	107	96	93%
FTSE US	668	574	93%	596	546	96%	615	568	95%	640	580	93%	633	576	93%
FTSE US All caps	2,323	862	86%	2,130	967	89%	1,975	958	89%	1,972	983	88%	1,964	1,735	93%
FTSE Eurofirst 100	99	90	85%	100	93	86%	100	93	85%	100	94	89%	100	93	88%
FTSE Eurotop 100	107	92	91%	105	95	91%	107	94	90%	106	94	91%	108	95	92%
S&P 500	500	454	N/A	500	469	N/A	500	471	N/A	501	467	N/A	505	470	N/A
S&P 100	100	92	N/A	100	96	N/A	100	95	N/A	101	96	N/A	102	95	N/A
S&P EURO	178	170	N/A	180	175	N/A	177	169	N/A	177	171	N/A	185	173	N/A
STOXX 50	50	45	92%	50	47	94%	50	47	95%	50	47	96%	50	48	96%
STOXX 600	600	537	92%	600	554	93%	600	559	92%	600	563	94%	600	559	93%

Table A1.12 Correlation matrix between the variables used in the research

This table shows the Pearson coefficient between the variables in the first column and the first row. The * and ** indicate statistical significance at the 5%, 1%, levels, respectively.

	ESG Score	Combined Score	Market Value USD	P. Rank ESG Score	P. Rank Com. Score	P. Rank Market Value	ROA	Liabilities to Assets	Capital expenditures
ESG Score	1.000**								
Combined Score	0.802**	1.000**							
Market Value USD	0.332**	0.066**	1.000**						
Rankpct ESG Score	0.994**	0.808**	0.311**	1.000**					
Rankpct Com. Score	0.797**	0.981**	0.069**	0.813**	1.000**				
Rankpct Market Value	0.477**	0.268**	0.493**	0.471**	0.260**	1.000**			
ROA	0.047**	0.042**	0.076**	0.049**	0.046**	0.169**	1.000**		
Liabilities to Assets	0.175**	0.105**	0.079**	0.174**	0.105**	0.187**	-0.106**	1.000**	
Capital expenditures	-0.052**	-0.040**	-0.025**	-0.051**	-0.037**	-0.087**	0.008**	-0.173**	1.000**

Table A1.13 **Clustering results for the inclusion processes of indices**

The first column reports the indices analyzed while the second and third columns report values of the principal component analysis of the variables used. The following columns show the results of the clustering algorithms: k-means (KM), agglomerative clustering (AG) and spectral clustering (SC) from n=2 to 5; and mean shift (MS) and affinity propagation (AP).

	n clusters		**n=2**			**n=3**			**n=4**			**n=5**				
	algorithm		**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**MS**	**AP**
Index Name	**PCA1**	**PCA2**														
FTSE4Good Global	0.04	-0.16	0	0	0	2	2	2	2	2	0	1	2	1	1	0
FTSE4Good Developed 100	-0.21	-0.04	1	0	1	1	1	1	1	3	1	2	1	0	0	1
FTSE4Good US	-0.00	-0.14	0	0	0	2	2	2	2	2	0	1	2	1	1	0
FTSE4Good US 100	-0.07	-0.10	1	0	1	2	2	2	2	2	0	1	2	1	1	0
FTSE4Good Europe	0.04	-0.13	0	0	0	2	2	2	2	2	0	1	2	1	1	0
FTSE Global	0.33	0.01	0	1	0	0	0	0	0	1	3	3	0	3	3	2
FTSE Global 100	-0.14	0.04	1	0	1	1	1	1	1	3	1	2	1	0	0	1
FTSE US	0.07	0.16	0	1	0	0	0	0	3	0	2	0	4	2	2	3
FTSE US All caps	0.22	0.04	0	1	0	0	0	0	0	1	3	3	0	3	3	2
FTSE Eurofirst 100	-0.14	0.05	1	0	1	1	1	1	1	3	1	2	1	0	0	1
FTSE Eurotop 100	-0.15	0.01	1	0	1	1	1	1	1	3	1	2	1	0	0	1
S&P 500	0.01	0.06	0	1	0	2	0	0	3	0	2	4	3	4	2	3
S&P 100	-0.09	0.11	1	0	1	1	1	1	1	3	1	2	1	0	0	1
S&P EURO	0.10	-0.01	0	1	0	2	0	0	3	0	2	4	3	2	2	3
STOXX 50	-0.20	0.04	1	0	1	1	1	1	1	3	1	2	1	0	0	1
STOXX 600	0.21	0.05	0	1	0	0	0	0	0	1	3	3	0	3	3	2

Table A1.14 Clustering results for the exclusion processes of indices

The first column reports the indices analyzed while the second and third columns report values of the principal component analysis of the variables used. The following columns show the results of the clustering algorithms: k-means (KM), agglomerative clustering (AG) and spectral clustering (SC) from n=2 to 5; and mean shift (MS) and affinity propagation (AP).

	n clusters		**n=2**			**n=3**			**n=4**			**n=5**				
	algorithm		**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**MS**	**AP**
Index Name	**PCA1**	**PCA2**														
FTSE4Good Global	0.18	0.02	1	0	1	0	2	2	1	2	1	1	2	0	2	0
FTSE4Good Developed 100	-0.04	-0.00	0	1	0	2	0	1	3	0	0	2	1	2	0	1
FTSE4Good US	0.16	0.04	1	0	1	0	2	2	1	2	1	1	2	0	2	0
FTSE4Good US 100	0.03	0.07	1	1	1	1	0	0	2	0	1	3	1	0	1	3
FTSE4Good Europe	0.09	0.02	1	0	1	1	1	2	2	1	1	3	0	0	1	3
FTSE Global	0.07	-0.05	1	0	1	1	1	1	2	1	2	3	0	3	1	3
FTSE Global 100	-0.08	-0.03	0	1	0	2	0	1	3	0	0	2	4	2	0	1
FTSE US	-0.02	-0.07	0	0	0	1	1	1	2	1	2	4	0	3	1	3
FTSE US All caps	0.05	-0.06	1	0	1	1	1	1	2	1	2	3	0	3	1	3
FTSE Eurofirst 100	-0.08	0.10	0	1	0	2	0	0	0	3	3	0	3	1	3	2
FTSE Eurotop 100	-0.09	0.03	0	1	0	2	0	0	0	3	3	0	3	1	0	2
S&P 500	-0.01	-0.00	0	1	0	1	0	1	2	0	3	3	1	4	1	3
S&P 100	-0.11	0.03	0	1	0	2	0	0	0	3	3	0	3	1	0	2
S&P EURO	-0.05	-0.00	0	1	0	2	0	1	3	0	3	2	1	1	0	2
STOXX 50	-0.10	-0.05	0	1	0	2	0	1	3	0	0	2	4	2	0	1
STOXX 600	0.00	-0.03	0	1	0	1	0	1	2	0	2	3	1	3	1	3

Table A1.15 Clustering results for the inclusion and exclusion processes of indices
The first column reports the indices analyzed while the second and third columns report values of the principal component analysis of the variables used. The following columns show the results of the clustering algorithms: k-means (KM), agglomerative clustering (AG) and spectral clustering (SC) from n=2 to 5; and mean shift (MS) and affinity propagation (AP).

	n clusters		**n=2**			**n=3**			**n=4**			**n=5**				
	algorithm		**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**KM**	**AG**	**SC**	**MS**	**AP**
Index Name	**PCA1**	**PCA2**														
FTSE4Good Global	0.11	0.21	0	0	0	1	2	2	2	2	0	2	0	1	2	0
FTSE4Good Developed 100	-0.21	0.06	1	1	1	2	1	1	1	0	3	3	4	2	0	2
FTSE4Good US	0.06	0.21	0	0	0	1	2	2	2	2	0	2	0	1	2	0
FTSE4Good US 100	-0.06	0.13	1	0	1	1	2	2	2	2	0	2	0	1	2	0
FTSE4Good Europe	0.08	0.13	0	0	0	1	2	2	2	2	0	2	0	1	2	0
FTSE Global	0.33	-0.06	0	0	0	0	0	0	0	3	2	1	3	0	1	1
FTSE Global 100	-0.16	-0.04	1	1	1	2	1	1	1	0	3	3	4	2	0	2
FTSE US	0.06	-0.18	0	0	0	0	0	0	3	1	1	4	1	4	1	3
FTSE US All caps	0.23	-0.07	0	0	0	0	0	0	0	3	2	1	3	0	1	1
FTSE Eurofirst 100	-0.17	-0.04	1	1	1	2	1	1	1	0	3	0	2	3	0	2
FTSE Eurotop 100	-0.18	-0.01	1	1	1	2	1	1	1	0	3	0	4	3	0	2
S&P 500	0.00	-0.05	0	0	0	0	0	0	3	1	1	4	1	4	1	3
S&P 100	-0.14	-0.12	1	1	1	2	1	1	1	0	3	0	2	3	0	2
S&P EURO	0.07	-0.04	0	0	0	0	0	0	3	1	1	4	1	4	1	3
STOXX 50	-0.22	-0.04	1	1	1	2	1	1	1	0	3	3	4	2	0	2
STOXX 600	0.19	-0.10	0	0	0	0	0	0	0	3	2	1	3	0	1	1

Chapter 2.

In search of inclusive ESG ratings

Synopsis

Studies have used ESG ratings to measure CSR; however, these ratings provide an overall measure that does not consider the capabilities of companies to fulfil social and environmental standards. We propose a cross-sectional regression to estimate inclusive ESG ratings in order to assess companies based on their capabilities. This regression identifies the virtuous behavior of companies, that is, the ESG excess relative to other companies of similar size, country, and industry. We use both the conventional and the inclusive ESG ratings to analyze the relationship between CSR and the financial performance that is measured by accounting and market variables. Our results show that inclusive ratings have a negative influence on financial performance, while the conventional ratings, as in other studies, show mixed results. The inclusive ratings could help socially responsible investors maximize their welfare by identifying virtuous companies.

1. Introduction

Does corporate social responsibility (CSR) lead to superior corporate financial performance (CFP)? Studies have tried to answer this question for decades. However, their empirical evidence is mixed (Margolis et al., 2007; Wang et al., 2016), and their approaches face two main problems: the endogeneity between CSR and CFP (Flammer, 2015; Soytas et al., 2019) and the suitability of environmental social and governance (ESG) ratings for measuring CSR (Berrone et al., 2017; Chatterji et al., 2016; Lyon and Montgomery, 2015). The aim of this research is to propose a method that solves the endogeneity problem and allows the derivation of an inclusive score that considers firm-specific capabilities to meet environmental and social standards. Thus, using the conventional and the inclusive ESG ratings, we shed light on the CSR-CFP relationship.

The development of regression-based models is common in the financial literature. Some examples are the Capital Asset Pricing Model and the Fama-French models in portfolio management studies (Fama and French, 2015, 1993; Sharpe, 1964), or the Jones model and the modified Jones model in earnings management studies (Dechow et al., 1995; Jones, 1991). In the first group, the intercept of the model shows the manager's ability to generate value; and in the second group, the error term of the regression shows the company's discretionary accrual. Both group of models are widely used, and both focus on a single dimension: the same company in different periods (portfolio management), or different companies in the same period (earnings management).

Recent literature on CSR applies a similar approach, where the firm's optimal CSR is the estimated value of a regression (Lys et al., 2015) or where the error term of a regression indicates the abnormal CSR (Naughton et al., 2019). However, these studies merge different dimensions (individual and time) and lack a solid theoretical basis. This research provides a theoretical background for a cross-sectional regression model in which we consider the error term as a proxy of the firm's virtuous behavior.

The literature has suggested that the level of CSR in companies depends on several factors with size being the most cited (Aguinis and Glavas, 2012; Ali et al., 2017; McWilliams and Siegel, 2001; Udayasankar, 2008). Recently, Drempetic et al. (2020) have criticized the method used by ESG raters to score companies because they give an advantage to large ones without providing useful information to socially responsible (SR) investors to make the correct decisions according to their beliefs. This statement can also be applied to other determinants of CSR, such as the company's country origin (Demirbag et al., 2017; Dyck et al., 2019). It is easier for large companies to obtain higher ESG scores than for small companies. Similarly, it is easier for companies in developed countries to obtain higher ESG scores because the environmental and social standards tend to be stricter than in emerging countries. However, a higher score of large companies from developed countries does not mean superior virtuous behavior. Our methodology addresses these biases by estimating inclusive ESG ratings.

Inclusive ratings are valuable to society in general and for SR investors in particular. Concepts such as inclusive education (UNESCO, 2013) or inclusive growth (OECD, 2014) are becoming increasingly important. Currently, society seeks to not exclude anyone on the basis of gender, race, economic situation, or cognitive ability. However, current ESG ratings have some biases that may unfairly exclude some companies from SR investment. Unless we adopt an inclusive framework, the definition of best-in-class provided by EUROSIF (2018) could be rephrased as the strategy that allows investors to pick large companies of developed countries in a particular industrial sector. We adopt an inclusive framework to study the CSR-CFP link from a different perspective than previous literature.

Despite the large number of studies analyzing the relationship between CSR and CFP, there is not a consensus on its existence, and if existence is the case, whether it is positive or negative (Nuber et al., 2020; Zhao and Murrell, 2016). The empirical studies usually measure CSR with ESG ratings, that is, conventional ratings (Odziemkowska and Henisz, 2020). However, we measure CSR as the company's ESG excess relative to its size, country, and industry, that is, inclusive ratings. Therefore, our study contributes to the literature by analyzing the link between a company's virtuous behavior and CFP. For this purpose, we estimate the inclusive rating for each ESG score for all companies in the Refinitiv ESG rating database from 2010 to 2019. Hence, the study is free of survivorship bias.

Our results show that the company's virtuous behavior measured by the inclusive ratings leads to lower CFP regardless of the use of accounting or market measures. However, we find inconclusive results when we perform the same analysis by measuring CSR with conventional ESG ratings. The conventional ratings show a positive, negative, or neutral influence on the measures of accounting performance depending on the control variables, the fixed effects and the proxies for CFP considered. This variability in the results does not arise with our inclusive ESG ratings which always show a negative influence on the accounting performance indicators. Similarly, the difference in the return of high and low portfolios created on the basis of conventional ratings does not show significant abnormal returns, while for the inclusive ratings we find statistically significant negative abnormal returns. These findings are not robust for the governance pillar. However, this is in line with the literature because this pillar overlaps with traditional corporate governance issues, which are materially different from CSR (Liang and Renneboog, 2017). Therefore, our study provides a methodology for estimating an indicator of CSR that is not influenced by company size, country and industry.

Advocates of CSR usually argue that "doing good leads to doing well", but if this statement were true, what rational company would not do good? We show that "doing good"—being virtuous—has a cost. Therefore, virtuous companies should be rewarded by SR flows, because SR investors are the ones who are willing to sacrifice returns to invest in companies that meet high environmental and social standards (Gutsche and Ziegler, 2019; Hart and Zingales, 2017; Pástor et al., 2021; Renneboog et al., 2008). However, SR investors cannot notice many of these companies because ESG providers only capture a global measure of CSR that does not consider the power of the company to fulfil social and environmental standards. Therefore, our methodology can improve the decision-making of SR investors by providing comparable ESG indicators for companies with different sizes, from different countries, and in different industries.

The paper is organized as follows: In Section 2, we review the determinants of the CSR and introduce the hypotheses, Section 3 presents the data, while Section 4 presents the method. In Section 5, we explain the empirical results, and Section 6 concludes.

2. Determinants of CSR and hypotheses

McWilliams and Siegel (2001) hypothesize that the firms' level of CSR depends on its size and its levels of diversification, research and development, advertising, and government sales as well as consumer income, labor market conditions, and its stage in the industry life cycle. Other studies indicate that regulations, strategic policies, and the legal origin of countries are also important determinants of CSR (Demirbag et al., 2017; Di Giuli and Kostovetsky, 2014; Liang and Renneboog, 2017). In addition, the literature tends to present an instrumental conception of CSR: a tool for marketing strategy (Mishra and Modi, 2016; Varadarajan and Menon, 1988), an instrument to improve corporate reputation (Brammer and Pavelin, 2006; Edmans, 2012), or a method to create a competitive advantage (Hart, 1995; Porter and Kramer, 2006; Seo et al., 2021).

The instrumental view of CSR, as well as the determinants of CSR, make it difficult for ESG ratings to fulfill their role, that is, to provide SR investors accurate information about the firm's responsible behavior (Chatterji et al., 2009). In order to achieve a more accurate measurement of CSR, companies must be rated from an inclusive perspective that considers their capabilities to meet environmental and social standards. This approach would soften the criticism of greenwashing and social-washing (Basu et al., 2022; Flammer, 2021).

We consider country, industry, and size to evaluate the companies from an inclusive perspective since these variables are often used to segment the sample in studies that address the CSR-CFP relationship. Awaysheh et al. (2020) analyze the link between CSR and CFP by benchmarking companies against industry peers in each year to identify the best-in- and worst-in-class. Similarly, Badía et al., (2020) analyze this relationship by creating high- and low-rated ESG portfolios for different geographical areas and Minutolo et al. (2019) carry out the analysis by segmenting their sample into size quartiles.

The descriptive statistics from empirical studies show substantial differences in ESG ratings across countries, and to a lesser extent, across industries (Awaysheh et al., 2020; Capelle-Blancard and Petit, 2015; Dyck et al., 2019; Ferrell et al., 2016). We assume that this heterogeneity among countries is due to some determinants of CSR, such as the labor market's conditions or the country's regulations concerning ESG issues. Similarly, public scrutiny or marketing intensity are more similar inside the same industry. Therefore, to obtain accurate measures of CSR, the elimination of the influence of structural factors on ESG ratings, such as country or industry, is important.

The literature has also established that large companies engage in more CSR activities than small ones. First, large companies have more (slack) resources to deal with sustainability issues than small companies (Hörisch et al., 2015). Second, large companies are more visible, and therefore, they are more likely to be more responsible (Udayasankar, 2008). This visibility puts greater pressure on them to invest more in environmentally friendly technologies and to adhere to an appropriate level of CSR (Chiu and Sharfman, 2009; Etzion, 2007). Third, CSR activities lead to fixed costs that are less important to large companies (Ziegler and Schröder, 2010). However, we underline that the higher levels of CSR in large companies are due to their size and not because they are more virtuous than small companies.

The influence of size on CSR led Orlitzky (2001) to analyze whether it was the real determinant of the relation between CSR and CFP, but he concluded that the covariation was only partially explained by the size factor. The typical approach in the CSR-CFP literature is to regress accounting based measures of CFP on measures of CSR, or to analyze whether portfolios of securities that are based on ESG standards outperform conventional ones (Barauskaite and Streimikiene, 2021).

The literature on the risk-adjusted returns of sustainability indices show mixed results depending on the index or geographic area analyzed (Cunha et al., 2020; Ziegler and Schröder, 2010). Similarly, studies on whether high-rated portfolios based on ESG ratings outperform low-rated portfolios (Badía et al., 2020; Halbritter and Dorfleitner, 2015) or whether SR funds outperform conventional ones (Bauer et al., 2005; Hong and Kacperczyk, 2009) also show inconclusive results. The conclusions of event studies also differ. Hawn et al. (2018) show that investors punish firms that are added to the Dow Jones Sustainability Index, while Flammer (2021) concludes that the market responds positively to the announcement of green bonds issues. Krüger (2015) shows that the market responds strongly negatively to negative CSR-related events and weakly to positive ones. Recently, Hwang et al. (2021) also show that the revelation of higher SR ownership is associated with a negative return.

Waddock and Graves (1997) focus on accounting based measures rather than market ones and conclude that CSR leads to superior financial performance. However, McWilliams and Siegel (2000) suggest that their correlation was a misspecification. More recently, Zhao and Murrell (2016) replicate the Waddock and Graves' study and conclude that doing good does not necessarily lead to doing well. Other studies propose a curvilinear relationship between CSR and CFP (Barnett and Salomon, 2012; Nuber et al., 2020). Hence, the literature on accounting based measures is also inconclusive (Hussain et al., 2018). Given that the literature shows mixed results in the analysis of the CSR-CFP relationship, our first hypothesis is as follows:

Hypothesis 1. The CSR, as measured by ESG ratings, has a mixed influence on CFP.

Friedman (1970) states that the social responsibility of business is to increase its profits, but his shareholder theory holds that a company's main responsibility is to its shareholders. Is it not the responsibility of companies to satisfy the nonfinancial utility of SR shareholders? SR shareholders care less about financial performance since they derive nonfinancial utility from investing in companies that meet high standards of CSR (Gutsche and Ziegler, 2019; Renneboog et al., 2008). Recently, Pástor et al. (2021) propose a model where green assets have low expected returns in equilibrium, and Hart and Zingales (2017) state that companies should maximize shareholder welfare rather than market value. Therefore, the more virtuous a company is, the more it tries to maximize the nonfinancial utility of its shareholders by internalizing the negative externalities of their activity to a greater degree than other companies of similar size, country, and industry. This higher internalization leads to higher costs for these companies. Hence, hypothesis 2 is as follows:

Hypothesis 2. The virtuous behavior of the company, as measured by inclusive ESG ratings, negatively influences its CFP.

3. Data and inclusive ESG ratings construction

The aim of this research is to construct inclusive ratings for the entire ESG Refinitiv database and subsequently analyze the relationship between CSR and CFP. For each company we need information on its country, industry, and size to estimate the inclusive ratings. Thus, we use the Refinitiv geographical classification for country and the ICB classification for industry. The company's size is measured by the market value because it is the best variable to capture its economic slack and visibility. Specifically, we use the timeseries of market value in USD dollars to homogenize our sample composed of companies from different countries.

Table 2.1 Description of the research variables.

This table gives detailed descriptions of the variables used in this research. The first column shows the name of the variable; the second column shows the source of the variables, Own or Refinitiv; the third column gives a description of the variable; and the last column shows the Refinitiv code of the variable.

Name	Source	Description	Code
ESG Score	Refinitiv	An overall company score based on the self-reported information in the environmental, social, and corporate governance pillars.	TRESGS
Environment Pillar Score	Refinitiv	Refinitiv's Environment Pillar Score is the weighted average relative rating of a company based on the reported environmental information and the resulting three environmental category scores.	ENSCORE
Social Pillar Score	Refinitiv	Refinitiv's Social Pillar Score is the weighted average relative rating of a company based on the reported social information and the resulting four social category scores.	SOSCORE
Governance Pillar Score	Refinitiv	Refinitiv's Governance Pillar Score is the weighted average relative rating of a company based on the reported governance information and the resulting three governance category scores.	CGSCORE
ESG Inclusive	Own	The excess of the overall ESG score relative to other companies of similar size, country, and industry.	
Environmental Inclusive	Own	The excess of the environmental score relative to other companies of similar size, country, and industry.	
Social Inclusive	Own	The excess of social score relative to other companies of similar size, country, and industry.	
Governance Inclusive	Own	The excess of governance score relative to other companies of similar size, country, and industry.	
Total Return Index	Refinitiv	Theoretical growth in value of a share over a specified period, assuming that dividends are re-invested.	RI
Market Value in USD	Refinitiv	Is the share price multiplied by the number of ordinary shares (automatically downloaded in USD)	X(MV)~U$
Market Value Consolidated	Refinitiv	The consolidated market value of a company in USD: sum of the market value of the listed shares when one company has different emissions.	X(MVC)~U$
Percentile Rank MV	Own	Percentile rank of the average market value of each company analyzed in a given year.	
Geographical Classification of Company	Refinitiv	Returns a geographical classification of company by specific two-digit alpha code.	GEOG
Industry Name	Refinitiv	Industry of the company according to the ICB classification	TR3N
Exchange Rate Middle	Refinitiv	Exchange rate between bid and ask rate	ER
Currency of Document	Refinitiv	Represents the ISO currency code which corresponds to the currency in which the company's financial statements are presented.	WC06099
Company Exchange Rate	Own	The average of the daily exchange rate to US dollars between company fiscal years	
Date of Fiscal Year End	Refinitiv	Represents the year, month, and day the company closes its books at the end of its fiscal period.	WC05350
Return on Assets	Refinitiv	(Net Income – Bottom Line + ((Interest Expense on Debt-Interest Capitalized) * (1-Tax Rate))) / Average of Last Year's and Current Year’s Total Assets * 100	WC08326
Return on Equity Total %	Refinitiv	Profitability Ratio, Annual & Interim Item: All Industries: (Net Income – Bottom Line - Preferred Dividend Requirement) / Average of Last Year's and Current Year’s Common Equity * 100	WC08301
Total Assets	Refinitiv	Represents the sum of total current assets; long term receivables; investment in unconsolidated subsidiaries; other investments, net property, plant, and equipment; and other assets.	WC02999
Total Liabilities	Refinitiv	Represents all short- and long-term obligations expected to be satisfied by the company	WC03351
Net Sales or Revenues	Refinitiv	Represents gross sales and other operating revenue less discounts, returns, and allowances.	WC01001
Common Equity	Refinitiv	Represents common shareholders' investment in a company.	WC03501
Capital Expenditures (Additions to Fixed Assets)	Refinitiv	Represents the funds used to acquire fixed assets other than those associated with acquisitions.	WC04601

Once the inclusive ESG ratings are estimated, we test the relation between CSR and CFP using accounting and market measures of CFP. As accounting measures, we use the return on assets (ROA) and the return on equity (ROE), while for market ones we use the company's daily returns in USD dollars. We also considered different accounting variables such as log of net sales, log of total assets, total liabilities to equity, long debt to assets, and capital expenditures as control variables. Net sales and total assets are converted to USD using the daily average of the exchange rate of each currency for each fiscal year. Thus, we homogenize the information for each company in our sample. Table 2.1 lists the variables used in the research with their descriptions.

We construct the inclusive ratings using the ESG Refinitiv database that replaced the ASSET4® Equal Weighted Ratings to analyze the period from 2010 to 2019. Specifically, we analyze listed and delisted companies to avoid survivorship bias. Each year we excluded from our analysis companies belonging to a country with less than four ESG-rated companies. Thus, we create the inclusive ratings by regressing each year and each conventional ESG rating on country, industry, and size according to Equation 1.

$$ESG_i = \alpha_c \times D_c + \alpha_s \times D_s + \beta\, PctRank\_MV_i + \varepsilon_i \qquad \textbf{(1)}$$

where *ESG* denotes each rating analyzed—overall, environmental, social, and governance—*i* denotes the company, *c* the country, and *s* the industry. Thus, D_c and D_s are country and industry dummies, while α_c and α_s capture the effect of the country and industry on the rating. The *PctRank_MV* is the percentile rank of the average market value of each company *i* in a given year relative to the other companies in that year. We use the percentile rank because it is the same methodology used by our provider to obtain the ratings. Finally, ε_i is the company's ESG excess relative to size, country, and industry of the other companies. Therefore, the error term of this cross-sectional regression is our inclusive ESG rating and approximates the company's virtuous behavior in a given year.

Table 2.2 shows for each year and ESG rating the R^2 of the regression used to calculate our inclusive rating. The last row shows the number of companies analyzed in each regression. The number of observations increases because our provider has been increasing the number of companies covered. The R^2 of the governance pillar is the lowest because the approach of this research is related to CSR and not corporate governance, but we have analyzed all pillars for the sake of exhaustiveness.

Table 2.2 Goodness of the model used to estimate inclusive ESG ratings.
This table provides the R^2 of each cross-sectional regression used to obtain the inclusive ESG ratings on the basis of equation 1. Specifically, rows 2 to 5 show the R^2 for each inclusive rating—overall, environmental, social, and governance—while the last two rows show the number of countries and companies considered in each regression.

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
ESG Inclusive	38.1%	37.1%	34.8%	35.6%	34.7%	36.6%	39.6%	40.5%	39.3%	41.8%
Environ. Inclusive	41.2%	39.9%	38.0%	38.2%	38.3%	42.3%	46.8%	48.1%	45.2%	47.4%
Social Inclusive	39.0%	38.2%	36.1%	36.8%	36.0%	35.7%	36.7%	38.7%	38.2%	41.2%
Governance Inclusive	12.0%	12.6%	11.8%	11.9%	10.1%	11.3%	13.0%	12.8%	13.1%	14.0%
#Countries	44	45	45	45	47	50	52	54	54	54
Observations	3,815	3,947	4,003	4,116	4,372	5,240	6,006	6,543	7,377	8,070

To ensure that we are capturing the correct causality relationship—from CSR to the accounting performance measures—we follow the approach proposed by Servaes and Tamayo (2013). We merge the variables for the same year when the fiscal year ends in December, and we merge the ESG data of a given year with the accounting variables of the following year for those companies with a fiscal year-end prior to December. Similar to the literature, we also winsorize the outliers to prevent the well-known potential biases (Awaysheh et al., 2020; Drempetic et al., 2020). Therefore, we winsorize ROA, ROE, total liabilities to assets, total liabilities to equity, and capital expenditures to assets at the 2.5 and 97.5 percentiles. Table A2.8 provides some statistics about our sample.

4. Methodology

4.1 Accounting based measures

First, we examine the relationship between CSR and CFP using accounting measures. Similar to previous literature, we estimate a regression in which CSR, CFP, and control variables are measured concurrently (Awaysheh et al., 2020; Garcia and Orsato, 2020; Hussain et al., 2018; Minutolo et al., 2019; Servaes and Tamayo, 2013). Studies often consider three control variables—size, leverage, and capital intensity—but they do not agree on how to measure them. The same statement applies for the fixed effects considered and the proxy used for CFP. Therefore, in our analysis we consider different proxies for CFP, for the control variables, and for the fixed effects as can be seen in equations 2 and 3:

$$CFP_{it} = \beta_0 + \beta_1 Rating_{it} + \beta_2 log(Sales)_{it} + \beta_3 \frac{Total\ liabilities_{it}}{Assets_{it}} + \beta_4 \frac{Capital\ expenditures_{\ it}}{Assets_{\ it}} + Dummies\ Fixed\ Effects + \varepsilon_{it} \quad \textbf{(2)}$$

$$CFP_{it} = \beta_0 + \beta_1 Rating_{it} + \beta_2 log(Assets)_{it} + \beta_3 \frac{Total\ liabilities_{it}}{Equity_{it}} + \beta_4 \frac{Capital\ expenditures_{\ it}}{Assets_{\ it}} + Dummies\ Fixed\ Effects + \varepsilon_{it} \quad \textbf{(3)}$$

where CFP_{it} is measured by the ROA and the ROE of company *i* in year *t*, respectively; $Rating_{it}$ refers to each rating analyzed for each company (overall/overall inclusive, environmental/ environmental inclusive, social/ social inclusive, governance/ governance inclusive). The log net sales and log total assets are proxies for size; total liabilities to assets and total liabilities to equity are proxies for capital structure; and capital expenditures to assets are a proxy for the capital intensity of the company. The dummies are added to control for the unobservable heterogeneity that can produce strong differences among the CFP of different companies, years, countries, or industries. We perform equations 2 and 3 without introducing any fixed effects, by adding: year dummies, year + industry dummies, year + industry + country dummies, and year + company dummies.

Table 2.3 shows the correlation matrix among our variables because a relevant limitation in the CSR-CFP literature is that CSR is endogenous with respect to CFP; that is, companies invest in CSR to enhance their profitability and value, or only well-performing companies can afford to invest in CSR (Flammer, 2015; Garcia and Orsato, 2020; Liang and Renneboog, 2017). The correlation between conventional ratings and CFP is positive and statistically significant while this correlation is negative and significant for the inclusive ratings (except for the governance pillar). Thus, our inclusive ratings, as opposed to conventional ones, do not suffer from the endogeneity problems described in the literature. Additionally, Table A2.9 of the appendix shows the variance inflation factor for each rating and control variable in equations 2 and 3. The variance inflation factor is always below five, hence there is not multicollinearity among the independent variables.

Table 2.3 Correlation Matrix
This table shows the Pearson correlation coefficients between the variables in the first column and the first row. The * and ** indicate statistical significance at the 5%, 1%, levels, respectively.

	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(0) ROA	1.00**														
(1) ROE	0.87**	1.00**													
(2) ESG Conv.	0.08**	0.12**	1.00**												
(3) Environmental Conv.	0.08**	0.11**	0.86**	1.00**											
(4) Social Conv.	0.06**	0.10**	0.90**	0.73**	1.00**										
(5) Governance Conv.	0.06**	0.09**	0.69**	0.41**	0.41**	1.00**									
(6) ESG Incl.	-0.07**	-0.04**	0.78**	0.61**	0.67**	0.64**	1.00**								
(7) Environmental Incl.	-0.08**	-0.05**	0.63**	0.75**	0.51**	0.31**	0.81**	1.00**							
(8) Social Incl.	-0.07**	-0.05**	0.67**	0.50**	0.78**	0.33**	0.86**	0.66**	1.00**						
(9) Governance Incl.	-0.02**	-0.01	0.53**	0.24**	0.27**	0.94**	0.69**	0.32**	0.35**	1.00**					
(10) Log Sales	0.15**	0.18**	0.37**	0.43**	0.25**	0.22**	0.11**	0.11**	0.09**	0.07**	1.00**				
(11) Log Assets	0.01*	0.08**	0.34**	0.39**	0.24**	0.20**	0.09**	0.10**	0.07**	0.05**	0.90**	1.00**			
(12) Liabilities to Equity	-0.18**	0.01	0.11**	0.06**	0.10**	0.06**	0.07**	0.09**	0.07**	0.03**	0.15**	0.33**	1.00**		
(13) Liabilities to Assets	-0.18**	0.04**	0.18**	0.13**	0.16**	0.12**	0.13**	0.12**	0.11**	0.08**	0.23**	0.31**	0.75**	1.00**	
(14) Capital Expenditures	0.07**	0.01*	-0.03**	0.04**	-0.04**	0.02**	-0.02**	-0.00	-0.01**	-0.00	-0.03**	-0.10**	-0.25**	-0.19**	1.00**

4.2 Market based measures

We also analyze the relationship between CSR and CFP using market measures. Based on the company's score, we create high and low ESG portfolios for different cutoffs. Specifically, each year we identify the companies that are in the top and bottom 20%, 10%, and 5% according to the ratings analyzed. Subsequently, we calculate the daily average return of companies in the bottom (low ESG portfolio), of companies in the top (high ESG portfolio), and the difference between both returns (high-minus-low portfolio).[5] We also obtain the abnormal return of the portfolio (Alpha) for each portfolio (low, high, high-minus-low), for each cutoff (5%, 10%, 20%), and for each rating (overall/overall inclusive, environmental/ environmental inclusive, social/ social inclusive, governance/ governance inclusive) following equation 4. This methodology is widely used to study whether portfolios that are created on the basis of any criteria, such as ESG ratings, have abnormal returns (Azevedo et al., 2021).

$$r_{pt} = \alpha_p + \beta_{MKT}MKT_t + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{RMW}RMW_t + \beta_{CMA}CMA_t + \varepsilon_{pt} \tag{4}$$

where r_{pt} is the daily dollar excess return over the one-month US Treasury bill rate of portfolio p in period t; and α_p shows the excess return over the risk factors: MKT, SMB, HML, RMW and CMA obtained from the Fama and French data library (Fama & French, 1993, 2015).

5. Results

5.1 Conventional ESG ratings and accounting based measures

Table 2.4 shows the results of the influence of each conventional ESG ratings on the CFP measured by ROA and ROE. The regressions that do not introduce country or firm effects offer a positive influence regardless of the control variables. Instead, when we introduce the country effects, the results are assorted depending on the control variables used. The regressions that control for unobserved heterogeneity among companies mainly indicate the absence of the relationship. Hence, our results show that the CSR, represented by ESG ratings, has a mixed influence on CFP. The different control

[5] The return data is winsorized at 1 and 99 percentiles to obtain the daily return of each portfolio. The quotes of companies that did not change over a period of 126 days and quotes with a price index in Refintiv of less than one were not considered

variables may partially explain the contradictory results in the previous literature. In fact, our results show that the relationship between CSR and CFP is volatile, as it can change rapidly with a small modification of the model. Hence, we provide evidence of a mixed influence of CSR measured by ESG ratings on CFP as stated in our first hypothesis.

Our results are in line with Awaysheh et al. (2020) who find that the significant relationship between CSR and operating performance disappears when they control for endogeneity and with Zhao and Murrell (2016) who underline that there is no strong evidence to support the argument that CSR will necessarily lead to an increase in CFP.

5.2 Inclusive ESG ratings and accounting based measures

Table 2.5 shows the results of the influence of each inclusive rating on the CFP. The table shows that the influence is negative and statistically significant for the overall rating, environment rating, and social rating regardless of the dependent variable, the control variables, and the type of effects considered. This negative influence is not so robust for the governance pillar. These results are in line with Krüger (2015) who argues that corporate governance does not necessarily require monetary payments, while improving the welfare of other stakeholders usually requires expenditures. Our findings show that the virtuous behavior of a company negatively influences the CFP. Hence, we find evidence supporting our second hypothesis.

5.3 Conventional ESG ratings and market based measures

Table 2.6 shows the alpha and the R2 of equation 4 for each conventional score and portfolio analyzed. The portfolios created on the basis of the overall, environmental, and social ratings do not show any statistically significant alpha. We only find that the alpha is slightly significant for the overall score in the high-minus-low portfolio at the 5% cutoff. On the other hand, the portfolios created on the basis of the governance rating show a statistically significant alpha in the high-minus-low ESG portfolio for each cutoff. Once again, we show that the governance pillar has different results. Then, as in the accounting measures, we conclude that the CSR measured by conventional ESG ratings does not influence the market measures of CFP.

Table 2.4 The influence of conventional ratings on CFP

This table shows the results on the influence of each conventional rating (overall rating in Panel A, environmental in Panel B, social in Panel C, and governance in Panel D) on the CFP by the dependent variable used (ROA and ROE), the equation used (2 or 3), and the type of effects considered (year, industry, country, and firm). Each cell of the table shows the coefficient and the significance of the conventional rating and the R2 of the regression in percentage. The * and ** indicate statistical significance at the 5% and 1% levels, respectively.

Year effects		No	Yes	Yes	Yes	Yes
Industry Effects		**No**	**No**	**Yes**	**No**	**No**
Country Effects		**No**	**No**	**No**	**Yes**	**No**
Firm Effects		**No**	**No**	**No**	**No**	**Yes**
Panel A: ESG Conventional						
ROA	**Eq. 2**	0.021** 7.45%	0.022** 8.02%	0.027** 12.49%	-0.017** 19.90%	-0.006* 71.12%
	Eq. 3	0.031** 4.59%	0.033** 5.37%	0.035** 9.83%	0.022** 12.77%	-0.001 68.96%
ROE	**Eq. 2**	0.062** 3.64%	0.066** 4.20%	0.079** 8.46%	-0.012** 14.39%	-0.003 66.42%
	Eq. 3	0.099** 1.78%	0.103** 2.51%	0.109** 7.33%	0.071** 10.02%	0.005 65.18%
Panel B: Environmental Conventional						
ROA	**Eq. 2**	0.007** 7.25%	0.007** 7.79%	0.011** 12.20%	-0.017** 20.02%	-0.005* 71.12%
	Eq. 3	0.019** 4.41%	0.020** 5.13%	0.023** 9.64%	0.012** 12.68%	0.000 68.96%
ROE	**Eq. 2**	0.024** 3.36%	0.026** 3.88%	0.037** 8.10%	-0.023** 14.45%	-0.003 66.42%
	Eq. 3	0.057** 1.40%	0.059** 2.07%	0.067** 6.94%	0.038** 9.83%	0.006 65.18%
Panel C: Social Conventional						
ROA	**Eq. 2**	0.016** 7.41%	0.018** 8.00%	0.021** 12.44%	-0.017** 19.94%	-0.003 71.12%
	Eq. 3	0.021** 4.38%	0.024** 5.17%	0.025** 9.61%	0.012** 12.65%	-0.000 68.96%
ROE	**Eq. 2**	0.046** 3.55%	0.051** 4.11%	0.060** 8.35%	-0.018** 14.41%	-0.007 66.42%
	Eq. 3	0.067** 1.42%	0.073** 2.16%	0.078** 6.99%	0.043** 9.83%	-0.003 65.18%
Panel D: Governance Conventional						
ROA	**Eq. 2**	0.015** 7.38%	0.016** 7.92%	0.019** 12.36%	-0.001 19.78%	-0.002 71.12%
	Eq. 3	0.020** 4.30%	0.020** 5.02%	0.022** 9.47%	0.015** 12.73%	-0.002 68.96%
ROE	**Eq. 2**	0.045** 3.52%	0.046** 4.03%	0.054** 8.24%	0.007 14.39%	0.004 66.42%
	Eq. 3	0.064** 1.32%	0.065** 1.97%	0.069** 6.76%	0.045** 9.90%	0.004 65.18%

Table 2.5 The influence of inclusive ratings on CFP

This table shows the results on the influence of each inclusive rating (overall rating in Panel A, environmental in Panel B, social in Panel C, and governance in Panel D) on the CFP by the dependent variable used (ROA and ROE), the equation used (2 or 3), and the type of effects considered (year, industry, country, and firm). Each cell of the table shows the coefficient and the significance of the inclusive rating and the R^2 of the regression in percentage. The * and ** indicate statistical significance at the 5% and 1% levels, respectively.

Year effects		No	Yes	Yes	Yes	Yes
Industry Effects		**No**	**No**	**Yes**	**No**	**No**
Country Effects		**No**	**No**	**No**	**Yes**	**No**
Firm Effects		**No**	**No**	**No**	**No**	**Yes**
Panel A: ESG Inclusive						
ROA	**Eq. 2**	-0.030** 7.57%	-0.029** 8.09%	-0.026** 12.35%	-0.038** 20.36%	-0.042** 71.24%
	Eq. 3	-0.029** 4.35%	-0.029** 5.04%	-0.027** 9.39%	-0.031** 12.97%	-0.051** 69.15%
ROE	**Eq. 2**	-0.074** 3.63%	-0.073** 4.13%	-0.062** 8.13%	-0.091** 14.96%	-0.103** 66.55%
	Eq. 3	-0.056** 0.99%	-0.055** 1.62%	-0.050** 6.31%	-0.064** 9.93%	-0.119** 65.35%
Panel B: Environmental Inclusive						
ROA	**Eq. 2**	-0.027** 7.75%	-0.027** 8.27%	-0.024** 12.53%	-0.033** 20.61%	-0.029** 71.23%
	Eq. 3	-0.024** 4.45%	-0.024** 5.14%	-0.023** 9.48%	-0.026** 13.08%	-0.034** 69.12%
ROE	**Eq. 2**	-0.066** 3.81%	-0.065** 4.30%	-0.058** 8.29%	-0.080** 15.18%	-0.069** 66.52%
	Eq. 3	-0.054** 1.14%	-0.053** 1.76%	-0.049** 6.43%	-0.060** 10.11%	-0.077** 65.31%
Panel C: Social Inclusive						
ROA	**Eq. 2**	-0.028** 7.63%	-0.028** 8.16%	-0.026** 12.44%	-0.033** 20.38%	-0.029** 71.20%
	Eq. 3	-0.027** 4.40%	-0.027** 5.10%	-0.026** 9.46%	-0.029** 13.02%	-0.036** 69.10%
ROE	**Eq. 2**	-0.068** 3.68%	-0.068** 4.17%	-0.061** 8.19%	-0.080** 14.96%	-0.078** 66.53%
	Eq. 3	-0.056** 1.06%	-0.056** 1.69%	-0.052** 6.38%	-0.062** 10.00%	-0.090** 65.32%
Panel D: Governance Inclusive						
ROA	**Eq. 2**	-0.006** 7.22%	-0.006** 7.75%	-0.004* 12.09%	-0.010** 19.85%	-0.013** 71.14%
	Eq. 3	-0.006** 4.01%	-0.006** 4.72%	-0.005** 9.11%	-0.007** 12.60%	-0.017** 69.01%
ROE	**Eq. 2**	-0.017** 3.29%	-0.017** 3.79%	-0.012** 7.87%	-0.026** 14.47%	-0.029** 66.44%
	Eq. 3	-0.009* 0.78%	-0.008* 1.41%	-0.006 6.13%	-0.012** 9.67%	-0.035** 65.21%

5.4 Inclusive ESG ratings and market based measures

Table 2.7 shows the constant and the R2 of equation 4 for each inclusive score and portfolio analyzed. The portfolios created on the basis of the overall, environmental, and social ratings show statistically significant alphas in the high-minus-low portfolio. These findings show that the companies that score high in our inclusive ratings have worse financial performance than those that score low. Again, these results support our second hypothesis. Similar to conventional ratings, portfolios created on the basis of the governance pillar show significant negative alphas in the high-minus-low ESG portfolio for each cutoff.

Our study shows a different influence of conventional ratings and our inclusive ones on the company's CFP. The CSR measured by conventional ratings do not affect CFP that is measured with accounting and market measures, while the virtuous behavior of the company negatively influences its CFP. These results support the idea that companies with better environmental and social performance than their peers in terms of size, country, and industry suffer a decrease in their CFP. Our results are in line with recent theoretical studies (Hart and Zingales, 2017; Pástor et al., 2021).

Table 2.6 High and Low ESG portfolios based on conventional ratings

This table shows the constant (alpha) and R^2 of equation 4 for each conventional rating (overall rating in Panel A, environmental in Panel B, social in Panel C, and governance in Panel D), cutoff, and portfolio analyzed. The first row shows the cutoff that was used to construct the portfolio (20%, 10%, 5%), and the second row shows the type of ESG portfolio (low, high, and high-minus-low). The intersection between the rows and columns shows the coefficient, the standard error in parenthesis and the significance of the alphas, and the R^2 value. The * and ** indicate statistical significance at the 5% and 1% levels, respectively.

Cut	20%			10%			5%		
Portfolio	Low	High	High-Low	Low	High	High-Low	Low	High	High-Low
Panel A: Overall Conventional									
Alfa	0.008 (0.010)	-0.001 (0.011)	-0.011 (0.007)	0.007 (0.011)	-0.004 (0.012)	-0.014 (0.008)	0.015 (0.012)	-0.006 (0.012)	-0.023* (0.009)
R2	56.77%	58.52%	15.79%	46.84%	58.41%	19.86%	37.48%	57.41%	25.70%
Panel B: Environmental Conventional									
Alfa	0.006 (0.009)	-0.002 (0.012)	-0.009 (0.007)	0.007 (0.009)	-0.003 (0.013)	-0.012 (0.008)	0.009 (0.010)	-0.003 (0.014)	-0.014 (0.010)
R2	64.99%	53.72%	8.83%	64.35%	52.81%	8.67%	61.75%	52.07%	9.65%
Panel C: Social Conventional									
Alfa	0.008 (0.011)	-0.001 (0.011)	-0.011 (0.008)	0.011 (0.013)	-0.004 (0.012)	-0.017 (0.010)	0.014 (0.014)	-0.003 (0.012)	-0.019 (0.012)
R2	43.17%	61.57%	23.34%	28.53%	60.51%	29.41%	19.05%	59.49%	32.33%
Panel D: Governance Conventional									
Alfa	0.010 (0.009)	0.002 (0.010)	-0.010** (0.003)	0.014 (0.009)	-0.002 (0.010)	-0.018** (0.004)	0.015 (0.009)	-0.004 (0.011)	-0.021** (0.005)
R2	64.37%	62.43%	29.40%	63.25%	60.52%	25.16%	62.08%	57.24%	19.00%

Table 2.7 High and Low ESG portfolios based on inclusive ratings.
This table shows the constant (alpha) and R^2 of equation 4 for each inclusive rating (overall rating in Panel A, environmental in Panel B, social in Panel C, and governance in Panel D), cutoff, and portfolio analyzed. The first row shows the cutoff that was used to construct the portfolio, and the second row shows the type of ESG portfolio (low, high, and high-minus-low). The intersection between the rows and columns shows the coefficient, the standard error in parenthesis and the significance of the alphas, and the R^2 value. The * and ** indicate statistical significance at the 5% and 1% levels, respectively.

Cut		20%			10%			5%	
Portfolio	Low	High	High-Low	Low	High	High-Low	Low	High	High-Low
Panel A: Overall Inclusive									
Alfa	0.016 (0.009)	-0.004 (0.010)	-0.022** (0.003)	0.023* (0.010)	-0.006 (0.010)	-0.031** (0.004)	0.023* (0.010)	-0.007 (0.011)	-0.032** (0.005)
R2	62.33%	60.72%	19.62%	61.01%	58.82%	16.75%	58.02%	57.17%	12.28%
Panel B: Environmental Inclusive									
Alfa	0.019* (0.009)	-0.004 (0.010)	-0.025** (0.003)	0.023* (0.010)	-0.006 (0.010)	-0.032** (0.004)	0.023* (0.010)	-0.009 (0.010)	-0.035** (0.005)
R2	65.26%	62.06%	16.02%	61.13%	64.91%	23.35%	54.29%	66.00%	27.24%
Panel C: Social Inclusive									
Alfa	0.014 (0.010)	-0.003 (0.010)	-0.019** (0.002)	0.016 (0.010)	-0.006 (0.010)	-0.024** (0.004)	0.018 (0.011)	-0.008 (0.011)	-0.028** (0.005)
R2	61.48%	61.77%	13.83%	60.42%	59.43%	7.92%	57.06%	54.16%	6.86%
Panel D: Governance Inclusive									
Alfa	0.012 (0.009)	0.001 (0.010)	-0.014** (0.002)	0.018 (0.009)	0.000 (0.010)	-0.020** (0.003)	0.020* (0.009)	-0.007 (0.010)	-0.029** (0.004)
R2	64.13%	64.20%	29.42%	64.31%	63.78%	29.05%	67.58%	63.38%	22.76%

6. Discussion and conclusion

Traditionally, studies have quantified the degree of CSR in companies with ESG ratings. However, ESG ratings provide a global indicator of CSR that does not consider the specific capabilities of companies to meet environmental and social standards. In this study, we consider the specific capabilities of companies by creating inclusive ESG ratings which show their virtuous behaviors. Thus, this study contributes to the literature by proposing a method to obtain more accurate measures of CSR.

Specifically, we propose a cross-sectional regression in which the estimated value is the rating that a company should have due to its size, country, and industry; while the error term is the ESG excess of the company relative to the size, country, and industry of the other companies. Thus, the error term is the inclusive ESG rating. This rating is useful for managers, investors, regulators, and researchers because it provides a comparable indicator among companies. Our results show that the influence of the ESG ratings on the measures of accounting performance is inconclusive, while the inclusive ESG ratings have a robust negative influence on ROA and ROE. Specifically, the influence of conventional ESG ratings on these indicators changes depending on the control variables and the fixed effects considered. We find similar results when we analyze the return difference between high rated and low rated portfolios. There are no significant abnormal returns when the portfolios are created on the basis of conventional ratings. However, when portfolios are created based on inclusive ratings, significantly negative abnormal returns are observed.

Advocates of CSR often argue that "doing good leads to doing well". We argue that the problem is not in the financial performance, but in the meaning of doing good. Focusing on the empirical studies, doing good is closely related to high ESG ratings. However, companies in certain geographic areas and of certain size usually obtain the best ESG ratings. In this study, we argue that doing good, in a virtuous sense, should not be related, among other things, with size, country, or industry. We show that doing good does not lead to doing well.

References

Aguinis, H., & Glavas, A. (2012). What we know and don’t know about corporate social responsibility: A review and research agenda. Journal of Management, 38(4), 932–968. https://doi.org/10.1177/0149206311436079

Ali, W., Frynas, J. G., & Mahmood, Z. (2017). Determinants of corporate social responsibility (CSR) disclosure in developed and developing countries: A literature review. Corporate Social Responsibility and Environmental Management, 24(4), 273–294. https://doi.org/https://doi.org/10.1002/csr.1410

Awaysheh, A., Heron, R. A., Perry, T., & Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. Strategic Management Journal, 41(6), 965–987. https://doi.org/10.1002/smj.3122

Azevedo, V., Kaserer, C., & Campos, L. M. S. (2021). Investor sentiment and the time-varying sustainability premium. Journal of Asset Management, 22(7), 600–621. https://doi.org/10.1057/s41260-021-00233-1

Badía, G., Cortez, M. C., & Ferruz, L. (2020). Socially responsible investing worldwide: Do markets value corporate social responsibility? Corporate Social Responsibility and Environmental Management, 27(6), 2751–2764. https://doi.org/https://doi.org/10.1002/csr.1999

Barauskaite, G., & Streimikiene, D. (2021). Corporate social responsibility and financial performance of companies: The puzzle of concepts, definitions and assessment methods. Corporate Social Responsibility and Environmental Management, 28(1), 278–287. https://doi.org/https://doi.org/10.1002/csr.2048

Barnett, M. L., & Salomon, R. M. (2012). Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. Strategic Management Journal, 33(11), 1304–1320. https://doi.org/https://doi.org/10.1002/smj.1980

Basu, S., Vitanza, J., Wang, W., & Ross, X. (2022). Walking the walk ? Bank ESG disclosures and home mortgage lending. SSRN. https://doi.org/http://dx.doi.org/10.2139/ssrn.3807685

Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. Journal of Banking & Finance, 29(7), 1751–1767. https://doi.org/https://doi.org/10.1016/j.jbankfin.2004.06.035

Berrone, P., Fosfuri, A., & Gelabert, L. (2017). Does greenwashing pay off? Understanding the relationship between environmental actions and environmental legitimacy. Journal of Business Ethics, 144(2), 363–379. https://doi.org/10.1007/s10551-015-2816-9

Brammer, S. J., & Pavelin, S. (2006). Corporate reputation and social performance: The importance of fit. Journal of Management Studies, 43(3), 435–455. https://doi.org/https://doi.org/10.1111/j.1467-6486.2006.00597.x

Capelle-Blancard, G., & Petit, A. (2015). The weighting of CSR dimensions: One size does not fit all. Business & Society, 56(6), 919–943. https://doi.org/10.1177/0007650315620118

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. Strategic Management Journal, 37(8), 1597–1614. https://doi.org/https://doi.org/10.1002/smj.2407

Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility? Journal of Economics & Management Strategy, 18(1), 125–169. https://doi.org/https://doi.org/10.1111/j.1530-9134.2009.00210.x

Chiu, S.-C., & Sharfman, M. (2009). Legitimacy, visibility, and the antecedents of corporate social performance: An investigation of the instrumental perspective. Journal of Management, 37(6), 1558–1585. https://doi.org/10.1177/0149206309347958

Cunha, F. A. F. de S., de Oliveira, E. M., Orsato, R. J., Klotzle, M. C., Cyrino Oliveira, F. L., & Caiado, R. G. G. (2020). Can sustainable investments outperform traditional benchmarks? Evidence from global stock markets. Business Strategy and the Environment, 29(2), 682–697. https://doi.org/10.1002/bse.2397

Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. The Accounting Review, 70(2), 193–225. Retrieved from http://www.jstor.org/stable/248303

Demirbag, M., Wood, G., Makhmadshoev, D., & Rymkevich, O. (2017). Varieties of CSR: Institutions and Socially Responsible Behaviour. International Business Review, 26(6), 1064–1074. https://doi.org/https://doi.org/10.1016/j.ibusrev.2017.03.011

Di Giuli, A., & Kostovetsky, L. (2014). Are red or blue companies more likely to go green? Politics and corporate social responsibility. Journal of Financial Economics, 111(1), 158–180. https://doi.org/https://doi.org/10.1016/j.jfineco.2013.10.002

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. Journal of Business Ethics, 167(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1

Dyck, A., Lins, K. V, Roth, L., & Wagner, H. F. (2019). Do institutional investors drive corporate social responsibility? International evidence. Journal of Financial Economics, 131(3), 693–714. https://doi.org/https://doi.org/10.1016/j.jfineco.2018.08.013

Edmans, A. (2012). The link between job satisfaction and firm value, with implications for corporate social responsibility. Academy of Management Perspectives, 26(4), 1–19. https://doi.org/10.5465/amp.2012.0046

Etzion, D. (2007). Research on organizations and the natural environment, 1992-present: A review. Journal of Management, 33(4), 637–664. https://doi.org/10.1177/0149206307302553

EUROSIF, (2018). European SRI study 2018. Available at: http://www.eurosif.org/wp-content/uploads/2018/11/European-SRI-2018-Study.pdf

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56. https://doi.org/https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1–22. https://doi.org/https://doi.org/10.1016/j.jfineco.2014.10.010

Ferrell, A., Liang, H., & Renneboog, L. (2016). Socially responsible firms. Journal of Financial Economics, 122(3), 585–606. https://doi.org/https://doi.org/10.1016/j.jfineco.2015.12.003

Flammer, C. (2015). Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. Management Science, 61(11), 2549–2568. https://doi.org/10.1287/mnsc.2014.2038

Flammer, C. (2021). Corporate green bonds. Journal of Financial Economics, 142(2), 499–516. https://doi.org/https://doi.org/10.1016/j.jfineco.2021.01.010

Friedman, M. The social responsibility of business is to increase its profits, The New York Times Magazine, 13 September 1970

Garcia, A. S., & Orsato, R. J. (2020). Testing the institutional difference hypothesis: A study about environmental, social, governance, and financial performance. Business Strategy and the Environment, 29(8), 3261–3272. https://doi.org/https://doi.org/10.1002/bse.2570

Gutsche, G., & Ziegler, A. (2019). Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. Journal of Banking & Finance, 102, 193–214. https://doi.org/https://doi.org/10.1016/j.jbankfin.2019.03.007

Halbritter, G., & Dorfleitner, G. (2015). The wages of social responsibility — where are they? A critical review of ESG investing. Review of Financial Economics, 26, 25–35. https://doi.org/https://doi.org/10.1016/j.rfe.2015.03.004

Hart, O., & Zingales, L. (2017). Companies should maximize shareholder welfare not market value. SSRN, (August). https://doi.org/http://dx.doi.org/10.2139/ssrn.3004794

Hart, S. L. (1995). A natural-resource-based view of the firm. Academy of Management Review, 20(4), 986–1014. https://doi.org/10.5465/amr.1995.9512280033

Hawn, O., Chatterji, A. K., & Mitchell, W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). Strategic Management Journal, 39(4), 949–976. https://doi.org/https://doi.org/10.1002/smj.2752

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. Journal of Financial Economics, 93(1), 15–36. https://doi.org/https://doi.org/10.1016/j.jfineco.2008.09.001

Hörisch, J., Johnson, M. P., & Schaltegger, S. (2015). Implementation of sustainability management and company size: A knowledge-based view. Business Strategy and the Environment, 24(8), 765–779. https://doi.org/https://doi.org/10.1002/bse.1844

Hussain, N., Rigoni, U., & Cavezzali, E. (2018). Does it pay to be sustainable? Looking inside the black box of the relationship between sustainability performance and financial performance. Corporate Social Responsibility and Environmental Management, 25(6), 1198–1211. https://doi.org/https://doi.org/10.1002/csr.1631

Hwang, C. Y., Titman, S., & Wang, Y. (2021). Investor tastes, corporate behavior, and stock returns: An analysis of corporate social responsibility. Management Science. https://doi.org/10.1287/mnsc.2021.4179

Jones, J. J. (1991). Earnings management during import relief investigations. Journal of Accounting Research, 29(2), 193–228. https://doi.org/10.2307/2491047

Krüger, P. (2015). Corporate goodness and shareholder wealth. Journal of Financial Economics, 115(2), 304–329. https://doi.org/https://doi.org/10.1016/j.jfineco.2014.09.008

Liang, H. ao, & Renneboog, L. (2017). On the foundations of corporate social responsibility. The Journal of Finance, 72(2), 853–910. https://doi.org/https://doi.org/10.1111/jofi.12487

Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. Organization & Environment, 28(2), 223–249. https://doi.org/10.1177/1086026615575332

Lys, T., Naughton, J. P., & Wang, C. (2015). Signaling through corporate accountability reporting. Journal of Accounting and Economics, 60(1), 56–72. https://doi.org/https://doi.org/10.1016/j.jacceco.2015.03.001

Margolis, J. D., Elfenbein, H. A., & Walsh, J. P. (2007). Does it pay to be good...And does it matter? A meta-analysis of the relationship between corporate social and financial performance. SSRN. https://doi.org/10.2139/ssrn.1866371

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? Strategic Management Journal, 21(5), 603–609. https://doi.org/10.1002/(SICI)1097-0266(200005)21:5<603::AID-SMJ101>3.0.CO;2-3

McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. The Academy of Management Review, 26(1), 117–127. https://doi.org/10.2307/259398

Minutolo, M. C., Kristjanpoller, W. D., & Stakeley, J. (2019). Exploring environmental, social, and governance disclosure effects on the S&P 500 financial performance. Business Strategy and the Environment, 28(6), 1083–1095. https://doi.org/https://doi.org/10.1002/bse.2303

Mishra, S., & Modi, S. B. (2016). Corporate social responsibility and shareholder wealth: The role of marketing capability. Journal of Marketing, 80(1), 26–46. https://doi.org/10.1509/jm.15.0013

Naughton, J. P., Wang, C., & Yeung, I. (2019). Investor sentiment for corporate social performance. The Accounting Review, 94(4), 401–420. https://doi.org/10.2308/accr-52303

Nuber, C., Velte, P., & Hörisch, J. (2020). The curvilinear and time-lagging impact of sustainability performance on financial performance: Evidence from Germany. Corporate Social Responsibility and Environmental Management, 27(1), 232–243. https://doi.org/https://doi.org/10.1002/csr.1795

Odziemkowska, K., & Henisz, W. J. (2020). Webs of influence: Secondary stakeholder actions and cross-national corporate social performance. Organization Science, 32(1), 233–255. https://doi.org/10.1287/orsc.2020.1380

OECD, (2014). All on board: Making inclusive growth happen.

Orlitzky, M. (2001). Does firm size comfound the relationship between corporate social performance and firm financial performance? Journal of Business Ethics, 33(2), 167–180. https://doi.org/10.1023/A:1017516826427

Pástor, Ľ., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable investing in equilibrium. Journal of Financial Economics, 142(2), 550–571. https://doi.org/https://doi.org/10.1016/j.jfineco.2020.12.011

Porter, M. E., & Kramer, M. R. (2006). Strategy and society: The link between competitive advantage and corporate social responsibility. Harvard Business Review, 84(12), 78–92.

Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. Journal of Banking & Finance, 32(9), 1723–1742. https://doi.org/https://doi.org/10.1016/j.jbankfin.2007.12.039

Seo, H., Luo, J., & Kaul, A. (2021). Giving a little to many or a lot to a few? The returns to variety in corporate philanthropy. Strategic Management Journal, 42(9), 1734–1764. https://doi.org/https://doi.org/10.1002/smj.3309

Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. Management Science, 59(5), 1045–1061. https://doi.org/10.1287/mnsc.1120.1630

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk*. The Journal of Finance, 19(3), 425–442. https://doi.org/https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Soytas, M. A., Denizel, M., & Durak Usar, D. (2019). Addressing endogeneity in the causal relationship between sustainability and financial performance. International Journal of Production Economics, 210, 56–71. https://doi.org/https://doi.org/10.1016/j.ijpe.2019.01.016

Udayasankar, K. (2008). Corporate social responsibility and firm size. Journal of Business Ethics, 83(2), 167–175. https://doi.org/10.1007/s10551-007-9609-8

UNESCO, (2013, June). Education sector technical notes: Inclusive education.

Varadarajan, P. R., & Menon, A. (1988). Cause-related marketing: A coalignment of marketing strategy and corporate philanthropy. Journal of Marketing, 52(3), 58–74. https://doi.org/10.1177/002224298805200306

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance-financial performance link. Strategic Management Journal, 18(4), 303–319. https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.0.CO;2-G

Wang, Q., Dou, J., & Jia, S. (2016). A meta-analytic review of corporate social responsibility and corporate financial performance: The moderating effect of contextual factors. Business & Society, 55(8), 1083–1121. https://doi.org/10.1177/0007650315584317

Zhao, X., & Murrell, A. J. (2016). Revisiting the corporate social performance-financial performance link: A replication of Waddock and Graves. Strategic Management Journal, 37(11), 2378–2388. https://doi.org/https://doi.org/10.1002/smj.2579

Ziegler, A., & Schröder, M. (2010). What determines the inclusion in a sustainability stock index?: A panel data analysis for european firms. Ecological Economics, 69(4), 848–856. https://doi.org/https://doi.org/10.1016/j.ecolecon.2009.10.009

Appendix

Table A2.8 Descriptive statistics

This table shows some descriptive statistics about of the variables used in the research. The first column shows the type of variable, and the second column shows the name of the variable. The following columns list some descriptive statistics for the average, standard deviation, minimum, quartiles, and maximum value for each variable. The number of observations for each variable is 48,551.

		mean	std	min	25%	50%	75%	max
Financial Perform.	**ROA**	4.840	7.760	-21.660	1.650	4.720	8.530	22.770
	ROE	9.840	19.008	-56.330	4.480	10.530	17.800	56.460
ESG Ratings	**ESG Conventional**	42.210	20.590	0.100	25.670	40.300	57.790	94.750
	Environ. Conv.	33.030	28.980	0.000	4.230	27.810	57.210	99.250
	Social Conv.	42.660	23.460	0.050	23.980	40.160	60.240	98.550
	Governance Conv.	48.390	22.460	0.100	30.460	48.610	66.420	99.280
	ESG Inclusive	0.155	16.147	-66.362	-10.897	0.189	11.441	62.970
	Environ. Incl.	0.073	21.772	-77.727	-15.220	-1.376	14.898	80.672
	Social Inclusive	0.038	18.325	-72.182	-12.713	-0.644	12.797	69.372
	Governance Incl.	0.332	21.077	-60.359	-15.738	0.947	16.723	61.326
Control Variables	**Log Sales**	6.590	1.225	-0.158	5.863	6.426	7.109	11.346
	Log Assets	6.969	1.164	3.162	6.200	6.743	7.503	11.676
	Liabilities to Equity	2.713	3.659	0.095	0.674	1.319	2.751	16.509
	Liabilities Assets	0.556	0.223	0.099	0.399	0.558	0.718	0.941
	C. Expen.	0.043	0.044	0.000	0.010	0.030	0.060	0.190

Table A2.9 Variance inflation factor

This table shows the variance inflation factor (VIF) of the regressors in equations 2 and 3. The first row shows the equation for which the VIF is obtained. The second row shows the name of the regressors. Note that the rating analyzed is indicated in the first column.

	Equation 2				**Equation 3**			
	Rating	**Log Sales**	**Liabilities to Assets**	**Capital Expen.**	**Rating**	**Log Assets**	**Liabilities to Equity**	**Capital Expen.**
ESG Conv.	1.168	1.193	1.109	1.038	1.129	1.254	1.187	1.068
Environmental Conv.	1.236	1.280	1.098	1.042	1.195	1.334	1.192	1.073
Social Conv.	1.082	1.114	1.108	1.038	1.061	1.180	1.187	1.068
Governance Conv.	1.058	1.100	1.103	1.039	1.041	1.165	1.187	1.069
ESG Incl.	1.026	1.065	1.109	1.038	1.009	1.129	1.189	1.068
Environmental Incl.	1.024	1.066	1.107	1.038	1.015	1.130	1.192	1.068
Social Incl.	1.017	1.062	1.105	1.038	1.007	1.127	1.189	1.068
Governance Incl.	1.010	1.060	1.102	1.038	1.002	1.125	1.187	1.068

Chapter 3.

The influence of public attention on corporate social responsibility

Synopsis

Using the legitimacy theory as foundation, this study examines the role of public attention towards companies on corporate social responsibility (CSR). Previous studies use the search volume index (SVI) of Google Trends as a measure of attention. However, the SVI is a relative indicator and is not comparable among companies. To overcome this limitation, we propose carry out pairwise comparisons of the individual SVIs in Google Trends to identify the most searched companies. By repeating this routine several times for a sample of S&P 500 companies, we create a yearly measure of public attention from 2012 to 2020 that ranks companies according to the number of web searches that they receive. We find a positive relation between public attention and CSR performance but not between public attention and CSR disclosure. Following a quasi-experimental approach, we find that companies react to a *shock* of public attention by improving their CSR performance, but not their CSR disclosure.

1. Introduction

The literature has always taken for granted the relation between the public attention and a company's corporate social responsibility (CSR). However, empirical studies that address this relation are scarce and the proxies for measuring attention are not always suitable. The seminal study by Da et al. (2011) introduced Google Trends as a proxy for measuring attention, and it has become well established in the research. In this paper, we use Google Trends for the first time to study the relation between public attention and CSR. Furthermore, we propose an innovative method of using Google Trends to adapt this source of information to our case study.

The relation between public attention and CSR is underpinned by the legitimacy theory. The legitimacy theory states that companies need a generalized perception that their actions are desirable and appropriate within a socially constructed system of norms and values to justify their own existence (Suchman, 1995). Under this framework, CSR is viewed as a tool for increasing and securing company legitimacy (Baldini et al., 2018; Cormier and Magnan, 2015; Hörisch et al., 2015). Several authors argue that companies react to greater public attention, such as increased visibility due to greater exposure to public scrutiny, by increasing their commitment to CSR (Berrone et al., 2017; Chiu and Sharfman, 2009; Shabana et al., 2016). Although the literature assumes this relation, the corresponding empirical evidence is scarce, and better proxies for measuring attention and the pressure for legitimacy should be analyzed. This study advances the literature by focusing on this relation using web searches as proxy for attention. Despite the growing importance of environmental and social risks, little is known about the factors that drive companies to mitigate those risks. We examine whether being in the public eye leads companies to improve their environmental and social records.

From a theoretical perspective, companies react to legitimacy pressures by increasing their level of CSR in order to align their actions with societal values and norms. However, from the empirical perspective, the scholarly interest has not focused on the analysis of long and representative time periods due to the difficulties in finding a good indicator for attention and legitimacy pressure. For example, Shabana et al. (2016) analyze 224 firms of the Fortune 500 list in four different years (1992, 1997, 2002 and 2006) and conclude that the media coverage of the firms does not influence the probability of publishing CSR reports. Chiu and Sharfman, (2009) analyze 124 companies of the S&P 500 from 1999 to 2001. They measure CSR with KLD ratings and Fortune survey about most admired companies and the level of their public attention with a mix of six different variables. Their results show that CSR performance is an instrumental response by managers to legitimacy pressure towards the firm based on its visibility. Campbell and Slack (2006) show that companies' rate of charitable donations responds positively to public visibility measured with student surveys, while Schreck and Raithel, (2015) find a U-shaped relationship between visibility and sustainability reporting on a sample of 280 companies in 2009.

This paper overcomes the limitations of previous studies (1) by proposing a concise and generalizable indicator of public attention based on web searches; (2) by analyzing a long period, 2012 to 2020; and (3) by selecting a robust sample of companies based on the constituents of the S&P 500. We use Google Trends to measure public attention although we can find examples of alternative metrics: the number of times a company appears in newspapers (Fang and Peress, 2009), the number of times news about a company is read on a Bloomberg terminal (Ben-Rephael et al., 2017), or the number of comments in investor forums and social network groups (Dong and Gil-Bazo, 2020; Jiang et al., 2022). However, these metrics are overly biased towards the attention of institutional and retail investors while we need a more general measure of attention. The number of web searches is an appropriate metric for this purpose because the visibility and the public scrutiny towards companies leave their mark on the network. Moreover, the web searches as a method for measuring legitimacy pressure has inexplicably remained unexplored in the literature.

Studies have widely used Google Trends for measuring attention. For example, Dimpfl and Jank (2016) study the public attention to the term “dow” and its influence on the volatility of the Dow Jones Index, and Colaco et al. (2017) study the influence of the attention to the company name on the valuation of the initial public offering. Other studies use Google Trends to predict the volatility and the prices of petroleum (Han et al., 2017; Qadan and Nama, 2018); the volatility and the prices of cryptocurrencies (Liu and Tsyvinski, 2021; Urquhart, 2018); the returns of sustainability indices (El Ouadghiri et al., 2021), and assets prices in general (Choi et al., 2020; Da et al., 2015). However, this literature always uses the same measure of public attention: the search volume index (SVI). The SVI is the relative evolution of web queries over a given time horizon, where 100 means the maximum number of web queries in that time horizon, and the rest of the values are relativized from that 100.

The SVI used in other studies has two constraints: (1) If the time horizon requested to Google Trends changes (e.g., if years are included or excluded), then the time-series values of the SVI for the unmodified dates may vary considerably; (2) The individual SVI of different search topics are only comparable when they are requested together from Google Trends. While the literature uses a relative indicator of public attention that is not comparable among companies, we develop an indicator that does. Specifically, instead of using the individual time series of the SVI, we detect which companies receive more attention in period t by comparing the SVIs requested together from Google Trends. As a result, we construct a ranking of public attention that orders companies according to the number of web searches. In summary, the joint query to Google Trends of several search topics (companies) identifies which companies are the most searched for. However, Google only allows a comparison up to five search topics that makes it impossible to directly obtain a ranking of public attention to companies. Therefore, we perform pairwise comparisons following the quicksort algorithm to create a yearly ranking of public attention for the companies of the S&P 500 from 2012 to 2020.

This new measure allows us to address our main research question: the relation between public attention and CSR. First, we analyze whether there is a correlation between public attention and the performance and disclosure of CSR; and second, following a quasi-experimental approach, we study whether there is a causal relation in which public attention affects the performance and disclosure of CSR.

In order to measure the CSR performance of a company we use the Refinitiv environmental, social, and governance (ESG) ratings. Although the literature has used several proxies for CSR, such as CSR awards (Hou, 2019), carbon emissions (Boermans and Galema, 2019; Humphrey and Li, 2021), and membership in a sustainability index (Cunha et al., 2020; Hawn et al., 2018); CSR is a multidimensional construct and aggregate indicators covering different dimensions are a good proxy (Carroll, 1999). In fact, this is the most widely used approach in the literature (Dai et al., 2021; Hwang et al., 2021; Kim et al., 2021). Specifically, we analyze the overall ESG ranting, the environmental rating, and the social rating. On the other hand, the CSR disclosure of the company is measured according to the data available in each indicator used by Refinitiv to obtain the aggregate ratings. Furthermore, for robustness purposes, we apply the same analysis to the governance rating: the public attention should not influence this rating because the governance overlaps with traditional corporate governance issues that are materially different from the CSR issues (Krüger, 2015). Therefore, if our empirical design is correct, we should find a positive influence of public attention on the environmental and social pillars but not on the governance one.

We use two techniques to study the relation between public attention and CSR. We follow a traditional approach using fixed effects linear regression models, but we also test our hypothesis with a quasi-experimental framework. First, from our pool of firms we identify those that have experienced a large increase in public attention from *t-1* to *t*. Then, following the approach of (Flammer, 2021) and (Hartzmark and Sussman, 2019), we use the nearest neighbor algorithm to construct our contrafactual and draw accurate conclusions on how firms react to the prior increase in their public attention in *t+1*.

First, our study provides strong empirical evidence that companies exposed to greater public attention have higher CSR performance. These results are explained by the legitimacy theory in which companies with higher visibility use CSR as a potential source of legitimacy (Udayasankar, 2008). Second, following a quasi-experimental framework, we also provide evidence that companies improve their CSR performance after a "shock" of public attention. These results are similar to those of Chiu and Sharfman (2009) who conclude that CSR performance is an instrumental response by managers to legitimacy pressure. In addition, the influence of our control variables is in line with previous literature. We show a positive influence of firm size (Drempetic et al., 2020), return on assets (Flammer, 2015; Waddock and Graves, 1997) and research and development expenses (McWilliams and Siegel, 2000; Surroca et al., 2010) on CSR performance. However, our results show that the CSR disclosure of the company is not influenced by public attention. These findings are similar to those of Shabana et al., (2016) who conclude that the media coverage of the company does not influence the probability of publishing CSR reports. Our research shows that companies react to legitimation pressures by improving their CSR performance rather than their disclosure. Therefore, companies focus on improving the indicators they already disclose in the face of these pressures rather than increase their transparency.

One of the contributions of our research is the method used to construct the indicator of public attention to study the relation between legitimacy pressure and CSR. The literature has used the SVI of Google Trends to obtain the relative interest towards a search topic. SVIs from different search topics are not comparable with each other, while our method allows us to obtain an indicator of public attention comparable among different search topics (companies in our study). This method opens up new lines of research because until now, the focus had been on the time dimension, while our study also focusses on the individual dimension.

Our empirical findings also contribute to a flourishing literature that analyzes the factors influencing companies' commitment to CSR. Demirbag et al. (2017) and Liang and Renneboog (2017) focus on the institutional aspect and show that the country's legal origin is an important driver of CSR. However, isolating the legal origin effect from the effect of regulation and competitive strategies of countries and regions is not easy (Hart, 1995; Porter and Kramer, 2006). Company size is one of the most-examined determinants in the literature (Orlitzky, 2001; Udayasankar, 2008). For example, Drempetic et al. (2020) show that size positively influences the company's CSR performance and disclosure, while Schreck and Raithel (2015) identify an inverted U-shaped function between CSR reporting and company size. The influence of the CEOs' preferences on CSR ratings has also attracted the attention of scholars. Chin et al. (2013) and Di Giuli and Kostovetsky (2014) analyze the effect of the CEOs' political ideology on CSR and O'Sullivan et al. (2021) analyze the relation between traumatic experiences early in CEOs' lives and CSR. Recent studies show that shareholder activism increases the voluntary disclosure of environmental risks (Flammer et al., 2021) and the issuance of green bonds is followed by an increase in the company's environmental performance (Flammer, 2021). Our results contribute to this literature by showing that legitimacy pressure, as measured by web searches, has a positive effect on CSR performance and no effect on CSR disclosure.

These insights have implications for practitioners and regulators, whereby being in the public eye influences the corporate behavior of companies. Public scrutiny and signaling could be an effective strategy to put pressure on companies to improve their environmental and social records.

The paper is organized as follows: In Section 2, we describe our data and the econometric approach; Section 3 shows the empirical results, and Section 4 concludes.

2. Data and methodology

2.1 Sample and data

Measuring attention

Our sample is based on the monthly constituents of the S&P 500. We select this index for three reasons: (1) it is the most, or one of the most, important ones in the world; (2) the number of constituents ensures a satisfactory sample size; and (3) the companies in this index are important enough to leave a trace on the web. Our sample period starts in 2012 because Google Insights for Search was merged into Google Trends in 2012 and ends in 2020. For each year of the sample period, we select the companies that had been in the index for 12 months because our measure of attention is also created in yearly basis.

Google Trends distinguishes between search terms and search topics. The search term is only the set of words typed in the search engine and therefore depends on the language. On the other hand, the search topic is an entity recognized by Google Trends that does not depend on the language and has a name and a unique ID with a short description. For example, for a given search term "Apple", Google Trends suggests search topics with the same name but different IDs and descriptions: Apple whose description is "Fruit" and whose ID is "/m/014j1m"; or Apple whose description is "Technology company" and whose ID is "/m/0k8z". Thus, the first search topic ("/m/014j1m") comprises searches for manzana or pomme (apple in Spanish or French), while the second comprises searches related to Apple the multinational company. If we want to match the company called APPLE INC in Refinitiv with the search topic Apple, the technology company in Google Trends, we should check the similarity between both names and use the description to ensure that we are analyzing the public attention of technological company and not that of the fruit. Therefore, the name and description of the search topic enable an efficient merging between the data provided by Google Trends and the data provided by Refinitiv.

We conclude that two companies are the same in Refinitiv and Google Trends when the similarity between the name of the company according to Refinitiv and the name according to Google Trends is equal to or higher than 50% in the normalized Levenshtein distance and the cosine similarity indicators.[6] In addition, as in the previous

[6] These indicators measure the degree of similarity between two strings.

example, the Google Trends description is required to contain one of the following business-related words: *airport operator, bank, business, company, conglomerate, corporation, enterprise, holding, manufacturer, organization, real estate,* or *stock.* We merge Google Trends and Refinitiv data using the legal name of the company in Refinitiv-Worldscope; and if the machining is unsuccessful, we use the name of the company as stored in Refinitiv-Datastream.

Table 3.1 Descriptive statistics about the public attention ranking
This table shows the top five (1, 2, 3, 4, 5) and the bottom five (-1, -2, -3, -4, -5) companies according to the number of web searches each year from 2012 to 2020. For a given year, *n* is the number of companies in the ranking, and *N* is the number of comparisons performed to create the ranking. The string in brackets is the name according to Google Trends and the other is the name according to Refinitiv (as is stored in the Datastream database).

	2012	2013	2014	2015	2016	2017	2018	2019	2020
1	ebay (eBay)	ebay (eBay)	amazon.com (Amazon.com)	amazon.com (Amazon.com)	amazon.com (Amazon.com)	amazon.com (Amazon.com)	amazon.com (Amazon.com)	amazon.com (Amazon.com)	amazon.com (Amazon.com)
2	amazon.com (Amazon.com)	amazon.com (Amazon.com)	ebay (eBay)	ebay (eBay)	ebay (eBay)	ebay (eBay)	netflix (Netflix)	twitter (Twitter)	twitter (Twitter)
3	apple (Apple)	apple (Apple)	apple (Apple)	apple (Apple)	apple (Apple)	netflix (Netflix)	ebay (eBay)	netflix (Netflix)	netflix (Netflix)
4	walmart (Walmart)	walmart (Walmart)	walmart (Walmart)	walmart (Walmart)	walmart (Walmart)	walmart (Walmart)	apple (Apple)	apple (Apple)	walmart (Walmart)
5	microsoft (Microsoft Corporation)	ford motor (Ford Motor Company)	netflix (Netflix)	netflix (Netflix)	netflix (Netflix)	apple (Apple)	walmart (Walmart)	walmart (Walmart)	apple (Apple)
-1	xl group (XL Group Ltd)	prec.castparts (Precision Castparts Corporation (PCC) Airfoils, LLC)	cnx resources (CNX Resources)	prec.castparts (Precision Castparts Corporation (PCC) Airfoils, LLC)	l3harris technologies (L3Harris Technologies)	l3harris technologies (L3Harris Technologies)	scana (Scana)	svb financial group (SVB Financial Group)	raytheon technologies (Raytheon BBN Technologies)
-2	cnx resources (CNX Resources)	cnx resources (CNX Resources)	prec.castparts (Precision Castparts Corporation (PCC) Airfoils, LLC)	xl group (XL Group Ltd)	xl group (XL Group Ltd)	xl group (XL Group Ltd)	l3harris technologies (L3Harris Technologies)	cimarex en. (Cimarex Energy)	svb financial group (SVB Financial Group)
-3	prec.castparts (Precision Castparts Corporation (PCC) Airfoils, LLC)	xl group (XL Group Ltd)	l3harris technologies (L3Harris Technologies)	cnx resources (CNX Resources)	scana (Scana)	scana (Scana)	cimarex en. (Cimarex Energy)	raytheon technologies (Raytheon BBN Technologies)	pinnacle west cap. (Pinnacle West Capital)
-4	l3harris technologies (L3Harris Technologies)	l3harris technologies (L3Harris Technologies)	xl group (XL Group Ltd)	scana (Scana)	altaba (Altaba)	raytheon technologies (Raytheon BBN Technologies)	raytheon technologies (Raytheon BBN Technologies)	loews (Loews Corporation)	regency centers (Regency Centers)
-5	scana (Scana)	scana (Scana)	hudson city banc. (Hudson City Bancorp)	l3harris technologies (L3Harris Technologies)	cimarex en. (Cimarex Energy)	cimarex en. (Cimarex Energy)	jefferies financial group (Jefferies Financial Group)	regency centers (Regency Centers)	loews (Loews Corporation)
n	389	390	391	382	383	385	388	390	392
N	3,684	4,784	3,574	3,549	4,047	3,983	3,528	4,077	3,609

After this matching, we construct our yearly ranking of public attention by pairwise comparisons of the SVI because only five search topics (companies) can be requested from Google Trends at the same time. Therefore, we compare one company against another to find out which company receives more web searches. We perform this routine several times using the Quicksort algorithm to order companies according to the number of webs searches. The use of a sorting algorithm reduces the number of comparisons and requests to Google Trends needed to create our ranking. In each request, we compare the SVI with no geographic restrictions of two companies for a given year. Finally, we obtain our variable of analysis (*PublicAttention*) by transforming each year the ranking (1 to N) into percentile rank (0 to 1), where one is the company that receives the most web searches and zero the least.

Table 3.1 shows the top five and the bottom five companies in the ranking for each year. This table also shows the number of comparisons performed to create the ranking. Specifically, our *PublicAttention* variable comprises 3,490 yearly observations, and each year we consider between 382 and 392 companies to create the ranking.

Further data sources

We use the Refinitiv ESG ratings to measure the CSR performance of the company. For exhaustiveness, we analyze the overall ESG rating (*ESGScore*) and each pillar: environmental (*EnvScore*), social (*SocScore*) and governance (*GovScore*). On the other hand, we measure the CSR disclosure of the company as follows: each year we obtain the company's disclosure in each ESG pillar as the number of available items divided by the number of total items considered by Refinitiv to construct its ratings (*EnvDisclosure, SocDisclosure,* and *GovDisclosure*). Refinitiv considers more than 500 ESG measures collected from company disclosure that aggregates to 186 metrics and then constructs the ESG ratings. Therefore, the measures available in each pillar are a good proxy for company disclosure.

Regarding the control variables, the base 10 logarithm of the total assets measures the size (*LogAssets*) of the company, the ratio between debt and common equity measures the financial structure of the company (*Leverage*), the return on assets measures the profitability of the assets (*ROA*) of the company, and the ratio between the research and development (R&D) expenditures and the total assets measures the R&D intensity of the company (*RD_Intensity*). The R&D expenditures are not usually capitalized on the balance sheet, but they are computed as an expense. Therefore, similar to other studies when R&D is missing we set it equal to zero (Aouadi and Marsat, 2018; Servaes and Tamayo, 2013). We winsorize *Leverage, RD_intensity* and *ROA* at the 2.5 and 97.5 percentiles to avoid problems with outliers. Table 3.2 has the definitions of all the variables used in the research and indicates the source of the variable, and Table A3.10 in the appendix shows the descriptive statistics for these variables.

Table 3.2. Summary of the main variables

This table gives detailed descriptions of the variables used in this research. The first column shows the name of the variable, the second column gives a description and the source of the variable (Own, Google Trends, Refinitiv), and the third column shows the frequency of the variable.

Name	Definition	Freq.
Variables for the matching between Refitniv and Google Trends		
ID WorldScope	Represents the nine-digit identifier used by WorldScope to identify companies/securities in the database. (Source: Refinitiv-WorldScope; Code: WC06035).	Static
ID Datastream	This is the unique six-digit identification code for every stock, allocated by Datastream. (Source: Refinitiv-Datastream; Code: DSCD)	Static
ID Google Trends	Code that identifies the search name and its description. This code appears in the URL when making the request to Google Trends. (Source: Google Trends using the python library pytrends).	Static
Company Legal Name	Represents the legal name of the company as reported in the 10-K for U.S. companies and the annual report for non-U.S. companies. (Source: Refinitiv-WorldScope; Code: WC06001).	Static
Datastream Name	This is the name of the equity/company or equity list, as stored on Datastream's databases. (Source: Refinitiv-Datastream; Code: NAME).	Static
Google Trends Name	Name of the search topic (company) in Google Trends. (Source: Google Trends using the python library pytrends).	Static
Google Trends Description	Description of the search name. (Source: Google Trends using the python library pytrends).	Static
Dependent Variables		
ESGScore	An overall company score based on the self-reported information based on the environmental, social and corporate governance pillars. (Source: Refinitiv-ASSET4; Code: TRESGS).	Yearly
EnvScore	Weighted average relative rating of a company based on the reported environmental information and the resulting three environmental category scores. (Source: Refinitiv-ASSET4; Code: ENSCORE)	Yearly
SocScore	Weighted average relative rating of a company based on the reported social information and the resulting four social category scores. (Source: Refinitiv-ASSET4; Code: SOSCORE)	Yearly
GovScore	Weighted average relative rating of a company based on the reported governance information and the resulting three governance category scores. (Source: Refinitiv-ASSET4; Code: CGSCORE)	Yearly
EnvDisclosure	Number of indicators available for the company in the environmental pillar divided by the total number of indicators considered to obtain the score. (Source: Own elaboration from Refinitiv-ASSET 4)	Yearly
SocDisclosure	Number of indicators available for the company in the social pillar divided by the total number of indicators considered to obtain the score. (Source: Own elaboration from Refinitiv-ASSET 4)	Yearly
GovDisclosure	Number of indicators available for the company in the governance pillar divided by the total number of indicators considered to obtain the score. (Source: Own elaboration from Refinitiv-ASSET 4)	Yearly
Exogenous Variables		
PublicAttention	From 0 to 1, where 1 is the company that receives the most web searches in a given year and 0 the least. (Source: Own elaboration from the search volumne index of Google Trends).	Yearly
Industry	Name of the industry under which the equity is classified according to the FTSE/DJ Industry Classification Benchmark (ICB). (Source: Refinitiv-Datastream; Code: ICBIN).	Static
LogAssets	Represents the base 10 logarithm of the sum of total current assets. (Source: Own elaboration from Refinitiv-WorldScope; Code: WC02999)	Yearly
Leverage	Ratio between total liabilities and common equity. (Source: Own elaboration from Refinitiv-WorldScope; Codes: WC03351, WC03501)	Yearly
ROA	Return on Assets (Profitability Ratio). (Refinitiv-WorldScope; Code: WC08326)	Yearly
RD_Intensity	Represents the ratio between all direct and indirect costs related to the creation and development of new processes, techniques, applications and products with commercial possibilities and the company total assets. (Source: Own elaboration from Refinitiv-WorldScope; Codes: WC01201, WC02999)	Yearly

Table 3.2 clusters the variables in three groups. First, the variables used to perform the matching between Refinitiv and Google Trends; the second group comprises our dependent variables, that is, our measures of CSR performance and CSR disclosure; and the third group comprises our variable of interest, *PublicAttention*, and our set of control variables.

Table 3.3 Correlation matrix and variance inflation factor
This table shows the Pearson coefficient between the variables (1) to (12), where (1) to (7) are each of the dependent variables in equation 1 and variables (8) to (12) are the exogenous variables. The last row shows the variance inflation factor (VIF) between the exogenous variables (8 to 12). The *, and ** indicate statistical significance at the 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) ESGScore	1.00**											
(2) EnvScore	0.84**	1.00**										
(3) SocScore	0.88**	0.72**	1.00**									
(4) GovScore	0.65**	0.34**	0.32**	1.00**								
(5) EnvDisclosure	0.66**	0.69**	0.58**	0.33**	1.00**							
(6) SocDisclosure	0.60**	0.58**	0.55**	0.34**	0.77**	1.00**						
(7) GovDisclosure	0.37**	0.34**	0.32**	0.24**	0.49**	0.66**	1.00**					
(8) PublicAttention	0.24**	0.19**	0.30**	0.05**	0.08**	0.13**	0.04*	1.00**				
(9) LogAssets	0.33**	0.28**	0.30**	0.18**	0.26**	0.34**	0.21**	0.30**	1.00**			
(10) Leverage	0.07**	0.03	0.07**	0.03*	0.01	0.06**	0.02	0.12**	0.42**	1.00**		
(11) ROA	0.01	0.03	0.04*	-0.04*	-0.03	-0.06**	-0.03	0.12**	-0.37**	-0.26**	1.00**	
(12) RD_Intensity	0.09**	0.08**	0.14**	-0.02	0.16**	0.09**	0.02	0.12**	-0.19**	-0.15**	0.26**	1.00**
VIF								4,47	5,78	1,76	2,58	1,44

2.2 Econometric approach

Public Attention and CSR link

We analyze the influence of public attention on the performance and disclosure of CSR by using equation 1:

$$Score_{it} = \beta_0 + \beta_1 PublicAttention_{it} + \beta_2 logAssets_{it} + \beta_3 Leverage_{it} + \beta_4 ROA_{it} + \beta_4 RD\_Intensity_{it} + \varepsilon_{it} \quad \textbf{(1)}$$

where $Score_{it}$ refers to the *ESGScore, EnvScore, SocScore, GovScore, EnvDisclosure, SocDisclosure*, and *GovDisclosure* of each company *i* in each year *t*. *PublicAttention* is our proxy for public attention, *LogAssets* is a proxy for company size, *Leverage* is a proxy for the capital structure of the company, *RD_Intensity* controls for the company investment in R&D and ε is the error term. Our control variables are in line with the studies that analyze the drivers of CSR (Flammer, 2021; O'Sullivan et al., 2021).

Additionally, Table 3.3 presents the correlation matrix among the variables in equation 1 and the variance inflation factor for the independent variables.[7]

Public Attention and subsequent CSR

We also analyze how public attention affects subsequent firm CSR performance and disclosure. From an empirical perspective, the ideal experiment would be to randomly assign firms to a control group and to a treated group and compare their CSR performance and disclosure after the "treatment" (e.g., an intervention that increases the public attention to ta company). However, such an ideal experiment would be difficult and excessively expensive to implement in the field.

We proxy this "treatment" by identifying the firms that receive a shock of public attention. To this end we calculate the differences of the series of *PublicAttention* ($PublicAttention_t$ – $PublicAttention_{t-1}$) for each firm and detect 10% of the largest increases. These firms are the "treated" group and suffered a shock of public attention on *t* compared to *t-1*. Therefore, we expect that their level of CSR will increase in *t+1*. Next, we build a plausible counterfactual of how the CSR of the firm would evolve in the absence of the public attention shock. Thus, we match a control company that is as similar as possible to the treated one the year of the shock. Our pool of companies for the matching procedure is the remaining 90% of the observations, that is, those companies that do not suffer a shock of public attention. To construct the matched control group, we use the nearest neighbor algorithm based on seven company characteristics: the score analyzed in *t* ($Score_t$), the score increases in *t* ($\Delta Score_t = Score_t - Score_{t-1}$), the *PublicAttention, LogAssets, Leverage, ROA, and RD_Intensity* also in *t.* The seven variables are standardized (removing the mean and scaling to unit variance) to prevent variables with larger scales from dominating the matching procedure.[8]

Table 3.4 reports the mean of the matching variables for each CSR performance score analyzed: *ESGScore, EnvScore, SocScore and GovScore*. Table 3.5 provides the same information for the *EnvDisclosure, SocDisclosure,* and the *GovDisclsoure*. The *Score* variable as a matching metric ensures that treated and control companies have

[7] The variance inflation factor is below five for all the variables except for the LogAssets. However, this variable is not our variable of interest, and the variance inflation factor is far from 10, which is usually the cut-off for multicollinearity problems. Hence, there is no evidence for strong multicollinearity in our model specification.
[8] The metric of distance used in the matching procedure is the Minkowski distance with p=2, which is equivalent to the Euclidean metric.

similar levels of CSR the year of the shock. Similarly, the *ΔScore* ensures that the "pre-trend" is comparable. The use of *ROA* rules out concerns that the treated firms might have more slack resources. The *LogAssets* and *Leverage* ensure that both groups have the same access to capital markets. Moreover, the matching based on the year ensures that both groups were facing the same conditions of business environments. Both tables show the difference in mean t-test between the matching variables. We show that the treated and control firms are very similar because the null hypothesis of equal means cannot be rejected. Overall, this table confirms that treated and control firms are very similar that provides reliability to our counterfactual. After the matching procedure we perform equation 2 to test whether companies improve their CSR after a shock of legitimacy pressure:

$$\Delta Score_{it+1} = \alpha + Shock_{it} + X_{it} + \varepsilon_{it} \tag{2}$$

where $\Delta Score_{it+1}$ refers to the increase in the score of the company *i* in the year *t+1*, $Shock_{it}$ is a dummy variable ("treatment variable") that is one if the company suffered a shock of public attention in year *t* and zero otherwise, the X_{it} are the set of control variables used in the matching procedure, and ε is the error term.

Table 3.4 Matching for CSR performance

This table presents the descriptive statistics comparing the means of the treated group (shock of public attention) and the matched control firms. The first row shows each score analyzed, and the first column shows the matching variables, in which *Score* refers to the score indicated in the first row. The p-val columns show the p-value of the t-test for the difference in means between the two groups.

	ESGScore			**EnvScore**			**SocScore**			**GovScore**		
	Shock	**Control**	**p-val**	**Shock**	**Control**	**p-val**	**Shock**	**Control**	**p-val**	**Shock**	**Control**	**p-val**
Score	56.990	58.980	0.25	49.309	51.795	0.36	57.327	58.379	0.60	61.288	62.234	0.64
ΔScore	1.899	1.845	0.92	2.499	2.236	0.75	2.912	2.038	0.25	0.159	0.231	0.95
PublicAttention	0.470	0.479	0.70	0.470	0.479	0.66	0.470	0.476	0.80	0.470	0.471	0.98
LogAssets	7.364	7.334	0.50	7.364	7.329	0.44	7.364	7.306	0.18	7.364	7.341	0.63
Leverage	2.482	2.433	0.88	2.482	2.441	0.91	2.482	2.322	0.63	2.482	2.598	0.73
ROA	7.830	7.625	0.75	7.830	7.674	0.80	7.830	7.621	0.74	7.830	7.293	0.39
RD_Intensity	0.016	0.014	0.58	0.016	0.015	0.87	0.016	0.014	0.51	0.016	0.015	0.69
#	189	189		189	189		189	189		189	189	

Table 3.5 Matching for CSR disclosure

This table presents descriptive statistics comparing the means of the treated group (shock of public attention) and the matched control firms. The first row shows each score analyzed, and the first column shows the matching variables, in which *Score* refers to the score indicated in the first row. The p-val columns show the p-value of the t-test for the difference in means between the two groups.

	EnvDisclosure			SocDisclosure			GovDisclosure		
	Shock	**Control**	**p-val**	**Shock**	**Control**	**p-val**	**Shock**	**Control**	**p-val**
Score	0.542	0.543	0.91	0.461	0.461	0.96	0.764	0.769	0.42
ΔScore	0.013	0.011	0.66	0.009	0.006	0.36	0.011	0.009	0.71
PublicAttention	0.466	0.468	0.93	0.466	0.476	0.63	0.466	0.468	0.92
LogAssets	7.380	7.319	0.14	7.380	7.358	0.60	7.380	7.368	0.76
Leverage	2.495	2.606	0.74	2.495	2.576	0.81	2.495	2.386	0.74
ROA	7.694	7.503	0.74	7.694	7.443	0.67	7.694	7.169	0.36
RD_Intensity	0.015	0.015	0.86	0.015	0.013	0.39	0.015	0.015	0.79
#	207	207		207	207		207	207	

3. Results

3.1 Public attention and CSR performance

The results of equation 1 are reported in Table 3.6. Our variable of interest, *PublicAttention,* has a significant and positive influence on the CSR performance of the company (*ESGScore, EnvScore,* and *SocScore*). The positive and statistically significant influence of public attention on CSR remains regardless of the fixed effects considered. As a placebo check, we have applied the same analysis to the governance score to ensure that this relationship is not accidental. The absence of relationship between *PublicAttention* and *GovScore* support our results and the empirical design of the study because the governance pillar overlaps with traditional corporate governance issues that are materially different from the CSR issues (Liang & Renneboog, 2017). While some authors remove the influence of the governance pillar in their aggregate indicators of CSR (Dyck et al., 2019; Krüger, 2015) we have preferred to analyze all three ESG pillars for robustness purposes.

Table 3.6 Relation between CSR performance and public attention

This table presents the estimates from equation 1. The first row shows the dependent variable and the second row shows the model used. Model (1) does not include fixed effects, model (2) includes year dummies, model (3) includes year and industry dummies, and model (4) includes year and firm dummies. For each independent variable this table gives the coefficients, the significance, and the standard errors in parentheses. *, and ** indicate statistical significance at the 5% and 1% levels, respectively.

Dep Var	**ESGScore**				**EnvScore**				**SocScore**				**GovScore**			
	(1)	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	-32.02**	-30.27**	-26.13**	-36.83**	-75.31**	-72.47**	-91.49**	-94.92**	-36.03**	-33.74**	-32.36**	-47.20**	7.44	7.68	26.37**	57.93**
	(4.435)	(4.364)	(4.654)	(10.040)	(6.951)	(6.936)	(7.033)	(13.798)	(4.993)	(4.900)	(5.269)	(11.566)	(5.426)	(5.437)	(5.938)	(17.122)
Public Attention	7.06**	7.51**	13.07**	10.38*	6.82**	7.33**	13.63**	24.74**	11.96**	12.49**	17.01**	12.09*	-0.52	-0.24	7.35**	-3.90
	(1.066)	(1.039)	(1.186)	(4.372)	(1.670)	(1.651)	(1.793)	(6.008)	(1.200)	(1.166)	(1.343)	(5.036)	(1.304)	(1.294)	(1.514)	(7.455)
LogAssets	11.45**	10.58**	11.26**	9.42**	16.32**	15.36**	19.93**	14.29**	11.70**	10.66**	11.71**	9.82**	7.26**	6.73**	5.16**	3.20
	(0.613)	(0.601)	(0.637)	(1.428)	(0.960)	(0.956)	(0.963)	(1.963)	(0.690)	(0.675)	(0.721)	(1.645)	(0.750)	(0.749)	(0.813)	(2.435)
Leverage	-0.29**	-0.27**	-0.03	-0.03	-0.63**	-0.62**	0.15	0.11	-0.23*	-0.22*	-0.03	-0.02	-0.23*	-0.22*	-0.03	-0.18
	(0.085)	(0.083)	(0.083)	(0.061)	(0.133)	(0.131)	(0.126)	(0.084)	(0.096)	(0.093)	(0.094)	(0.070)	(0.104)	(0.103)	(0.106)	(0.104)
ROA	0.25**	0.23**	0.19**	0.02	0.41**	0.38**	0.35**	0.03	0.30**	0.27**	0.18**	-0.01	0.08	0.07	0.13*	0.01
	(0.051)	(0.050)	(0.049)	(0.035)	(0.080)	(0.079)	(0.074)	(0.049)	(0.057)	(0.056)	(0.055)	(0.041)	(0.062)	(0.062)	(0.062)	(0.060)
RD_Intensity	55.46**	52.76**	60.90**	12.73	77.52**	74.05**	95.68**	46.41	83.74**	80.37**	88.65**	-5.63	2.13	1.07	-4.28	-4.63
	(8.511)	(8.291)	(10.057)	(23.110)	(13.337)	(13.178)	(15.198)	(31.761)	(9.580)	(9.309)	(11.386)	(26.623)	(10.412)	(10.329)	(12.833)	(39.410)
Year Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Ind Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Firm Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
#	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322	3,322
R2	15.49%	20.09%	28.57%	86.30%	11.67%	14.09%	30.57%	88.99%	17.26%	22.15%	29.25%	85.95%	3.29%	5.19%	11.08%	69.53%

The other coefficients of our regression are also consistent with other studies. Drempetic et al. (2020) also show that the size of the company has a positive effect on the CSR performance of the company. Three of the four models show that R&D intensity and the return on assets influence the company's CSR performance. However, in the model that controls for firm unobserved heterogeneity shows no significant relation. These findings are in line with the existing literature. McWilliams and Siegel (2000) underline the importance of adding the R&D intensity in the study of the relationship between CSR performance and financial performance to avoid omitted variable bias. This variable is highly correlated with CSR performance (Awaysheh et al., 2020); thus, our results also support this correlation. On the other hand, the literature also underlines the endogeneity problem about the direction of the relation between CSR performance and financial performance (Flammer, 2015; Soytas et al., 2019; Zhao and Murrell, 2016). That is, "doing good leads to doing well" or "doing well leads to doing good" (Barnett et al., 2020; Krüger, 2015). Our study does not explicitly address this issue, but our results suggest that "doing well leads to doing good".

Table 3.7 Relation between CSR disclosure and public attention.
This table presents the estimates from equation 1. The first row shows the dependent variable and the second row shows the model used. Model (1) does not include fixed effects, model (2) includes year dummies, model (3) includes year and industry dummies, and model (4) includes year and firm dummies. For each independent variable this table gives the coefficients, the significance, and the standard errors in parentheses. *, and ** indicate statistical significance at the 5% and 1% levels, respectively.

	EnvDisclosure				**SocDisclosure**				**GovDisclosure**			
	(1)	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	8.76* (3.693)	9.52** (3.538)	13.66** (3.825)	-1.42 (14.144)	18.81** (2.698)	19.01** (2.548)	21.06** (2.819)	3.48 (11.519)	50.48** (4.378)	50.19** (4.052)	52.31** (4.549)	13.34 (18.716)
Public Attention	-0.16 (0.888)	-0.15 (0.843)	3.44** (0.974)	-2.65 (6.074)	1.29* (0.649)	1.25* (0.607)	3.04** (0.718)	-6.38 (4.946)	1.01 (1.053)	1.01 (0.965)	2.36* (1.159)	-11.17 (8.037)
LogAssets	5.63** (0.510)	5.55** (0.487)	6.10** (0.523)	6.72** (2.005)	3.21** (0.373)	3.23** (0.351)	3.42** (0.385)	6.47** (1.633)	2.84** (0.605)	2.75** (0.558)	2.79** (0.622)	9.73** (2.653)
Leverage	-0.28** (0.069)	-0.23** (0.066)	-0.04 (0.067)	-0.07 (0.085)	-0.16** (0.051)	-0.12* (0.047)	-0.01 (0.050)	-0.05 (0.069)	-0.20* (0.082)	-0.14 (0.075)	-0.11 (0.080)	-0.07 (0.113)
ROA	0.17** (0.042)	0.13** (0.040)	0.14** (0.040)	0.10* (0.050)	0.12** (0.031)	0.09** (0.029)	0.10** (0.029)	0.10* (0.041)	0.21** (0.050)	0.16** (0.046)	0.15** (0.048)	0.17* (0.066)
RD_Intensity	51.79** (7.082)	54.84** (6.716)	70.26** (8.298)	-23.45 (32.419)	16.51** (5.174)	19.18** (4.836)	21.91** (6.116)	-4.80 (26.402)	4.07 (8.395)	8.54 (7.691)	10.59 (9.870)	-50.03 (42.898)
Year Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Ind Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Firm Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
#	3,461	3,461	3,461	3,461	3,461	3,461	3,461	3,461	3,461	3,461	3,461	3,461
R2	5.12%	14.98%	21.91%	53.71%	3.28%	15.81%	19.01%	41.38%	1.21%	17.37%	18.14%	39.94%

3.2 Public Attention and CSR disclosure

Table 3.7 shows the results of the influence of public attention on the company's CSR disclosure (*EnvDisclosure, SocDisclosure,* and *GovDisclosure*). Our variable of interest is not statistically significantly different from zero in most of the models used to test this relation. Hence, we do not find evidence of a link between public attention and a company's CSR disclosure. However, the positive relation between company size and CSR disclosure supports other studies that indicate a positive relation between size and environmental, social, or governance disclosure (Ali et al., 2017; Tamimi and Sebastianelli, 2017). These findings are similar to Shabana et al. (2016) who conclude that the media coverage of the company does not influence the probability of publishing CSR reports.

Table 3.8 Shock of public attention and subsequent CSR performance

This table presents the estimates from equation 2. The first row shows the dependent variable, and the second row shows the model used. Model (1) does not include fixed effects, model (2) includes year and industry dummies, and model (3) includes year and industry dummies and the matching variables of Table 3.3. For each independent variable this table gives the coefficients, the significance, and the standard errors in parentheses. *, and ** denotes statistical significance at the 5% and 1% levels, respectively.

Dep Var	**ΔESGScore$_{t+1}$**			**ΔEnvScore$_{t+1}$**			**ΔSocScore$_{t+1}$**			**ΔGovScore$_{t+1}$**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Intercept	1.00* (0.450)	-0.95 (1.722)	-0.80 (7.112)	1.42* (0.629)	-0.33 (2.169)	-7.07 (9.708)	1.22* (0.562)	-1.71 (2.020)	-18.03* (8.794)	1.75 (0.907)	-8.48* (3.416)	25.22* (12.664)
Shock$_t$	1.60* (0.636)	1.66** (0.639)	1.41* (0.617)	2.29* (0.890)	2.41** (0.900)	1.98* (0.857)	1.89* (0.795)	2.03* (0.795)	1.76* (0.779)	-1.27 (1.282)	-1.36 (1.291)	-1.28 (1.167)
Score$_t$			-0.12** (0.021)			-0.12** (0.019)			-0.11** (0.023)			-0.22** (0.033)
ΔScore$_t$			-0.05 (0.058)			0.11* (0.055)			-0.01 (0.055)			-0.21** (0.053)
Public Attention$_t$			-0.05 (1.648)			0.62 (2.374)			0.35 (2.177)			2.00 (3.235)
LogAssets$_t$			0.98 (0.970)			1.84 (1.337)			2.99* (1.210)			-2.57 (1.695)
Leverage$_t$			-0.13 (0.103)			-0.23 (0.140)			0.04 (0.130)			-0.00 (0.194)
ROA$_t$			0.06 (0.064)			-0.03 (0.089)			0.21** (0.080)			-0.08 (0.122)
RD Intensity$_t$			-4.26 (15.271)			38.16* (19.246)			11.66 (17.542)			-65.29* (27.842)
Year Effs	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind Effs	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
#	378	378	378	378	378	378	378	378	378	378	378	378
R2	1.65%	7.12%	16.28%	1.73%	6.14%	17.57%	1.49%	8.40%	15.79%	0.26%	6.38%	25.76%

3.3 Public attention shock and subsequent CSR performance

In Table 3.8 we present the results of equation 2. We find that *ESGScore*, *EnvScore* and *SocScore* go up after the public attention shock. Specifically, the *ESGScore* increases around 1.6 points, the EnvScore around 2.3 points, and the *SocScore* around 2 points. As a robustness check, we show that the *GovScore* is not influenced by the public attention shock. These findings are consistent with the legitimation pressure argument. Companies that face growing visibility and public scrutiny improve their CSR performance to create a generalized perception that their actions are desirable and appropriate within a socially constructed system of norms and values. Table 3.8 also shows that companies that have a high score at *t* have more difficulties in improving their score at *t+1*.

Table 3.9 Shock of public attention and subsequent CSR disclosure

This table presents the estimates from equation 2. The first row shows the dependent variable, and the second row shows the model used. Model (1) does not include fixed effects, model (2) includes year and industry dummies, and model (3) includes year and industry dummies and the matching variables of Table 3.3. For each independent variable this table gives the coefficients, the significance, and the standard errors in parentheses. *, and ** denotes statistical significance at the 5% and 1% levels, respectively.

Dep Var	**ΔEnvDisclosure$_{t+1}$**			**ΔSocDisclosure$_{t+1}$**			**ΔGovDisclosure$_{t+1}$**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Intercept	-0.03** (0.011)	0.06 (0.039)	0.30* (0.150)	-0.04** (0.009)	0.01 (0.032)	0.30* (0.138)	-0.07** (0.016)	0.00 (0.055)	0.32 (0.241)
Shock$_t$	-0.00 (0.015)	-0.01 (0.013)	-0.01 (0.013)	0.01 (0.013)	0.00 (0.011)	0.01 (0.011)	0.02 (0.023)	0.02 (0.019)	0.02 (0.020)
Score$_t$			-0.07 (0.089)			0.02 (0.173)			0.09 (0.206)
ΔScore$_t$			0.05 (0.191)			0.09 (0.283)			-0.22 (0.239)
Public Attention$_t$			0.11** (0.035)			0.05 (0.031)			0.09 (0.053)
LogAssets$_t$			-0.03 (0.021)			-0.04* (0.018)			-0.05 (0.028)
Leverage$_t$			0.00 (0.002)			-0.00 (0.002)			0.00 (0.003)
ROA$_t$			-0.00 (0.001)			-0.00 (0.001)			-0.00 (0.002)
RD Intensity$_t$			0.30 (0.301)			0.06 (0.265)			-0.23 (0.436)
Year Effs	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind Effs	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
#	414	414	414	414	414	414	414	414	414
R2	0.01%	30.73%	32.91%	0.04%	32.49%	34.03%	0.16%	32.51%	33.76%

3.4 Public attention shock and subsequent CSR disclosure

In Table 3.9, we present the results from equation 2 for the *EnvDisclosure, SocDisclosure and GovDisclosure*. Once again, we do not find any evidence supporting the idea that public attention can influence the subsequent CSR disclosure.

4. Discussion and conclusions

The existing measures of retail attention based on the Google Trends' SVI provide a relative indicator of attention that is not comparable among different companies. That is, the individual SVI is not able to identify which companies are most visible. In this paper, we have proposed a new method of using Google Trends to overcome the limitations of the SVI. By performing several comparisons of the individual SVIs, we rank companies according to the number of web searches to measure attention. With this indicator and using the legitimacy theory as a foundation, we shed light on the relation between public attention and CSR. While the studies that analyze this relationship have used indirect measures of attention and short time periods or small samples, we use a direct measure of attention for S&P 500 companies over the period 2012 to 2020 (Ali et al., 2017; Aouadi and Marsat, 2018; Hou, 2019; Rosati and Faria, 2019; Udayasankar, 2008). Thus, for the first time in the literature, we proxy the visibility and the legitimation pressure of the company using the number of web searches in Google.

First, as is stated by the literature, we show a positive relation between public attention and the CSR performance of the company. Firms with higher visibility use CSR as a source of legitimacy to align their actions with societal values and norms. Moreover, following a quasi-experimental approach, we show that companies improve their CSR performance after a shock of attention in the previous year. This finding confirms that companies react to an increase in public scrutiny by improving their CSR performance. These analyses are carried out with traditional CSR measures, such as environmental performance and social performance. As a robustness check, the same analysis is performed on the governance performance, and it shows no relation with public attention. This result removes concerns about spurious relationships and confirms our main findings because there is no evidence supporting the idea that public scrutiny influences the company's corporate governance. On the other hand, when we analyze the CSR disclosure of the company, we do not find any evidence suggesting a relationship between public attention and CSR disclosure. Our findings show that companies in order to legitimize their activity will increase their environmental and social performance (CSR performance) rather than increase their transparency (CSR disclosure).

Our research contributes to the literature on the factors that influence the CSR of companies. The analysis of these factors is important because socially responsible investors derive a non-financial utility from investing in companies that meet high environmental and social standards (Gutsche and Ziegler, 2019; Renneboog et al., 2008), and there is a growing demand for socially responsible products (Ammann et al., 2019; Bauer et al., 2021). However, socially responsible investors may overweight some companies that are not the best to satisfy their non-financial utility due to some biases in ESG ratings. For example, Drempetic et al. (2020) criticize the relation between company size and ESG ratings, and other authors show that developed European countries obtain the best ESG ratings (Dyck et al., 2019; Liang and Renneboog, 2017). Our research shows that companies under greater public scrutiny obtain higher ESG ratings. These correlations may show that ESG ratings are not the best indicators to help socially responsible investors in their investment decisions. The methodology used for measuring public attention also may contribute to several disciplines that use Google Trends as a source of information, such as finance (Choi et al., 2020; Da et al., 2015), marketing (Chandrasekaran et al., 2018; Hu et al., 2014), or even medicine (Arora et al., 2019; Flanagan et al., 2021).

References

Ali, W., Frynas, J.G., Mahmood, Z., 2017. Determinants of corporate social responsibility (CSR) disclosure in developed and developing countries: A literature review. Corp. Soc. Responsib. Environ. Manag. 24, 273–294. https://doi.org/https://doi.org/10.1002/csr.1410

Ammann, M., Bauer, C., Fischer, S., Müller, P., 2019. The impact of the Morningstar Sustainability Rating on mutual fund flows. Eur. Financ. Manag. 25, 520–553. https://doi.org/https://doi.org/10.1111/eufm.12181

Aouadi, A., Marsat, S., 2018. Do ESG Controversies Matter for Firm Value? Evidence from International Data. J. Bus. Ethics 151, 1027–1047. https://doi.org/10.1007/s10551-016-3213-8

Arora, V.S., McKee, M., Stuckler, D., 2019. Google Trends: Opportunities and limitations in health and health policy research. Health Policy (New. York). 123, 338–341. https://doi.org/https://doi.org/10.1016/j.healthpol.2019.01.001

Awaysheh, A., Heron, R.A., Perry, T., Wilson, J.I., 2020. On the relation between corporate social responsibility and financial performance. Strateg. Manag. J. 41, 965–987. https://doi.org/https://doi.org/10.1002/smj.3122

Baldini, M., Maso, L.D., Liberatore, G., Mazzi, F., Terzani, S., 2018. Role of Country- and Firm-Level Determinants in Environmental, Social, and Governance Disclosure. J. Bus. Ethics 150, 79–98. https://doi.org/10.1007/s10551-016-3139-1

Barnett, M.L., Henriques, I., Husted, B.W., 2020. Beyond Good Intentions: Designing CSR Initiatives for Greater Social Impact. J. Manage. 46, 937–964. https://doi.org/10.1177/0149206319900539

Bauer, R., Ruof, T., Smeets, P., 2021. Get real! Individuals prefer more sustainable investments. Rev. Financ. Stud. 34, 3976–4043. https://doi.org/10.1093/rfs/hhab037

Ben-Rephael, A., Da, Z., Israelsen, R.D., 2017. It Depends on Where You Search: Institutional Investor Attention and Underreaction to News. Rev. Financ. Stud. 30, 3009–3047. https://doi.org/10.1093/rfs/hhx031

Berrone, P., Fosfuri, A., Gelabert, L., 2017. Does greenwashing pay off? Understanding the relationship between environmental actions and environmental legitimacy. J. Bus. Ethics 144, 363–379. https://doi.org/10.1007/s10551-015-2816-9

Boermans, M.A., Galema, R., 2019. Are pension funds actively decarbonizing their portfolios? Ecol. Econ. 161, 50–60. https://doi.org/https://doi.org/10.1016/j.ecolecon.2019.03.008

Campbell, D., Slack, R., 2006. Public visibility as a determinant of the rate of corporate charitable donations. Bus. Ethics A Eur. Rev. 15, 19–28. https://doi.org/https://doi.org/10.1111/j.1467-8608.2006.00425.x

Carroll, A.B., 1999. Corporate Social Responsibility: Evolution of a Definitional Construct. Bus. Soc. 38, 268–295. https://doi.org/10.1177/000765039903800303

Chandrasekaran, D., Srinivasan, R., Sihi, D., 2018. Effects of offline ad content on online brand search: insights from super bowl advertising. J. Acad. Mark. Sci. 46, 403–430. https://doi.org/10.1007/s11747-017-0551-8

Chin, M.K., Hambrick, D.C., Treviño, L.K., 2013. Political Ideologies of CEOs: The Influence of Executives' Values on Corporate Social Responsibility. Adm. Sci. Q. 58, 197–232. https://doi.org/10.1177/0001839213486984

Chiu, S.-C., Sharfman, M., 2009. Legitimacy, visibility, and the antecedents of corporate social performance: An investigation of the instrumental perspective. J. Manage. 37, 1558–1585. https://doi.org/10.1177/0149206309347958

Choi, D., Gao, Z., Jiang, W., 2020. Attention to Global Warming. Rev. Financ. Stud. 33, 1112–1145. https://doi.org/10.1093/rfs/hhz086

Colaco, H.M.J., De Cesari, A., Hegde, S.P., 2017. Retail Investor Attention and IPO Valuation. Eur. Financ. Manag. 23, 691–727. https://doi.org/https://doi.org/10.1111/eufm.12113

Cormier, D., Magnan, M., 2015. The Economic Relevance of Environmental Disclosure and its Impact on Corporate Legitimacy: An Empirical Investigation. Bus. Strateg. Environ. 24, 431–450. https://doi.org/https://doi.org/10.1002/bse.1829

Cunha, F.A.F. de S., de Oliveira, E.M., Orsato, R.J., Klotzle, M.C., Cyrino Oliveira, F.L., Caiado, R.G.G., 2020. Can sustainable investments outperform traditional benchmarks? Evidence from global stock markets. Bus. Strateg. Environ. 29, 682–697. https://doi.org/10.1002/bse.2397

Da, Z., Engelberg, J., Gao, P., 2015. The Sum of All FEARS Investor Sentiment and Asset Prices. Rev. Financ. Stud. 28, 1–32. https://doi.org/10.1093/rfs/hhu072

Da, Z., Engelberg, J., Gao, P., 2011. In Search of Attention. J. Finance 66, 1461–1499. https://doi.org/https://doi.org/10.1111/j.1540-6261.2011.01679.x

Dai, R., Liang, H., Ng, L., 2021. Socially responsible corporate customers. J. financ. econ. 142, 598–626. https://doi.org/https://doi.org/10.1016/j.jfineco.2020.01.003

Demirbag, M., Wood, G., Makhmadshoev, D., Rymkevich, O., 2017. Varieties of CSR: Institutions and Socially Responsible Behaviour. Int. Bus. Rev. 26, 1064–1074. https://doi.org/https://doi.org/10.1016/j.ibusrev.2017.03.011

Di Giuli, A., Kostovetsky, L., 2014. Are red or blue companies more likely to go green? Politics and corporate social responsibility. J. financ. econ. 111, 158–180. https://doi.org/https://doi.org/10.1016/j.jfineco.2013.10.002

Dimpfl, T., Jank, S., 2016. Can Internet Search Queries Help to Predict Stock Market Volatility? Eur. Financ. Manag. 22, 171–192. https://doi.org/https://doi.org/10.1111/eufm.12058

Dong, H., Gil-Bazo, J., 2020. Sentiment stocks. Int. Rev. Financ. Anal. 72, 101573. https://doi.org/https://doi.org/10.1016/j.irfa.2020.101573

Drempetic, S., Klein, C., Zwergel, B., 2020. The influence of firm size on the ESG score: Corporate sustainability ratings under review. J. Bus. Ethics 167, 333–360. https://doi.org/10.1007/s10551-019-04164-1

Dyck, A., Lins, K. V, Roth, L., Wagner, H.F., 2019. Do institutional investors drive corporate social responsibility? International evidence. J. financ. econ. 131, 693–714. https://doi.org/https://doi.org/10.1016/j.jfineco.2018.08.013

El Ouadghiri, I., Guesmi, K., Peillex, J., Ziegler, A., 2021. Public Attention to Environmental Issues and Stock Market Returns. Ecol. Econ. 180, 106836. https://doi.org/https://doi.org/10.1016/j.ecolecon.2020.106836

Fang, L., Peress, J., 2009. Media Coverage and the Cross-section of Stock Returns. J. Finance 64, 2023–2052. https://doi.org/https://doi.org/10.1111/j.1540-6261.2009.01493.x

Flammer, C., 2021. Corporate green bonds. J. financ. econ. 142, 499–516. https://doi.org/https://doi.org/10.1016/j.jfineco.2021.01.010

Flammer, C., 2015. Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. Manage. Sci. 61, 2549–2568. https://doi.org/10.1287/mnsc.2014.2038

Flammer, C., Toffel, M.W., Viswanathan, K., 2021. Shareholder activism and firms' voluntary disclosure of climate change risks. Strateg. Manag. J. 42, 1850–1879. https://doi.org/https://doi.org/10.1002/smj.3313

Flanagan, R., Kuo, B., Staller, K., 2021. Utilizing Google Trends to Assess Worldwide Interest in Irritable Bowel Syndrome and Commonly Associated Treatments. Dig. Dis. Sci. 66, 814–822. https://doi.org/10.1007/s10620-020-06290-7

Gutsche, G., Ziegler, A., 2019. Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. J. Bank. Financ. 102, 193–214. https://doi.org/https://doi.org/10.1016/j.jbankfin.2019.03.007

Han, L., Lv, Q., Yin, L., 2017. Can investor attention predict oil prices? Energy Econ. 66, 547–558. https://doi.org/https://doi.org/10.1016/j.eneco.2017.04.018

Hart, S.L., 1995. A natural-resource-based view of the firm. Acad. Manag. Rev. 20, 986–1014. https://doi.org/10.5465/amr.1995.9512280033

Hartzmark, S.M., Sussman, A.B., 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. J. Finance 74, 2789–2837. https://doi.org/https://doi.org/10.1111/jofi.12841

Hawn, O., Chatterji, A.K., Mitchell, W., 2018. Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). Strateg. Manag. J. 39, 949–976. https://doi.org/https://doi.org/10.1002/smj.2752

Hörisch, J., Johnson, M.P., Schaltegger, S., 2015. Implementation of sustainability management and company size: A knowledge-based view. Bus. Strateg. Environ. 24, 765–779. https://doi.org/https://doi.org/10.1002/bse.1844

Hou, T.C.-T., 2019. The relationship between corporate social responsibility and sustainable financial performance: firm-level evidence from Taiwan. Corp. Soc. Responsib. Environ. Manag. 26, 19–28. https://doi.org/https://doi.org/10.1002/csr.1647

Hu, Y., Du, R.Y., Damangir, S., 2014. Decomposing the Impact of Advertising: Augmenting Sales with Online Search Data. J. Mark. Res. 51, 300–319. https://doi.org/10.1509/jmr.12.0215

Humphrey, J.E., Li, Y., 2021. Who goes green: Reducing mutual fund emissions and its consequences. J. Bank. Financ. 126, 106098. https://doi.org/https://doi.org/10.1016/j.jbankfin.2021.106098

Hwang, C.Y., Titman, S., Wang, Y., 2021. Investor tastes, corporate behavior, and stock returns: An analysis of corporate social responsibility. Manage. Sci. https://doi.org/10.1287/mnsc.2021.4179

Jiang, L., Liu, J., Peng, L., Wang, B., 2022. Investor Attention and Asset Pricing Anomalies*. Rev. Financ. 26, 563–593. https://doi.org/10.1093/rof/rfab032

Kim, S., Lee, G., Kang, H.-G., 2021. Risk management and corporate social responsibility. Strateg. Manag. J. 42, 202–230. https://doi.org/https://doi.org/10.1002/smj.3224

Krüger, P., 2015. Corporate goodness and shareholder wealth. J. financ. econ. 115, 304–329. https://doi.org/https://doi.org/10.1016/j.jfineco.2014.09.008

Liang, H. ao, Renneboog, L., 2017. On the foundations of corporate social responsibility. J. Finance 72, 853–910. https://doi.org/https://doi.org/10.1111/jofi.12487

Liu, Y., Tsyvinski, A., 2021. Risks and Returns of Cryptocurrency. Rev. Financ. Stud. 34, 2689–2727. https://doi.org/10.1093/rfs/hhaa113

McWilliams, A., Siegel, D., 2000. Corporate social responsibility and financial performance: Correlation or misspecification? Strateg. Manag. J. 21, 603–609.

https://doi.org/10.1002/(SICI)1097-0266(200005)21:5<603::AID-SMJ101>3.0.CO;2-3

O'Sullivan, D., Zolotoy, L., Fan, Q., 2021. CEO early-life disaster experience and corporate social performance. Strateg. Manag. J. 42, 2137–2161. https://doi.org/https://doi.org/10.1002/smj.3293

Orlitzky, M., 2001. Does firm size comfound the relationship between corporate social performance and firm financial performance? J. Bus. Ethics 33, 167–180. https://doi.org/10.1023/A:1017516826427

Porter, M.E., Kramer, M.R., 2006. Strategy and society: The link between competitive advantage and corporate social responsibility. Harv. Bus. Rev. 84, 78–92.

Qadan, M., Nama, H., 2018. Investor sentiment and the price of oil. Energy Econ. 69, 42–58. https://doi.org/https://doi.org/10.1016/j.eneco.2017.10.035

Renneboog, L., Ter Horst, J., Zhang, C., 2008. Socially responsible investments: Institutional aspects, performance, and investor behavior. J. Bank. Financ. 32, 1723–1742. https://doi.org/https://doi.org/10.1016/j.jbankfin.2007.12.039

Rosati, F., Faria, L.G.D., 2019. Business contribution to the Sustainable Development Agenda: Organizational factors related to early adoption of SDG reporting. Corp. Soc. Responsib. Environ. Manag. 26, 588–597. https://doi.org/https://doi.org/10.1002/csr.1705

Schreck, P., Raithel, S., 2015. Corporate Social Performance, Firm Size, and Organizational Visibility: Distinct and Joint Effects on Voluntary Sustainability Reporting. Bus. Soc. 57, 742–778. https://doi.org/10.1177/0007650315613120

Servaes, H., Tamayo, A., 2013. The impact of corporate social responsibility on firm value: The role of customer awareness. Manage. Sci. 59, 1045–1061. https://doi.org/10.1287/mnsc.1120.1630

Shabana, K.M., Buchholtz, A.K., Carroll, A.B., 2016. The Institutionalization of Corporate Social Responsibility Reporting. Bus. Soc. 56, 1107–1135. https://doi.org/10.1177/0007650316628177

Soytas, M.A., Denizel, M., Durak Usar, D., 2019. Addressing endogeneity in the causal relationship between sustainability and financial performance. Int. J. Prod. Econ. 210, 56–71. https://doi.org/https://doi.org/10.1016/j.ijpe.2019.01.016

Suchman, M.C., 1995. Managing Legitimacy: Strategic and Institutional Approaches. Acad. Manag. Rev. 20, 571–610. https://doi.org/10.5465/amr.1995.9508080331

Surroca, J., Tribó, J.A., Waddock, S., 2010. Corporate responsibility and financial performance: the role of intangible resources. Strateg. Manag. J. 31, 463–490. https://doi.org/https://doi.org/10.1002/smj.820

Tamimi, N., Sebastianelli, R., 2017. Transparency among S&P 500 companies: an analysis of ESG disclosure scores. Manag. Decis. 55, 1660–1680. https://doi.org/10.1108/MD-01-2017-0018

Udayasankar, K., 2008. Corporate social responsibility and firm size. J. Bus. Ethics 83, 167–175. https://doi.org/10.1007/s10551-007-9609-8

Urquhart, A., 2018. What causes the attention of Bitcoin? Econ. Lett. 166, 40–44. https://doi.org/https://doi.org/10.1016/j.econlet.2018.02.017

Waddock, S.A., Graves, S.B., 1997. The corporate social performance-financial performance link. Strateg. Manag. J. 18, 303–319. https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.0.CO;2-G

Zhao, X., Murrell, A.J., 2016. Revisiting the corporate social performance-financial performance link: A replication of Waddock and Graves. Strateg. Manag. J. 37, 2378–2388. https://doi.org/https://doi.org/10.1002/smj.2579

Appendix

Table A3.10 Descriptive statistics

This table shows descriptive statistics including the average, standard deviation, minimum, quartiles and maximum value for each variable used in the study.

	mean	std	min	25%	50%	75%	max
ESGScore	57.75	17.59	2.46	45.46	59.68	70.89	94.04
EnvScore	50.75	27.00	0.00	30.29	54.55	73.86	98.55
SocScore	59.08	19.99	0.26	44.46	60.90	74.69	98.12
GovScore	60.45	20.07	0.45	46.59	63.30	76.03	99.41
EnvDisclosure	0.51	0.14	0.00	0.44	0.52	0.60	0.80
SocDisclosure	0.44	0.10	0.00	0.42	0.45	0.49	0.63
GovDisclosure	0.73	0.17	0.00	0.68	0.79	0.81	0.87
PublicAttention	0.50	0.29	0.00	0.25	0.50	0.75	1.00
LogAssets	7.35	0.57	5.99	6.94	7.27	7.67	9.53
Leverage	2.75	3.72	-5.40	0.93	1.67	3.16	16.14
ROA	7.34	6.30	-6.36	3.32	6.34	10.62	23.70
RD_Intensity	0.02	0.03	0.00	0.00	0.00	0.02	0.14

Chapter 4.

The limited role of sustainability in mutual fund investor decisions: A machine learning approach

Synopsis

Despite the growth in the supply of socially responsible investment products, the weight of environmental, social and governance (ESG) factors in the decisions of mutual fund investors remains under-researched. We conducted a study relating fund flows to past returns, ESG performance and other financial variables using data from 4,237 US mutual funds from 2015 to 2021. First, we aimed to assess the importance of ESG performance in investment decisions. Next, we studied whether ESG performance is increasingly important or has reached its limit. Finally, we developed decisional models to predict the flows raised by each investment fund, its financial return and ESG performance. We used logistic regression, neural networks, random forest and gradient boosting decision trees. We found that the investors consider ESG performance, but the factors that matter most are past growth, mutual fund fees and past returns. Our models predicted the money raised by the funds, obtaining accuracy rates of around 70%. In addition to confirming that "past financial return does not guarantee future financial return," we found that "past ESG performance guarantees future ESG performance," which may be of interest to socially responsible investors.

1. Introduction

Conventional portfolio theory states that investments should be made based on risk-adjusted financial returns (Fama, 1970; Mansour et al., 2019; Markowitz, 1952; Zopounidis et al., 2015), and this is how rational investors should make decisions in financial markets (Miller and Modigliani, 1961). However, many investment decisions are driven by other motivations and investors' cognitive biases, as behavioral finance theories proved (Kahneman and Tversky, 1979; Thaler, 1980). Socially responsible investing (SRI) provides another example of going beyond the risk-return trade-off, as these investors incorporate environmental, social and governance (ESG) concerns into their decisions (Bilbao-Terol et al., 2012; Calvo et al., 2016; Pedersen et al., 2021). Therefore, portfolio selection progressively requires multi-criteria decision support methods (Aouni et al., 2018; Chen et al., 2021; Li et al., 2022). Our study aimed to analyze mutual fund investment decisions by comparing the importance of past returns (and other variables) versus ESG performance, to study whether ESG performance is increasingly important, and to develop decision models that predict whether a mutual fund will attract money, as well as its financial return and social performance.

The paper addresses four research questions. First, we explored the factors that mutual fund investors take into account. It is particularly interesting to study whether investors consider ESG performance to be more or less important than past fund returns. The debate on financial market decisions has a long pedigree. The hypothesis of efficient financial markets with rational investors dominated financial theory (Fama, 1970). However, prospect theory showed the imperfections of financial markets and the inconsistencies of decision-makers (Barberis, 2013). The nonrationality of decision-makers favored the development of financial decision support systems, which have the advantage of not being affected by human emotions (Bhandari et al., 2008). Previous research identified the importance of past returns as a factor that investors look for when choosing a mutual fund (Guercio and Tkac, 2008; Sirri and Tufano, 1998). Other studies examined relevant aspects of investors' decisions, such as mutual fund fees (Servaes and Sigurdsson, 2022), risk aversion (Dorn and Huberman, 2010), herd behavior (Nofsinger and Sias, 1999), and investment style (Cremers et al., 2019). Some researchers studied the influence of social aspects on investment decisions (Bauer et al., 2021; Bollen, 2007; Reboredo and Otero, 2021; Renneboog et al., 2011), concluding that socially responsible investors may behave differently from other investors.

Environmentally and socially conscious investors are even more willing to sacrifice financial returns to invest in sustainable investment products than their counterparts (Gutsche and Ziegler, 2019; Riedl and Smeets, 2017). This may explain why SRI flows are less sensitive to past negative returns than conventional ones (Renneboog et al., 2011). Moreover, some studies suggest that the volatility of SR investments is significantly lower than that of conventional assets (Albuquerque et al., 2018; Bollen, 2007). Previous studies paid attention to the relationship between labeling a fund as sustainable and the flows it receives (Ammann et al., 2019; Hartzmark and Sussman, 2019). Despite these studies, the weight of social considerations in mutual fund investment decisions remains unclear. Our study contributes to the growing literature by comparing the relative importance of past returns and other financial variables versus ESG performance. We found that investors take ESG into account, but past returns trump sustainability in mutual fund investment decisions. Mutual fund fees and past growth are also more important than ESG performance.

Society's concern for ESG issues has taken hold and investors have also begun to value sustainability when choosing mutual funds (Bauer et al., 2021). Our second research question was whether the increase in ESG concerns that society has been experiencing in recent years translated into a greater weight of ESG in investment decisions. Motivation could be social, but also financial if there is a positive relationship between social and financial performance. The mantra "doing good leads to doing well" is heard so often that it seems reasonable to expect many investors to include ESG criteria when selecting funds. However, empirical studies showed mixed results (Badía et al., 2020; Flammer, 2021; Galema et al., 2008; Halbritter and Dorfleitner, 2015; Hawn et al., 2018; Hong and Kacperczyk, 2009; Krüger, 2015; Muñoz et al., 2014). Flammer (2021) concluded that the market responds positively to the announcement of green bond issuance, while Hawn et al. (2018) found that investors punished companies that were added to the Dow Jones Sustainability Index. Krüger (2015) showed that the market responds strongly negatively to negative social-related events and weakly negatively to positive ones. Similar inconclusive results were also obtained in studies that analyzed whether SRI funds outperform conventional ones (Galema et al., 2008; Hong and Kacperczyk, 2009; Muñoz et al., 2014) or whether high-rated portfolios constructed based on ESG outperformed low-rated portfolios (Badía et al., 2020; Halbritter and Dorfleitner, 2015). Our study contributes to this literature by finding that

ESG performance has not played an increasingly important role in explaining the investment decisions of US mutual fund investors from 2015 to 2021.

Our third research question aimed to develop a decision model to predict the flows raised by each mutual fund. Previous studies identified key factors for investment decisions (Ammann et al., 2019; Guercio and Tkac, 2008; Reboredo and Otero, 2021), but they are not predictive models. These studies used regression analysis and found statistically significant relationships, but they do not provide information on the relevant performance measures in forecast verification. Other studies used advanced techniques for selecting funds (Chen and Ren, 2022; Deboeck, 1998; Vo et al., 2019), but their objective was not to identify the factors that explain the subscription or redemption of funds carried out by investors. Our empirical study examined US equity mutual funds from 2015 to 2021, using as a dependent variable the increases in fund flows. We used logistic regression as a baseline model to predict the flows raised by each mutual fund, and various machine learning tools (random forest, gradient boosting decision trees and neural networks) because of their ability to predict with remarkable accuracy in highly nonlinear ways. We performed a temporal validation of the models splitting the data into two periods, the first being the training sample and the second the test sample. Our study contributes to decision modeling research by predicting the future flows collected by the funds, with accuracy rates ranging around 70%.

Our fourth research question was whether investors who invested in the funds that received the most flows made the right decision (in financial and social terms). To do this, we compared their financial returns with that of the average mutual fund. Previous research obtained inconclusive results (Feng et al., 2014; Zheng, 1999). Some studies found a “smart money” effect so that investors moved their money into funds that subsequently performed well (Zheng, 1999). However, other studies found that only institutional investors showed a “smart money” effect while individual investors showed a “dumb money” effect (Feng et al., 2014), the latter in line with the efficient market hypothesis that states that it is impossible to predict movements in stock prices (Fama, 1970). We found that the funds that received the most flows obtained slightly higher returns than the others did, at least in the short term. Finally, we developed decision models to predict both the financial returns and ESG performance. Complementing the well-established statement that “past (financial) performance does not guarantee future results,” our study contributes by finding that “investing in a fund that meets ESG criteria guarantees that the fund will continue to perform well socially," because most mutual funds that obtain a high ESG score retain it, at least in the short term. Individual investors would do well to support their financial choices by using decisional systems to avoid behavioral biases and increase returns, both in financial and social terms.

2. Literature review and model development

The efficient market hypothesis and prospect theory stand out among the theories that can help to understand how investors make financial decisions. The efficient market hypothesis proposed that current stock prices fully reflect available information about the value of the firm; hence, the past cannot be used to predict the future in any meaningful way (Fama, 1965). In other words, there is no way to beat the market using this information, because stock prices follow a random walk rather than a predictable path; hence, a rational investor should not analyze past information. The efficient market hypothesis is based on two assumptions: investors are fully rational decision-makers who do not behave erratically, and there are no information asymmetries. According to the efficient market hypothesis, it does not matter what you buy; therefore, a quite rational position would be to buy mutual funds with the lowest fees. In fact, many investors consider mutual fund fees when it comes to investment (Servaes and Sigurdsson, 2022), which is fully justified given the negative relation between fees and fund performance (Gil-Bazo and Ruiz-Verdú, 2009). Passive funds, which replicate a benchmark index to match its performance, are becoming increasingly popular, as opposed to active funds, which require frequent trading to try to outperform the benchmark index (Cremers et al., 2019). The more actively the fund is managed, the more trades it undertakes and the higher the costs it incurs. Conversely, passive management has low fees because it performs fewer transactions than active management. Assets under management by US passive funds exceeded those of active funds for the first time in September 2019 (Gittelsohn, 2019). Considering the above, it is expected that mutual fund fees and investment strategy will be factors explaining fund flows.

However, the efficient market hypothesis is a theoretical model that does not explain some market anomalies, such as the possibility of beating the market by identifying undervalued companies (Basu, 1977). Prospect theory was developed to explain how people decide, not how they should decide (Kahneman and Tversky, 1979). Prospect theory is central to behavioral finance by explaining that the cause of these anomalies is the behavioral biases of financial decision-makers. One of the most important biases is the overreaction of investors to information, who react disproportionately to new information and cause the stock price to change in an unjustified way (De Bondt and Thaler, 1985). Other cognitive biases explain why investors prefer funds that were profitable in the past. For example, the extrapolation bias consists of believing that past performance is the best indicator for predicting future performance (Chen et al., 2007). Investors are heavily influenced by past returns in their purchase decisions (Barber and Odean, 2013) but some of them experience the opposite effect – the Gambler's Fallacy – and think that a trend will reverse (Huber et al., 2010). Although it has been shown that past returns do not guarantee future returns (Malkiel, 2005), and the opposite may be true (De Bondt and Thaler, 1985), decision-makers tend to invest in mutual funds that have had above-average returns (Guercio and Tkac, 2008; Sirri and Tufano, 1998). For all these reasons, it is expected that past performance will be a factor in explaining fund flows.

Another factor that can explain why a fund receives flows from investors is the fund's past growth. Herd behavior may be an explanatory factor, as the tendency to imitate what other investors do is well documented in the capital markets (Nofsinger and Sias, 1999). Many investors mimic the behavior of other investors so that increases in fund size can explain future fund growth. It is expected that the increases in size experienced by the fund will explain future fund flows. However, the size of the fund limits its growth because the larger the fund, the greater the difficulties in continuing to grow (Chen et al., 2004). Investors exhibit different patterns in the face of risk. While high risk-taking may reveal narcissism in some investors (Campbell et al., 2004), risk-averse investors show a stronger tendency to invest in mutual funds as a way to ensure that their portfolios are highly diversified (Dorn and Huberman, 2010). Given the risk aversion of fund investors, low volatility is expected to be a factor behind fund flows.

Including social aspects in decision-making means adding a constraint to the decisional model, so that an investor seeking sustainable investments would have lower financial returns, all other things being equal. However, from the stakeholder theory approach (Freeman, 1984), it can be argued that those companies that stand out for their

ESG performance signal high managerial quality, which can translate into favorable financial performance and may reduce the high costs that emerge during corporate social crises or environmental disasters (Renneboog et al., 2008). Nevertheless, literature reviews found little evidence that the risk-adjusted returns of ESG funds differ substantially from conventional funds (Plagge and Grim, 2020; Renneboog et al., 2008). Bollen (2007) studied the behavior of social investors, finding that they are more loyal than other investors, which is explained because they seem to derive utility from being exposed to the social attribute. Investors value sustainable investments and, indeed, react to the availability of sustainability ratings (Ammann et al., 2019). Whether or not social investors pay a premium for ethics, it seems undeniable that there are investors willing to invest in ESG funds, therefore it is expected that ESG performance will be a factor explaining fund flows.

Taking into account all of the above factors, we modeled decisions on mutual funds as a function of the past return, risk, ESG performance, fund size, previous size increase, investment strategy, and management fees.

3. Methodology

3.1. Sample and data collection

We analyzed the share classes of equity mutual funds domiciled and commercialized in the United States from January 2015 to December 2021. The start date was chosen because Morningstar began reporting ESG ratings of fund portfolios in January 2015. Morningstar is one of the largest providers of information for mutual fund investors. We chose listed and delisted share classes to avoid survivorship bias. Our initial sample comprised 20,184 share classes belonging to 5,330 funds. However, not all share classes had complete data on the study variables. After data cleaning, we analyzed 14,497 share classes from 4,237 funds. Table 4.1 shows the financial and nonfinancial variables used and their definition.

Our study had three groups of dependent variables: the flow in the next N months ($Flow_{t+n}$), the return in the next N months ($Return_{t+n}$), and the sustainability score in the next N months ($ESGscore_{t+n}$). *ESGscore* measures the overall environmental, social, and governance performance according to Morningstar, and captures the scores obtained by the fund's portfolio holdings. Morningstar changed its methodology as of September 2019 and now the interpretation is the opposite: A high value means high social risk. Therefore, we used min–max normalization to homogenize the time series of this variable. In addition, Morningstar communicates ESG information with a one-month lag. Therefore, we delayed all portfolio scores by one month.

Table 4.1 Description of the variables used

Variables	*Definition*
Independent variables	
ESGscore	Asset-weighted average of the company ESG scores (environmental, social, and governance) for the covered holdings in a portfolio. Morningstar changed the calculation method from September 2019, measuring the degree to which a company may be at risk driven by social factors. The data were transformed to maintain consistency in the series (Source: Morningstar; Code: portfolio corporate sustainability score).
Yield	Fund cumulative net return in the previous 12 months (Source: Own elaboration from Morningstar; Code: Return).
Volatility	Standard deviation of the previous 12 months' return (Source: Own elaboration from Morningstar; Code: Return).
Alpha	The previous 12-month excess return generated by the fund, relative to the Fama–French five-factor model, as per Equation 3 (Fama & French, 2015).
Flow	The percentage change in the total net assets (TNA) of a share class (see Equation 1).
logTNA	The logarithm in base 10 of the total net assets of the individual share classes (Source: Morningstar; Code: net assets – share class (monthly)).
Fees	The percentage of fund assets used to pay for operating expenses and management fees, including 12b-1 fees, administrative fees, and all other asset-based costs incurred by the fund, except brokerage costs (Source: Morningstar; Code: annual report net expense ratio).
Turnover	Fund's trading activity, which is computed by taking the lesser of purchases or sales and dividing by average monthly net assets (Source: Morningstar; Code: turnover ratio %).
Dependent variables	
$Flow_{t+n}$	The percentage change in the total net assets (TNA) of a share class over the next N months (N = 1, 3, 6 or 12).
$DFlow_{t+n}$	A dummy variable obtained by transforming the cumulative flow ($Flow_{t+n}$), where 1 indicates that the cumulative flow was greater than the median and 0 otherwise.
$Return_{t+n}$	Financial return over the next N months (N = 1, 3, 6 or 12).
$ESGscore_{t+n}$	ESG score over the next N months (N = 1, 3, 6 or 12).

We first obtained the money flow of share class *i* (*Flow*) following Equation 1 (Bollen, 2007; Guercio and Tkac, 2008).

$$Flow_{it} = \frac{TNA_{it} - TNA_{it-1}(1 + Return_{it})}{TNA_{it-1}} \quad (1)$$

where TNA_{it} is the total net assets of a share class *i* in month *t*, and $Return_{it}$ measures the net revaluation suffered by the assets of share class *i* in month *t*.

Regarding the independent variables, *Yield* shows the cumulative net return of the fund in the last 12 months. *Volatility* measures the risk of the investment in the last 12 months. The return adjusted for risk (*Alpha*) provides a measure of the fund's outperformance or underperformance and was obtained as the excess return over the last 12 months on the Fama–French five-factor model following Equation 2 (Fama & French, 2015).

$$r_{it} = Alpha_i + \beta_{MKT}MKT_t + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{RMW}RMW_t + \beta_{CMA}CMA_t + \varepsilon_{it} \quad (2)$$

where r_{it} is the return of share class *i* in period *t*, and $Alpha_i$ shows the excess return over the risk factors: market effect (*MKT*), size effect (*SMB*), value effect (*HML*), profitability (*RMW*), and investment style (*CMA*).

Next, the fees of each share class were measured by calculating the percentage of share class assets used to pay operating expenses and management fees (*Fees*). The turnover ratio measured the fund's trading activity and was calculated by taking the lower of purchases or sales and dividing it by the average monthly net assets (*Turnover*) (Elton et al., 2010). A high value of this variable reflects an investment strategy that involves more trading than holding, which increases costs for investors. To limit the influence of extreme outliers we winsorized the financial variables each month at the 2.5th and 97.5th percentile. Finally, we calculated the logarithm of the TNA of the share class as a measure of size (*LogTNA*).

3.2. Preliminary analysis

Table 4.2 shows the results of a Pearson correlation analysis and Table 4.3 shows descriptive statistics of the variables. The correlation coefficient between flows and the independent variables was very low, as was the correlation between future financial return and the independent variables. By contrast, the correlation between past and future ESG performance variables was very high, ranging from 0.99 for the following month to 0.83 for the 12 months, meaning that the funds maintain ESG scores over time. As for the independent variables, the correlation coefficient between *Alpha* and *Yield* was 0.38. Although multicollinearity may not affect predictive power, the effect of each independent variable on the dependent variable could be miscalculated (Myers, 1990). We opted to remove *Alpha* from the regression analysis and use only *Yield* as a measure of return, because it is observable and easily understood by investors.

Table 4.2 Pearson's correlation matrix between variables

	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(0) ESGscore	1.00																			
(1) Yield	0.08	1.00																		
(2) Volatility	0.11	-0.08	1.00																	
(3) Alpha	-0.04	0.38	-0.06	1.00																
(4) Flow	0.01	0.09	-0.05	0.09	1.00															
(5) logTNA	0.01	0.08	0.00	0.06	-0.03	1.00														
(6) Fees	-0.08	-0.09	-0.05	-0.05	-0.09	-0.39	1.00													
(7) Turnover	-0.09	-0.05	0.00	-0.03	-0.01	-0.20	0.23	1.00												
(8) Flow$_{t+1}$	0.01	0.09	-0.04	0.09	0.33	-0.06	-0.10	-0.01	1.00											
(9) Flow $_{t+3}$	0.02	0.07	-0.03	0.09	0.26	-0.07	-0.10	-0.01	0.29	1.00										
(10) Flow $_{t+6}$	0.02	0.06	-0.01	0.07	0.21	-0.08	-0.10	-0.01	0.22	0.26	1.00									
(11) Flow $_{t+12}$	0.03	0.03	0.02	0.04	0.16	-0.09	-0.11	-0.00	0.15	0.17	0.20	1.00								
(12) Return $_{t+1}$	0.03	-0.05	0.20	0.05	-0.02	0.01	-0.03	-0.01	0.03	0.03	0.02	0.02	1.00							
(13) Return $_{t+3}$	0.01	-0.07	0.16	0.01	-0.01	0.01	-0.03	-0.00	-0.01	0.03	0.02	0.03	-0.05	1.00						
(14) Return $_{t+6}$	0.06	-0.00	0.08	-0.04	-0.01	0.01	-0.02	-0.01	-0.00	-0.01	0.04	0.03	0.05	-0.07	1.00					
(15) Return $_{t+12}$	0.04	-0.06	0.08	-0.06	-0.01	0.01	-0.02	-0.01	-0.01	-0.01	-0.01	0.04	-0.10	-0.12	0.00	1.00				
(16) ESGscore $_{t+1}$	0.99	0.08	0.13	-0.03	0.01	0.01	-0.08	-0.09	0.01	0.02	0.02	0.03	0.03	0.02	0.04	0.03	1.00			
(17) ESGscore $_{t+3}$	0.97	0.07	0.17	-0.01	0.01	0.01	-0.09	-0.09	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.04	0.98	1.00		
(18) ESGscore $_{t+6}$	0.93	0.04	0.23	-0.00	0.00	0.01	-0.09	-0.08	0.01	0.01	0.01	0.02	0.04	0.02	0.04	0.08	0.94	0.96	1.00	
(19) ESGscore $_{t+12}$	0.83	-0.11	0.26	0.02	-0.01	0.00	-0.09	-0.08	-0.01	-0.00	0.00	0.02	0.03	0.05	0.06	0.05	0.85	0.88	0.93	1.00

Table 4.3 Descriptive statistics of the variables

	mean	std	min	25%	50%	75%	max	#
ESGScore	58.25	11.49	0.00	51.37	58.28	66.34	99.68	760,285
Yield	11.82%	17.36%	-29.94%	-0.09%	9.66%	20.05%	114.56%	760,285
Volatility	4.08%	1.69%	0.88%	2.79%	3.87%	5.00%	11.23%	760,285
Alpha	-0.09%	0.74%	-2.76%	-0.46%	-0.11%	0.22%	4.71%	760,285
Flow	-0.23%	4.73%	-18.98%	-1.73%	-0.44%	0.63%	27.66%	760,285
logTNA	7.62	1.20	4.60	6.84	7.72	8.50	9.96	760,285
Fees	1.15%	0.50%	0.08%	0.82%	1.09%	1.43%	2.43%	760,285
Turnover	57.48%	45.29%	4.00%	25.47%	45.00%	76.00%	227.75%	760,285
$Flow_{t+1}$	-0.28%	4.69%	-18.98%	-1.74%	-0.45%	0.60%	27.66%	760,285
$Flow_{t+3}$	-0.32%	4.65%	-18.98%	-1.76%	-0.47%	0.57%	27.66%	724,957
$Flow_{t+6}$	-0.39%	4.59%	-18.98%	-1.79%	-0.50%	0.53%	27.66%	675,403
$Flow_{t+12}$	-0.49%	4.45%	-18.98%	-1.83%	-0.55%	0.47%	27.66%	588,320
$Return_{t+1}$	0.86%	4.54%	-25.66%	-1.41%	1.10%	3.33%	19.69%	760,285
$Return_{t+3}$	0.88%	4.59%	-25.66%	-1.39%	1.15%	3.37%	19.69%	724,957
$Return_{t+6}$	0.95%	4.62%	-25.66%	-1.28%	1.22%	3.45%	19.69%	675,403
$Return_{t+12}$	1.10%	4.62%	-25.66%	-0.92%	1.38%	3.51%	19.69%	588,320
$ESGscore_{t+1}$	58.35	11.54	0.00	51.42	58.34	66.52	99.68	760,285
$ESGscore_{t+3}$	58.43	11.61	0.63	51.42	58.37	66.68	99.68	747,706
$ESGscore_{t+6}$	58.38	11.63	-0.00	51.31	58.23	66.68	99.68	713,744
$ESGscore_{t+12}$	58.40	11.60	-0.00	51.17	58.09	66.80	99.68	645,815

4. Results

4.1. Factors explaining the purchase of funds

Our first research question aimed to study the factors that explain why an investor buys a fund, and in particular to compare the relative importance of past returns versus ESG performance. Table 4 presents the results of exploratory analysis and an independent T-test to compare the differences between funds that received flows above the median and those that did not. As expected, investors subscribed to funds with a higher past return, better ESG performance and smaller fund size but which had increased in size, had lower volatility, lower fees, and lower turnover ratio. The mean differences were statistically significant for all variables. Focusing on the flows received in the following month, *Flow* showed the largest mean differences: Funds, where flows increased, had previously grown by 1.05%, while funds whose flows decreased had previously decreased by -1.52%. *Fees* also showed highly significant differences in means, 1.08% in funds where flows increased versus 1.22% in funds where flows decreased. Differences in *Yield* were also high: 12.84% versus 10.81%. *Volatility* and *Turnover* showed relevant differences. Substantively modest mean differences were observed for *ESGscore*: 58.61 in funds where flows increased and 57.90 in funds where flows decreased. The results were similar if the dependent variable measured the flows in the following 3, 6, and 12 months.

Table 4.4 T-test of mean differences

This table shows the T-test of mean differences between funds that increased their flows over the next 1, 3, 6, and 12 months above the median ($DFlow_{t+n}$=1) and those that did not ($DFlow_{t+n}$=0). *, and ** denotes statistical significance at the 5% and 1% levels, respectively.

	$DFlow_{t+1}$			$DFlow_{t+3}$			$DFlow_{t+6}$			$DFlow_{t+12}$		
	D=1	**D=0**	**Test**	**D=1**	**D=0**	**Test**	**D=1**	**D=0**	**Test**	**D=1**	**D=0**	**Test**
ESGscore	58.61	57.90	27.00**	58.48	57.75	27.10**	58.31	57.56	27.96**	57.91	57.19	25.79**
Yield	12.84%	10.81%	51.29**	12.20%	10.41%	44.79**	10.95%	9.67%	32.48**	8.26%	7.62%	20.27**
Volatility	4.05%	4.11%	-15.87**	4.02%	4.08%	-14.53**	3.99%	4.04%	-11.14**	3.81%	3.83%	-3.73**
Alpha	-0.02%	-0.16%	83.37**	-0.02%	-0.15%	78.43**	-0.03%	-0.15%	63.65**	-0.06%	-0.13%	36.55**
Flow	1.05%	-1.52%	245.39**	0.90%	-1.34%	206.44**	0.80%	-1.18%	172.19**	0.62%	-0.96%	127.21**
logTNA	7.44	7.79	-128.80**	7.45	7.80	-127.88**	7.46	7.82	-125.18**	7.48	7.84	-117.47**
Fees	1.08%	1.22%	-126.19**	1.08%	1.23%	-125.48**	1.08%	1.23%	-124.17**	1.08%	1.23%	-118.04**
Turnover	56.71%	58.25%	-14.82**	56.48%	58.18%	-16.03**	56.31%	58.08%	-16.20**	56.26%	57.82%	-13.32**
#	380,123	380,162		362,460	362,497		337,683	337,720		294,144	294,176	

Next, we performed the regression model of Equation (3) to study the factors that explain investors' flows and to compare the importance of these factors.

$$Flow_{it+n} = \alpha_i + \beta_1 \times ESGscore_{it} + \beta_2 \times Yield_{it} + \beta_3 \times Volatility_{it} + \beta_4 \times Flow_{it} + \beta_5 \times logTNA_{it} + \beta_6 \times Fees_{it} + \beta_7 \times Turnover_{it} + \varepsilon_{it} \quad (3)$$

Table 4.5 shows the results of 3 regression models for each of the four time periods analyzed for the dependent variable *Flow* (N= 1, 3, 6, and 12 months). Model 1 used *ESGscore* as the independent variable. The goodness of fit of the model was low (adjusted R2 was 0.0002 for $Flow_{t+1}$). Model 2 used financial variables as the independent variables, and the adjusted R2 was 0.1256 for $Flow_{t+1}$. Model 3 added the *ESGscore* variable, i.e., it is the full model and the adjusted R-squared remained the same. The results were very similar when the flows of the following periods were considered. Thus, regarding the first research question, although ESG performance obtained a statistically significant coefficient, it does not explain much of the variance. The table shows the standardized beta coefficients, which are useful for our purpose of knowing the weights of the criteria used by investors, as they eliminate the problem of dealing with different units of measurement. The relative sizes of these coefficients indicate the comparative influence of the independent variables in the model. *ESGscore* standardized beta coefficients ranged from 0.01 to 0.09, *Yield* coefficients from 0.16 to 0.26, *Flow* ranged from 0.60 to 1.47, and *Fees* ranged from -0.47 to -0.65. Investors may take ESG performance into account, but past returns, past growth, and mutual fund fees are much more important.

Table 4.5 Regression analysis for the flows over the next 1, 3, 6, and 12 months
*, and ** denotes statistical significance at the 5% and 1% levels, respectively.

	Flow $_{t+1}$			Flow $_{t+3}$			Flow $_{t+6}$			Flow $_{t+12}$		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Intercept	-0.28** (0.005)	-0.28** (0.005)	-0.28** (0.005)	-0.32** (0.005)	-0.32** (0.005)	-0.32** (0.005)	-0.39** (0.006)	-0.37** (0.005)	-0.37** (0.005)	-0.48** (0.006)	-0.42** (0.006)	-0.41** (0.006)
ESGscore	0.07** (0.005)		0.01* (0.005)	0.08** (0.006)		0.03** (0.005)	0.10** (0.006)		0.06** (0.006)	0.13** (0.006)		0.09** (0.006)
Yield		0.26** (0.005)	0.26** (0.005)		0.22** (0.005)	0.22** (0.005)		0.21** (0.006)	0.21** (0.006)		0.14** (0.009)	0.16** (0.009)
Volatility		-0.12** (0.005)	-0.12** (0.005)		-0.07** (0.005)	-0.08** (0.005)		-0.01* (0.005)	-0.02** (0.005)		0.14** (0.006)	0.13** (0.006)
Flow		1.47** (0.005)	1.47** (0.005)		1.12** (0.005)	1.12** (0.005)		0.86** (0.005)	0.86** (0.005)		0.60** (0.006)	0.60** (0.006)
logTNA		-0.44** (0.006)	-0.44** (0.006)		-0.51** (0.006)	-0.51** (0.006)		-0.57** (0.006)	-0.56** (0.006)		-0.62** (0.006)	-0.61** (0.006)
Fees		-0.47** (0.006)	-0.47** (0.006)		-0.54** (0.006)	-0.54** (0.006)		-0.60** (0.006)	-0.59** (0.006)		-0.65** (0.006)	-0.65** (0.006)
Turnover		0.00 (0.005)	0.00 (0.005)		-0.01 (0.005)	-0.00 (0.005)		-0.00 (0.006)	0.00 (0.006)		0.02** (0.006)	0.03** (0.006)
#	760,285	760,285	760,285	724,957	724,957	724,957	675,403	675,403	675,403	588,320	588,320	588,320
Adj R2	0.0002	0.1256	0.1256	0.0003	0.0869	0.0870	0.0004	0.0662	0.0664	0.0007	0.0500	0.0503

4.2. The evolution of investors' concern about ESG issues

Our second research question aimed to study the evolution of investors' concern about ESG performance, compared to all other factors. We analyzed the evolution of the importance of each independent variable by performing a rolling regression with a rolling window of 12 months from December 2015 based on Equation 3, totaling 61 regressions. Rolling regression is a time series modeling technique used to analyze how the coefficients of variables change over time. We obtained a time series of regression coefficients, which varied over time but were always within a range. Figure 4.1 shows the evolution of the standardized beta coefficients during the period analyzed. However, the interpretation of the coefficients of a multivariate regression can be misleading in the presence of multicollinearity. Therefore, the use of general dominance weights is recommended to complement the analysis (Jung and Suh, 2019), which indicates the importance of the variable for the goodness of fit of the model (Tonidandel and LeBreton, 2011). Figure 4.2 shows the evolution of the general dominance weights. No major variations were observed during the period analyzed. Both the analysis of the standardized coefficients and the analysis of the general dominance weights revealed that ESG performance was never highly relevant for investors, although its importance fluctuated, which suggests that sentiment toward sustainability varies over time (Pástor et al., 2021; Pedersen et al., 2021). The relative importance of the past flow is notable in explaining the next month's flow but gradually decreases in explaining the flows as the distance in time increases.

4.3. A decision model for predicting mutual fund flows

Our third research question aimed to develop decision models to predict the flows raised by each mutual fund. In this study, we used the dummy dependent variable $DFlow_{t+n}$ to calculate accuracy and other absolute performance measures. We adopted a viewpoint that resembles a real case of an observer trying to predict which funds will increase their flows, using one year's past information from a set of independent variables. We divided the sample into a training sample and a test sample. The training sample included for each share class all the information of the dependent and independent variables, in different periods, from January 2015 to December 2018. The test sample included data from January 2019 to December 2021. Therefore, the test made it possible to perform a temporal validation of the results, which is very convenient.

Predictions were performed with logistic regression (LR), random forest (RF), extreme gradient boosting (XGBoost), and multilayer perceptron neural network (MLP). We used cross-validation for training, splitting the training sample into 5 K-folds. We used the *scikit-learn* machine learning library to build and test the RF and MLP models (Pedregosa et al., 2011) and the *xgboost* package for XGBoost (Chen and Guestrin, 2016). For RF, the following hyperparameters were optimized: n_estimators (100,200); max_depht (5, 10); max_features (1, 0.333, 0.666); and bootstrap (True). For XGBoost, the following hyperparameters were optimized: max_depth (1, 3, 5); subsample (0.5, 0.75, 1) learning_rate (0.005, 0.025, 0.05, 0.1,0.3); n_estimators (1000), early_stopping_rounds (5); objective='reg:logistic'. For MLP, the following hyperparameters were optimized: hidden_layer_sizes [(10, 10), (25, 10), (25, 25), (50, 25), (50, 50), (10, 25), (25, 50)]; max_iter (1500), n_iter_no_change (5). The rest of the hyperparameters used were those selected by default by both the *scikit-learn* and *xgboost* libraries.

Figure 4.1 Evolution of the beta standardized coefficients of the rolling regression

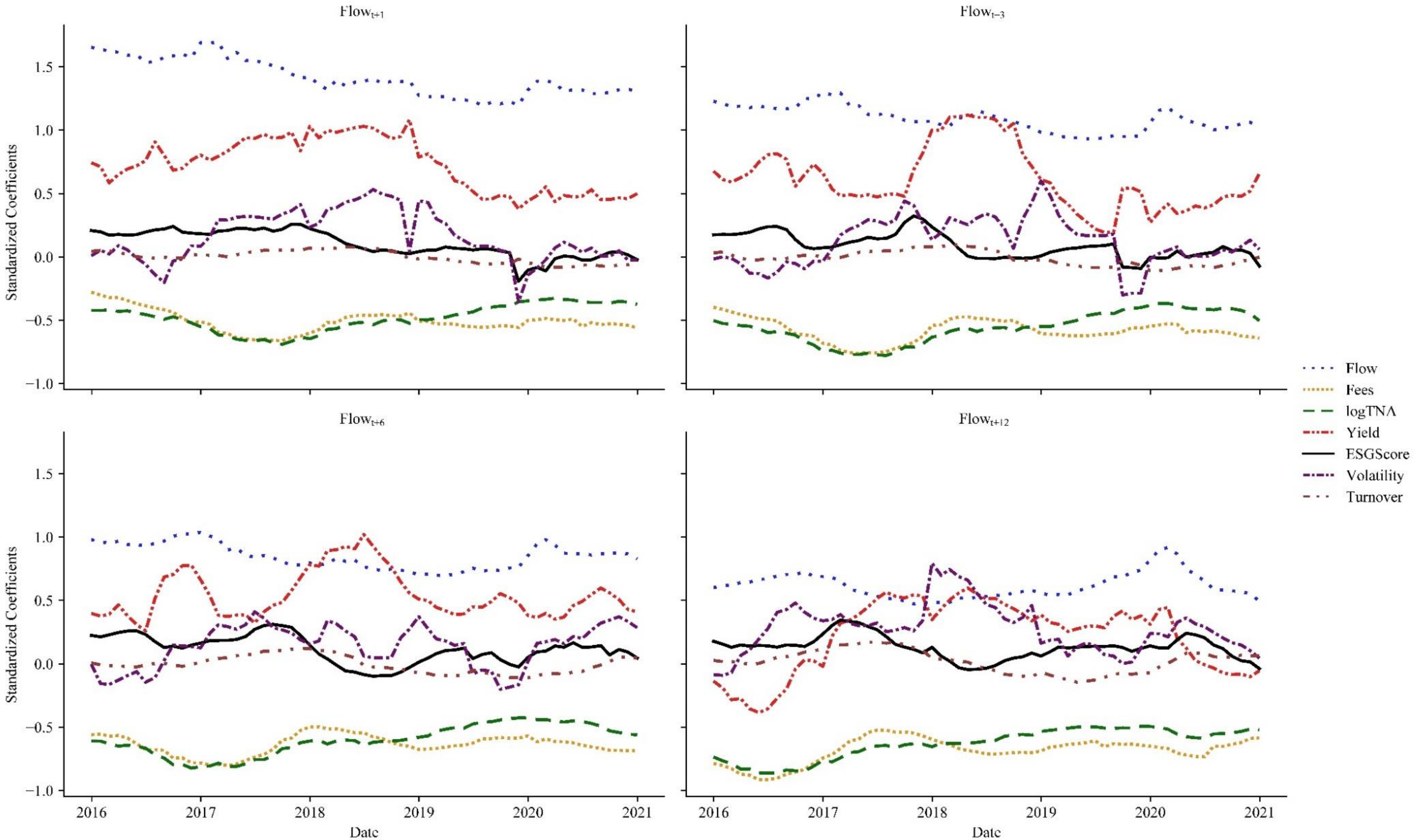


Figure 4.2 Evolution of the contribution of each independent variable to the R2 of the rolling regression using dominance analysis

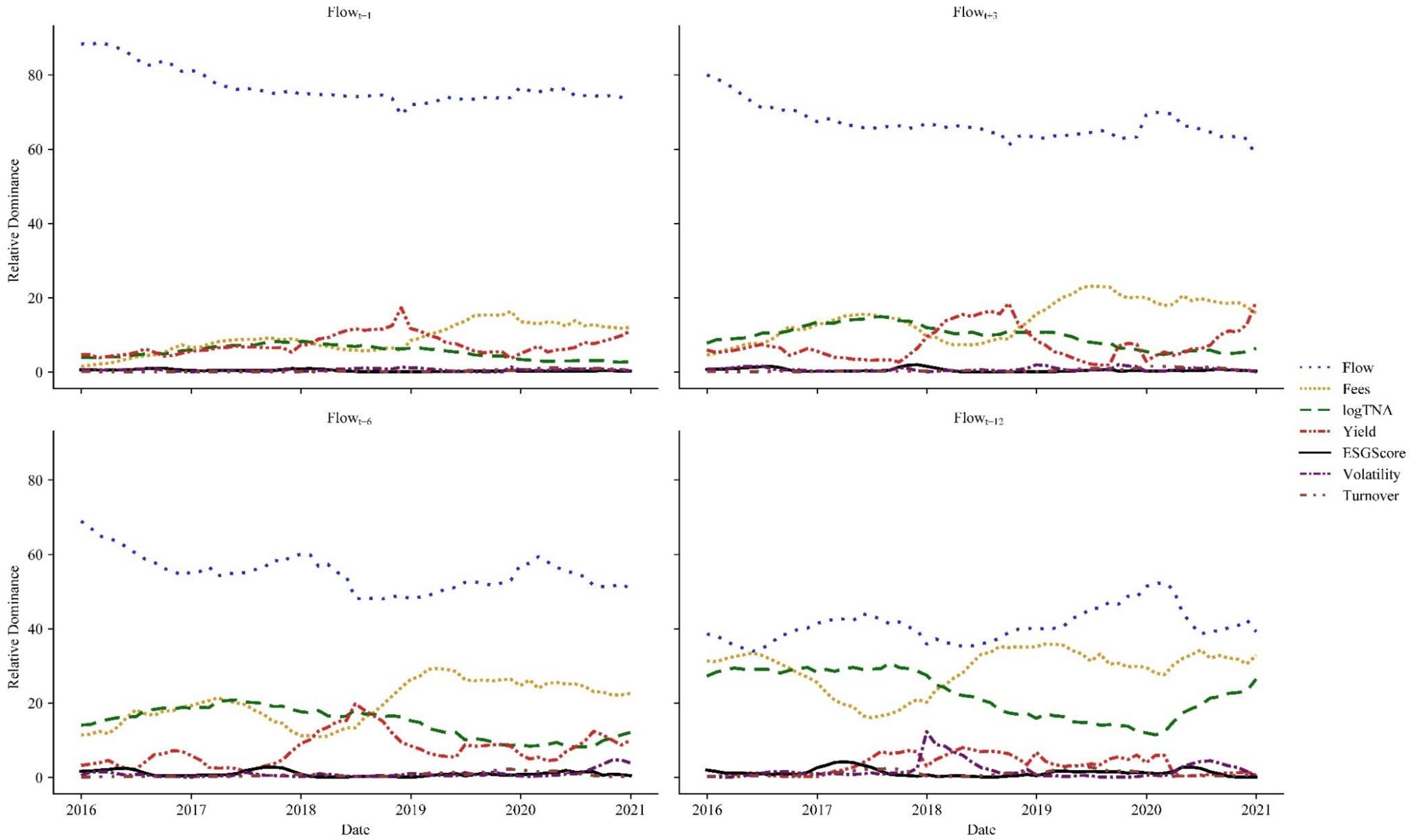


Table 4.6 shows several performance measures (accuracy, sensitivity, specificity, precision, F-score, and area under the curve (AUC)) for each model and dependent variable analyzed (fund flows over the next 1, 3, 6, and 12 months). When considering the AUC criterion, XGBoost obtained the best results, slightly better than RF. When considering the accuracy criterion, RF obtained the best results compared to the other techniques and successfully predicted the future flows collected by the funds with an accuracy ranging from 65.5% to 70.5% for the test sample. The prediction accuracy of XGBoost ranked from 65.4% to 69.6% and that of MLP ranked from 63.1% to 66.5%. LR performed the worst with a prediction accuracy ranging from 60.5% to 60.9%.

Different explainable machine learning approaches can be used to interpret the results of black-box techniques such as RF, XGBoost, and MLP (Carta et al., 2022; Moreira et al., 2021). Table 4.7 shows the results of the SHapley Additive exPlanations (SHAP) over the test sample and the permutation feature importance (PFI) techniques. SHAP uses a game-theory-based approach to calculate individual contributions of the variables in the prediction model. The SHAP values show the contribution of the variable to the output of the model for a given share class *i* in period *t*. However, the calculation of SHAP values is computationally demanding, thus we only calculated them for 5% of the test sample randomly selected. PFI randomly shuffles the values of each variable in the model to assess its effect on model performance. Thus, the PFI value shows the importance of the variable on the model's accuracy.

When using the SHAP technique, *Flow* (0.341 on average over the four periods) and *LogTNA* (0.147) were the most important variables, followed by *Fees* (0.090), *Yield* (0.019), and *Volatility* (0.019). The least important variables were *ESGscore* (0.015) and *Turnover* (0.009). The results of the PFI analysis were consistent with the previous one. Note how the relative importance of some variables decreases as time increases. For example, current returns (*Yield*) influence decisions made one month later, but have little influence on decisions made one year later.

Table 4.6 Performance of different techniques in predicting mutual fund flows over the next 1, 3, 6, and 12 months

		$DFlow_{t+1}$				$DFlow_{t+3}$				$DFlow_{t+6}$				$DFlow_{t+12}$			
		LR	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**
Training Sample	*Accuracy*	67.54%	72.65%	73.13%	72.90%	65.95%	71.33%	70.62%	70.59%	64.66%	68.25%	68.50%	68.72%	63.45%	67.88%	66.54%	66.68%
	Sensitivity	63.68%	70.73%	73.00%	72.72%	62.10%	71.30%	70.74%	70.47%	61.06%	68.64%	70.28%	71.34%	60.48%	68.86%	68.21%	67.77%
	Specificity	71.41%	74.58%	73.27%	73.07%	69.79%	71.35%	70.50%	70.70%	68.26%	67.87%	66.73%	66.10%	66.42%	66.90%	64.87%	65.59%
	Precision	69.01%	73.56%	73.20%	72.98%	67.27%	71.34%	70.57%	70.63%	65.79%	68.11%	67.87%	67.79%	64.30%	67.53%	66.00%	66.32%
	F-score	66.24%	72.12%	73.10%	72.85%	64.58%	71.32%	70.65%	70.55%	63.34%	68.37%	69.05%	69.52%	62.33%	68.19%	67.09%	67.04%
	AUC	73.49%	79.07%	80.00%	79.57%	71.59%	78.45%	76.94%	77.04%	70.12%	74.40%	74.77%	75.02%	68.51%	74.47%	72.57%	72.64%
Test Sample	*Accuracy*	60.81%	70.53%	69.56%	66.47%	60.50%	68.76%	68.77%	64.40%	60.64%	67.27%	67.31%	64.07%	60.94%	65.51%	65.41%	63.15%
	Sensitivity	83.97%	66.94%	77.93%	82.47%	82.49%	71.03%	69.51%	82.93%	79.71%	65.50%	69.42%	82.01%	75.85%	69.20%	70.20%	78.62%
	Specificity	37.65%	74.11%	61.19%	50.48%	38.50%	66.48%	68.03%	45.87%	41.57%	69.04%	65.19%	46.13%	46.03%	61.83%	60.62%	47.68%
	Precision	57.39%	72.11%	66.75%	62.48%	57.29%	67.94%	68.50%	60.50%	57.70%	67.90%	66.60%	60.35%	58.42%	64.45%	64.06%	60.04%
	F-score	68.18%	69.43%	71.91%	71.10%	67.62%	69.45%	69.00%	69.96%	66.94%	66.68%	67.98%	69.53%	66.01%	66.74%	66.99%	68.08%
	AUC	68.63%	76.75%	76.78%	74.48%	67.62%	74.82%	74.91%	72.55%	67.26%	73.07%	73.22%	71.53%	66.73%	71.00%	71.09%	70.06%

Table4.7 Results of the variable importance analysis

This table shows the results of the variable importance for each machine-learning method. Panel A: SHapley Additive exPlanations (SHAP) values. Panel B: permutation feature importance (PFI) values.

Panel A	**Dflow$_{t+1}$**				**Dflow$_{t+3}$**				**Dflow$_{t+6}$**				**Dflow$_{t+12}$**			
	LR	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**
ESGscore	0.038	0.000	0.014	0.036	0.037	0.014	0.005	0.014	0.040	0.001	0.009	0.028	0.030	0.040	0.033	0.050
Yield	0.121	0.002	0.055	0.108	0.106	0.018	0.013	0.089	0.088	0.001	0.004	0.055	0.039	0.007	0.002	0.017
Volatility	0.056	0.000	0.041	0.046	0.057	0.016	0.003	0.072	0.054	0.000	0.005	0.069	0.049	0.016	0.026	0.070
logTNA	0.143	0.076	0.128	0.086	0.168	0.131	0.132	0.114	0.188	0.095	0.135	0.129	0.224	0.186	0.192	0.183
Fees	0.133	0.021	0.091	0.105	0.159	0.072	0.067	0.123	0.208	0.040	0.075	0.147	0.223	0.125	0.125	0.166
Turnover	0.004	0.001	0.011	0.008	0.007	0.010	0.005	0.011	0.008	0.001	0.007	0.027	0.008	0.023	0.013	0.022
Flow	0.119	0.432	0.324	0.249	0.109	0.371	0.383	0.228	0.109	0.425	0.374	0.218	0.089	0.284	0.281	0.178

Panel B	**Dflow$_{t+1}$**				**Dflow$_{t+3}$**				**Dflow$_{t+6}$**				**Dflow$_{t+12}$**			
	LR	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**	**LR**	**RF**	**XGBoost**	**MLP**
ESGscore	-0.01%	-0.00%	-0.19%	-0.27%	0.04%	0.03%	-0.04%	-0.13%	0.01%	-0.01%	-0.01%	-0.03%	-0.01%	-0.00%	-0.07%	0.04%
Yield	0.75%	-0.01%	0.17%	0.28%	0.79%	0.15%	-0.04%	0.31%	0.30%	0.01%	-0.02%	0.23%	0.14%	-0.00%	-0.00%	0.12%
Volatility	-0.30%	0.00%	0.20%	-0.03%	-0.28%	0.09%	0.01%	-0.26%	-0.23%	0.00%	0.00%	0.00%	-0.05%	-0.02%	-0.06%	0.00%
logTNA	3.17%	2.17%	4.29%	2.89%	3.58%	3.63%	3.44%	3.28%	4.28%	2.87%	3.80%	4.10%	5.35%	4.60%	4.57%	4.77%
Fees	3.94%	0.22%	2.62%	2.47%	4.58%	1.79%	1.54%	3.15%	5.22%	0.76%	1.99%	3.83%	6.11%	3.21%	3.36%	4.61%
Turnover	0.01%	0.00%	-0.00%	-0.02%	0.04%	-0.01%	0.03%	0.04%	0.03%	0.01%	0.03%	0.14%	0.06%	-0.00%	0.02%	-0.03%
Flow	5.37%	17.47%	14.48%	11.83%	4.31%	13.97%	14.24%	9.24%	3.48%	13.41%	12.31%	8.10%	2.45%	8.88%	8.81%	6.15%

Table 4.8 RF and SHAP values by class type

This Accuracies using RF and SHAP values for the subsamples obtained from Morningstar's fund share class classification (test sets).

Subsample		$DFlow_{t+1}$	$DFlow_{t+3}$	$DFlow_{t+6}$	$DFlow_{t+12}$	AvgObs Train	AvgObs Test	Average of the SHAP values
Institutional	No	71.18%	69.54%	68.27%	66.59%	295,941	170,123	Flow(0.397), logTNA(0.094), Fees(0.039), Turnover(0.004), ESGscore(0.004), Yield(0.002), Volatility(0.002)
	Yes	69.49%	66.91%	65.49%	63.78%	133,048	88,128	Flow(0.375), logTNA(0.100), Fees(0.048), Turnover(0.012), ESGscore(0.009), Yield(0.005), Volatility(0.003)
Sustainable Fund	No	70.36%	68.45%	67.08%	65.09%	385,977	246,354	Flow(0.394), logTNA(0.110), Fees(0.054), ESGscore(0.010), Turnover(0.007), Volatility(0.006), Yield(0.005)
	Yes	75.24%	73.58%	72.99%	72.29%	16,168	11,779	Flow(0.222), Fees(0.131), logTNA(0.079), ESGscore(0.025), Turnover(0.023), Yield(0.008), Volatility(0.005)
Share Class Type	A	69.08%	66.99%	65.51%	61.94%	61,739	35,159	Flow(0.370), logTNA(0.111), Fees(0.039), Yield(0.019), Turnover(0.019), ESGscore(0.004), Volatility(0.004)
	Adv	68.32%	64.85%	63.11%	59.29%	13,038	7,795	Flow(0.279), logTNA(0.101), Fees(0.072), Volatility(0.040), Yield(0.040), ESGscore(0.033), Turnover(0.023)
	B	82.96%	83.22%	82.68%	81.86%	11,400	2,679	Flow(0.214), Fees(0.082), logTNA(0.027), Volatility(0.019), ESGscore(0.015), Turnover(0.014), Yield(0.010)
	C	78.85%	78.09%	77.15%	75.02%	55,477	30,760	Flow(0.220), logTNA(0.153), Fees(0.012), Yield(0.012), Turnover(0.004), Volatility(0.003), ESGscore(0.002)
	D	58.47%	52.46%	59.06%	55.49%	1,168	729	Flow(0.276), logTNA(0.107), Turnover(0.065), ESGscore(0.064), Fees(0.059), Yield(0.036), Volatility(0.033)
	Inst	69.37%	67.00%	65.49%	63.32%	113,584	70,513	Flow(0.362), logTNA(0.103), Fees(0.065), ESGscore(0.017), Turnover(0.012), Yield(0.008), Volatility(0.004)
	Inv	70.17%	68.60%	66.21%	65.01%	22,140	12,206	Flow(0.354), logTNA(0.153), Fees(0.053), ESGscore(0.013), Turnover(0.012), Yield(0.009), Volatility(0.006)
	M	70.71%	66.20%	68.15%	68.16%	3,758	1,921	Flow(0.263), logTNA(0.120), Fees(0.098), Turnover(0.066), ESGscore(0.043), Yield(0.028), Volatility(0.015)
	N	69.31%	67.69%	66.74%	64.31%	5,705	3,213	Flow(0.332), logTNA(0.137), Fees(0.082), Yield(0.033), Volatility(0.029), ESGscore(0.022), Turnover(0.016)
	No Load	66.94%	65.48%	62.99%	58.96%	19,520	11,183	Flow(0.445), ESGscore(0.028), logTNA(0.027), Yield(0.026), Fees(0.023), Turnover(0.022), Volatility(0.022)
	Other	69.76%	67.36%	66.33%	64.73%	33,965	22,933	Flow(0.363), Fees(0.107), logTNA(0.074), Turnover(0.014), ESGscore(0.006), Yield(0.004), Volatility(0.004)
	Retirement	68.87%	67.85%	66.94%	65.71%	75,712	52,407	Flow(0.278), logTNA(0.219), Fees(0.049), ESGscore(0.008), Yield(0.003), Turnover(0.002), Volatility(0.002)
	S	71.66%	71.11%	69.79%	69.53%	10,557	6,367	logTNA(0.302), Flow(0.204), Fees(0.031), Turnover(0.022), Volatility(0.016), Yield(0.014), ESGscore(0.011)
	T	81.61%	88.07%	90.11%	90.28%	1,118	382	logTNA(0.118), Flow(0.075), Yield(0.030), Volatility(0.028), Turnover(0.027), Fees(0.018), ESGscore(0.018)

Not all investors can access all types of funds as some are reserved for institutional investors. The behavior of an individual investor managing a small amount of money may differ from that of an institutional investor managing a large pension fund. Table 4.8 shows Morningstar's classification of funds, based on the type of investor (institutional or individual), type of fund (sustainable or non-sustainable fund), and type of fees and minimum investment required (different share classes). The table shows for each subsample the accuracy and the relevant variables according to the RF technique and SHAP. The accuracy of the prediction increased significantly when segmented by fund type. Overall, the importance of the variables was maintained in each of the samples. When predicting fund flows purchased by individual investors, the accuracy of the model scored 1.69 points above the accuracy of institutional funds, up to 71.18%. Accuracy in predicting sustainable fund flows exceeded that of nonsustainable funds by 4.88 points, to 75.24%. The highest accuracy was obtained when using class B funds (funds that have lower investment minimums and carry a deferred-load sales charge), which reached 82.96%, and the minority class T (tax-deferral vehicle), which reached 81.61%. By contrast, the accuracy of class D funds (typically carried by broker-sold fund shops) barely reached 58.47%. The variables hardly changed their position in the relative importance ranking.

4.4. The right decisions

In this subsection, we studied whether investors' decisions were successful. First, we analyzed whether funds that received flows above the median performed better than those that received flows below the median in terms of financial return. For this purpose, we performed a T-test and a Kruskal–Wallis test to compare the return of both groups, which is shown in Table 4.9. To carry out this study, we accumulated the flows received, as well as the returns, at 3, 6, and 12 months. Funds that received flows above the median obtained higher returns over the next months than those that received flows below the median. The differences were statistically significant but rather small in magnitude. In the 1-month case, the average returns were 0.89% versus 0.84%. When considering 12 months, cumulative returns were 14.85% and 13.05%, respectively. The same study was carried out with the ESG performance. We obtained the average ESG scores at 3, 6, and 12 months. Funds that received flows above the median obtained higher ESG scores over the next months than those that received flows below the median. The differences were statistically significant, but also very small in magnitude.

In the 1-month case, the means were 58.70 versus 57.99. By incorporating ESG aspects, this finding could go beyond the "smart money effect," meaning that investors can predict the performance of mutual funds and invest accordingly (Feng et al., 2014; Zheng, 1999), and be considered a case of "smart and virtuous effect."

Table 4.9 T-test of mean differences

This table shows the T-test of mean differences between funds that increased their flows in the next N months above the median ($DFlow_{t+n}$=1) and those that did not ($DFlow_{t+n}$=0). $Return_{t+n}$ (cumulative) measures the cumulative return of the fund in the next N months. $ESGscore_{t+n}$ (average) measures the mean ESG score in the next N months

		$DFlow_{t+1}$			**$DFlow_{t+3}$ (cumulative)**			**$DFlow_{t+6}$ (cumulative)**			**$DFlow_{t+12}$ (cumulative)**		
		D=1	**D=0**	**Test**	**D=1**	**D=0**	**Test**	**D=1**	**D=0**	**Test**	**D=1**	**D=0**	**Test**
$Return_{t+n}$ (cumulative)	**Mean**	0.89%	0.84%	4.9**	2.81%	2.51%	16.8**	6.40%	5.68%	28.2**	14.85%	13.05%	39.3**
	Median	1.12%	1.09%	15.0**	3.32%	3.05%	280.7**	6.29%	5.60%	780.5**	13.88%	12.16%	1648.2**
$ESGscore_{t+n}$ (average)	**Mean**	58.70	57.99	26.6**	58.60	57.89	26.5**	58.39	57.68	25.2**	57.95	57.29	21.2**
	Median	58.74	57.94	726.6**	58.64	57.84	736.4**	58.37	57.56	696.2**	57.81	57.09	503.5**

Next, we ran a model taking $Return_{t+n}$ as the dependent variable and those shown in Table 4.4 as independent variables. Table 10 shows the results of the regressions. Most of the variables obtained statistically significant coefficients. However, the adjusted R2 was 0.04 for the following month's return, which indicates low goodness of fit, which even decreased for the following periods. In the period under investigation, past ESG performance ($ESGscore_t$) was positively associated with the future financial return ($Return_{t+n}$), but the predictive power was very small. Analyzing the standardized coefficients of the regression, it was found that the only variable that explains the return is volatility – the well-known relationship between profitability and risk. The association between past financial return ($Yield_t$) and future financial return ($Return_{t+n}$) was sometimes positive and sometimes negative.

However, the objective of some socially responsible investors may be to achieve ESG performance and therefore the right decision will be to choose funds that will achieve the highest ESG score in the near future. We ran several models taking future *ESGscore* as the dependent variable. Table 11 shows the results of the regressions. The adjusted R2 of the model ranged from 0.97 (one month later) to 0.73 (one year later), indicating high goodness of fit. Although several variables had statistically significant coefficient values, the variable with the highest predictive power was *ESGscore*. In other words, past ESG scores predicted future ESG scores.

Table 4.10 Regression analysis for the return over the next 1, 3, 6, and 12 months

	Return$_{t+1}$			Return $_{t+3}$			Return $_{t+6}$			Return $_{t+12}$		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	0.86**	0.86**	0.86**	0.89**	0.89**	0.89**	0.96**	0.97**	0.97**	1.11**	1.08**	1.10**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
ESGscore	0.13**		0.04**	0.06**		-0.03**	0.29**		0.22**	0.22**		0.13**
	(0.005)		(0.005)	(0.005)		(0.005)	(0.006)		(0.006)	(0.006)		(0.007)
Yield		-0.18**	-0.19**		-0.28**	-0.28**		0.02**	0.02**		-0.27**	-0.24**
		(0.005)	(0.005)		(0.005)	(0.005)		(0.006)	(0.006)		(0.009)	(0.009)
Volatility		0.88**	0.87**		0.70**	0.70**		0.38**	0.34**		0.32**	0.31**
		(0.005)	(0.005)		(0.005)	(0.005)		(0.006)	(0.006)		(0.007)	(0.007)
Flow		-0.03**	-0.03**		0.00	0.00		-0.05**	-0.05**		0.00	-0.00
		(0.005)	(0.005)		(0.005)	(0.005)		(0.006)	(0.006)		(0.006)	(0.006)
logTNA		-0.01	-0.01		0.02**	0.01*		0.02*	0.02**		0.04**	0.04**
		(0.006)	(0.006)		(0.006)	(0.006)		(0.006)	(0.006)		(0.007)	(0.007)
Fees		-0.10**	-0.10**		-0.11**	-0.12**		-0.09**	-0.08**		-0.09**	-0.08**
		(0.006)	(0.006)		(0.006)	(0.006)		(0.006)	(0.006)		(0.007)	(0.007)
Turnover		-0.05**	-0.05**		0.00	0.00		-0.00	0.02**		-0.01	-0.00
		(0.005)	(0.005)		(0.006)	(0.006)		(0.006)	(0.006)		(0.006)	(0.006)
#	760,285	760,285	760,285	724,957	724,957	724,957	675,403	675,403	675,403	588,320	588,320	588,320
Adj R2	0.0008	0.0413	0.0414	0.0002	0.0294	0.0294	0.0036	0.0076	0.0098	0.0020	0.0084	0.0090

Table 4.11. Regression analysis for the ESGscore over the next 1, 3, 6, and 12 months

	ESGScore $_{t+1}$			ESGScore $_{t+3}$			ESGScore $_{t+6}$			ESGScore $_{t+12}$		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	58.35**	58.35**	58.35**	58.55**	58.46**	58.55**	58.86**	58.48**	58.90**	59.52**	58.54**	59.97**
	(0.002)	(0.013)	(0.002)	(0.003)	(0.013)	(0.003)	(0.005)	(0.013)	(0.005)	(0.008)	(0.015)	(0.008)
ESGscore	11.42**		11.38**	11.24**		11.14**	10.92**		10.72**	9.94**		9.79**
	(0.002)		(0.002)	(0.003)		(0.003)	(0.005)		(0.005)	(0.008)		(0.008)
Yield		0.96**	0.05**		0.89**	0.14**		0.75**	0.35**		-0.47**	1.10**
		(0.013)	(0.002)		(0.013)	(0.003)		(0.014)	(0.005)		(0.021)	(0.011)
Volatility		1.56**	0.23**		2.01**	0.71**		2.63**	1.37**		2.94**	2.52**
		(0.013)	(0.002)		(0.013)	(0.003)		(0.013)	(0.005)		(0.015)	(0.008)
Flow		0.03	-0.01**		0.02	-0.03**		0.00	-0.06**		-0.01	-0.16**
		(0.013)	(0.002)		(0.013)	(0.003)		(0.013)	(0.005)		(0.014)	(0.007)
logTNA		-0.41**	-0.01**		-0.42**	-0.03**		-0.44**	-0.07**		-0.44**	-0.16**
		(0.014)	(0.002)		(0.014)	(0.004)		(0.015)	(0.005)		(0.015)	(0.008)
Fees		-0.75**	-0.03**		-0.79**	-0.10**		-0.81**	-0.20**		-0.95**	-0.42**
		(0.014)	(0.002)		(0.015)	(0.004)		(0.015)	(0.005)		(0.015)	(0.008)
Turnover		-0.86**	-0.01**		-0.87**	-0.02**		-0.87**	-0.06**		-0.86**	-0.12**
		(0.013)	(0.002)		(0.014)	(0.004)		(0.014)	(0.005)		(0.014)	(0.008)
#	760,285	760,285	760,285	747,706	747,706	747,706	713,744	713,744	713,744	645,815	645,815	645,815
Adj R2	0.9782	0.0359	0.9786	0.9325	0.0467	0.9363	0.8588	0.0670	0.8735	0.6885	0.0833	0.7325

5. Discussion and conclusions

This study aimed to analyze mutual fund investor decisions, particularly to compare the importance of ESG performance with past returns and other financial variables. We modeled fund flows as a function of the past return, ESG performance, volatility, fund size, past growth, turnover ratio, and managerial fees. Although there is a great deal of interest in SRI (Ammann et al., 2019; Vo et al., 2019), we found that ESG concerns are not as important in predicting investment decisions as past performance, past growth, and managerial fees. We used statistical and machine learning models to predict future flows raised by the funds, future short-term returns, and ESG scores.

5.1. Main contributions

We found that past growth was the most important variable, which can be explained by herd behavior (Nofsinger and Sias, 1999). Investors are attracted to funds that have grown in the past. It seems that the mutual fund industry is influenced by the Matthew effect (Merton, 1968), as funds with higher inflows grow even more. Fund fee is the second relevant aspect for investors, which is also consistent with the low turnover ratio values found. These findings may be associated with the growth of passive management (Cremers and Petajisto, 2009), which can be interpreted as a sign of rationality in financial decision-making, being underpinned by the efficient market hypothesis. The importance of past return for investors is well known (Guercio and Tkac, 2008; Sirri and Tufano, 1998). They often suffer from extrapolation bias, as past return does not guarantee future return (Malkiel, 2005). Therefore, our study corroborates the existence of biases identified by prospect theory (De Bondt and Thaler, 1985; Kahneman and Tversky, 1979). As expected, volatility matters, which is explained by the fact that mutual fund clients like diversification, as opposed to those who buy a few stocks on their own (Dorn and Huberman, 2010). The relationship between fund size and fund growth is negative, confirming that larger funds have greater difficulty in continuing to grow (Chen et al., 2004).

In particular, our first research question was to study the importance of ESG aspects for the decision-makers. Following previous studies, we found that individual investors take ESG criteria into account (Ammann et al., 2019; Plagge and Grim, 2020; Renneboog et al., 2008). However, we found that the weight of ESG scores in mutual fund purchasing decisions is small and the predictive ability of ESG variables is very low. It is particularly relevant that the high interest that industry and academia seem to show in ESG performance does not correspond to the low interest currently shown by investors. In addition, our second research question was to study whether ESG concerns are becoming increasingly important and we studied the evolution of the importance of each variable by analyzing standardized regression coefficients and general dominance weights. With this study, we contributed by showing that the (low) importance of ESG scores was maintained during the period under analysis (2015 to 2021).

Our third research question was to develop predictive models using logistic regression and machine learning techniques (RF, MLP, and XGBoost). Previous studies investigated the determinants of fund flows (Guercio and Tkac, 2008; Reboredo and Otero, 2021) but they are not predictive models and do not provide performance measures, which is a contribution of our study. Our models were tested using intertemporal validation. RF obtained the best performance, with an accuracy of around 70% in the sample test. The importance of each variable was examined by analyzing the SHAP and PFI values. The type of mutual fund and to whom it is directed can affect decision-making. For example, in funds labeled as sustainable, the model's accuracy was as high as 75.24%. The model fitted better in the subsample of funds targeting individual investors rather than institutional investors, who may consider other aspects.

The fourth research question analyzed the outcome of the decisions made by investors. We found that the investments in funds that received the most flows obtained slightly higher financial returns than the rest, at least in the period considered. The existence of a "smart money" effect, meaning the ability of mutual fund investors to predict the short-term performance of funds and invest accordingly, is a widely debated topic, with no conclusive results (Feng et al., 2014; Zheng, 1999). We found that funds that received flows above the median obtained higher returns (both financial and social) over the next months than those that received flows below the median. Therefore, our study confirms a statistically significant "smart and virtuous money" effect, although of small magnitude. This is a contribution to the debate, which would need to be confirmed by further studies using a longer sample period.

In the period under investigation, we did not find a clear relationship between past returns and future returns. Our model was not able to predict financial returns, as the maximum adjusted R2 was 0.04, a finding that supports the efficient market hypothesis (Fama, 1970; Markowitz, 1952). However, the objectives of some investors are not limited to the search for financial but also social returns. Our study contributes to the literature by finding that past ESG performance explains future ESG performance. The adjusted R2 ranged from 0.98 in the model that tries to explain the fund's sustainability score in the following month to 0.73 one year later. Investors who choose to acquire funds that meet ESG criteria are not guaranteed a financial return (like other investments), but at least the association between past ESG performance and future ESG performance is very strong, so the social return is largely guaranteed. The explanation for this “good money” effect is simple: There are no abrupt changes in ESG scores and the fund that performs well in the rankings continues to do so in subsequent periods.

5.2. Limitations and future research

Morningstar’s ESG scores began in 2015. The lack of ESG scores for mutual funds until recently limits the robustness of the study’s findings. The period analyzed was not long enough to draw robust conclusions on whether ESG concerns decreased or increased largely. Our study does not address the outcome of long-term decisions, but rather the time frame covers 1, 3, 6, and 12 months. Future studies are needed to understand the impact of decisions over the long term, in different financial periods, and financial markets other than the US. Another limitation of the study refers to the accuracy of the models in predicting flows – about 70% – which is not very high. Much remains unexplained, which calls for future studies that include other types of nonfinancial variables. Accuracy increased when segmented by type of fund, which gives us a clue as to where to focus the subsequent studies. In this regard, not only the financial management of the fund may be important. Other factors may explain why a fund attracts investors, such as the sales force efforts, investment in advertising, popularity in financial social networks, a high ranking on Internet search engines, and the current media attention (Sirri & Tufano, 1998).

5.3. Practical implications

The study contributes to the identification of practical implications for fund managers, regulators, and investors. Being able to predict flows is important because it helps to understand how investors make decisions. The study helps fund managers to better understand what their clients look for. As ESG performance is not as relevant as expected, perhaps they should change the sales pitch. It does not appear to be a problem with access to information as Morningstar provides a free, intuitive 1-to-5 globe rating system on ESG performance. Perhaps investors think that there is a negative relationship between social and financial performance and they must choose between one or the other. The fund’s sales force should emphasize that high ESG performance implies less risk of a reputational crisis and that the performance of socially responsible funds does not differ statistically from that obtained by conventional mutual funds (Hamilton et al., 1993). In fact, we found a positive and statistically significant association between past ESG performance and future financial performance. However, the association was very weak, with negligible predictive power. It is useful for the regulator to know the reasons for fund flows. The model can help the supervisor to detect and predict trends in flow movements. The persistence of investor biases calls for increased financial literacy. Individual investors focus on irrelevant aspects (past financial return is often not predictive of future financial return) and place little value on social aspects (past ESG performance is predictive of future ESG performance). However, it is difficult to get investors to avoid biases, as they are predisposed to listen to the sirens' songs (Buffett, 2016). One solution is to let a decision support system make the financial decisions.

References

Albuquerque, R., Koskinen, Y., Zhang, C., 2018. Corporate social responsibility and firm risk: Theory and empirical evidence. Manage. Sci. 65, 4451–4469. https://doi.org/10.1287/mnsc.2018.3043

Ammann, M., Bauer, C., Fischer, S., Müller, P., 2019. The impact of the Morningstar Sustainability Rating on mutual fund flows. Eur. Financ. Manag. 25, 520–553. https://doi.org/https://doi.org/10.1111/eufm.12181

Aouni, B., Doumpos, M., Pérez-Gladish, B., Steuer, R.E., 2018. On the increasing importance of multiple criteria decision aid methods for portfolio selection. J. Oper. Res. Soc. 69, 1525–1542. https://doi.org/10.1080/01605682.2018.1475118

Badía, G., Cortez, M.C., Ferruz, L., 2020. Socially responsible investing worldwide: Do markets value corporate social responsibility? Corp. Soc. Responsib. Environ. Manag. 27, 2751–2764. https://doi.org/https://doi.org/10.1002/csr.1999

Barber, B.M., Odean, T., 2013. Chapter 22 - The behavior of individual investors, in: Constantinides, G.M., Harris, M., Stulz, R.M.B.T.-H. of the E. of F. (Eds.), . Elsevier, pp. 1533–1570. https://doi.org/https://doi.org/10.1016/B978-0-44-459406-8.00022-6

Barberis, N.C., 2013. Thirty years of prospect theory in economics: A review and assessment. J. Econ. Perspect. 27, 173–196. https://doi.org/10.1257/jep.27.1.173

Basu, S., 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. J. Finance 32, 663–682. https://doi.org/https://doi.org/10.1111/j.1540-6261.1977.tb01979.x

Bauer, R., Ruof, T., Smeets, P., 2021. Get real! Individuals prefer more sustainable investments. Rev. Financ. Stud. 34, 3976–4043. https://doi.org/10.1093/rfs/hhab037

Bhandari, G., Hassanein, K., Deaves, R., 2008. Debiasing investors with decision support systems: An experimental investigation. Decis. Support Syst. 46, 399–410. https://doi.org/https://doi.org/10.1016/j.dss.2008.07.010

Bilbao-Terol, A., Arenas-Parra, M., Cañal-Fernández, V., 2012. A fuzzy multi-objective approach for sustainable investments. Expert Syst. Appl. 39, 10904–10915. https://doi.org/https://doi.org/10.1016/j.eswa.2012.03.034

Bollen, N.P.B., 2007. Mutual fund attributes and investor behavior. J. Financ. Quant. Anal. 42, 683–708.

Buffett, W.E., 2016. Berkshire Hathaway's 2016 annual shareholder letter. Retrieved from https//www.berkshirehathaway.com/letters/2016ltr.pdf.

Calvo, C., Ivorra, C., Liern, V., 2016. Fuzzy portfolio selection with non-financial goals: Exploring the efficient frontier. Ann. Oper. Res. 245, 31–46. https://doi.org/10.1007/s10479-014-1561-2

Campbell, W.K., Goodie, A.S., Foster, J.D., 2004. Narcissism, confidence, and risk attitude. J. Behav. Decis. Mak. 17, 297–311. https://doi.org/https://doi.org/10.1002/bdm.475

Carta, S., Consoli, S., Podda, A.S., Recupero, D.R., Stanciu, M.M., 2022. Statistical arbitrage powered by Explainable Artificial Intelligence. Expert Syst. Appl. 206, 117763. https://doi.org/https://doi.org/10.1016/j.eswa.2022.117763

Chen, G., Kim, K.A., Nofsinger, J.R., Rui, O.M., 2007. Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. J. Behav. Decis. Mak. 20, 425–451. https://doi.org/https://doi.org/10.1002/bdm.561

Chen, J., Hong, H., Huang, M., Kubik, J.D., 2004. Does fund size erode mutual fund performance? The role of liquidity and organization. Am. Econ. Rev. 94, 1276–1302. https://doi.org/10.1257/0002828043052277

Chen, J., Luo, C., Pan, L., Jia, Y., 2021. Trading strategy of structured mutual fund based on deep learning network. Expert Syst. Appl. 183, 115390. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115390

Chen, R., Ren, J., 2022. Do AI-powered mutual funds perform better? Financ. Res. Lett. 47, 102616. https://doi.org/https://doi.org/10.1016/j.frl.2021.102616

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Cremers, K.J.M., Fulkerson, J.A., Riley, T.B., 2019. Challenging the conventional wisdom on active management: A review of the past 20 years of academic literature on actively managed mutual funds. Financ. Anal. J. 75, 8–35. https://doi.org/10.1080/0015198X.2019.1628555

Cremers, K.J.M., Petajisto, A., 2009. How active is your Ffund manager? A new measure that predicts performance. Rev. Financ. Stud. 22, 3329–3365. https://doi.org/10.1093/rfs/hhp057

De Bondt, W.F.M., Thaler, R., 1985. Does the stock market overreact? J. Finance 40, 793–805. https://doi.org/https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

Deboeck, G., 1998. Picking mutual funds with self-organizing maps, in: Springer Finance. pp. 39–58. https://doi.org/10.1007/978-1-4471-3913-3_3

Dorn, D., Huberman, G., 2010. Preferred risk habitat of individual investors. J. financ. econ. 97, 155–173. https://doi.org/https://doi.org/10.1016/j.jfineco.2010.03.013

Elton, E.J., Gruber, M.J., Blake, C.R., Krasny, Y., Ozelge, S.O., 2010. The effect of holdings data frequency on conclusions about mutual fund behavior. J. Bank. Financ. 34, 912–922. https://doi.org/https://doi.org/10.1016/j.jbankfin.2009.10.002

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. J. Finance 25, 383–417. https://doi.org/10.2307/2325486

Fama, E.F., 1965. The behavior of stock-market prices. J. Bus. 38, 34–105.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. financ. econ. 116, 1–22. https://doi.org/https://doi.org/10.1016/j.jfineco.2014.10.010

Feng, X., Zhou, M., Chan, K.C., 2014. Smart money or dumb money? A study on the selection ability of mutual fund investors in China. North Am. J. Econ. Financ. 30, 154–170. https://doi.org/https://doi.org/10.1016/j.najef.2014.09.004

Flammer, C., 2021. Corporate green bonds. J. financ. econ. 142, 499–516. https://doi.org/https://doi.org/10.1016/j.jfineco.2021.01.010

Freeman, R.E., 1984. Strategic management: A stakeholder approach. Bost. Pitman.

Galema, R., Plantinga, A., Scholtens, B., 2008. The stocks at stake: Return and risk in socially responsible investment. J. Bank. Financ. 32, 2646–2654. https://doi.org/https://doi.org/10.1016/j.jbankfin.2008.06.002

Gil-Bazo, J., Ruiz-Verdú, P., 2009. The relation between price and performance in the mutual fund industry. J. Finance 64, 2153–2183.

Gittelsohn, J., 2019. End of era: Passive equity funds surpass active in epic shift. Bloom. Sept. 11.

Guercio, D. Del, Tkac, P.A., 2008. Star power: The effect of morningstar ratings on mutual fund flow. J. Financ. Quant. Anal. 43, 907–936.

Gutsche, G., Ziegler, A., 2019. Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. J. Bank. Financ. 102, 193–214. https://doi.org/https://doi.org/10.1016/j.jbankfin.2019.03.007

Halbritter, G., Dorfleitner, G., 2015. The wages of social responsibility — where are they? A critical review of ESG investing. Rev. Financ. Econ. 26, 25–35. https://doi.org/https://doi.org/10.1016/j.rfe.2015.03.004

Hamilton, S., Jo, H., Statman, M., 1993. Doing well while doing Good? The investment performance of socially responsible mutual funds. Financ. Anal. J. 49, 62–66.

Hartzmark, S.M., Sussman, A.B., 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. J. Finance 74, 2789–2837. https://doi.org/https://doi.org/10.1111/jofi.12841

Hawn, O., Chatterji, A.K., Mitchell, W., 2018. Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). Strateg. Manag. J. 39, 949–976. https://doi.org/https://doi.org/10.1002/smj.2752

Hong, H., Kacperczyk, M., 2009. The price of sin: The effects of social norms on markets. J. financ. econ. 93, 15–36. https://doi.org/https://doi.org/10.1016/j.jfineco.2008.09.001

Huber, J., Kirchler, M., Stöckl, T., 2010. The hot hand belief and the gambler's fallacy in investment decisions under risk. Theory Decis. 68, 445–462. https://doi.org/10.1007/s11238-008-9106-2

Jung, Y., Suh, Y., 2019. Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. Decis. Support Syst. 123, 113074. https://doi.org/https://doi.org/10.1016/j.dss.2019.113074

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263–291. https://doi.org/10.2307/1914185

Krüger, P., 2015. Corporate goodness and shareholder wealth. J. financ. econ. 115, 304–329. https://doi.org/https://doi.org/10.1016/j.jfineco.2014.09.008

Li, Y., Fu, K., Zhao, Y., Yang, C., 2022. How to make machine select stocks like fund managers? Use scoring and screening model. Expert Syst. Appl. 196, 116629. https://doi.org/https://doi.org/10.1016/j.eswa.2022.116629

Malkiel, B.G., 2005. Reflections on the efficient market hypothesis: 30 years later. Financ. Rev. 40, 1–9. https://doi.org/https://doi.org/10.1111/j.0732-8516.2005.00090.x

Mansour, N., Cherif, M.S., Abdelfattah, W., 2019. Multi-objective imprecise programming for financial portfolio selection with fuzzy returns. Expert Syst. Appl. 138, 112810. https://doi.org/https://doi.org/10.1016/j.eswa.2019.07.027

Markowitz, H., 1952. Portfolio selection. J. Finance 7, 77–91. https://doi.org/10.2307/2975974

Merton, R.K., 1968. The matthew effect in science: The reward and communication systems of science are considered. Science (80-.). 159, 56–63.

Miller, M.H., Modigliani, F., 1961. Dividend policy, growth, and the valuation of shares. J. Bus. 34, 411–433.

Moreira, C., Chou, Y.-L., Velmurugan, M., Ouyang, C., Sindhgatta, R., Bruza, P., 2021. LINDA-BN: An interpretable probabilistic approach for demystifying black-box predictive models. Decis. Support Syst. 150, 113561. https://doi.org/https://doi.org/10.1016/j.dss.2021.113561

Muñoz, F., Vargas, M., Marco, I., 2014. Environmental mutual funds: Financial performance and managerial abilities. J. Bus. Ethics 124, 551–569. https://doi.org/10.1007/s10551-013-1893-x

Myers, R.H., 1990. Classical and modern regression with applications. Belmont, CA Duxbury Press.

Nofsinger, J.R., Sias, R.W., 1999. Herding and feedback trading by institutional and individual investors. J. Finance 54, 2263–2295. https://doi.org/https://doi.org/10.1111/0022-1082.00188

Pástor, Ľ., Stambaugh, R.F., Taylor, L.A., 2021. Sustainable investing in equilibrium. J. financ. econ. 142, 550–571. https://doi.org/https://doi.org/10.1016/j.jfineco.2020.12.011

Pedersen, L.H., Fitzgibbons, S., Pomorski, L., 2021. Responsible investing: The ESG-efficient frontier. J. financ. econ. 142, 572–597. https://doi.org/https://doi.org/10.1016/j.jfineco.2020.11.001

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Plagge, J.-C., Grim, D.M., 2020. Have investors paid a performance price? Examining the behavior of ESG equity funds. J. Portf. Manag. 46, 123 LP – 140. https://doi.org/10.3905/jpm.2020.46.3.123

Reboredo, J.C., Otero, L.A., 2021. Are investors aware of climate-related transition risks? Evidence from mutual fund flows. Ecol. Econ. 189, 107148. https://doi.org/https://doi.org/10.1016/j.ecolecon.2021.107148

Renneboog, L., Ter Horst, J., Zhang, C., 2011. Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds. J. Financ. Intermediation 20, 562–588. https://doi.org/10.1016/j.jfi.2010.12.003

Renneboog, L., Ter Horst, J., Zhang, C., 2008. Socially responsible investments: Institutional aspects, performance, and investor behavior. J. Bank. Financ. 32, 1723–1742. https://doi.org/https://doi.org/10.1016/j.jbankfin.2007.12.039

Riedl, A., Smeets, P., 2017. Why do investors hold socially responsible mutual funds? J. Finance 72, 2505–2550. https://doi.org/https://doi.org/10.1111/jofi.12547

Servaes, H., Sigurdsson, K., 2022. The costs and benefits of performance fees in mutual funds. J. Financ. Intermediation 50, 100959. https://doi.org/https://doi.org/10.1016/j.jfi.2022.100959

Sirri, E.R., Tufano, P., 1998. Costly search and mutual fund flows. J. Finance 53, 1589–1622. https://doi.org/https://doi.org/10.1111/0022-1082.00066

Thaler, R., 1980. Toward a positive theory of consumer choice. J. Econ. Behav. Organ. 1, 39–60. https://doi.org/https://doi.org/10.1016/0167-2681(80)90051-7

Tonidandel, S., LeBreton, J.M., 2011. Relative importance analysis: A useful supplement to regression analysis. J. Bus. Psychol. 26, 1–9. https://doi.org/10.1007/s10869-010-9204-3

Vo, N.N.Y., He, X., Liu, S., Xu, G., 2019. Deep learning for decision making and the optimization of socially responsible investments and portfolio. Decis. Support Syst. 124, 113097. https://doi.org/https://doi.org/10.1016/j.dss.2019.113097

Zheng, L., 1999. Is money smart? A study of mutual fund investors’ fund selection ability. J. Finance 54, 901–933. https://doi.org/https://doi.org/10.1111/0022-1082.00131

Zopounidis, C., Galariotis, E., Doumpos, M., Sarri, S., Andriosopoulos, K., 2015. Multiple criteria decision aiding for finance: An updated bibliographic survey. Eur. J. Oper. Res. 247, 339–348. https://doi.org/https://doi.org/10.1016/j.ejor.2015.05.032

Main conclusions

There is an increasing demand for investments to produce not only financial returns, but also social and environmental returns (Bauer et al., 2021). Sustainability indices play an important role as a reference for investors who are interested in companies that meet high environmental and social standards. The providers of these indices steer investment flows through their decisions to list or delist companies from their products (Petry et al., 2019). In this context, it is important to validate and develop alternative index classifications beyond the sustainability labels used by the index providers. To the best of our knowledge, the first chapter is the first study to use cluster analysis to compare the listing and delisting processes of sustainability indices to those of conventional indices in order to validate the "sustainability" label of five FTSE4Good indices. Future research could apply our approach to classify the huge number of indices that exist. This chapter also shows that the size criterion prevails over the environmental, social, and governance (ESG) criteria in the inclusion and exclusion processes of sustainability indices. We conclude that sustainability indices should reduce the influence of size to achieve a real differentiation from the conventional indices.

The literature has shown that large companies have higher ESG ratings (Schreck & Raithel, 2015; Udayasankar, 2008). This phenomenon leads some authors to wonder whether ESG suppliers are providing appropriate information to investors that maximizes their nonfinancial utility (Drempetic et al., 2020). The second chapter of this thesis adapts the methodology used to detect earnings management—discretionary accruals (Dechow et al., 1995; Jones, 1991)—to measure the virtuous behavior of companies. Specifically, we analyze all companies included in the ESG Refinitiv database; and for each score and year, we regress the score on the sizes, countries, and industries of companies. We proxy the virtuous behavior of any company with the extra ESG commitment that represents the error term in the regression. The methodology proposed in Chapter 2 is useful for developing inclusive ESG ratings that do not only consider how "good" the company is, but also the capabilities the company has to be "good". Our results show that the virtuous behavior of a company is negatively associated with its financial performance. However, these weaker financial results are offset by the company's additional commitment to environmental and social issues. In other words, we identify companies that sacrifice financial profitability to yield environmental and social returns for their shareholders. As opposed to the traditional approach of maximizing market value, our results are in line with recent studies arguing that companies should maximize shareholder welfare (Hart & Zingales, 2017).

Several studies have analyzed the drivers of CSR, but the role of public attention on CSR has remained unexplored (Flammer, 2021; Liang & Renneboog, 2017; O'Sullivan et al., 2021). From a theoretical perspective, companies react to greater public attention, such as greater exposure to public scrutiny and visibility, by increasing their commitment to CSR to align their actions with societal values and norms (Baldini et al., 2018; Suchman, 1995). However, the empirical evidence supporting this relationship is weak due to the difficulties in finding a good indicator of attention. The third chapter of the thesis overcomes this limitation by proposing a new method that uses Google Trends to measure public attention. While other studies have used the search volume index that is not comparable among companies, we manage to identify the most searched companies by using pairwise comparisons in Google Trends of the individual search volume index. Thus, we create a measure of public attention that ranks companies according to the number of web searches. Our results show a positive relationship between public attention and CSR performance. We also follow a quasi-experimental approach to study a causal relationship. From our pool of firms, we detect those that have experienced a large increase in public attention compared to the previous year. Then, we use the nearest neighbor algorithm to identify a contrafactual for each company. This method allows us to make robust conclusions on how firms react to this increase in their public attention. We conclude that companies react to a shock from public attention by improving their CSR performance. Thus, the third chapter not only sheds light on the relationship between public attention and CSR, but also proposes a new method of measuring public attention that can be used in several disciplines by using Google Trends as a source of information (Chandrasekaran et al., 2018; Choi et al., 2020; Flanagan et al., 2021).

The growing interest for sustainable investments has been confirmed in some studies that show a positive relation between some sustainability indicators and money flows of mutual funds (Ammann et al., 2019; Hartzmark & Sussman, 2019). However, Chapter 4 shows that the ESG score of the fund is not as important in predicting fund flows as past performance, past growth, or managerial fees. We use several models (linear and logistic regressions, random forest, gradient boosting, and neural networks) to predict the future flows raised by the funds. The dominance analysis and the permutation feature importance show the scarce contribution of the fund's ESG score in explaining fund flows. Similarly, the analyses of standardized regression coefficients and Shapley additive explanations show that the nonfinancial factors have a limited effect on the fund flows. The great interest shown by industry and academia on nonfinancial information does not seem to be reflected in investors’ preferences. Our results also contribute to the literature by finding that past ESG performance of the fund explains future ESG performance.

References

Ammann, M., Bauer, C., Fischer, S., & Müller, P. (2019). The impact of the Morningstar Sustainability Rating on mutual fund flows. European Financial Management, 25(3), 520–553. https://doi.org/https://doi.org/10.1111/eufm.12181

Baldini, M., Maso, L. D., Liberatore, G., Mazzi, F., & Terzani, S. (2018). Role of Country- and Firm-Level Determinants in Environmental, Social, and Governance Disclosure. Journal of Business Ethics, 150(1), 79–98. https://doi.org/10.1007/s10551-016-3139-1

Bauer, R., Ruof, T., & Smeets, P. (2021). Get real! Individuals prefer more sustainable investments. The Review of Financial Studies, 34(8), 3976–4043. https://doi.org/10.1093/rfs/hhab037

Chandrasekaran, D., Srinivasan, R., & Sihi, D. (2018). Effects of offline ad content on online brand search: insights from super bowl advertising. Journal of the Academy of Marketing Science, 46(3), 403–430. https://doi.org/10.1007/s11747-017-0551-8

Choi, D., Gao, Z., & Jiang, W. (2020). Attention to Global Warming. The Review of Financial Studies, 33(3), 1112–1145. https://doi.org/10.1093/rfs/hhz086

Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. The Accounting Review, 70(2), 193–225. http://www.jstor.org/stable/248303

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. Journal of Business Ethics, 167(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1

Flammer, C. (2021). Corporate green bonds. Journal of Financial Economics, 142(2), 499–516. https://doi.org/https://doi.org/10.1016/j.jfineco.2021.01.010

Flanagan, R., Kuo, B., & Staller, K. (2021). Utilizing Google Trends to Assess Worldwide Interest in Irritable Bowel Syndrome and Commonly Associated Treatments. Digestive Diseases and Sciences, 66(3), 814–822. https://doi.org/10.1007/s10620-020-06290-7

Gutsche, G., & Ziegler, A. (2019). Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. Journal of Banking & Finance, 102, 193–214. https://doi.org/https://doi.org/10.1016/j.jbankfin.2019.03.007

Hart, O., & Zingales, L. (2017). Companies should maximize shareholder welfare not market value. SSRN, August. https://doi.org/http://dx.doi.org/10.2139/ssrn.3004794

Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. The Journal of Finance, 74(6), 2789–2837. https://doi.org/https://doi.org/10.1111/jofi.12841

Jones, J. J. (1991). Earnings management during import relief investigations. Journal of Accounting Research, 29(2), 193–228. https://doi.org/10.2307/2491047

Liang, H. ao, & Renneboog, L. (2017). On the foundations of corporate social responsibility. The Journal of Finance, 72(2), 853–910. https://doi.org/https://doi.org/10.1111/jofi.12487

O'Sullivan, D., Zolotoy, L., & Fan, Q. (2021). CEO early-life disaster experience and corporate social performance. Strategic Management Journal, 42(11), 2137–2161. https://doi.org/https://doi.org/10.1002/smj.3293

Petry, J., Fichtner, J., & Heemskerk, E. (2019). Steering capital: the growing private authority of index providers in the age of passive asset management. Review of International Political Economy, 28(1), 152–176. https://doi.org/10.1080/09692290.2019.1699147

Reboredo, J. C., & Otero, L. A. (2021). Are investors aware of climate-related transition risks? Evidence from mutual fund flows. Ecological Economics, 189, 107148. https://doi.org/https://doi.org/10.1016/j.ecolecon.2021.107148

Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. Journal of Banking & Finance, 32(9), 1723–1742. https://doi.org/https://doi.org/10.1016/j.jbankfin.2007.12.039

Schreck, P., & Raithel, S. (2015). Corporate Social Performance, Firm Size, and Organizational Visibility: Distinct and Joint Effects on Voluntary Sustainability Reporting. Business & Society, 57(4), 742–778. https://doi.org/10.1177/0007650315613120

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. Academy of Management Review, 20(3), 571–610. https://doi.org/10.5465/amr.1995.9508080331

Udayasankar, K. (2008). Corporate social responsibility and firm size. Journal of Business Ethics, 83(2), 167–175. https://doi.org/10.1007/s10551-007-9609-8

Resumen y conclusiones
(Summary in Spanish)

Tal y como recoge el Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de Doctorado, las tesis doctorales que quieran optar a la Mención Internacional deben incluir los principales contenidos de la misma en dos lenguas oficiales para la comunicación científica.

Dado que el idioma de redacción de la tesis es el inglés, a continuación, se presenta un resumen de la tesis doctoral desarrollada en español, con el objetivo de que la misma pueda ser considerada para la obtención de la Mención de Internacional.

Resumen

Esta tesis doctoral está motivada por la creciente importancia que los mercados financieros otorgan a la información no financiera. Actualmente, 1 de cada 3 euros bajo gestión profesional en los Estados Unidos sigue algún tipo de estrategia sostenible frente a los uno de cada ocho de 2010 (USIF, 2010; USIF, 2020). Además, la mitad de los 28 billones de inversión necesarios para lograr el objetivo de cero emisiones de carbono para 2050 no son rentables (EUROSIF, 2021). En este contexto, los mercados de capitales juegan un papel crucial para financiar la transición ecológica. Esta tesis doctoral pretende contribuir a este objetivo estudiando la idoneidad de los índices de sostenibilidad para guiar aquellos flujos de inversión que siguen criterios ambientales sociales y de gobernanza (ASG); la capacidad de los ratings ASG para identificar compañías comprometidas con temas ambientales y sociales; los factores que empujan a las compañías a mejorar sus registros ambientales y sociales; así como hasta qué punto los inversores valoran los ratings ASG de los fondos de inversión a la hora de tomar sus decisiones de inversión.

El primer capítulo de la tesis doctoral analiza como los índices de sostenibilidad FTSE4Good aplican criterios ASG a la hora de incluir o excluir compañías de su cesta de componentes. Estudios previos se han centrado en comparar el desempeño financiero de estos índices (Cunha et al., 2020), o en estudiar si los mercados financieros valoran la inclusión o exclusión de una compañía en estos índices (Hawn et al., 2018). Sin embargo, se sabe poco sobre cómo estos índices aplican los criterios ASG en su proceso de inclusión y exclusión en comparación con los índices convencionales. Además, una de las mayores preocupaciones que plantean los productos sostenibles son las sobras de *greenwasing* (Berrone et al., 2017; Lyon & Montgomery, 2015). Es decir, los proveedores de índices bursátiles pueden utilizar la etiqueta de sostenibilidad como herramienta de marketing sin que realmente sus índices sostenibles sigan criterios diferentes a los empleados por sus homólogos convencionales. El primer capítulo de esta tesis doctoral aborda esta problemática utilizando modelos probit y análisis cluster. Los resultados muestran que el factor más importante empleado por los índices de sostenibilidad al incluir o excluir compañías es su capitalización de mercado en lugar de su desempeño ASG. También, el análisis cluster muestra que los criterios de inclusión y exclusión aplicados por algunos índices de sostenibilidad no difiere de los criterios

aplicados por índices convencionales –especialmente en el proceso de exclusión–. Este capítulo sugiere que los proveedores de índices sostenibles deben disminuir la influencia que la capitalización de mercado tiene en su proceso de inclusión y exclusión, y de este modo lograr una diferenciación más clara respecto con sus índices convencionales.

Estudios recientes critican la excesiva influencia que el tamaño de la compañía tiene a la hora de conseguir un buen rating ASG (Drempetic et al., 2020). Además, frecuentemente son las compañías de países desarrollados, como las compañías europeas, las que obtienen los mejores ratings ASG (Demirbag et al., 2017; Liang & Renneboog, 2017). El segundo capítulo analiza todas las compañías con un rating ASG en la base de datos Refinitiv para desarrollar una metodología que permita obtener ratings ASG inclusivos. Estos ratings inclusivos evitarían que empresas pequeñas o ciertos países queden excluidos de los flujos de dinero socialmente responsable. Los ratings inclusivos que proponemos valoran a las empresas de acuerdo con su capacidad para cumplir con estándares ambientales y sociales. Concretamente, proponemos una regresión en corte transversal que capture el comportamiento virtuoso de la empresa como el exceso ASG de la compañía en relación a su valor de mercado, país e industria. Este capítulo contribuye a la literatura previa al estudiar la relación entre la responsabilidad social corporativa de la empresa (RSC) y su desempeño financiero desde una perspectiva inclusiva. Los resultados muestran que aquellas compañías que cumplen con estándares ambientales y sociales más elevados que sus homólogas tienen un peor desempeño financiero. Además, la metodología propuesta, al proporcionar un indicador comparable entre empresas que mide su comportamiento virtuoso, es útil para gestores, inversores, reguladores e investigadores.

El tercer capítulo de esta tesis doctoral examina el papel que la atención pública hacia la compañía tiene en su política de RSC. Bajo el marco teórico de la teoría de la legitimidad, la RSC de la empresa es vista como una herramienta para aumentar y salvaguardar su legitimad a ojos de la sociedad (Baldini et al., 2018; Cormier & Magnan, 2015; Hörisch et al., 2015). Las compañías, ante incrementos de su escrutinio público y visibilidad, reaccionarían aumentando su compromiso con la RSC. Sin embargo, los estudios empíricos que abordan esta relación son escasos y las variables utilizadas para medir la atención pública no son siempre adecuadas. El tercer capítulo aborda esta problemática proponiendo un nuevo método de utilizar Google Trends para obtener un ranking anual de atención pública para las compañías del S&P 500. Este ranking ordena a las empresas de acuerdo con el número de búsquedas web que han recibido. Primero, nuestros resultados muestran una relación positiva entre atención pública y el desempeño en RSC. Luego, siguiendo un enfoque cuasiexperimental y aplicando técnica de macheo, mostramos que las compañías mejoran su desempeño en RSC después de un "shock" de atención pública. Por lo tanto, concluimos que el escrutinio público y la señalización sería una estrategia efectiva para presionar a las compañías a que mejoren sus estándares ambientales y sociales. Más allá de los enfoques convencionales (Choi et al., 2020; Da et al., 2011), este capítulo también proporciona un nuevo método de utilizar Google Trends útil para investigadores y profesionales.

El último capítulo de esta tesis doctoral estudia la influencia de los factores no financieros en las decisiones de inversión. Estudios basados en encuestas o en experimentos de elección muestran que los inversores están dispuestos a renunciar a retornos financieros a cambio de invertir en productos sostenibles (Gutsche & Ziegler, 2019; Riedl & Smeets, 2017). En esta línea, algunos autores demuestran que los inversores invierten más en fondos con altas calificaciones de sostenibilidad que en fondos con bajas calificaciones (Ammann et al., 2019; Hartzmark & Sussman, 2019). Estos estudios, revelan una relación positiva y significativa de los factores de sostenibilidad en las decisiones de inversión. Sin embargo, la influencia que los factores no financieros tienen en los flujos de inversión en comparación con la influencia de los factores financieros ha quedado sin explorar. En este capítulo abordamos esta cuestión mediante regresiones y el uso de técnicas de aprendizaje automático (redes neuronales, *random forest* y *gradient boosting*). Nuestros modelos predicen los flujos con una precisión del 70%. El análisis de dominancia y la importancia de la característica de permutación muestra que los factores no financieros tienen un impacto limitado en la bondad de ajuste del modelo. De forma similar, el análisis de regresiones con coeficientes estandarizados y *Shapley additive explanations* muestran que las variables no financieras tienen un impacto limitado en los flujos del fondo. De este modo, este capítulo concluye que los inversores consideran el desempeño ESG, pero los factores que realmente importan son el crecimiento pasado, las comisiones de gestión y la rentabilidad pasada.

Conclusiones

Cada vez más los inversores exigen que las inversiones rindan no sólo beneficios financieros, sino también sociales y medioambientales (Bauer et al., 2021). Los índices de sostenibilidad desempeñan un importante papel como referencia para aquellos inversores que desean invertir en empresas que cumplen con altos estándares medioambientales y sociales. De este modo, los creadores de estos índices dirigen los flujos de inversión mediante sus decisiones de incluir o excluir empresas de sus índices de sostenibilidad (Petry et al., 2019). En este contexto, es importante validar y desarrollar clasificaciones alternativas más allá de las etiquetas utilizadas por estos proveedores. Hasta dónde sabemos, este capítulo es el primer estudio que utiliza análisis cluster para comparar el proceso de inclusión y exclusión de índices de sostenibilidad y validar la etiqueta de sostenibilidad de cinco índices FTSE4Good. Otros estudios pueden seguir nuestro enfoque para proponer clasificaciones alternativas al gran número de índices que existe. Este capítulo también muestra que la capitalización de mercado prevalece sobre los criterios ASG en el proceso de inclusión y exclusión de los índices de sostenibilidad. Concluimos que la influencia del valor de mercado debe disminuir para lograr que los índices de sostenibilidad logren una diferenciación real respecto con los índices convencionales.

Varios estudios muestran que las grandes empresas tienen también ratings ASG más elevados (Schreck & Raithel, 2015; Udayasankar, 2008). Este fenómeno lleva a algunos autore a plantearse si los proveedores ASG ofrecen la información apropiada para que los inversores puedan maximizar su utilidad no financiera (Drempetic et al., 2020). El segundo capítulo de esta tesis adapta la metodología utilizada para detectar prácticas de contabilidad creativa —ajustes discrecionales (Dechow et al., 1995; Jones, 1991)— para medir el comportamiento virtuoso de las empresas. Concretamente, analizamos todas las compañías en la base de datos ASG de Refitniv, y para cada rating y año, hacemos una regresión de la puntuación ASG en función del tamaño de la empresa, el país y el sector. Medimos el comportamiento virtuoso de cualquier empresa como el compromiso adicional de ASG que representa el término de error de esta regresión. La metodología propuesta en el capítulo 2 es útil para desarrollar rating ASG inclusivos que no consideren únicamente lo “buena” que es una empresa, sino también las capacidades que tiene una empresa de ser “buena”. Nuestros resultados muestran que el comportamiento virtuoso de la empresa está asociado negativamente con su desempeño financiero. Sin embargo, estos resultados financieros más débiles se ven compensados por el compromiso adicional que tiene la empresa con cuestiones medioambientales y sociales. En otras palabras, en este trabajo identificamos empresas que estarían sacrificando rentabilidad financiera para obtener rentabilidad medioambiental y social para sus accionistas. Frente al enfoque tradicional de maximizar el valor de mercado nuestros resultados están en línea con estudios recientes que argumentan que las compañías tienen que maximizar el bienestar de sus accionistas (Hart & Zingales, 2017).

Varias investigaciones han estudiado los determinantes de la RSC, pero la influencia de la atención pública en las políticas de RSC de la empresa ha permanecido sin analizar (Flammer, 2021; Liang & Renneboog, 2017; O'Sullivan et al., 2021). Desde una perspectiva teórica, las empresas reaccionan a una mayor atención pública, como una mayor exposición al escrutinio público y visibilidad, aumentando su compromiso con la RSC para alinear sus acciones con los valores y normas sociales (Baldini et al., 2018; Suchman, 1995). Sin embargo, la evidencia empírica que apoya esta relación es frágil debido a las dificultades que existen para encontrar un buen indicador de atención pública. El tercer capítulo de la tesis supera esta limitación proponiendo una nueva forma de usar Google Trends. Mientras estudios previos utilizan el índice de volumen de búsquedas, el cual no es comprable entre empresas, nosotros logramos identificar que empresas son más buscadas mediante comparación por pares de los volúmenes de búsqueda individuales. De este modo creamos una medida de atención que ordena a las compañías de acuerdo con su número de búsquedas en Google. Nuestros resultados muestran una relación positiva entre atención pública y el desempeño en RSC. También seguimos un enfoque cuasiexperimental para estudiar la existencia de una relación causal. Del conjunto de empresas analizadas detectamos aquellas compañías que han experimentado un gran aumento de atención respecto al año previo. Entonces, utilizando el algoritmo del vecino más cercano identificamos un contrafactual para cada empresa. Esto nos permite obtener conclusiones precisas de como las compañías reaccionan ante un incremento de su atención pública. Nuestros resultados muestran que las empresas reaccionan ante un shock de atención pública aumentando su desempeño en RSC. El tercer capítulo de la presente tesis doctoral arroja luz sobre la relación entre atención pública y RSC. Al mismo tiempo también se propone un nuevo método de medir la atención pública que puede ser utilizado en muchas otras disciplinas que utilizan Google Trends como fuente de información (Chandrasekaran et al., 2018; Choi et al., 2020; Flanagan et al., 2021).

El creciente interés por las inversiones sostenibles ha sido confirmado por diversos estudios que muestran una relación positiva entre indicadores de sostenibilidad y flujos de dinero hacia los fondos de inversión (Ammann et al., 2019; Hartzmark & Sussman, 2019). Sin embargo, el capítulo 4 de esta tesis muestra que la puntuación ESG del fondo no es tan importante a la hora de predecir los flujos del fondo como la rentabilidad pasada, el crecimiento pasado o las comisiones de gestión. En este capítulo se utilizan varios métodos para pedir los flujos futuros obtenidos por los fondos de inversión (regresiones logísticas y lineales, *random forest*, *gradient boosting*, y redes neuronales). El análisis de dominancia y la importancia de la característica de permutación muestra que el ESG score del fondo tienen una importancia residual a la hora de predecir los flujos de los fondos. De forma similar, el análisis de coeficientes estandarizados y los *Shapley additive explanations* muestran que los factores no financieros tienen un impacto limitado en los flujos de inversión. El gran interés que la industria y el mundo académico muestran por la información no financiera parece no tener su reflejo en las preferencias de inversión. Nuestros resultados también contribuyen a la literatura al constatar que el desempeño ASG pasado del fondo explican su desempeño ASG futuro.

Bibliografía

Ammann, M., Bauer, C., Fischer, S., & Müller, P. (2019). The impact of the Morningstar Sustainability Rating on mutual fund flows. European Financial Management, 25(3), 520–553. https://doi.org/https://doi.org/10.1111/eufm.12181

Baldini, M., Maso, L. D., Liberatore, G., Mazzi, F., & Terzani, S. (2018). Role of Country- and Firm-Level Determinants in Environmental, Social, and Governance Disclosure. Journal of Business Ethics, 150(1), 79–98. https://doi.org/10.1007/s10551-016-3139-1

Bauer, R., Ruof, T., & Smeets, P. (2021). Get real! Individuals prefer more sustainable investments. The Review of Financial Studies, 34(8), 3976–4043. https://doi.org/10.1093/rfs/hhab037

Berrone, P., Fosfuri, A., & Gelabert, L. (2017). Does greenwashing pay off? Understanding the relationship between environmental actions and environmental legitimacy. Journal of Business Ethics, 144(2), 363–379. https://doi.org/10.1007/s10551-015-2816-9

Chandrasekaran, D., Srinivasan, R., & Sihi, D. (2018). Effects of offline ad content on online brand search: insights from super bowl advertising. Journal of the Academy of Marketing Science, 46(3), 403–430. https://doi.org/10.1007/s11747-017-0551-8

Choi, D., Gao, Z., & Jiang, W. (2020). Attention to Global Warming. The Review of Financial Studies, 33(3), 1112–1145. https://doi.org/10.1093/rfs/hhz086

Cormier, D., & Magnan, M. (2015). The Economic Relevance of Environmental Disclosure and its Impact on Corporate Legitimacy: An Empirical Investigation. Business Strategy and the Environment, 24(6), 431–450. https://doi.org/https://doi.org/10.1002/bse.1829

Cunha, F. A. F. de S., de Oliveira, E. M., Orsato, R. J., Klotzle, M. C., Cyrino Oliveira, F. L., & Caiado, R. G. G. (2020). Can sustainable investments outperform traditional benchmarks? Evidence from global stock markets. Business Strategy and the Environment, 29(2), 682–697. https://doi.org/https://doi.org/10.1002/bse.2397

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. The Journal of Finance, 66(5), 1461–1499. https://doi.org/https://doi.org/10.1111/j.1540-6261.2011.01679.x

Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. The Accounting Review, 70(2), 193–225. http://www.jstor.org/stable/248303

Demirbag, M., Wood, G., Makhmadshoev, D., & Rymkevich, O. (2017). Varieties of CSR: Institutions and socially responsible behaviour. International Business Review, 26(6), 1064–1074. https://doi.org/https://doi.org/10.1016/j.ibusrev.2017.03.011

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. Journal of Business Ethics, 167(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1

EUROSIF, (2018). European SRI study 2018. Available at: http://www.eurosif.org/wp-content/uploads/2018/11/European-SRI-2018-Study.pdf

Flammer, C. (2021). Corporate green bonds. Journal of Financial Economics, 142(2), 499–516. https://doi.org/https://doi.org/10.1016/j.jfineco.2021.01.010

Flanagan, R., Kuo, B., & Staller, K. (2021). Utilizing Google Trends to Assess Worldwide Interest in Irritable Bowel Syndrome and Commonly Associated Treatments. Digestive Diseases and Sciences, 66(3), 814–822. https://doi.org/10.1007/s10620-020-06290-7

Gutsche, G., & Ziegler, A. (2019). Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. Journal of Banking & Finance, 102, 193–214. https://doi.org/https://doi.org/10.1016/j.jbankfin.2019.03.007

Hart, O., & Zingales, L. (2017). Companies should maximize shareholder welfare not market value. SSRN, August. https://doi.org/http://dx.doi.org/10.2139/ssrn.3004794

Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. The Journal of Finance, 74(6), 2789–2837. https://doi.org/https://doi.org/10.1111/jofi.12841

Hawn, O., Chatterji, A. K., & Mitchell, W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). Strategic Management Journal, 39(4), 949–976. https://doi.org/https://doi.org/10.1002/smj.2752

Hörisch, J., Johnson, M. P., & Schaltegger, S. (2015). Implementation of sustainability management and company size: A knowledge-based view. Business Strategy and the Environment, 24(8), 765–779. https://doi.org/https://doi.org/10.1002/bse.1844

Jones, J. J. (1991). Earnings management during import relief investigations. Journal of Accounting Research, 29(2), 193–228. https://doi.org/10.2307/2491047

Liang, H. ao, & Renneboog, L. (2017). On the foundations of corporate social responsibility. The Journal of Finance, 72(2), 853–910. https://doi.org/https://doi.org/10.1111/jofi.12487

Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. Organization & Environment, 28(2), 223–249. https://doi.org/10.1177/1086026615575332

O'Sullivan, D., Zolotoy, L., & Fan, Q. (2021). CEO early-life disaster experience and corporate social performance. Strategic Management Journal, 42(11), 2137–2161. https://doi.org/https://doi.org/10.1002/smj.3293

Petry, J., Fichtner, J., & Heemskerk, E. (2019). Steering capital: the growing private authority of index providers in the age of passive asset management. Review of International Political Economy, 28(1), 152–176. https://doi.org/10.1080/09692290.2019.1699147

Riedl, A., & Smeets, P. (2017). Why do investors hold socially responsible mutual funds? The Journal of Finance, 72(6), 2505–2550. https://doi.org/https://doi.org/10.1111/jofi.12547

Schreck, P., & Raithel, S. (2015). Corporate Social Performance, Firm Size, and Organizational Visibility: Distinct and Joint Effects on Voluntary Sustainability Reporting. Business & Society, 57(4), 742–778. https://doi.org/10.1177/0007650315613120

Social Investment Forum Foundation, 2010. Report on Responsible Investing Trends in the United States 2010. http://www.socialinvest.org

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. Academy of Management Review, 20(3), 571–610. https://doi.org/10.5465/amr.1995.9508080331

Udayasankar, K. (2008). Corporate social responsibility and firm size. Journal of Business Ethics, 83(2), 167–175. https://doi.org/10.1007/s10551-007-9609-8

US SIF Foundation, 2020. Report on US Sustainable and Impact Investing Trends 2020. https://www.ussif.org/