Journal of Cleaner Production 330 (2022) 129862

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

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

Pablo Vilas, Laura Andreu *, José Luis Sarto

Department of Accounting and Finance, University of Zaragoza, Zaragoza, Spain

ARTICLE INFO

Handling Editor: Zhifu Mi

Keywords:
Clustering
Corporate social performance (CSP)
Corporate social responsibility (CSR)
ESG
Socially responsible investment (SRI)
Sustainability indices

ABSTRACT

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.

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) (see, e.g., Montiel, 2008; Liang and Renneboog, 2017; Abbas, 2020). However, there is no single definition of CSR (see, e.g., 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 (see, e.g., 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

Abbreviations: ESG, Environmental, Social, Governance; CSR, Corporate Social Responsibility; CSP, Corporate Social / Sustainability Performance; SR, Socially responsible.

* Corresponding author.
E-mail addresses: pvilasn@unizar.es (P. Vilas), landreu@unizar.es (L. Andreu), jlsarto@unizar.es (J.L. Sarto).

https://doi.org/10.1016/j.jclepro.2021.129862
Received 3 March 2021; Received in revised form 28 October 2021; Accepted 24 November 2021
Available online 25 November 2021

because they are some of the most important ones, and other studies have widely used the FTSE4Good series (see, e.g., 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 (see, e.g., Borgers and Pownall, 2014; Gutsche and Ziegler, 2019).

However, most studies focus on comparing the financial performance of conventional and sustainability indices (see, e.g., 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 (see, e.g., 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 (see, e.g., Pineiro-Chousa et al., 2019; Arribas et al., 2021). However, these studies only analyze sustainability indices, 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 (see, e.g., 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 (see, e.g., 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 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

[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 (see, e.g., 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.

Table 1
Review of studies that use inclusion, permanence or exclusion from sustainability indices as a proxy for CSP.

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

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.

Table 2
Indices analyzed by geographic area and data supplier.

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

This table lists the 16 indices analyzed in this study and their geographic areas, the index type, the data suppliers, and the Refinitiv ticker.

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 (see, e.g., 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.

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 2 provides more information about the analyzed indices such as the geographic area or the index supplier.

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 IA.1 of the Internet 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 7378 companies. The number of companies analyzed has increased over time which reflects the growth in the ESG rating industry (see, e.g., 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 (see, e.g., 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 (see, e.g., 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 IA.2 of the Internet 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 IA.3 of the Internet 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 (see, e.g., Ferrell 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.

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.

Fig. 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, Fig. 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 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:

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

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; $ROA_{iy\text{-}1}$ 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

[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.

[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.

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.

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 IA.4 of the

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

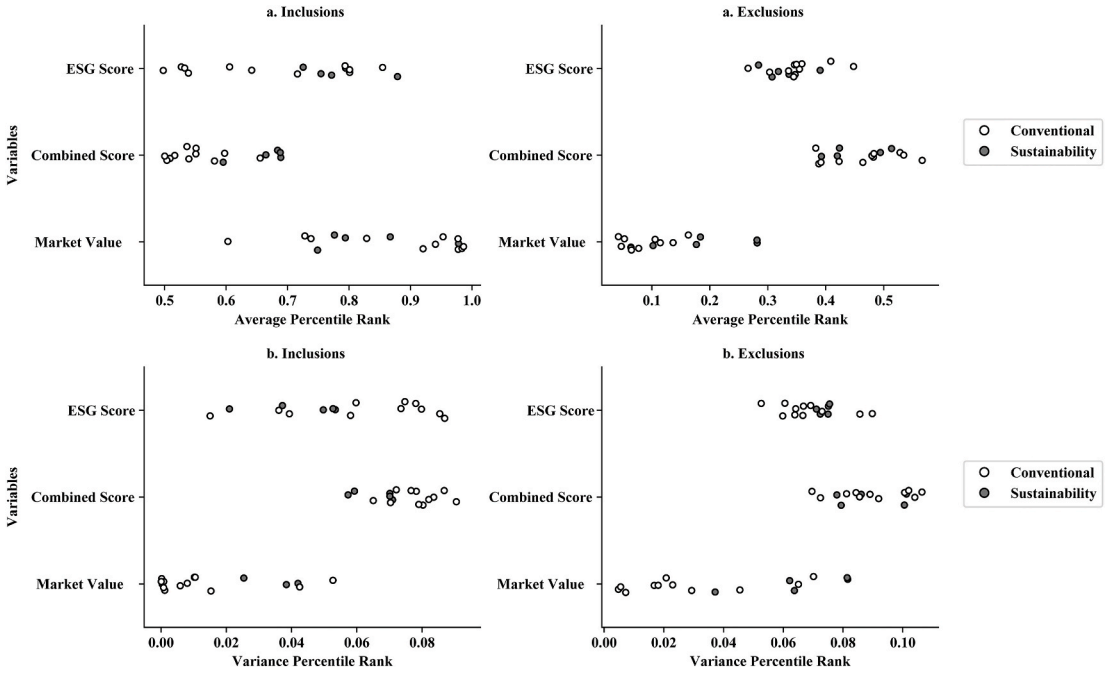

Fig. 1. Average and variance of the percentile rank of inclusions and exclusions. This figure shows the average and variance of the percentile rank for exclusions and inclusions in the three variables analyzed (ESG score, ESG combined score, and market value) for conventional and sustainability indices.

Table 3
Comparison of the companies included in the index versus the universe of companies that could be included.

	#		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	1037	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	5884	0.527	0.499	1.5	0.551	0.498	2.9**
FTSE Eurofirst 100	30	9929	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**

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.

Internet 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 (see, e.g., 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, Fig. 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 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.

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 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.

Table 4
Comparison of the companies that remain in the index versus those excluded from the index.

	#		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	1975	0.39	0.506	−4.1**	0.494	0.5	−0.2
FTSE4Good US	102	3621	0.318	0.505	−6.4**	0.424	0.502	−2.7**
FTSE4Good US 100	98	2163	0.337	0.507	−5.7**	0.513	0.499	0.5
FTSE4Good Europe	124	8780	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	3970	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	1080	0.266	0.506	−4.2**	0.566	0.498	1.2
FTSE Eurotop 100	56	1072	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	1577	0.355	0.503	−3.0**	0.534	0.499	0.7
S&P EURO	40	4772	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**

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.

Table 5
Comparison between the CSP and size criteria for inclusions and exclusions in terms of average.

	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**

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.

Table 6
Comparison between the CSP and size criteria for inclusions and exclusions in terms of variance.

	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**

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: $\sigma_S^2 = \sigma_{CSP}^2$) between size (S) and the two CSP variables. The * and ** indicate statistical significance at the 5%, 1%, levels, respectively.

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 (see, e.g., Hasan, et al., 2018; Yen et al., 2019). Thus, we only accept Hypothesis 1B for three sustainability indices.

4.2. *Hypothesis 2: does the influence of CSP dominate the influence of size in the inclusion and exclusion processes of sustainability indices?*

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

Table 7
Probit regression for analyzing the influence of CSP and size on the index inclusions.

	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 (0.202)	−0.79**	−2.61**	0.84*	0.91*	0.03	0.54 (0.359)	−1.32**	0.96*	−0.49**
	(0.099)	(0.273)	(0.197)	(0.234)	(0.160)	(0.169)		(0.266)	(0.533)	(0.394)	(0.374)	(0.171)		(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 (0.266)	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)		(0.163)	(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 (0.007)	0.01*	−0.00	0.01**	0.00**	0.01 (0.006)	0.00	−0.00	0.01	−0.02	0.01**	0.00 (0.009)	−0.00	0.03 (0.017)	0.00
	(0.002)		(0.004)	(0.004)	(0.002)	(0.001)		(0.003)	(0.002)	(0.007)	(0.013)	(0.003)		(0.006)		(0.002)
Liabilities to Assets	0.18*	0.19 (0.231)	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 (0.542)	0.13
	(0.080)		(0.160)	(0.192)	(0.146)	(0.059)	(0.207)	(0.178)	(0.119)	(0.425)	(0.421)	(0.137)	(0.365)	(0.243)		(0.092)
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	5994	9662	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

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.

(exclusions) is higher (lower) in terms of the CSP than in terms of size. Table 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.

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 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.

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 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 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

Table 8
Probit regression for analyzing the influence of CSP and size on the index exclusions.

	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	2124	3736	2224	8870	265,313	4195	12,625	30,513	1224	1318	26,277	1678	4790	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

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.

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.

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 Fig. 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 IA.5 of the Internet Appendix and are plotted in Fig. 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.

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 IA.6 of the Internet Appendix and are plotted in Fig. 3. All algorithms confirm that the criteria of the

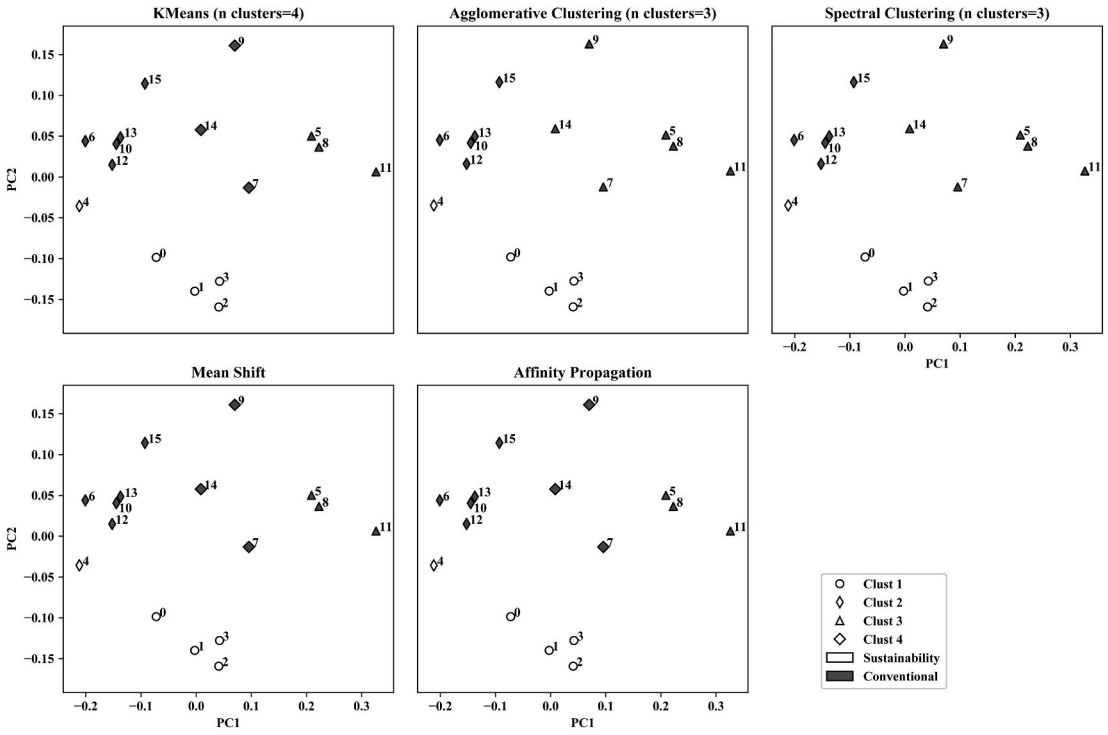

Fig. 2. Best Clustering algorithms that find differences between the inclusion processes of conventional and sustainability indices
This figure shows those clustering algorithms that group some sustainability indices separately from conventional ones using the information in Table IA.5 of the Internet 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.

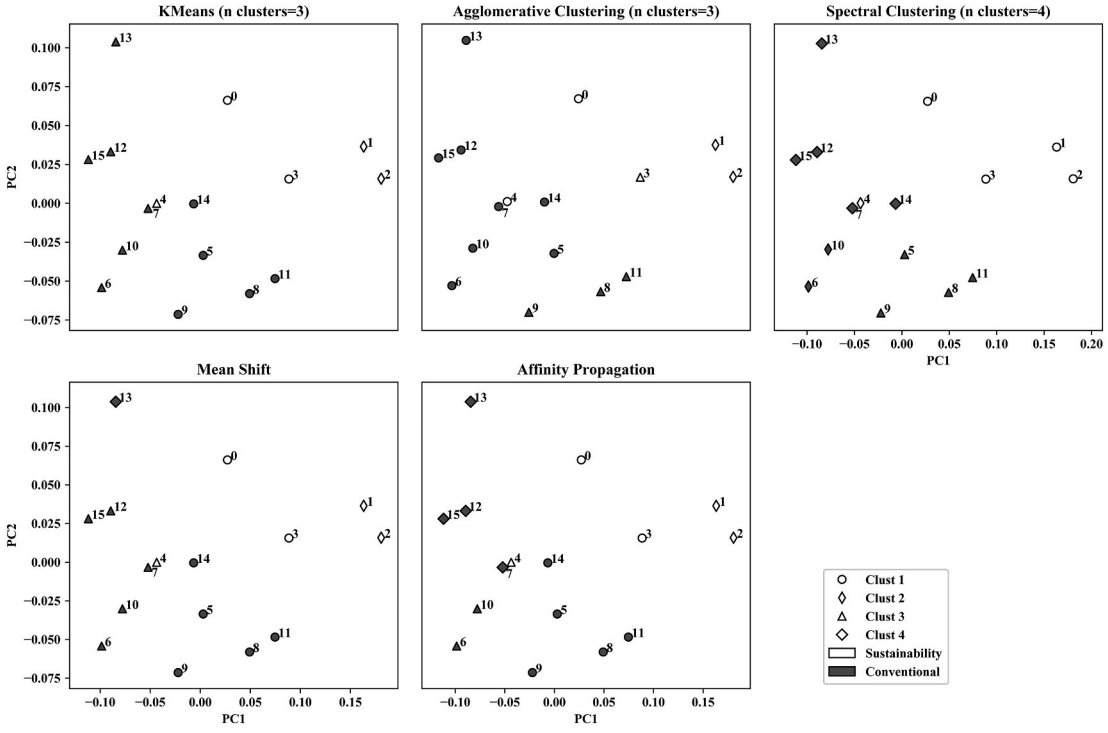

Fig. 3. Best Clustering algorithms that find differences between the exclusion processes of conventional and sustainability indices
This figure shows those clustering algorithms that group some sustainability indices separately from conventional ones using the information in Table IA.6 of the Internet 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.

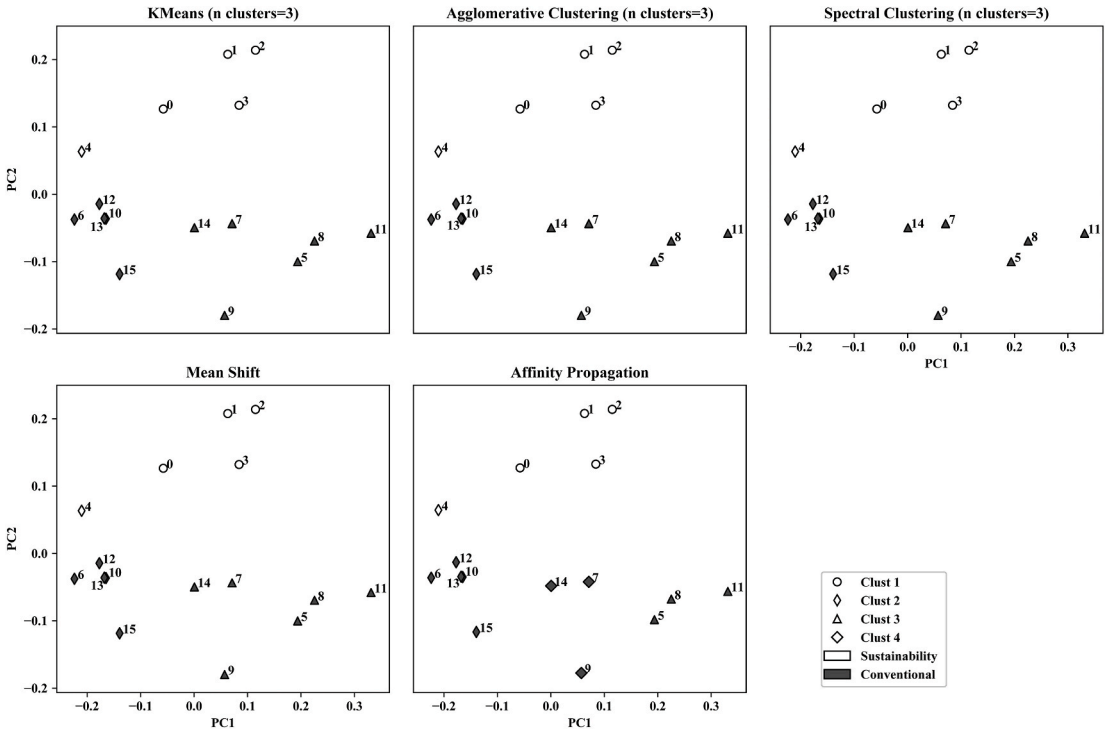

Fig. 4. Best Clustering algorithms that find differences between the inclusion and exclusion processes of conventional and sustainability indices
This figure shows those clustering algorithms that group some sustainability indices separately from conventional ones using the information in Table IA.7 of the Internet 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.

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 Fig. 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.

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 IA.7 of the Internet Appendix and are plotted in Fig. 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, Fig. 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.

CRediT authorship contribution statement

Pablo Vilas: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft. **Laura Andreu:** Conceptualization, Methodology, Writing – review & editing, Supervision. **José Luis Sarto:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the participants at the NIPE Seminar held in April 2021 in University of Minho and at the 11th Portuguese Finance Network Conference for their insightful comments. This work was supported by MCIN/10.13039/50110001103 and by ERDF A way of making Europe (Grant RTI2018-093483-B-I00) and by the Government of Aragon (S38_20R) and by ERDF (Grant S38_20R).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2021.129862.

References

Abbas, J., 2020. Impact of total quality management on corporate green performance through the mediating role of corporate social responsibility. J. Clean. Prod. 242 https://doi.org/10.1016/j.jclepro.2019.118458.
Aldenderfer, M.S., Blashfield, R.K., 1984. Cluster Analysis. SAGE Publications, Inc. https://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. J. Clean. Prod. 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. Account. 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. Int. J. Sustain. Dev. World Ecol. 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. Int. Rev. Appl. Econ. 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. Strat. Manag. J. 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 Trans. Pattern Anal. Mach. Intell. 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. J. Oper. Res. Soc. 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. Appl. Econ. 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. J. Bank. 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. J. Behav. Exp. Finan. 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. Bus. Soc. 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). Strat. Manag. J. 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. J. Bus. 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. J. Bus. 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. J. Bus. 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. Bus. Strat. Environ. 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. J. Environ. Manag. 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. J. Manag. 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. J. Bus. 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. Strat. Manag. J. 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. Ecol. Econ. 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. J. Financ. Econ. 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. Bus. Polit. 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. Corp. Soc. Responsib. Environ. Manag. 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. Bus. Strat. Environ. 23 (6), 361–374. https://doi.org/10.1002/bse.1786.
Garriga, E., Melé, D., 2004. Corporate social responsibility theories: mapping the territory. J. Bus. 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. Bus. Ethics Eur. Rev. 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. J. Bank. 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. J. Bus. 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, 6. National Bureau of Economic Research Working Paper Series. https://doi.org/10.3386/w18476.
Hopwood, B., Mellor, M., O'Brien, G., 2005. Sustainable development: mapping different approaches. Sustain. Dev. 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. Bus. Strat. Environ. 24 (8), 765–779. https://doi.org/10.1002/bse.1844.
Jain, A.K., 2010. Data clustering: 50 years beyond k-means. Pattern Recogn. Lett. 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 Min. Knowl. Discov. 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. J. Bank. 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. J. Bus. 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. Bus. Ethics Eur. Rev. 25 (1), 1–24. https://doi.org/10.1111/beer.12079.
Krüger, P., 2015. Corporate goodness and shareholder wealth. J. Financ. Econ. 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. Res. Int. Bus. 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. J. 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? Strat. Manag. J. 21 (5), 603–609. https://doi.org/10.1002/(SICI)1097-0266(200005)21:5<603::AID-SMJ101>3.0.CO;2-3.
Milligan, G.W., Cooper, M.C., 1988. A study of standardization of variables in cluster analysis. J. Classif. 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. Bus. Strat. Environ. 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. Bus. Strat. Environ. 24 (8), 819–835. https://doi.org/10.1002/bse.1848.
Montiel, I., 2008. Corporate social responsibility and corporate sustainability: separate pasts, common futures. Organ. Environ. 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. Corp. Soc. Responsib. Environ. Manag. 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. J. Environ. Econ. Manag. 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? J. Bus. 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. J. Clean. Prod. 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. Int. J. Approx. Reason. 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. Rev. Int. Polit. Econ. (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. Bus. Strat. Environ. 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. J. Bank. 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. Sustain. Account. Manag. Pol. J. 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. J. Bus. Finance Account. 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. Int. J. Prod. Econ. 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 Q. Rev. 113–131. March. 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. J. Bus. 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., Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. 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. Strat. Manag. J. 18 (4), 303–319. https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.0.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. Corp. Soc. Responsib. Environ. Manag. 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. Strat. Manag. J. 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. Environ. Resour. Econ. 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. Ecol. Econ. 69 (4), 848–856. https://doi.org/10.1016/j.ecolecon.2009.10.009.