2018/34

DP

Bertrand Candelon, Jean-Baptiste Hasse and Quentin Lajaunie

SRI: Truths and lies

CORE
CENTER FOR OPERATIONS RESEARCH AND ECONOMETRICS

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: immaq-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/discussion-papers.html

SRI: Truths and Lies ☆

Bertrand Candelon[a], Jean-Baptiste Hasse[b], Quentin Lajaunie[c]

[a] *Université Catholique de Louvain, CORE Louvain Finance, Belgium*
[b] *Aix-Marseille University (Aix-Marseille School of Economics), CNRS and EHESS, France*
[c] *Paris Dauphine University, PSL Research University, LEDa-CGEMP, France*

Abstract

This paper proposes a skeptical look at the socially responsible investing (SRI) industry. Building upon a new database for European and American domestic equity mutual funds, it investigates whether there is a discrepancy between what is said (e.g., names or labels) and what is done (investments of mutual funds holdings) about SRI. It turns out that the correspondence between *de jure* and *de facto* SRI is weak. Additionally, using a novel non-linear factor-augmented panel model, it is found that the *de facto* ethical positioning only matters for the funds' financial performance. Both results shed new light on the SRI industry and pave the way for a new regulation framework.

Keywords: Socially Responsible Investing (SRI); Environmental, Social and Governance (ESG) Criteria ; Ethical Mutual Funds; Performance Measurement

JEL Classification: G11, G14, G23

☆The authors thank Franz Fuerst, Yannick Le Pen, Arno Riedl and Paul Smeets for helpful comments as well as attendance at the seminars of the Université Catholique de Louvain and the Institut Louis Bachelier. This research was conducted as part of the research program entitled "Risk Management, Investment Strategies and Financial Stability" under the aegis of the Europlace Institute of Finance, a joint initiative with Insti7. The usual disclaimer applies.

Email addresses: candelonb@gmail.com (Bertrand Candelon), jb-hasse@hotmail.fr (Jean-Baptiste Hasse), quentin_lajaunie@hotmail.fr (Quentin Lajaunie)

1. Introduction

Socially responsible investing (SRI) currently constitutes a key issue for both investors and the mutual fund industry. According to the Social Investment Forum's (SIF) report (2016), the SRI market has obtained more than 20% of its total assets under professional management in the United States. In Europe, the same phenomenon is observed, as the number of socially responsible mutual funds grew by 12% between 2014 and 2016.[1] This significant growth in the SRI market has been driven by strong ethical concerns from investors regarding the environmental, social and governance (ESG) impacts of their investments. Broadly, socially responsible mutual funds invest in firms that encourage good corporate practices that promote environmental stewardship, consumer protection, human rights, and diversity. Although everybody understands the purpose of SRI, a precise definition of such mutual funds remains relatively subjective. In 2015, the French financial market regulator (Autorité des Marchés Financiers - AMF) defined SRI as *"a polymorphous and evolving concept that is sometimes difficult to understand"*.[2] This lack of transparency has led to the development of an industry aimed at providing SRI labels for mutual funds. In France, for example, the Novethic label was established in 2009 for European countries, and the FNG label was adopted in 2001 in German-speaking countries. Similarly, nonprofit organizations, such as the SIF, publish their own socially responsible asset managers lists, which are built from memberships in several SRI communities or adherence to ethical principles of investing (e.g., Principles for Responsible Investment - PRI) and accreditation to labels. In light of a growing supply and subjective definitions of SRI, mutual funds now also aim at obtaining

---

[1]The authors' calculations based on several sources: the Global Sustainable Investment review (2016), available at `www.gsi-alliance.org`; the European SRI Study (2016), available at `http://www.eurosif.org/`; and the US SRI trend report (2016), available at `https://www.ussif.org/`.

[2]Quote extracted from "Rapport de l'AMF sur l'investissement responsable dans la gestion collective", November 26, 2015.

these labels in addition to an ethically styled name as a communication strategy to attract investors. Indeed, they are presenting their membership in different SRI labels and ethical objectives as proof of their ethical investment policy, thus putting less emphasis on the real financial and extra-financial performance of the fund.

This paper proposes to shed new light on socially responsible mutual funds by going beyond these appearances driven by asset managers' marketing strategies and specialized audit agencies' labels. It investigates whether an SRI announcement (via self-presentation or label accreditation), broadly speaking, indeed signals that a mutual fund invests according to ethical concerns or whether it constitutes only a purely marketing position. To this aim, a 2-dimensional measure for the SRI characteristics of mutual funds is built. The first dimension concerns the announcements of asset managers or the accreditation (or not) by a specialized audit agency of the social responsibility of the mutual funds. This information provides us with the willingness of the asset managers to identify their mutual funds as socially responsible (via self-identification or via a labeling partner). The second dimension covers the realizations of these commitments. This aspect is measured via the newly available Morningstar Sustainability Rating database, which presents many advantages. In particular, it is continuous (scores are defined between 0 and 100), free of industry bias, normalized, and homogeneous for all European and US mutual funds. An analysis of the correspondence between both dimensions tells us whether SRI announcements provide sufficient information about the SRI identification of mutual funds or not. It also highlights the dimension that matters the most for an investor who has a genuine ethical objective. This new approach also extends the empirical literature on the performance of socially responsible mutual funds. Since the emergence of SRI, empirical studies have tried to estimate if there was a cost to being ethical and, if so, to evaluate it. The results turn out to be mixed, and the literature has hardly converged toward a consensus of assessing that SRI cannot be achieved without a cost for financial performance. As this debate finds its roots in the categorization and the various methodological implementations, this paper proposes a new framework to investigate

the *de jure* and *de facto* identity of mutual funds. Is there a difference between what is said and what is done? If that is the case, it would partially explain the divergence about the cost of ethics in the literature. From this new identification process, we empirically check the impact of *de jure* and *de facto* SRI on financial performance. To do so, we build a new database, which is exploited using a novel econometric methodology to avoid (i) a matching procedure between differently categorized funds and (ii) a two-step approach and its potential statistical biases.

Specifically, whereas earlier studies propose simple comparisons of the risk-adjusted measures, more recent studies consider factor-augmented models à la Fama and French (1993) or Carhart (1997). This paper implements both approaches, free from any assumption about categorization or list. First, it offers a comparison of the returns among different categories of socially responsible funds without any matching procedure between differently categorized funds. It also follows the second route, considering the most recent papers (Bauer et al. (2005); Renneboog et al., 2008; El Ghoul and Karoui, 2017) and the 4-factor model à la Carhart (1997). Nonetheless, to avoid a two-step approach and its potential statistical biases, the 4-factor model is estimated via a non-linear panel data model to allow the inclusion of the two dimensions of SRI.

The empirical part covers both US and European domestic equity mutual funds. To this aim, we build a new database of mutual fund returns that is robust to survival bias due to disappearing funds and the to look-ahead bias generated by the newly created funds. Hence, we end up with a novel database including approximately 600 European and 900 US domestic equity funds over the period $2013 - 2018$. The purpose of such a new database is to be able to exploit scores provided by Morningstar Sustainability Rating for every mutual fund. In addition to this *de facto* SRI measure, we also introduce a dummy variable as an indicator describing the identity of every mutual fund from the database. This *de jure* SRI measure is built from asset managers' announcements (self-presentation and label accreditation).

Based on our results, we observe that both in Europe and in the US, SRI features cannot

be exclusively summarized by asset managers' announcements or labels. Indeed, significant numbers of mutual funds exist that have low SRI scores, even though they have committed to be managed according to ethical concerns. Such a result clearly questions the pertinence of the SRI labels processed in the industry and thus calls for a harmonization of SRI scores such that this objective becomes directly applicable for investors. To go further, a performance analysis using a panel version of Carhart's 4-factor model indicates that announcements about ethics do not truly matter for the funds' financial performance. The exclusive ethical driver is constituted by the SRI score, i.e., funds holdings or *de facto* SRI. Such a result has several implications. For the mutual fund industry, it means that low performance cannot be justified solely by the announcement of ethical objectives solely via ethical positioning or adhesion to particular labels. It can only be an excuse if this ethical commitment is respected. This finding might also explain the different results reported in the literature, as some studies consider measures of *de jure* SRI to be *de facto* SRI scores, whereas others integrate only realized SRI scores.

The paper is organized as follows. Section 2 describes the database built from the Morningstar Sustainability Rating and its features. Section 3 evaluates the return performance of mutual funds according to corresponding SRI objectives. It also presents many robustness checks. Section 4 concludes.

2. A new SRI Database

To conduct a proper empirical analysis of SRI, the choice of database is crucial. As ethical standards differ among investors, asset managers and labeling organizations, the categorization of SRI and conventional funds is highly debatable. In a recent paper, Statman and Glushkov (2017) highlight such a difficulty in describing the differences between the databases used in the literature (e.g., Lipper's list, the Social Investment Forum (SIF) list and the Standard & Poor's ($S\&P$) list) and their consequences for empirical studies. In this paper, we choose to use a newly available database from Morningstar that scores mutual

funds with respect to the ethical quality of their investment holdings. This new database[3] presents three main advantages. First, more than 90% of existing mutual funds are rated, whether they are identified as socially responsible or not. Second, the SRI score of each mutual fund is a composite built from the aggregation of firm-level ratings and normalized. It thus provides us with scores that are free of industry bias (which is not the case for the firm-level data of MSCI KLD ESG; the data recently became available and were used by El Ghoul and Karoui, 2017). In addition, as the SRI score covers, on average, more than 90% of mutual funds exposure, it is possible to decompose it into sub-indices. In particular, the decomposition between the ESG score (computed from environmental, social and governance scores) and the Controversy score is a unique added value of the database.

2.1. The Morningstar database

We restrict our initial analysis to an examination of European and US domestic equity mutual funds. Morningstar lists $1,110$ mutual funds in Europe denominated in euros and $1,963$ mutual funds in the United States area denominated in dollars (large-cap funds with the oldest shares for both geographical areas). It covers the period $2013-2018$ at a monthly frequency. In addition to listing net returns, the database lists each fund's benchmark and geographic area. Several steps are applied to the original dataset to form a balanced and consistent database. First, 275 funds in Europe and 713 funds in the US are excluded, as their SRI scores are not available. Bloomberg is consulted to confirm the quality of information in the Morningstar database. It includes information on 786 mutual funds in Europe and $1,151$ mutual funds in the US. If a mutual fund is not available in the two databases, we exclude it. Then, when divergence between Morningstar and Bloomberg data is found, because no

[3]The use of the Morningstar Sustainability Rating database is a novelty, as to the best of our knowledge, only Hartzmark and Sussman (2018) are using it. However, they are working with pre-categorization (called "globes"), whereas we consider continuous rating. Our choice is motivated by avoiding potential non-linear effects in the model.

confirmation can be obtained, the fund is removed. Thereby, of the European mutual funds, 180 funds are excluded, as their benchmarks and/or geographic investment areas do not cover all of Europe but only some European countries. Similarly, of the US mutual funds, 264 funds are excluded. Our final database includes 606 funds in Europe and 887 in the United States. Each fund has a monthly SRI score that aggregates the ESG (environmental, social, and governance) score and the Controversy score.[4] This database is thus balanced and homogeneous and lists all European and US mutual funds with an SRI rating provided by Morningstar.

It is well known that such a database might be subject to potential survivor bias. This bias can be due to mutual funds disappearing from the market or merging with other funds. Neglecting such a bias would probably lead to an overestimation of the funds' performance. To the best of our knowledge, the only survivorship-corrected database that concerns US mutual funds and has been initiated by Carhart (1997) and updated is the CRSP database. As the objective of the paper consists of studying and comparing both the American and European markets, it requires a comparable database, particularly with respect to the treatment of survivorship.[5] First, we track every fund existing during our sample period, as in Brown and Goetzmann (1994), Carhart (1994; 1997) and Malkiel (1994). Then, following Elton et al. (1996) we use the risk-adjusted returns and perform a 4-factor CAPM (Carhart, 1997) using a single-index model. However, our approach differs from previous studies, as we complete missing returns not only at the end of the sample period but also for the missing returns of mutual funds at the beginning of the sample. Indeed, we take into account newly born funds as soon as they exist for at least two years. This decision is motivated by the fact that we will consider a balanced panel framework, and thus, we cannot afford to have missing

[4]Appendix 1 offers a description of the rating method developed by Morningstar.

[5]See Hanke et al. (2018) for the causes of the non comparability of different databases such as CRSP and Morningstar.

returns at the end or at the beginning of the sample. In addition, completely excluding these "new-born" funds would have reinforced the issue of selection bias.

2.2. Beyond the SRI classification

The Morningstar Sustainability Rating database provides us with a normalized and continuous SRI score. In addition, this rating is attributed ex-post, i.e., once the investments are realized. However, it is well known that some funds explicitly make commitments in favor of SRI investments or even obtain SRI label accreditation. This is in line with investors' behavior: As shown by Bauer and Smeets (2015) or Riedl and Smeets (2017), investing according to ethical concerns allow investors to be identified as socially responsible investors. This can take the form of marketing action and is a clear decision of the fund to refuse investing in sectors that are related to unethical industrial sectors. Thus, another classification can be made between funds that have made such commitments, which are labelled *de jure* ethical funds, while the others are considered conventional funds. To discriminate between *de jure* SRI and conventional mutual funds, we build a dummy variable to determine if asset managers claim that their portfolio follows ethical concerns or not. To this aim, we focus on mutual funds' names, looking for announced extra-financial objectives. Mutual funds' names are an excellent proxy of announcements about marketing strategy. Cooper et al. (2005) investigate the impact of changes in mutual funds' names driven by hot investment styles. They empirically show that investors are irrationally influenced by cosmetic effects. Similarly, Espenlaub et al. (2017) conduct a natural experiment and empirically show that mutual funds' names have a significant impact on investors. Using a regulation change (SEC Rule 35d-1) in July 2001, the authors highlight that superficial changes in mutual funds' names led to a significant increase in capital flows.

We follow Nofsinger and Varma (2014)'s methodology to classify conventional and socially responsible mutual funds. From a discrete selection process, they build a list of words related to SRI terminology: "social", "socially", "environment", "green", "sustainability", "sustainable", "ethics", "ethical", "faith", "religion", "Christian", "Islam", "Baptist" and

"Lutheran". Then, using the dictionary defined above, the authors keyword-search mutual funds' names to identify socially responsible mutual funds. We go further than Nofsinger and Varma (2014), using several lexical databases to broaden the SRI terminology. Our purpose is to build a more complete dictionary that enables us to look for words (nouns, adjectives or verbs) associated with SRI. First, we store a preliminary list of words from the terminology identified by the USSIF (2018) [6]: "community", "ethical", "green", "impact", "mission", "responsible", "socially", "sustainable", and "values". Then, we extend this initial list using the lexical database developed by Miller (1995) and Fellbaum (1998) and hosted/updated by Princeton University.[7] To the 9 initial words from USSIF (2018)'s terminology, we add 18 other words, which we present in Table 1. Our extended dictionary is then used to classify mutual funds via pattern search on mutual fund names. The rationale for using pattern search instead of keyword search like Nofsinger and Varma (2014) is to track words (nouns, adjectives and verbs) based on the same stem as keywords from our dictionary. The indicator variable is then built from a matching procedure, and we check the results using Bloomberg's description of mutual funds.

Table 2 reports the number of conventional and ethical funds as well as the descriptive statistics (their *de facto* SRI score mean and their standard deviation).

From the 606 European and 862 US funds, we detect 52/25 funds presenting themselves as ethical funds, resulting in 554/862 conventional funds for Europe and the US, respectively. We also investigate when this classification is robust to SRI labels. In comparing our classification for European funds with that proposed by Novethic[8], the largest European ethical label provider, we find that all 19 mutual funds possessing a label are classified as ethical,

[6]Terminology available at `https://www.ussif.org/sribasics`

[7]See Princeton University "About WordNet." WordNet. Princeton University. 2010.

[8]http://www.novethic.fr/labellisation-de-linvestissement-responsable.html

Table 1: SRI Terminology

Nofsinger and Varma (2014)	Extended dictionary
Baptist	Baptist
Christian	blue
environment	carbon
ethical	Catholic
ethics	Christian
faith	climate
green	community
Islam	durable
Lutheran	environment
religion	ESG
Social	ethical
socially	faith
sustainable	governance
sustainability	green
	human rights
	impact
	Islam
	Lutheran
	mission
	moral
	peace
	philosophy
	religion
	responsible
	social
	solidarity
	subsidiarity
	sustainable
	sustainability
	values

Table 2: Europe/US - Descriptive Statistics - Conventional vs *de jure* SRI mutual dunds

Europe

	Conventional funds	*de jure* SRI funds	Total
Number	554	52	606
Mean_{SRI_score}	55.18	57.61	55.39
σ_{SRI_score}	1.996	1.953	2.103

The United States

	Conventional funds	*de jure* SRI funds	Total
Number	862	25	887
Mean_{SRI_score}	45.85	48.19	45.92
σ_{SRI_score}	1.725	1.500	1.761

Notes: This table reports the number of funds included in our database and corresponding SRI rating means and standard errors.

whereas none of the conventional funds present such a feature. This classification can thus be interpreted as either an ethical objective or an ethical label. In the case of the US, such a comparison is not possible, as, to the best of our knowledge, it does not have ethical labels. It also appears that ethical funds represent a minority (approximately 9% in Europe and 3% in the United States), whereas conventional funds are numerous. However, it is striking to observe that on average, the corresponding SRI rates are almost similar (and not significantly different) across ethical and conventional funds. This suggests that some ethical funds have lower SRI scores than some conventional funds, whereas some conventional funds outperform ethical ones. This reveals a difference in ethical investments between what is announced and what is realized by such funds.

Figure 1 plots the distribution of SRI ratings for these two groups of funds in Europe and in the US. It is obvious that the peak of the distribution for socially responsible mutual funds is higher than that calculated for conventional ones. However, it also shows the presence of a very large overlap between the distributions, thus confirming that some ethical funds have a poor SRI score, whereas conventional funds have a relatively good SRI score. Such a stylized fact paves the way for a second dimension of ethics that addresses the realization

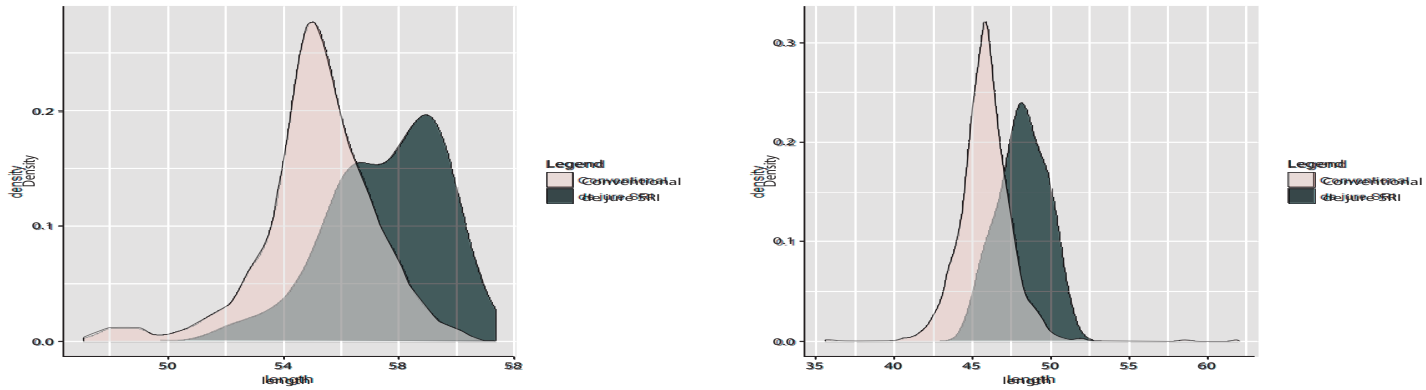

Figure 1: Density function of the Morningstar Sustainability Rating: Europe and the United States

of SRI investments that might differ from the announcements of the funds. To implement such a distinction, it is necessary to find a threshold (CT) above which a conventional fund is managed ethically as well as one (ET) below which an SRI mutual fund invests conventionally. To this aim, we consider a simple rule that is similar to the conditional value-at-risk (CoVaR) measure. The threshold CT is defined as the SRI score given to the lowest 10% of ethical funds, and ET is defined as the SRI score given to the highest 10% of conventional funds. Figure (2) illustrates this definition.

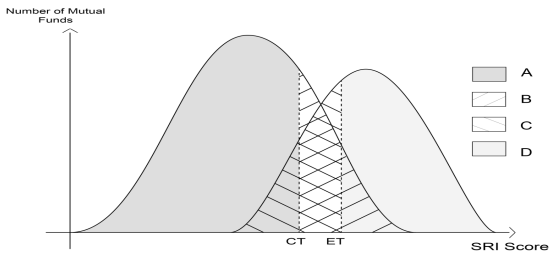

Figure 2: Scheme of the SRI classification from the density function of the Morningstar Sustainability Rating

2.3. Preliminary Analysis

For Europe, we find that $CT = 55.204$ and $ET = 57.598$, and for the US, we find that $CT = 46.056$ and $ET = 47.720$. Conditional on these threshold values, Tables 3 and 4 summarize our categorization of the funds according to their SRI objectives and realizations. For simplicity, we label them from A to D, as reported in Table 2.

Table 3: Europe/US - Categorization of funds according to SRI

Europe

	Conventional Real.	Ethical Real.
Conventional Obj.	A n=276 (49.82%)	B n=278 (50.18%)
Ethical Obj.	C n=23 (44.23%)	D n=29 (55.77%)

The United States

	Conventional Real.	Ethical Real.
Conventional Obj.	A n=510 (59.16%)	B n=352 (40.84%)
Ethical Obj.	C n=9 (36.00%)	D n=16 (64.00%)

We find that 50.18% (40.84%) of the European (US) conventional funds still present very high SRI scores, indicating that they respect ethical principles. On the contrary, and perhaps more interestingly, 44.23% (36.00%) of the European (US) ethical funds have a low ethical grade. This result indicates that 23 (9) ethical funds do not respect their commitments in terms of SRI investments. What is said does not seem to match what is done in terms of SRI. When looking specifically at the European mutual funds having a label, they are all classified in the C and D categories. Labels are coherent with announcements. Still, 4 of them belong to group C, highlighting that labels are only weak leading indicators of the effective respect given to ethical objectives. This industry should thus depart from considering ethical as a marketing positioning and instead integrate the effective realization of socially responsible investments. In addition, when considering the management fees of the mutual funds, we do not observe significant differences across categories. Instead, mutual funds with a label present significantly higher management fees, suggesting that they charge investors for the ethical label.

3. Evaluating the performance of socially responsible mutual funds

3.1. Literature review of socially responsible fund performance

The literature on the financial impacts of SRI focuses on the dichotomy between conventional and SRI-labelled mutual funds. The early studies of Hamilton et al. (1993), Goldreyer and Diltz (1999) and Statman (2000) compare risk-adjusted returns of SRI indices and mutual funds in the US in the 1990s and show that the impacts of ethical investment are not significantly different from zero. Luther and Matatko (1992; 1994) and Mallin et al. (1995) run the same experiments for the UK trusts for $1984 - 1990$ and $1986 - 1994$, respectively. They find similar results supporting the idea that ethical screening could introduce a bias about both the size of firms and the reference benchmark choice. From this conclusion, Goldreyer et al. (1999) move from the mean-variance framework to the CAPM model of Sharpe (1964) and Lintner (1965), as Kreander et al. (2005) and Renneboog et al. (2008a; 2008b) use the three-factor model of Fama and French (1992; 1993) to evaluate the performance of European and US mutual funds. Their results reach the same conclusions, namely, financial performance tends to be lower when investment is ethical, but their results are hardly significant. The introduction of the four-factor model of Carhart (1997) by Bauer et al. (2005), who study the financial performance of US, UK and German socially mutual funds, does not contrast the significance of the performance differences observed in previous studies. Statman and Glushkov (2009) and Hong and Kacperczyk (2009) choose to consider "sin stocks" in investment portfolios. A positive and significant excess performance of such funds would support the idea that ethics has a price. These seemingly different results obtained in the empirical literature highlight a methodological issue about the dual approach of conventional versus ethical investment. Consequently, the most recent empirical literature focuses on the lack of strict methodology for comparing performance. Indeed, early studies on the impact of ethics (Hamilton et al., 1993; Bauer et al., 2005) are based on a direct comparison between conventional and socially responsible mutual funds but do not assess the relevance of a comparison between such mutual funds. Since then, this methodology has been enriched fol-

lowing two major features. On the one hand, Statman (2000, 2006) and Schroder (2007) use financial indices rather than mutual funds to compare the financial performances of ethical and conventional benchmarks. On the other hand, a preliminary matching procedure among conventional, ethical (Mallin et al., 1995; Gregory et al., 1997; Kreander et al., 2005) and even "sin" (Hong and Kacperczyk, 2009; Humphrey and Tan, 2013; Borgers et al., 2015) mutual funds is introduced to improve the relevance of quantitative performance comparisons. In the most recent papers, the use of cross scores based on ESG criteria among data covering both conventional and ethical assets allows researchers to avoid statistical biases inherent to the duality of the mutual funds. El Ghoul and Karoui (2017) build an asset-weighted CSR score for mutual funds from their exposures by using firm-level data from MSCI ESG KLD STATS. This preliminary data treatment allows them to compare US domestic mutual funds performance on the same basis over the $2003 - 2011$ period. Borgers et al. (2015) use the same approach with US mutual fund holdings from 2004 to 2012 to measure the impact of social factors on financial performance.

3.2. A preliminary analysis of mutual fund performance

Given the funds' classification along the two dimensions of ethics, it is possible to run a first analysis of the performance, as in Hamilton et al. (1993), Goldreyer and Diltz (1999) and Statman (2000), to compare the characteristics of each category (A, B, C and D) of funds. Table 3 reports the descriptive statistics (mean return, median, standard error, Sharpe ratio, skewness and kurtosis) for each of these categories of funds. It appears that,in line with the existing studies in the literature, conventional funds (AB) outperformed SRI funds in terms of mean return for both the US and Europe. Such an outcome is due to the higher risks taken by these funds, as measured by the volatility or the Sharpe ratio. SRI funds present a negative skewness, which indicates a prominent left tail and explains the difference between the mean and the median. On the contrary, conventional funds present a positive skewness and, thus, a right tail, which indicates that some funds are over-performing. When we analyse each category of mutual funds (A, B, C or D) more deeply, we can observe

Table 4: Europe/US - Descriptive statistics for fund performance - 2013-2018

Europe

	AB	CD	A	B	C	D
Fund Return (Mean=μ)	7.21%	7.92%	7.44%	6.99%	8.40%	7.54%
Fund Risk (σ)	11.72%	11.10%	11.77%	11.67%	11.07%	11.12%
Fund Return (Median)	7.58%	7.91%	7.49%	7.78%	7.99%	7.62%
Sharpe ($\frac{\mu}{\sigma}$)	0.615	0.713	0.632	0.599	0.759	0.678
Skewness	-0.329	-0.284	-0.336	-0.322	-0.278	-0.288
Kurtosis	3.715	3.432	3.604	3.826	3.289	3.546

The United States

	AB	CD	A	B	C	D
Fund Return (Mean=μ)	13.50%	13.59%	13.98%	12.81%	13.56%	13.61%
Fund Risk (σ)	10.51%	10.37%	10.67%	10.26%	10.34%	10.39%
Fund Return (Median)	13.51%	13.82%	14.16%	12.88%	13.82%	13.51%
Sharpe ($\frac{\mu}{\sigma}$)	1.285	1.310	1.309	1.249	1.311	1.310
Skewness	-0.153	-0.172	-0.157	-0.147	-0.161	-0.178
Kurtosis	3.161	3.172	3.211	3.087	3.099	3.213

Notes: This table reports the annualized average/median returns of different types of funds. It also reports corresponding standard deviations, Sharpe ratios, and skewness and kurtosis scores.

important differences that justify this classification choice. For extreme categories (A and D), the previous argumentation holds; i.e., conventional funds that are conventional (A) have a higher return than do ethical funds that are *de facto* ethical (D). The difference between categories B (conventional funds that behave ethically) and C (ethical funds that behave conventionally) is in the opposite direction: the first category underperforms in terms of returns compared with the latter. It thus seems that the SRI score is a much more important factor of performance than the stated commitment. In other words, the *de facto* dimension dominates the *de jure* one.

3.3. Methodology

To offer a more extensive comparison of socially responsible mutual funds, we elaborate on the CAPM approach. Recent papers have largely estimated factor models, such as the traditional Fama-French (1993) model, which integrates 3 or more factors. A recently devel-

oped 4-factor risk-adjusted performance model has also been proposed by Carhart (1997). This model considers a market return index (r^m), the monthly premium of the book-to-market factor (r^{HML}), the monthly premium of the size factor (r^{SMB}) and momentum in stock markets (r^{MOM}). The model can thus be rewritten as

$$r_{i,t} - r_t^f = \alpha + \beta_{r^m}.(r^m - r_t^f) + \beta_{SMB}.r_t^{SMB} + \beta_{HML}.r_t^{HML} + \beta_{MOM}.r_t^{MOM} + \epsilon_{i,t}, \quad (1)$$

where r_i is the fund's i return, r^f is the monthly T-bill rate, and α is the net-of-fees annual risk-adjusted performance of fund i. The model is estimated independently for each fund i such that $\epsilon_{i,t}$ has i.i.d. white noise. The model can be estimated for a period of time and for a set of funds i (cross-sectional dimension) or for a particular fund i for a period of time $t = 1...T$ (time series dimension). El Ghoul and Karoui (2017) apply the latter strategy. In the first step, they estimate for each fund individually (1) to obtain an individual estimate of α the conditional return of the funds. The βs that represent the sensitivity to market factors remain a common factor for all funds. They introduce the SRI characteristics of the fund in the second step, in which they regress the estimated conditional return on the particular features of the fund.

To avoid a two-step approach, which can introduce statistical bias, we rely on recent studies (Petersen, 2009, Ando and Bai, 2014, to name but a few) that propose evaluating the performance of funds using a large-dimensional panel, i.e., considering in a single step both the time and the cross-sectional dimension. To this aim, a fixed effect factor η_i is added in order to take into account the potential unobserved heterogeneity. In addition, to distinguish between conventional and ethical funds, a non-linear panel is considered. More precisely, we

split the panel for these two types of funds and obtain the following model:

$$r_{i,t} - r_t^f = \mathbb{1}_{CONV}(\alpha_c + \beta_{c,r^m}.(r^m - r_t^f) + \beta_{c,SMB}.r_t^{SMB} + \beta_{c,HML}.r_t^{HML} + \beta_{c,MOM}.r_t^{MOM})$$
$$+ \mathbb{1}_{SRI}(\alpha_s + \beta_{s,r^m}.(r^m - r_t^f) + \beta_{s,SMB}.r_t^{SMB} + \beta_{s,HML}.r_t^{HML} + \beta_{s,MOM}.r_t^{MOM}) + \eta_i + \epsilon_{i,t}, \tag{2}$$

where $\mathbb{1}_{CONV}(.)$ is an index variable that takes a value of 1 if the fund is conventional and 0 otherwise, and $\mathbb{1}_{SRI}(.)$ is an index that takes a value of 1 if the fund is ethical and 0 otherwise. The subscript C refers to estimates associated with conventional funds, and the subscript s corresponds to ethical fund coefficients. Model (2) is estimated via GLM, and a Driscoll-Kraay (1998) correction is implemented to avoid bias due to cross-sectional dependence. It is thus possible to test whether a category of funds offers extra returns and whether it is more sensitive to a peculiar factor. In a sense, this approach is in line with papers that test for homogeneous breaks in slopes, such as those by Pesaran and Yamagata (2008) or Blomquist and Westerlund (2013). Here, the breaks are exogenous and are driven by economic motivations: ethical and conventional funds.

Furthermore, we follow Hansen (2006), and instead of considering model (2), we estimate a model integrating both the whole sample and the ethical funds subsample. The non-linear panel Carhart model takes the following form:

$$r_{i,t} - r_t^f = \alpha + \beta_{r^m}.(r^m - r_t^f) + \beta_{SMB}.r_t^{SMB} + \beta_{HML}.r_t^{HML} + \beta_{MOM}.r_t^{MOM}$$
$$+ \mathbb{1}_{SRI}(\alpha_s + \beta_{s,r^m}.(r^m - r_t^f) + \beta_{s,SMB}.r_t^{SMB} + \beta_{s,HML}.r_t^{HML} + \beta_{s,MOM}.r_t^{MOM}) + \eta_i + \epsilon_{i,t}. \tag{3}$$

This representation offers more precise estimates and straightforward interpretations. If a coefficient associated with a socially responsible mutual fund (denoted with an underscore s) is significant, then it would indicate a particular behavior of ethical funds. In the opposite case, it would suggest that they behave similarly to conventional funds. The model thus

separates ethical from conventional funds. However, as we stress in the previous section, some funds do not respect their commitments. Further, some ethical funds present low SRI scores, and some conventional funds present high SRI scores. We thus consider in model (3) the SRI score obtained from the final non-linear panel-augmented Carhart model:

$$r_{i,t} - r_t^f = \alpha + \beta_{r^m}.(r^m - r_t^f) + \beta_{SMB}.r_t^{SMB} + \beta_{HML}.r_t^{HML} + \beta_{MOM}.r_t^{MOM}$$
$$+\mathbb{1}_{SRI}(\alpha_s + \beta_{s,r^m}.(r^m - r_t^f) + \beta_{s,SMB}.r_t^{SMB} + \beta_{s,HML}.r_t^{HML} + \beta_{s,MOM}.r_t^{MOM}) + \alpha_{SRI}.SRI_i + \epsilon_{i,t}. \tag{4}$$

Let us note that in such a specification, the fixed term effect is omitted, as it would be highly correlated with the SRI rating if it is fixed over the period or presents a low variability. In this section, SRI is fixed over the given time period 2018. In the robustness sub-section 3.5, time-varying SRI notes are considered. In model (4) the estimated return of the mutual fund $\hat{\alpha}^*$ is now calculated as $\hat{\alpha} - \hat{\alpha}_{SRI}.\bar{SRI}$, where $\bar{SRI}$ is the average SRI note.

3.4. Empirical application

Through a preliminary analysis and to identify a benchmark, we estimate the basic linear Carhart model (1) for $2013 - 2018$ (i.e., $T = 72$) without considering any SRI dimensions. The market benchmark (r^m) is the MSCI Europe Index, and returns (r) are net fees. Table 5 reports the outcomes of the model (1) estimates.

It appears that all explanatory variables except momentum have a significant effect on the risk-adjusted performance of the funds. The $Adjusted - R^2$ value is also quite high (0.76 for Europe and 0.84 for the US). Such a result clearly rules out the efficient market hypothesis. Specifically, the results show that the market factor has almost a proportional impact on the funds' returns. SMB also affects returns positively but with less elasticity. By contrast, the

Table 5: Estimation of the panel version of Carhart's model (2013-2018)

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
$\hat{\alpha}$	-0.0018	0.0135	-0.0017	0.0011
$\hat{\beta}_{r^m}$	0.9803***	0.0185	1.0068***	0.0084
$\hat{\beta}_{SMB}$	0.1996***	0.0465	0.0711***	0.0158
$\hat{\beta}_{HML}$	-0.0268	0.0215	-0.0470**	0.0210
$\hat{\beta}_{MOM}$	0.0031	0.0239	0.0207	0.0217
$Adj.R^2$	0.7599		0.8466	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1) based on the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

value premium is negative but with a relatively small coefficient.[9]

The significance of annual risk-adjusted performance ($\hat{\alpha}$) is not significantly different from zero, corroborating the results of Fama and French (1993) or Carhart (1997), indicating that on average, mutual funds do not exhibit extra returns.

Table 6 gathers the results of the estimation of the 4-factor model augmented by a *de jure* dummy indicating whether or not the fund has committed to invest ethically. It appears that this dummy is never significantly different from 0, supporting the literature's findings (Bauer et al., 2005; Renneboog et al., 2008b). Thus, SRI announcements do not affect the performance of the fund.

In the next step, the non-linear augmented Carhart model represented by equation (4) can now be estimated. Remember that $\mathbb{1}_{SRI}$ is a dummy variable that takes a value of 1 (resp. 0) if the fund is ethical (resp. conventional); i.e., there are clear announcements concerning ethics (resp. there is no a clear objective in favor of SRI investments). This corresponds to the *de jure* dummy variable. The second variable, *SRI Score*, corresponds to the grade given by Morningstar. Such a variable is a proxy for the *de facto* dimension of the SRI because

[9]In his famous survey, Schwert (2003) confirms this fact, concluding that small-firm anomalies have almost disappeared in the most recent period.

it is independent of any announcement. The model is estimated by the generalized linear method (GLM), and the results are reported in Table 7.

The estimates obtained in the first part of the model (i.e., for the full model) are similar in both sign and amplitude to those obtained using the previous linear model (Table 5) for Europe and the United States. The only slight difference is the increase in the value of α when $Score_{SRI}$ is introduced. Such a result can be explained by the negative value of the *de facto* SRI score, suggesting that extra-financial constraints constitute a penalty for funds' performance (Bollen, 2007). In addition, when calculating the true α[10] in model (4), it turns out to be very close to 0.

The second part of Table 7 deals with the impact of announcements on the ethical behavior of the fund (proxied here by a regime characterized by the *de jure* dummy). The results between the US and Europe highlight several common features. First, the β_{s,r^m} is not significant, meaning that announcements about SRI have no significant impact on the market exposure of these funds. Second, neither the β_{MOM} factor nor the β_{MOM} for *de jure* SRI funds is significant, indicating the absence of persistence of the funds' returns. SRI thus does not impact the persistence of funds' returns. Finally, it appears that the SRI Score has a negative and significant effect on the funds' returns, supporting the idea that *de facto* socially responsible mutual funds have a return penalty (Bollen, 2007).

Still, we observe differences in funds' behavior. For the US, none of the *de jure* factors explain the funds' returns, confirming that announcements about ethics have no impact on mutual funds' performances. On the contrary, in Europe, SMB_{dejure} and HML_{dejure} are significant and positive and negative, respectively. Thus, *de jure* SRI could have an indirect impact on financial performance if investing in small businesses or value firms is considered ethical. This is consistent with the fact that in Europe, asset managers tend to combine ethics with investing in small firms. This difference between de jure SMB and HML between

[10] $\hat{\alpha}^* = \hat{\alpha} - \hat{\alpha}_{SRI}.\bar{SRI}$.

the EU and the US comes from the fact that in the US, the names of mutual funds must reflect the "real" strategy of funds, which is not the case in Europe. Such an observation can also explain why the number of *de jure* socially responsible mutual funds is relatively low in Europe. Finally, a difference can be observed when comparing the *de jure* and *de facto* α. We observe that *de jure* socially responsible funds present significantly lower average returns in Europe, whereas their performance in the US is identical to that of conventional funds. Still, regardless of the region, it appears that the amplitude of the *de facto* SRI score is much higher than that of the *de jure* dummy variable. The *de facto* socially responsible mutual funds exhibit a return penalty in the EU and the US, as supported by theory (Bollen, 2007), and are thus much more important than the ethical announcements.

Investors who are genuinely interested in ethical investing should not base their choice on labels or announcements from the funds. They should instead focus exclusively on the "ex-post" ethical measures. Finally, substantial concerns are raised about the mutual fund labels issued by specialized audit agencies, as they do not appear to be good leading indicators of real respect for ethical rules. Such a conclusion corroborates the findings regarding labels presented in the last section.

Table 6: Estimation of the panel version of Carhart's model (2013-2018) with dummy *de jure*

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
$\hat{\alpha}$	-0.0018	0.0069	-0.0017	0.0010
$\hat{\beta}_{r^m}$	0.9803***	0.0185	1.0068***	0.0084
$\hat{\beta}_{SMB}$	0.1996***	0.0465	0.0711***	0.0158
$\hat{\beta}_{HML}$	-0.0268	0.0215	-0.0470**	0.0210
$\hat{\beta}_{MOM}$	0.0031	0.0239	0.0207	0.0217
$\hat{D}ummyDeJure$	0.0011	0.0073	0.0041	0.0020
$Adj.R^2$	0.7545		0.8464	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1) based on the GLM method with an extra *de jure* dummy. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

Table 7: Estimation of the augmented non-linear panel version of Carhart's model (2013-2018)

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
		Full Sample		
$\hat{\alpha}$	0.0088**	0.0035	0.0094***	0.0028
$\hat{\beta}_{r^m}$	0.9638***	0.0178	1.0218***	0.0141
$\hat{\beta}_{SMB}$	0.1136***	0.0312	0.0896***	0.0259
$\hat{\beta}_{HML}$	-0.0953***	0.0315	-0.0361	0.0271
$\hat{\beta}_{MOM}$	-0.0192	0.0145	-0.0006	0.0140
		de jure SRI Mutual Funds		
$\hat{\alpha}_s$	-0.0013***	0.0003	-0.0004*	0.0003
$\hat{\beta}_{s,r^m}$	0.0181	0.0175	-0.0154	0.0143
$\hat{\beta}_{s,SMB}$	0.0940***	0.0336	-0.0190	0.0261
$\hat{\beta}_{s,HML}$	0.0750**	0.0345	-0.0112	0.0283
$\hat{\beta}_{s,MOM}$	0.0244	0.0188	0.0219*	0.0116
		de facto SRI Score		
$\hat{\alpha}_{SRI}$	-0.0118**	0.0000	-0.0170***	0.0000
$Adj.R^2$	0.7528		0.8418	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 4) based on the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

3.5. Robustness Checks

The following subsection presents a series of robustness checks. We begin by considering the stability of our results (i.e., by considering a smaller sample size). In a second experiment, we consider a time-varying SRI score. Then, the SRI scores are split into ESG and Controversy grades. Finally, the robustness of our results is analyzed while restricting mutual funds to Euro area funds. Finally, we categorize the funds according to size.

In a first exercise, a smaller sample score is considered for $2017 - 2018$, i.e., with 24 observations. Given the number of explanatory variables considered and the non-linear nature of our model (including 11 explanatory variables), this is the smallest sample to be considered without being subjected to severe finite-sample bias. Table 8 reports the estimation results.

These results are similar to those obtained for the whole sample, $2013 - 2018$, and the main previous findings still hold. First, socially responsible mutual funds do not perform significantly worse than conventional funds do in Europe and the United States. Second, SRI scores negatively affect the adjusted-risk returns of all mutual funds. Nevertheless, some differences can be observed in the European case. First, the HML factor is now not significantly different from zero. Similarly, the momentum factor is only significant at 90%, whereas the SMB factor has a stronger impact. This result signifies that the risk-adjusted returns in the recent period have been affected by the factors and might also mean that these factors are not stable over time, as recently proved by Pouillot (2016). It also supports the idea that the recent high performance of socially responsible funds in Europe is driven by their investment in new firms, which outperform others during their first years of existence.

In a second experiment, we consider time-varying SRI ratings. Even though a quick analysis of the Morningstar database would reveal that the the SRI ratings do not vary much over time, we estimate the non-linear panel-augmented Carhart model with a time-varying SRI score, which thus can be expressed as:

Table 8: Static (constant SRI ratings) estimation of the augmented non-linear panel version of Carhart's model (2017-2018)

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
		Full Sample		
$\hat{\alpha}$	0.0129	0.0127	0.0266	0.0275
$\hat{\beta}_{r^m}$	0.8951***	0.0316	0.9625***	0.0121
$\hat{\beta}_{SMB}$	-0.0549	0.0607	-0.0243	0.0244
$\hat{\beta}_{HML}$	-0.1095**	0.0165	-0.0696	0.0759
$\hat{\beta}_{MOM}$	0.1086***	0.0368	-0.0152	0.0608
		de jure SRI Mutual Funds		
$\hat{\alpha}_s$	-0.0007	0.0007	-0.0014	0.0011
$\hat{\beta}_{s,r^m}$	0.0367	0.0240	0.0066	0.0088
$\hat{\beta}_{s,SMB}$	0.1379***	0.0277	-0.0123	0.0119
$\hat{\beta}_{s,HML}$	0.0727**	0.0287	0.0943	0.0720
$\hat{\beta}_{s,MOM}$	0.0463*	0.0229	0.1012	0.0501
		de facto SRI Score		
$\hat{\alpha}_{SRI}$	-0.0213***	0.0001	-0.0524**	0.0001
$Adj.R^2$	0.7566		0.8381	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 4) derived from the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

$$r_{i,t} - r_t^f = \alpha + \beta_{r^m}.(r^m - r_t^f) + \beta_{SMB}.r_t^{SMB} + \beta_{HML}.r_t^{HML} + \beta_{MOM}.r_t^{MOM}$$
$$+\mathbb{1}_{SRI}(\alpha_s + \beta_{s,r^m}.(r^m - r_t^f) + \beta_{s,SMB}.r_t^{SMB} + \beta_{s,HML}.r_t^{HML} + \beta_{s,MOM}.r_t^{MOM}) + \alpha_{SRI}.SRI_{i,t} + \epsilon_{i,t}. \quad (5)$$

Because of data availability, only seven monthly historical SRI ratings are available for the US. Table 9 reports the results obtained.

Table 9: Dynamic (time-varying SRI ratings) estimation of the augmented non-linear panel version of Carhart's model (2017-2018)

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
Full Sample				
$\hat{\alpha}$	0.0118	0.0145	0.0071	0.0065
$\hat{\beta}_{r^m}$	0.9083***	0.0326	0.9277***	0.0331
$\hat{\beta}_{SMB}$	-0.0496	0.0620	0.0100	0.0394
$\hat{\beta}_{HML}$	-0.0989***	0.0172	-0.0358	0.0580
$\hat{\beta}_{MOM}$	0.1085***	0.0381	-0.0271	0.0357
de jure SRI Mutual Funds				
$\hat{\alpha}_s$	-0.0006	0.0007	-0.0003	0.0008
$\hat{\beta}_{s,r^m}$	0.0174	0.0221	0.0127	0.0340
$\hat{\beta}_{s,SMB}$	0.1348***	0.0293	-0.0140	0.0415
$\hat{\beta}_{s,HML}$	0.0659**	0.0281	0.0876	0.0607
$\hat{\beta}_{s,MOM}$	0.0496*	0.0231	0.0986	0.0385
de facto SRI Score				
$\hat{\alpha}_{SRI}$	-0.0196***	0.0001	-0.0154***	0.0001
$Adj.R^2$	0.7515		0.7980	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 5) derived from the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notaions ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

The results obtained using time-varying SRI scores are qualitatively similar to those obtained with static SRI scores and are reported in Table 7. First, these results signal that

considering static or time-varying scores does not matter for the result, as SRI scores are not very volatile. Such a result supports the long-term adherence to ethical objectives. Second, it turns out again that announcements and adherence to ethical labels thus do not harm funds' performance. On the contrary, we observe that the estimated coefficient of the *SRI Score* (which holds for all mutual funds) is negative and significantly different from zero.

In a fourth experiment, we divide the SRI into its two distinct components: the ESG score, which considers environmental, social and governance aspects, and the Controversy score, which evaluates risks associated with a controversial announcement of an investment. Both measures correspond to effective measures and not to announcements. The results of the estimation are reported in Table 10.

It appears that both the ESG and Controversy scores have negative and significant impacts on mutual fund adjusted risk performance. This confirms that both aspects of ethics have costs in terms of performance. Interestingly, the magnitude of the controversy score is twice as large as that of the ESG score, suggesting that it is of greater importance. Such a finding can be explained by the construction of the Controversy index, which relies on a 5-class categorization[11] before normalization on a $0 - 100$ scale.

4. Conclusion

This paper examines whether there is a difference between what is said and what is done in the SRI industry. To this aim, it analyses the socially responsible dimension of mutual funds along two dimensions. The first dimension addresses mutual funds' self-presentations and labels from specialized audit agencies (*de jure* or what is said about SRI), whereas the second one deals with the funds' holdings (*de facto* or what is done about SRI). This last aspect is measured via the newly available Morningstar Sustainability Rating database, which presents many advantages. In particular, it is continuous, free of industry bias, normalized

[11]The score ranges from 0, which means no controversy, to 5, which indicates high controversy.

Table 10: Estimation of the augmented non-linear panel version of Carhart's model (2013-2018) with ESG and Controversy scores

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
		Full Sample		
$\hat{\alpha}$	0.0121***	0.0004	0.0083***	0.0020
$\hat{\beta}_{r^m}$	0.9638***	0.0178	1.0218***	0.0141
$\hat{\beta}_{SMB}$	0.1136***	0.0312	0.0896***	0.0267
$\hat{\beta}_{HML}$	-0.0953***	0.0315	-0.0361	0.0297
$\hat{\beta}_{MOM}$	-0.0192	0.0145	-0.0007	0.0119
		de jure SRI Mutual Funds		
$\hat{\alpha}_s$	-0.0012***	0.0003	-0.0002	0.0002
$\hat{\beta}_{s,r^m}$	0.0181	0.0175	-0.0154	0.0144
$\hat{\beta}_{s,SMB}$	0.0940***	0.0336	-0.0190	0.0261
$\hat{\beta}_{s,HML}$	0.0750**	0.0345	-0.0112	0.0283
$\hat{\beta}_{s,MOM}$	0.0244	0.0188	0.0220*	0.0116
		de facto SRI Score		
$\hat{\alpha}_{ESGScore}$	-0.0139***	0.0001	0.0000	0.0000
$\hat{\alpha}_{ControversyScore}$	-0.0215***	0.0001	-0.0130***	0.0000
$Adj.R^2$	0.7527		0.8416	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 4) based on the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

and homogeneous for all European and US mutual funds. The performance of these funds is analysed using a panel-augmented Carhart (1997) model, in which conventional and socially responsible mutual funds are split into two different clusters. To achieve comparable analysis of returns, a new database is built for mutual funds' returns in Europe and in the US. The database is corrected for potential survivorship and entrance biases.

This paper enriches the literature on SRI and leads to several institutional and professional implications. First, it reveals the weak correspondence between what is said and what is done in the socially responsible mutual fund industry; announcements from asset managers and from specialized audit agencies appear to serve only as marketing positioning. Second, we observe a dichotomy between the affect on performance of what is said and what is done in SRI: the *de jure* SRI has no impact on a fund's financial performance, whereas *de facto* SRI has a significant negative effect on the performance. This last result is in line with the existing theoretical literature on the consequences of extra-financial constraints on mutual fund performance.[12] In addition, this dichotomy explains the puzzle often encountered in the empirical literature. The paper also stresses that the SRI positioning of a fund does not constitute a sufficient argument for justifying lower returns. Both results cast shadow on the socially responsible mutual funds industry. Indeed, investors who are buying shares of this type of funds praise many socially responsible objectives and give a high value to extra-financial returns. If they learn that the funds in which they invested are not significantly more ethical than the other ones, it would create a major controversy and generate a wave of defiance in the whole financial sector. Thus, this paper stresses the existing urgent need to regulate this sector of activity efficiently. One way could consist of creating a public

[12]Within a Markowitz (1952) mean-variance framework, adding an extra-financial constraint can only penalize the optimization program of the investor. See Bollen (2007) for a discussion on the substitutability property of the investor's utilities. For a behavioral approach, see also Levitt and List (2007) and Doskeland and Pedersen (2016) about investor's financial and moral utilities.

and specialized audit agency. This agency would only evaluate the *de facto* aspect of the socially responsible mutual funds in order to set the path and improve the quality of other SRI labels, which only consider *de jure* SRI aspects. Asymmetric information between funds' managers and investors would thus be reduced. It would also lead to harmonization of the evaluation of the SRI objectives of the funds. Favoring the transparency of the SRI market as well as the evaluation of extra-financial returns would generate confidence in this segment of funds. There is no doubt that it will be positive for the clients as well as the virtuous funds' managers.

References

Bauer, Rob, Kees Koedijk, and Roger Otten. "International evidence on ethical mutual fund performance and investment style." Journal of Banking & Finance 29.7 (2005): 1751-1767.

Bauer, Rob, and Paul Smeets. "Social identification and investment decisions." Journal of Economic Behavior & Organization 117 (2015): 121-134.

Blomquist, Johan, and Joakim Westerlund. "Testing slope homogeneity in large panels with serial correlation." Economics Letters 121.3 (2013): 374-378.

Bollen, Nicolas PB. "Mutual fund attributes and investor behavior." Journal of Financial and Quantitative Analysis 42.3 (2007): 683-708.

Borgers, Arian, Jeroen, Derwall, Kees, Koedijk and Jenke ter Horst "Do social factors influence investment behavior and performance? Evidence from mutual fund holdings." Journal of Banking & Finance 60 (2015): 112-126.

Carhart, Mark M. "On persistence in mutual fund performance." The Journal of Finance 52.1

(1997): 57-82.

Doskeland, Trond, and Lars Jacob Tynes Pedersen. "Investing with brain or heart? A field experiment on responsible investment." Management Science 62.6 (2016): 1632-1644.

Driscoll, John C., and Aart C. Kraay. "Consistent covariance matrix estimation with spatially dependent panel data." Review of economics and statistics 80.4 (1998): 549-560.

El Ghoul, Sadok, and Aymen Karoui. "Does corporate social responsibility affect mutual fund performance and flows?." Journal of Banking & Finance 77 (2017): 53-63.

Espenlaub, Susanne, Imtiaz ul Haq, and Arif Khurshed. "It's all in the name: Mutual fund name changes after SEC Rule 35d-1." Journal of Banking & Finance 84 (2017): 123-134.

Fama, Eugene F., and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." Journal of Financial Economics 33.1 (1993): 3-56.

Goldreyer, Elizabeth F., and J. David Diltz. "The performance of socially responsible mutual funds: incorporating sociopolitical information in portfolio selection." Managerial Finance 25.1 (1999): 23-36.

Gregory, Alan, John Matatko, and Robert Luther. "Ethical unit trust financial performance: small company effects and fund size effects." Journal of Business Finance and Accounting 24.5 (1997): 705-725.

Hamilton, Sally, Hoje Jo, and Meir Statman. "Doing well while doing good? The investment performance of socially responsible mutual funds." Financial Analysts Journal 49.6 (1993):

62-66.

Hartzmark, Samuel M., and Abigail B. Sussman. "Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows." (2018).

Hong, Harrison, and Marcin Kacperczyk. "The price of sin: The effects of social norms on markets." Journal of Financial Economics 93.1 (2009): 15-36.

Humphrey, Jacquelyn E., and David T. Tan. "Does it really hurt to be responsible?." Journal of Business Ethics 122.3 (2014): 375-386.

Kreander, Niklas, Gray, Robert H., Power, David M., and Sinclair, C. D. "Evaluating the performance of ethical and non-ethical funds: a matched pair analysis." Journal of Business Finance and Accounting 32.78 (2005): 1465-1493.

Levitt, Steven D., and John A. List. "What do laboratory experiments measuring social preferences reveal about the real world?." Journal of Economic perspectives 21.2 (2007): 153-174.

Lintner, John. "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets." The Review of Economics and Statistics (1965): 13-37.

Luther, Robert G., John Matatko, and Desmond C. Corner. "The investment performance of UK ethical Unit Trusts." Accounting, Auditing and Accountability Journal 5.4 (1992).

Luther, Robert G., and John Matatko. "The performance of ethical unit trusts: choosing an appropriate benchmark." The British Accounting Review 26.1 (1994): 77-89.

Mallin, Christine A., Brahim Saadouni, and Richard J. Briston. "The financial performance of ethical investment funds." Journal of Business Finance and Accounting 22.4 (1995): 483-496.

Nofsinger, John, and Abhishek Varma. "Socially responsible funds and market crises." Journal of Banking & Finance 48 (2014): 180-193.

Markowitz, Harry. "Portfolio selection." The Journal of Finance 7.1 (1952): 77-91.

Pesaran, M. Hashem, and Takashi Yamagata. "Testing slope homogeneity in large panels." Journal of Econometrics 142.1 (2008): 50-93.

Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang. "Socially responsible investments: Institutional aspects, performance, and investor behavior." Journal of Banking & Finance 32.9 (2008): 1723-1742.

Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang. "The price of ethics and stakeholder governance: The performance of socially responsible mutual funds." Journal of Corporate Finance 14.3 (2008): 302-322.

Riedl, Arno, and Paul Smeets. "Why do investors hold socially responsible mutual funds?." The Journal of Finance 72.6 (2017): 2505-2550.

Sharpe, William F. "Capital asset prices: A theory of market equilibrium under conditions of risk." The Journal of Finance 19.3 (1964): 425-442.

Schroder, Michael. "Is there a difference? The performance characteristics of SRI equity

indices." Journal of Business Finance and Accounting 34.12 (2007): 331-348.

Statman, Meir. "Socially responsible mutual funds (corrected)." Financial Analysts Journal 56.3 (2000): 30-39.

Statman, Meir. "Socially responsible indexes." The Journal of Portfolio Management 32.3 (2006): 100-109.

Statman, Meir, and Denys Glushkov. "The wages of social responsibility." Financial Analysts Journal 65.4 (2009): 33-46.

Appendix 1: Description of Morningstar Portfolio Sustainability Scores

Portfolio sustainability scores are obtained from Morningstar. This score reflects the quality of mutual funds relative to environmental, social and governance (ESG) parameters and is calculated as follows:

$$Portfolio.Sustainability.Score = Portfolio.ESG.Score - Portfolio.Controversy.Deduction \tag{6}$$

The calculation of the Portfolio ESG score is based on data provided by "Sustainalytics" [13], a leader in ESG asset rating. To attribute an ESG score to a company (or an asset), Sustainalytics compares a company to other companies of the same industry and uses different indicators on a $0-100$ scale. Morningstar then normalizes scores as follows:

$$Zc = \frac{ESG_i + \mu_{industry}}{\sigma_{industry}}, \tag{7}$$

$$ESG.Norm_i = 50 + 10Zc. \tag{8}$$

Morningstar determines the Portfolio ESG score from the weighted average of an asset's ESG score. For a fund to be graded, Morningstar requires the fund to have at least 50% of assets with an ESG score obtained from Sustainalytics (the percentage of assets scored is rescaled to 100%).

$$Portfolio.ESG.Score = \sum_{i=1}^{n} w_i ESG.Norm_i, \tag{9}$$

where:

$\sum_{i=1}^{n} w_i = 1$ and for each i, $w_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$ if $\sum_{i=1}^{n} x_i > 50\%$.

[13] http://www.sustainalytics.com

Portfolio controversy deduction is also obtained by Sustainalytics. Sustainalytics assesses companies involved in ESG-related incidents on a $0-100$ scale. This negatively contributes to the Portfolio sustainability score. For a fund to be graded, Morningstar requires that 50% of the fund's assets have a controversy score assigned by Sustainalytics (the percentage of assets scored is rescaled to 100%). Portfolio controversy deduction is calculated as follows:

$$Portfolio.Controversy.Score = \sum_{i=1}^{n} w_i.S.Contr_i. \tag{10}$$

Appendix 2: Panel version of Carhart's model with dummy *de jure* (labels)

This model with a dummy for funds that have been granted a label is only performed for European mutual funds, as there is no existing unified label in the US.

Estimation of the panel version of Carhart's model (2013-2018) with labels

	Europe	
	Estimates	*s.e.*
$\hat{\alpha}$	-0.0027	0.0067
$\hat{\beta}_{r^m}$	0.9765***	0.0192
$\hat{\beta}_{SMB}$	0.1754***	0.0492
$\hat{\beta}_{HML}$	-0.0170	0.0234
$\hat{\beta}_{MOM}$	0.0045	0.0250
$\hat{D}ummyDeJure$	0.0012	0.0068
$Adj.R^2$	0.7668	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1) based on the GLM method with an extra *de jure* dummy for labelised funds. Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.