179f

FINANCIAL PERFORMANCE OF GOVERNMENT BOND PORTFOLIOS BASED ON ENVIRONMENT, SOCIAL AND GOVERNANCE CRITERIA

Guillermo Badía[*]

Vicente Pina

Lourdes Torres

Faculty of Business and Economics, University of Zaragoza

Área temática: B) Valoración y Finanzas

Keywords: Socially responsible investing; Government bond portfolios; Performance evaluation; Market states.

** Corresponding author:*

FINANCIAL PERFORMANCE OF GOVERNMENT BOND PORTFOLIOS BASED ON ENVIRONMENT, SOCIAL AND GOVERNANCE CRITERIA

Abstract

In this paper we evaluate the financial performance of government bond portfolios, based on environment, social and governance criteria. Socially responsible investment sovereign bonds have undergone a remarkable growth in recent decades. In contrast to most previous studies, which apply sustainability ratings at the firm level, we use sustainability ratings at the country level. In addition, we investigate the influence of different market conditions on the performance of SRI government bond portfolios. Through several risk-adjusted measures, preliminary results show that there are no statistically significant differences between the financial performance of high- and low-ranked countries. However, high-ranked countries outperform low-ranked under any cut-off, generating economically significant differences. These findings evidence that an investor can satisfy social concerns without sacrificing financial performance.

1. Introduction

The growth of socially responsible investment (SRI) has been notable in recent decades. According to the 2016 Global Sustainable Investment Review, in 2016 there were $22.89 trillion of assets being professionally managed under responsible investment strategies worldwide, an increase of 25% since 2014. In 2016, 53% of the professionally managed assets in Europe used responsible investment strategies, this proportion in the US was 22% and, in Australia/New Zealand, 51% of funds under professional management was driven by SRI. For the Global Sustainable Investment Association (GSIA) 2016, sustainable investing is an investment approach that considers environment, social and governance (ESG) factors in portfolio selection and management. The SRI notion has been particularly linked to equity selection. Screening investment processes, which allow an investor to select or exclude firms from the available investment universe based on ESG criteria, have helped investors to align their personal beliefs and values with their investment decisions. In fact, rising individual awareness of environmental, social, and ethical concerns is nowadays strongly influencing the purchase decisions of investors (Mollet and Ziegler, 2014).

The interest of investors in finding sustainable firms has induced many firms to pay more attention to their behavior with respect to society and the environment. Firms have realized that improving their relationships with stakeholders may positively influence their own long-term opportunities. In fact, Preston and O'Bannon (1997) argue that satisfying the interests of different corporate stakeholders enhances a firm's reputation, resulting in a positive impact on its financial performance. Since corporate social responsibility (CSR) involves constantly assessing corporate influences and relationships with stakeholders and the environment, it allows management to recognize and react to evolving strategic opportunities and challenges. Derwall et al. (2011) support this argument pointing out that improving social or environmental performance can generate new market opportunities. This picture is also sustained by many empirical studies that evidence a positive effect of CSR on financial performance and valuation.[1]

[1] For an extensive review of the empirical studies in the field, see for example, the studies of Margolis and Walsh (2003), Orlitzky et al. (2003), Margolis et al. (2009), Lu et al. (2014) and Javed et al. (2016).

The financial benefits of ESG screening processes from the firm's perspective have been evaluated in previous research (e.g., Kempf and Ostoff, 2007; Statman and Glushkov, 2009; Renneboog et al., 2008; Borgers et al., 2013). These studies evaluate the financial performance of stock portfolios constructed from ESG screens. As Mollet et al. (2013) point out, portfolio analysis currently constitutes the most active methodology to exam the empirical effect of SRI on financial performance. This type of analyses is very valuable in an applied framework since, from an investor perspective, the critical issue is whether socially responsible stock selection adds or destroys value in terms of portfolio financial performance. However, although the concept of SRI was originally linked to stock selection, the proportion of portfolio investors applying SRI criteria to bonds has grown significantly in recent years. According to the European Sustainable Investment Forum (EUROSIF, 2016), equities represented over 30% of the SRI assets in December 2015, a significant decrease from the previous year's 50%. Meanwhile, there was a strong increase in bonds from the 40% registered in December 2013 to 64%. Both corporate bonds and sovereign bonds underwent a remarkable growth. The former rose from 21.3% to 51.17% of the bond allocation while the latter increased from 16.6% to 41.26%. In this regard, the financial implications of ESG screening processes on corporate bonds may be highly related to stock selections based on ESG concerns since corporate bonds are associated with firms. In fact, previous studies suggested this. For instance, Derwall and Koedijk (2009) and Leite and Cortez (2016), who evaluate the financial performance of mutual funds that invest in socially responsible fixed-income stocks, find that the average SRI bond fund performed similarly to conventional funds. These results are in line with most empirical studies about the performance of SRI funds which show that they tend to have a similar performance to their conventional peers (Revelli and Viviani, 2015).

As for sovereign bonds, since they are not related to firms, the financial implications of ESG screening processes on countries may bring new insights to the SRI field and help to build an in-depth understanding of SRI consequences for alternative assets to firms. Country ESG analysis may offer a view into an economy's underlying change drivers and provides investors with insights into a country's strengths and weaknesses. In this vein, despite the SRI sovereign bond market growth and the development of country ratings based on ESG factors in recent years, as Drut (2010) notes, the link between sovereign bond returns and state performance in terms of ESG concerns has been overlooked. In fact, to the best of our knowledge, no previous research has yet evaluated the financial performance of responsible sovereign bond investments. The main objective of this paper is to fill this gap. To this end, from a portfolio approach, we evaluate the financial performance of sovereign bond portfolios based on ESG criteria. In contrast to most previous studies, which apply sustainability ratings at the firm level, we use sustainability ratings at the country level. We employ the RobecoSAM's country sustainability ranking developed by RobecoSAM and Robeco. This ranking is a comprehensive framework for assessing countries' ESG performance. By focusing on ESG factors such as aging, competitiveness and environmental risks, this country sustainability ranking offers a view into a country's strengths and weaknesses. According to countries' ESG scores, we rank them and form a high and a low portfolio including ESG outperformers and underperformers, respectively. We also investigate the returns on a long-short portfolio (the difference portfolio), which is constructed by subtracting the low-ranked portfolio returns from the returns on the high-ranked stock portfolio. Since the country sustainability ranking is updated semi-annually, portfolios are constructed twice a year. Additionally, given that some recent literature identifies that different market phases (e.g., expansion and recession periods) affect the financial performance of SRI equity funds (Nofsinger and Varma, 2014; Becchetti et al., 2015), SRI fixed-income funds (Henke, 2016), and socially responsible stocks (Brzeszczyński and McIntosh, 2014; Carvalho and Areal, 2016), we investigate the influence of different market conditions on the performance of SRI government bond portfolios. In order to document changes in both risk and performance, we implement a conditional model in the spirit of Carvalho and Areal (2016) and Leite and Cortez (2016).

Previous research has shown that ESG factors are valuable for sovereign bonds. Through regression analysis, Capelle-Blancard et al. (2016) assess whether ESG performance can influence sovereign bond spreads. They find that countries with good ESG performance tend to have less default risk and thus lower bond spreads. Thus, the findings of Hoepner and Neher (2013) were reinforced since they found a negative and significant relationship between sovereign debt and a rating of national sustainability. We want to ascertain whether ESG factors are valuable from a portfolio management perspective and whether it is possible to form portfolios with sovereign bonds screening them according to sustainability country rankings without sacrificing financial performance. Drut (2010) assesses a feasible diversification portfolio problem associated with sovereign bond portfolios. He computes the efficient frontier of portfolios including sovereign bonds from 20 developed countries. His results show that sovereign bond portfolios with a high social responsibility score can be formed without significant loss of diversification. Hence, investors can build sovereign bond portfolios with a higher than average social responsibility rating without significantly renouncing the potential for diversification. Nonetheless, additional research needs to be conducted in order to determine whether sovereign bond portfolios based on ESG concerns can be constructed without sacrificing financial performance.

The driving force behind SRI demand has led firms to pay more attention to their CSR activities and policy. Consistently, investors screening countries according their sustainability scores could influence country ESG policy. With their investment decisions, investors may stimulate governments to take social and environmental aspects into account.

2. Literature Review

The growth of SRI and its consequences has stimulated empirical studies assessing financial behaviours. As Brammer et al. (2006), Osthoff (2015) and Auer (2016) stress, the majority of prior empirical studies have focused on the financial performance of investment funds. (e.g. Statman, 2006; Schröder, 2007; Renneboog et al., 2008; Cortez et al., 2012; Managi et al., 2012; Lean and Nguyen, 2014). In general, these studies find that there are no significant differences between the performance of SRI mutual funds and conventional funds (Leite and Cortez, 2016). This evidence encompasses different geographical areas. For instance, Gregory and Whittaker (2007) assess the performance of UK funds, Kreander et al. (2005) evaluate UK, Swedish, German and Dutch funds, Scholtens (2005) examine Dutch funds, Fernandez-Izquierdo and Mattalin-Saez (2008) analyse Spanish funds, Bauer et al. (2006)

study Australian funds, and Statman (2000), Bello (2005), and Utz and Wimmer (2014) focus on US funds[2]. In spite of the fact that all this attention is valuable from a practical point of view, some authors (e.g., Brammer et al., 2006; Kempf and Osthoff, 2007; Statman and Glushkov, 2009; Auer and Schuhmacher, 2016) have highlighted certain limitations related to fund studies. For instance, Brammer et al. (2006) and Kempf and Osthoff (2007) point out that confusing effects, such as fund manager performance and management fees, may make it difficult to show differences in mutual fund performance. Furthermore, the evidence of Wimmer (2013), Utz and Wimmer (2014), Humphrey et al. (2016), and Statman and Glushkov (2016) indicates that the label 'socially responsible' may be more akin to a marketing strategy, thus raising doubts among investors that an SRI fund is truly socially responsible. In this context, investors may find it difficult to know the extent to which an SRI fund is really considering social criteria in its selection process.

In parallel, a group of researchers have formed stock portfolios in order to analyse the financial performance of SRI. Some of them construct a high and a low portfolio, including ESG outperformers and underperformers, and investigate the returns on a long-short portfolio. Others compare a high portfolio covering the stocks with the highest ESG classifications to one of conventional investments. Perhaps the most controversial issue associated with SRI stock portfolios is the resulting financial performance. While some studies do not find significant financial differences between high and low sustainable firms or conventional benchmarks (e.g., Van de Velde et al., 2005; Galema et al., 2008, and Mollet et al., 2014), others support the positive financial performance of SRI (e.g., Derwall et al., 2005; Kempf and Osthoff, 2007; Eccles et al., 2014), and some find evidence of negative performance (Brammer et al., 2006; Auer and Schuhmacher, 2016).

As for the performance of SRI from a country portfolio approach, as we stress above, to our knowledge, there are no studies. Drut (2010) investigates the impact of socially responsible qualifications on the efficient frontier of sovereign portfolios, but he does not examine the financial performance of the SRI portfolios. Some studies have evaluated the performance of SRI fixed-income

[2] For more details on the financial performance of SRI, see for instance Renneboog et al (2008b), Margolis et al (2009), Capelle-Blancard and Monjon (2012), and Revelli and Viviani (2015).

funds although, as Leite and Cortez (2016) point out, these have also received little attention. Fixed-income funds invest in both corporate and government bonds. Goldreyer et al. (1999) find that US SRI fixed-income funds underperform conventional funds. However, years later, Derwall and Koedijk (2009) show that US SRI fixed-income funds perform similarly to conventional funds. Leite and Cortez (2016) evidence an empirical mix depending on the specific European market evaluated. For instance, whereas UK SRI funds underperform and German SRI funds outperform conventional funds, French SRI funds show similar performance to conventional funds.

It appears evident that the financial outcome of SRI has caused an intensive debate, evaluating it from different perspectives. However, in spite of the growth of portfolio investors applying SRI criteria to government bonds in recent years, country sustainability analysis from an investor – portfolio - approach has been overlooked.

3. Methods

3.1. Portfolio construction

Financial implications of social responsibility screens on a country level are evaluated following an ESG portfolio method. Each semester *p* two portfolios are formed. Countries are ranked according to their ESG overall score available at *p-1* and based on this ranking, different cut-off portfolio construction procedures are used, specifically: 10%, 20%, 30%, 40%, and %50%. Previous studies apply different cut-off percentages (see for instance Van de Velde et al., 2005; Kempf and Osthoff, 2007; Derwall et al., 2011; Halbritter and Dorfleitner, 2015; Auer, 2016). We use alternative cut-off percentages to check the robustness of our results. Thus, for instance at the 10% (the most demanding SRI high level), the high portfolio covers 10% of countries with the highest classifications, and the low portfolio collects 10% of the countries with the lowest classifications. Despite Drut (2010) shows that high socially responsible sovereign bond portfolios can be constructed without significant loss of diversification, rather than form a value-weighted portfolio, we construct an equally-weighted portfolio because of diversification improvements. Given the procedure to construct a value-weighted portfolio itself, standard deviation may be affected, since this type of portfolio is less diversified. Statman and Glushkov (2009) for instance, find that their value-weighted portfolio (top-bottom portfolio) has a higher standard deviation than their equally-weighted one. Therefore, if we formed a value-weighted portfolio, our conclusions on ESG concerns could be biased.

3.2. Financial performance measures

Previous studies have shown that the magnitude and even the sign of the long-run abnormal returns are sensitive to alternative measurement methodologies (e.g., Barber and Lyon, 1997; Fama, 1998; Loughran and Ritter, 2000). Hence, to determine the sensitivity of our results, we evaluate the financial performance applying several approaches.

3.2.1. *Sharpe ratio and significance tests*

We use the Sharpe ratio (1966) - the ratio of excess return to standard deviation - to compare the performance of alternative investment strategies. From two investment portfolios *i* and *j* whose excess returns over the risk-free rate at time *t* are r_{ti} and r_{tj} respectively, a total of *T* return pairs (r_{1i}, r_{1j}), …, (r_{Ti}, r_{Tj}) are observed. The difference between two Sharpe ratios is given by $\Delta = Sh_i - Sh_j = \mu_i / \sigma_i^2 - \mu_j / \sigma_j^2$, where μ and σ^2 are the sample mean and standard deviation respectively. As the value of the Sharpe ratio is really an estimate from historical return data, statistical inference is applied in order to compare the two indicators. For this purpose, prior studies (e.g., DeMiguel and Nogales, 2009; Gasbarro et al., 2007) used the test of Jobson and Korkie (1981) and the correction proposed by Memmel (2003). However, this test is not valid if the returns distribution is non-normal, or if the observations are correlated over time, both phenomena quite common in financial returns time series data. Recently, Ledoit and Wolf (2008), hereafter LW, propose a studentized time series bootstrap approach that works asymptotically and has satisfactory properties in finite samples. The literature (e.g., Hall, 1992; Lahiri, 2003) shows the enhanced inference accuracy of the studentized bootstrap over standard inference based on asymptotic normality. LW propose to test $H_0 : \Delta = Sh_i - Sh_j = 0$ by inverting a bootstrap confidence interval. A two-sided bootstrap confidence interval with nominal level 1-α for Δ (true difference between the Sharpe ratios) is constructed and if zero is not contained in the interval, then H_0 is rejected at nominal level α. Specifically, LW propose to construct a symmetric studentized time series bootstrap confidence interval. To this end, the two-sided distribution function of the studentized statistic is approximated through the bootstrap by $F\left(\left|\hat{\Delta} - \Delta\right| / s\left(\hat{\Delta}\right)\right) \approx F\left(\left|\hat{\Delta}^{¿} - \Delta\right| / s\left(\hat{\Delta}^{¿}\right)\right)$, where Δ is the true difference between the Sharpe ratios, $\hat{\Delta}$ is the estimated difference computed from the original data, $s\left(\hat{\Delta}\right)$ is a standard error for $\hat{\Delta}$ (also calculated from the original data), $\hat{\Delta}^{¿}$ is the estimated difference computed from bootstrap data, and $s\left(\hat{\Delta}^{¿}\right)$ is a standard error for $\hat{\Delta}^{¿}$ (also calculated from bootstrap data). Letting $z^{¿}_{|\cdot|,\lambda}$ be a λ quantile of $F\left(\left|\hat{\Delta}^{¿} - \Delta\right| / s\left(\hat{\Delta}^{¿}\right)\right)$, a bootstrap 1-α confident interval for Δ is given by $\hat{\Delta} \pm z^{¿}_{|\cdot|,1-\alpha} s\left(\hat{\Delta}\right)$. LW note that, with heavy-tailed data or data of a time series nature, this quantile will typically be somewhat larger than its standard normal counterpart (used in the traditional tests) in small to moderate samples, resulting in more conservative inferences. To generate the bootstrap data, we use the circular block bootstrap of Politis and Romano (1992), resampling blocks of pairs from the observed pairs (r_{ti}, r_{tj}), t=1,…, T, with replacement. Applying the studentized circular block bootstrap requires a choice of the block size b and LW propose to use the calibration procedure of Loh (1987), suggesting that M = 5000 bootstrap sequences is sufficient for reliable inference. The standard error $s(\hat{\Delta})$ is calculated through kernel estimation; specifically, the pre-whitened quadratic spectral kernel of Andrews and Monahan (1992). The standard error $s(\hat{\Delta}^{¿})$ is the natural standard error calculated from the bootstrap data, making use of a special block dependence structure. The bootstrap p-values are computed as $PV=(\tilde{d}^{¿,m} \geq d)+1/M+1$, where $d=|\hat{\Delta}|/s(\hat{\Delta})$, the original studentized test statistic, $\tilde{d}^{¿,m}=|\hat{\Delta}^{¿,m}+\hat{\Delta}|/s(\hat{\Delta}^{¿,m})$, denotes the centered studentized statistic computed form the mth bootstrap sample by $d^{¿,m}$, m=1,…, M, and M is the number of bootstrap resamples.

3.2.2. Multi-factor model

A parallel approach to evaluate portfolio performance involves computing alphas from multi-factor models. On a firm level, previous studies (e.g., Van de Velde et al., 2005; Brammer et al., 2006; Galema et al., 2008; Derwall et al., 2011; Edmans, 2011; and Humphrey et al., 2012) commonly examine performance using the four-factor Carhart (1997) model that captures the risk premiums associated with size and value versus growth (as in Fama and French, 1993) as well as momentum, motivated by Jegadeesh and Titman (1993). On a fixed-income approach, the four-factor Elton et al. (1995) fixed-income asset pricing model has been one of the most influential models (e.g., Derwall and Koedijk, 2009; Henke, 2016; and Leite and Cortez, 2016). Elton et al. (1995) develop a fixed-income asset pricing model which allows pricing expected return in bond markets. This model covers the exposure to the broad bond market, the high-yield market, the mortgage market, and the equity market, using factors formed by indices. Nevertheless, in comparison with the multitude of studies regarding equities and fixed-income, the specific international government bonds field is a less explored one. Recently, Zaremba and Czapkiewicz (2017) propose a four-factor model that includes influential factors on international government bonds (volatility, credit risk, value effect, and momentum) which can be used to evaluate investment portfolio performance in the spirit of Elton et al. (1995) and Carhart (1997). Specification model is as follow: (Eq.1)

$$R_t = \alpha + \beta_{VMS} VMS_t + \beta_{HML} HML_t + \beta_{UMD} UMD_t + \beta_{RMS} RMS_t + \varepsilon_t ,$$

Where R_t is the portfolio excess return on time *t*. Excess returns are computed using one-month US T-bills as risk-free rate. α is the intercept and represents the estimated financial performance of the portfolio. VMS_t, HML_t, UMD_t, and RMS_t are the explanatory factors representing government bond premia: volatile minus stable (VMS), the volatility premium offers compensation for the market risk of a bond as a result mostly from interest rate exchange; high minus low (HML), the value effect is a tendency of assets with high fundamentals to price ratios to outperform assets with low fundamentals to price ratios; up minus down (UMD), momentum effect is a tendency for assets that performed well (poorly) in the past to continue to outperform (underperform); and risky minus safe (RMS), credit risk premium rewards investors for risking the bankruptcy of the issuer. Portfolio factors based on one-way sorts are formed on tertile breakpoints. Returns on asset pricing factors are the returns on a long-short investment portfolio assuming a long (sort) position on highest (lowest) sorting variables. Betas are risk factor sensitivities associated to the different risk factors: volatility, value effect, momentum and credit risk, and $\varepsilon_{p,t}$ are the regression's residuals.

Factors are defined following to Zaremba and Czapkiewicz (2017) approach. Volatility risk is measured with the modified duration of the country bond portfolio. It expresses the measurable change in the value of a bond in response to a change in market-wide interest rates. Credit risk is proxied by budget deficit as a percent of GDP and net debt-to-GDP ratio. We calculate z-scores for the budget deficit and debt-to-GDP ratio for all 24 countries, then the averaged z-scores corresponding with the budget deficit and net debt are used as the credit values for each country. Given budget deficit and debt-to-GDP ratio data are annual, we use a cubic spline interpolation to generate monthly data.

Value effect is defined as the residuals from a regression of a country bond portfolio excess return on its duration and credit factors. This measure is closely related to the term spread used commonly in previous studies. Momentum effect is measured using six-month previous return skipping the most recent month.

3.2.3. Different market conditions' effect

To evaluate the influence of different market phases on financial performance, we use a conditional model in line with Nofsinger and Varma (2014), and Leite and Cortez (2016), which allows risk and performance to vary across different market phases. The conditional model incorporates two dummy variables in order to obtain different estimated coefficients in different market states, as follows: (Eq.2)

$$R_t = \alpha_{Bear} D_{Bear,t} + \alpha_{Bull} D_{Bull,t} + \beta_{1Bear} VMS_t D_{Bear,t} + \beta_{1Bull} VMS_t D_{Bull,t} + \beta_{2Bear} HML_t D_{Bear,t} + \beta_{2Bull} HML_t D_{Bull,t} +$$

Where $D_{Bear,t}$ is a dummy variable that takes value of one for bear market periods and zero otherwise and $D_{Bull,t}$ is a dummy variable that takes value of one for bull market periods and zero otherwise. α_{Bear} corresponds to the financial performance in bear markets and α_{Bull} in bull markets. β_{1Bear}, β_{2Bear}, β_{3Bear} and β_{4Bear} correspond to the factor loadings in bear periods, and β_{1Bull}, β_{2Bull}, β_{3Bull} and β_{4Bull} in bull periods. As Leite and Cortez (2016) point out, this procedure extends the model of Nofsinger and Varma (2014) by incorporating the dummy variables both for the alphas and for the risk factors, thereby enabling the analysis of financial performance and risk exposures in different market states.

To analyse the financial performance of SRI stock portfolios in different market states, we begin by identifying the different economic states across our sample period using the Pagan and Sossounov (2003), hereafter PS, approach. PS develop a statistical approach to determine peaks and troughs of a stock or bond market index, with the peaks and troughs representing relatively high and low points of a bond index series during a period of time. A peak is identified at *t* time if the event $PK = [\ln P_{t-8}, \ldots, \ln P_{t-1} < \ln P_t > \ln P_{t+1}, \ldots, \ln P_{t+8}]$ occurs, where P_t represents the quotation of the relevant bond index, and a trough at time *t* if the event $TH = [\ln P_{t-8}, \ldots, \ln P_{t-1} > \ln P_t < \ln P_{t+1}, \ldots, \ln P_{t+8}]$ occurs. Consistent with the literature, we qualify bear periods as those with a downtrend in the relevant bond market index of at least 20% from peak to trough. This process is recently used in financial studies such as Lee et al. (2013), and Carvalho and Areal (2016), among others.

4. Data

The RobecoSAM country sustainability ranking is used to classify the countries according their ESG performance. Robeco and RobecoSAM have jointly developed a comprehensive and systematic framework for determining country sustainability rankings. Sources used by RobecoSAM include international organizations such as the World Bank, the United Nations, and the International Labor Organization, as well as a variety of reputable government agencies, private institutions and NGOs. The framework forms the basis for incorporating environmental, social and governance risk analysis into the construction process for Robeco and RobecoSAM's sovereign debt portfolios and indices. RobecoSAM's country sustainability framework evaluates multiple countries on a broad range of ESG factors that are considered key risk and return drivers for investors[3]. It consists of 17 indicators, each of which is based on various data series, or sub-indicators, whereby each indicator is assigned a predefined weight out of the total framework. Based on the standardized scores countries receive for each of the indicators and their corresponding weights, a country sustainability score ranging from 1 to 10, with 10 being the highest, is calculated for each country. The resulting scores offer insights into the investment risks and opportunities associated with each country, and allow investors to better compare countries. The weighting scheme is reviewed periodically, reflecting RobecoSAM's view on the potential impact of each indicator on a country's risk profile.

In this study we assess 24 countries over the period June 2006 to December 2017[4]. Figures 1 and 2 show the top 5 and bottom 5 countries according the country sustainability ranking for the first (2006, 1st semester) and the last (2017, 2nd semester) period. The countries in both the top and the bottom remain the same in spite of the fact that more than ten years have passed

---

[3] See Appendix 1 for an extensive explanation of sustainability factors.

[4] Specific countries are: Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States.

between the two classifications. This evidence suggests that there may exist a noteworthy traditional and cultural component based on ethical values and respect for the environment and society. These figures may also highlight a limited capacity of previous policy initiatives to improve the ESG standards of low-rated countries. In this context, and given the behaviour of firms under CSR investor interests, the force of investors behind SRI demands may encourage the interest of policy-makers in ESG concerns.

Figure 1. Country sustainability ranking
Top 5 and bottom 5 countries in the 1st semester of 2006

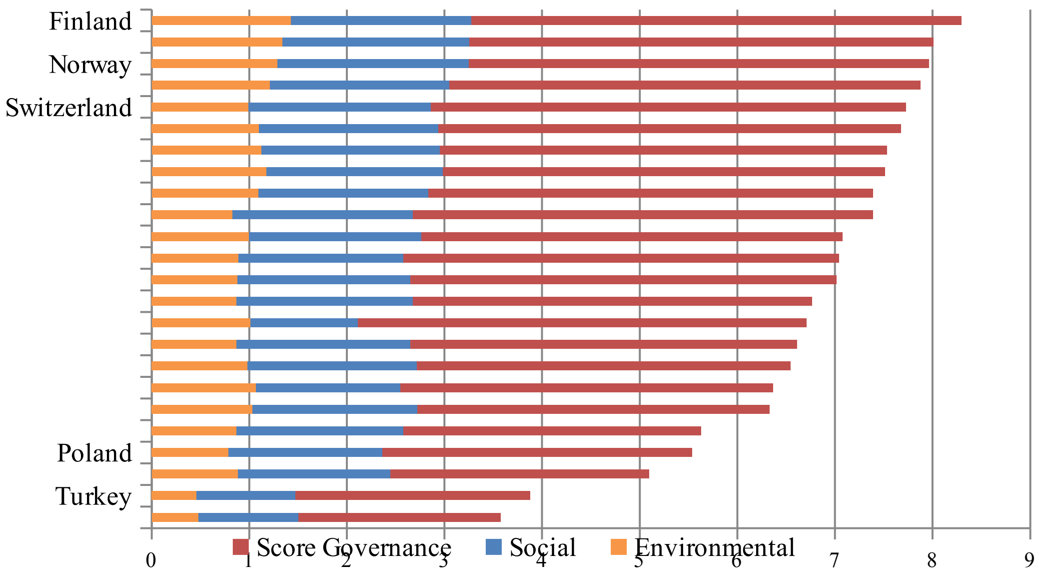
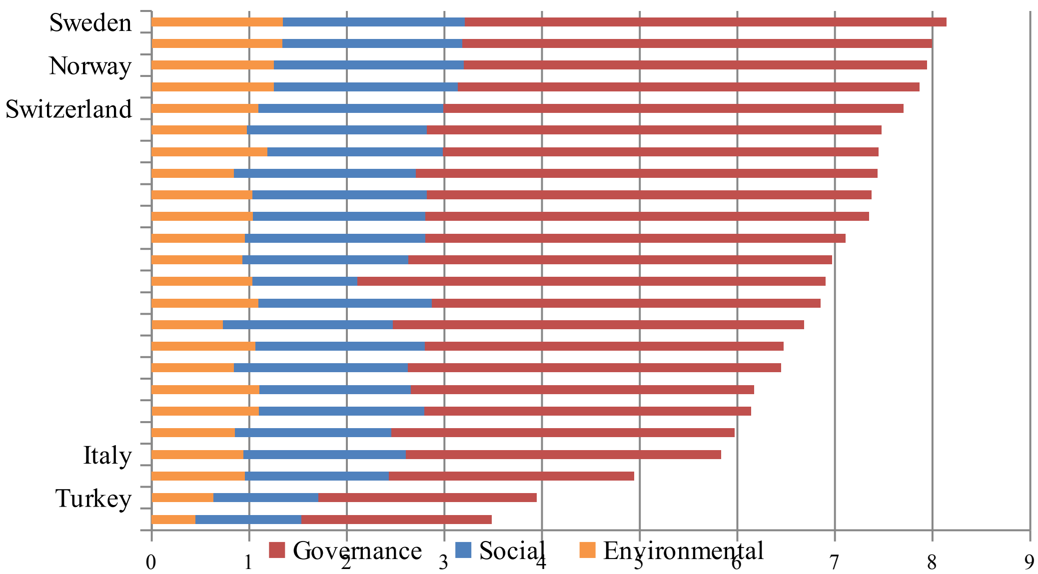


Figure 2. Country sustainability ranking
Top 5 and bottom 5 countries in the 2nd semester of 2017



The data on government bond monthly total returns come from FTSE Global Government Bond Indices ‘All maturities’, download from Datastream database in US dollars. Table 1 shows descriptive statistics for the high and low portfolio on several cut-off. The high portfolio shows higher average return than the low portfolio under any cut-off percentage. However, differences between average returns of both portfolios are not statistically significant. As for standard deviation, the high-ranked countries show a lower variability in terms of bond returns. Although standard deviation differences between the high and low portfolio are not significant, this suggests that risk affects to a large extent to low-ranked countries, whereas high-ranked feature a superior stability. These findings encourage the need to use risk-adjusted measures to evaluate the financial performance. T-Mean adds some information on the financial outcome of applying ESG screens to the investment universe. The high portfolio shows a higher t-Mean value on each cut-off. Descriptive statistics allow us to evidence what may be the financial outcome related to ESG screening process on countries. However, an extensive evaluation using risk-adjusted measures comes conveniently.

Table 1. Descriptive statistics

This table displays a summary statistic of the high and low portfolio according to several cut-off portfolio constructions procedures. Mean (SD) is the average return (standard deviation) of portfolios, T-mean is the ratio of mean to standard deviation, and *Diff* is the difference between high and low portfolios, including ESG outperformers and underperformers respectively. Parentheses encompass values of the t-test (F-test) on tests of equality of mean (variances). The full sample period is from June 2006 to December 2017.

	0.1	0.1	0.2	0.2	0.3	0.3	0.4	0.4	0.5	0.5
	High	Low	High	Low	High	Low	High	Low	High	Low
Mean	0.0026	0.0013	0.0028	0.0026	0.0039	0.0034	0.0039	0.0034	0.0038	0.0035
Diff	*0.0013*	*(0.3598)*	*0.0002*	*(0.0536)*	*0.0005*	*(0.1261)*	*0.0005*	*(0.1494)*	*0.0003*	*(0.0963)*
SD	0.0298	0.0310	0.0282	0.0319	0.0284	0.0324	0.0279	0.0287	0.0266	0.0271
Diff	*-0.0012*	*(1.0838)*	*-0.0038*	*(1.2842)*	*-0.0040*	*(1.2987)*	*-0.0008*	*(1.0605)*	*-0.0005*	*(1.0384)*
T-Mean	0.0863	0.0404	0.0988	0.0811	0.1365	0.1055	0.1413	0.1194	0.1436	0.1294
Diff	*0.0459*		*0.0177*		*0.0310*		*0.0218*		*0.0142*	

5. Empirical Results (*Preliminary*)

This section presents preliminary empirical results. Table 2 shows the results of applying the Sharpe ratio and the LW procedure to estimate the statistical significance of the difference between the Sharpe ratio of the high and low portfolios. It is striking that the high portfolio outperforms the low portfolio under any cut-off. The highest difference appears using the most demanding SRI high level, i.e., by only incorporating to the high portfolio those countries with best practices on socially responsible concerns and to the low portfolio those countries with worst practices. The Sharpe ratio estimate for the high-10% portfolio is 0.0624 and for the low-10% portfolio 0.0174, resulting in a difference of 0.0450. In spite of the fact that this difference is economically significant, the LW t-test is not statistically significant. The rest of long-short (Diff) portfolios show that high-ranked countries render better than low-ranked in terms of Sharpe ratio financial performance, but differences are not statistically significant according to the LW procedure. Next sections evaluate whether influential factors (volatility, credit risk, value effect, and momentum) and market phases affect the financial performance of government bond portfolios.

Table 2. Portfolio performance on the Sharpe ratio and the LW significant tests

This table shows financial performance of the high and low portfolio according to several cut-off. Portfolio financial performance is evaluated by means of the Sharpe ratio (SH), and the LW procedure is used to identify statistical significant differences between the Sharpe ratios of the portfolios. Diff is the Sharpe ratio difference between the high and low portfolio and LW t-test is the t-statistics according LW process. One-month US T-bills proxy for the risk-free rate. The full sample period is from June 2006 to December 2017.

	0.1	0.1	0.2	0.2	0.3	0.3	0.4	0.4	0.5	0.5
	High	Low	High	Low	High	Low	High	Low	High	Low
SH	0.0624	0.0174	0.0736	0.0590	0.1115	0.0837	0.1158	0.0949	0.1169	0.1034
Diff	0.0450		0.0146		0.0278		0.0209		0.0134	
LW *t-test*	*0.6395*		*0.2842*		*0.5355*		*0.4743*		*0.3298*	

[Results about multi-factor model and conditional model are under construction.]

Our proposal is to have a full version with extensive findings available at the meeting date.

References

Andrews, D. W., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 953-966.

Auer, B. R. (2016). Do socially responsible investment policies add or destroy European stock portfolio value? *Journal of Business Ethics, 135*(2), 381.

Auer, B. R., & Schuhmacher, F. (2016). Do socially (ir) responsible investments pay? new evidence from international ESG data. *The Quarterly Review of Economics and Finance, 59*, 51-62.

Barber, B. M., & Lyon, J. D. (1997). Firm size, book-to-market ratio, and security returns: A holdout sample of financial firms. *The Journal of Finance, 52*(2), 875-883.

Bauer, R., Otten, R., & Rad, A. T. (2006). Ethical investing in Australia: Is there a financial penalty? *Pacific-Basin Finance Journal, 14*(1), 33–48.

Becchetti, L., Ciciretti, R., Dalò, A., & Herzel, S. (2015). Socially responsible and conventional investment funds: Performance comparison and the global financial crisis. *Applied Economics, 47*(25), 2541-2562.

Borgers, A., Derwall, J., Koedijk, K., & Ter Horst, J. (2013). Stakeholder relations and stock returns: On errors in investors' expectations and learning. *Journal of Empirical Finance*, *22*, 159-175.

Brammer, S., Brooks, C., & Pavelin, S. (2006). Corporate social performance and stock returns: UK evidence from disaggregate measures. *Financial Management, 35*(3), 97-116.

Brzeszczyski, J., & McIntosh, G. (2014). Performance of portfolios composed of British SRI stocks. *Journal of Business Ethics, 120*(3), 335.

Capelle-Blancard, G., Crifo, P., Diaye, M. A., Scholtens, B., & Oueghlissi, R. (2016). Environmental, Social and Governance (ESG) performance and sovereign bond spreads: an empirical analysis of OECD countries. *Working paper*. Available at SSRN: abstract 2874262. Accessed in January 2018.

Carhart, M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52*(1), 57-82.

Carvalho, A., & Areal, N. (2016). Great places to work®: Resilience in times of crisis. *Human Resource Management, 55*(3), 479-498.

Cortez, M. C., Silva, F., & Areal, N. (2012). Socially responsible investing in the global market: The performance of US and European funds. *International Journal of Finance & Economics, 17*(3), 254-271.

DeMiguel, V., & Nogales, F. J. (2009). Portfolio selection with robust estimation. *Operations Research, 57*(3), 560-577.

Derwall, J., & Koedijk, K. (2009). Socially Responsible Fixed-Income Funds. *Journal of Business Finance and Accounting*, *36* (1), 210-229.

Derwall, J., Guenster, N., Bauer, R., & Koedijk, K. (2005). The Eco-Efficiency Premium Puzzle. *Financial Analysts Journal*, *61*(2), 51-63.

Derwall, J., Koedijk, K., & Ter Horst, J. (2011). A tale of values-driven and profit-seeking social investors. *Journal of Banking & Finance*, *35*(8), 2137-2147.

Drut, B. (2010). Sovereign Bonds and Socially Responsible Investment. *Journal of Business Ethics*, *92* (1), 131-145.

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*(11), 2835-2857.
Edmans, A. (2011). Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial Economics, 101*(3), 621-640.
Elton, E., Gruber, M., & Blake, C. (1995). Fundamental Economic Variables, Expected Returns, and Bond Fund Performance. *Journal of Finance*, *50* (4), 1229-1256.
European Sustainable Investment Forum – EUROSIF (2016). *European SRI Study 2016.* Accessed in July 2017, from www.eurosif.org.
Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56.
Fama, E. F., & French, K. R. (1998). Value versus growth: The international evidence. *The Journal of Finance, 53*(6), 1975-1999.
Fernandez-Izquierdo, A., & Matallin-Saez, J. C. (2008). Performance of ethical mutual funds in Spain: sacrifice or premium? *Journal of Business Ethics*, *81*(2), 247-260.
Galema, R., Plantinga, A., & Scholtens, B. (2008). The stocks at stake: Return and risk in socially responsible investment. *Journal of Banking & Finance, 32*(12), 2646-2654.
Gasbarro, D., Wong, W., & Kenton Zumwalt, J. (2007). Stochastic dominance analysis of iShares. *The European Journal of Finance, 13*(1), 89-101.
Global Sustainable Investment Alliance – GSIA. (2016). *Global sustainable investment review 2016.* Accessed in July 2017, available at www.gsi-alliance.org
Goldreyer, E., Ahmed, P., & Diltz, J. (1999). The Performance of Socially Responsible Mutual Funds: Incorporating Sociopolitical Information in Portfolio Selection. *Managerial Finance*, *25* (1), 23-36.
Gregory, A., & Whittaker, J. (2007). Performance and performance persistence of “ethical” unit trusts in the UK. *Journal of Business Finance and Accounting*, *34*(7–8), 1327–1344.
Halbritter, G., & Dorfleitner, G. (2015). The wages of social responsibility—where are they? A critical review of ESG investing. *Review of Financial Economics*, *26*, 25-35.
Hall, P. (1992). *The bootstrap and edgeworth expansion,* Springer, New York.
Henke, H. M. (2016). The effect of social screening on bond mutual fund performance. *Journal of Banking & Finance*, *67*, 69-84.
Hoepner, A. G., & Neher, A. L. (2013). Sovereign Debt and Sustainable Development Culture. *Working paper*. Available at SSRN: abstract 2295688. Accessed in January 2018.
Humphrey, J. E., Warren, G. J., & Boon, J. (2016). What is different about socially responsible funds? A holdings-based analysis. *Journal of Business Ethics, 138*(2), 263-277.
Javed, M., Rashid, M. A., & Hussain, G. (2016). When does it pay to be good–A contingency perspective on corporate social and financial performance: Would it work? *Journal of Cleaner Production, 133*, 1062-1073.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance, 48*(1), 65-91.
Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance, 36*(4), 889-908.
Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, *13*(5), 908-922.
Kreander, N., Gray, R. H., Power, D. M., & Sinclair, C. D. (2005). Evaluating the performance of ethical and non-ethical funds: A matched pair analysis. *Journal of Business Finance and Accounting*, *32*(7–8), 1465–1493.
Lahiri, S. (2003). *Resampling methods for dependent data*, Springer, New York.
Lean, H. H., & Nguyen, D. K. (2014). Policy uncertainty and performance characteristics of sustainable investments across regions around the global financial crisis. *Applied Financial Economics, 24*(21), 1367-1373.
Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance, 15*(5), 850-859.
Lee, J., Yen, P., & Chan, K. C. (2013). Market states and disposition effect: Evidence from Taiwan mutual fund investors. *Applied Economics, 45*(10), 1331-1342.
Leite, P., & Cortez, M. C. (2016). The Performance of European Socially Responsible Fixed-Income Funds. *Working paper*. Available at SSRN: abstract 2726094. Accessed in January 2018.
Loh, W. (1987). Calibrating confidence coefficients. *Journal of the American Statistical Association, 82*(397), 155-162.
Loughran, T., & Ritter, J. R. (2000). Uniformly least powerful tests of market efficiency. *Journal of Financial Economics, 55*(3), 361-389.
Lu, W., Chau, K., Wang, H., & Pan, W. (2014). A decade's debate on the nexus between corporate social and corporate financial performance: A critical review of empirical studies 2002–2011. *Journal of Cleaner Production, 79*, 195-206.
Managi, S., Okimoto, T., & Matsuda, A. (2012). Do socially responsible investment indexes outperform conventional indexes? *Applied Financial Economics, 22*(18), 1511-1527.
Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative science quarterly*, *48*(2), 268-305.
Margolis, J. D., Elfenbein, H. A., & Walsh, J. P. (2009). Does it pay to be good... and does it matter? A meta-analysis of the relationship between corporate social and financial performance. *Working paper*. Available at SSRN: abstract 1866371. Accessed in January 2018.
Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters, 1*, 21-23.
Mollet, J. C., & Ziegler, A. (2014). Socially responsible investing and stock performance: New empirical evidence for the US and European stock markets. *Review of Financial Economics, 23*(4), 208-216.
Mollet, J. C., von Arx, U., & Ilić, D. (2013). Strategic sustainability and financial performance: exploring abnormal returns. *Journal of business economics*, *83*(6), 577-604.

Nofsinger, J., & Varma, A. (2014). Socially responsible funds and market crises. *Journal of Banking & Finance, 48*, 180-193.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies, 24*(3), 403-441.

Osthoff, P. (2015). What matters to SRI investors? In T. Hebb, J. P. Hawley, A. G. Hoepner, A. L. Neher & D. Wood (Eds.), *The Routledge handbook of responsible investment*, 1st ed., 705-724. Routledge, New York.

Pagan, A. R., & Sossounov, K. A. (2003). A simple framework for analysing bull and bear markets. *Journal of Applied Econometrics, 18*(1), 23-46.

Politis, D. N., & Romano, J. P. (1992). A circular block-resampling procedure for stationary data. *Exploring the Limits of Bootstrap*, 263-270.

Preston, L. E., & O'bannon, D. P. (1997). The corporate social-financial performance relationship: A typology and analysis. *Business & Society*, *36*(4), 419-429.

Renneboog, L., Ter Horst, J., & Zhang, C. (2008). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance, 14*(3), 302-322.

Revelli, C., & Viviani, J. (2015). Financial performance of socially responsible investing (SRI): What have we learned? A meta-analysis. *Business Ethics: A European Review, 24*(2), 158-185.

Scholtens, B. (2005). Style and performance of Dutch social responsible investment funds. *Journal of Investing*, *14*(1), 63–72.

Schröder, M. (2007). Is there a difference? the performance characteristics of SRI equity indices. *Journal of Business Finance & Accounting, 34*(1-2), 331-348.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business, 39*(1), 119-138.

Statman, M. (2000). Socially responsible mutual funds. *Financial Analysts Journal, 56*(3), 30–39.

Statman, M. (2006). Socially responsible indexes. *The Journal of Portfolio Management, 32*(3), 100-109.

Statman, M., & Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, 65(4), 33-46.

Statman, M., & Glushkov, D. (2016). Classifying and measuring the performance of socially responsible mutual funds. *The Journal of Portfolio Management, 42*(2), 140-151.

Utz, S., & Wimmer, M. (2014). Are they any good at all? A financial and ethical analysis of socially responsible mutual funds. *Journal of Asset Management, 15*(1), 72-82.

Van de Velde, E. Vermeir, W. & Corten, F. (2005). Corporate social responsibility and financial performance. *Corporate Governance: The International Journal of Business in Society*, *5*(3), 129-138.

Wimmer, M. (2013). ESG-persistence in socially responsible mutual funds. *Journal of Management and Sustainability, 3*(1), 9-15.

Zaremba, A., & Czapkiewicz, A. (2017). The cross section of international government bond returns. *Economic Modelling*, *66*, 171-183.

Appendix 1

Environmental dimension: Environmental challenges pose a potential risk for investors, as environmental externalities can result in significant economic losses, while repairing environmental damage such as air and water pollution can generate considerable fiscal costs. Adequate investments towards preventing environmental problems limit such potential liabilities. Another important risk is related to the country's exposure to natural hazards such as floods, hurricanes or typhoons. In addition to evaluating a country's environmental vulnerabilities and policies, RobecoSAM also examines its energy dependency and energy policies. Countries that rely heavily on fossil fuel imports are vulnerable to abrupt and /or sharp external price movements or supply shortages. In addition to assessing the risks themselves, RobecoSAM specifically looks for evidence that policies for mitigating such risks have been put into place.

Social dimension: A weak social climate dominated by labor unrest, extreme inequality, or other social tensions is another potential investment risk. A delicate social climate can easily result in violent turmoil, disrupting important economic activity such as manufacturing or trade and/or paralyze policymaking. Strong social cohesion, on the other hand, supports orderly conflict resolution and facilitates the implementation of necessary reforms, thus contributing towards sustainable economic development.

Governance dimension: RobecoSAM looks at a broad range of data that takes into account a country's institutional framework, regulatory quality, rule of law, government efficiency, central bank independence and political stability, among other factors. Civil liberties, internal conflicts and corruption also reflect a country's governance profile. The corruption level, for instance, shows the extent to which public power is exercised to protect the interests of a small group at the expense of the economy and society at large. A recent study by Robeco demonstrates the added value of considering political risk when taking investment decisions for government bonds, over a time period of twenty five years.