The Impact of Socially Responsible Investing: What Can We Learn from Different Performance Measures?

Liu Wang[†] and Sylvia Maxfield[‡]

ABSTRACT

This study examines the impact of socially responsible investing on mutual fund performance. Using multiple performance measures, we find that different conceptualizations and measures of performance yield different empirical findings. In particular, ESG investing adds value when model-independent performance measures are utilized (e.g., pre-tax return, post-tax return, standard deviation, and the Sharpe ratio), while the role of ESG investing is not clear when model-dependent performance measures are considered (e.g., alpha, beta, the Treynor ratio, and the information ratio). The results are robust after controlling for other factors that may have a significant impact on fund performance, such as fund size, fund age, manager tenure, manager ownership, the fee structure, and the type of the fund. An interesting finding arising from this study is that the conflicting evidence on socially responsible investing may be, to a large extent, a result of various performance measures employed by different researchers. Because different quantifications of inputs may lead to different results, more robust and consistent measures of relative performance have to be developed and utilized.

Key Words: Socially Responsible Investment (SRI), Mutual Funds, Investment Performance, Performance Measures

[†] Corresponding author. Associate Professor of Finance, Providence College. Director of Research, Balanced Wealth Management. Address for correspondence: Department of Finance, School of Business, Providence College, Providence, RI 02918, USA. Tel: 401.865.1883. Email: lwang@providence.edu.

[‡] Professor of Management, Dean, School of Business, Providence College. Address for correspondence: School of Business, Providence College, Providence, RI 02918, USA. Tel: 401.865.1224. Email: maxfield@providence.edu.

1. INTRODUCTION

Over the past decade, the substantial rise of corporate social responsibility (CSR) practice has fueled research and practice on socially responsible investment (SRI), also called ethical investment or substantial investment. The main difference between SRI and conventional investments is that SRI screens assets based on a set of environmental, social, and governance (ESG) criteria.

Despite the growing popularity of socially responsible funds (SR funds) or ESG funds, evidence of their relative performance against conventional funds has been mixed. Some studies find that ESG investing results in inferior performance (e.g., Ferruz, Muñoz, & Vargas, 2012; Jones, van der Laan, Frost, & Loftus, 2008; Renneboog, Ter Horst, & Zhang, 2008a). Some studies find that ESG investing adds value (e.g., Derwall, Guenster, Bauer, & Koedijk, 2005; Kempf & Osthoff, 2007; Henke & Maehlmann, 2015; Nofsinger & Varma, 2014). Some studies fail to find any significant difference between SR funds and conventional funds (e.g., Bauer, Koedijk, & Otten, 2005; Bauer, Otten, & Tourani, 2006; Bello, 2005; Hamilton, Jo, & Statman, 1993; Kreander, Gray, Power, & Sinclair, 2005; Statman, 2000).

Among various potential reasons for the conflicting evidence, the classification of SR funds and the variation in measures used to capture fund performance may largely account for the mixed conclusions in the empirical investigation of SR funds. The main objective of this study, therefore, is to provide a more rigorous analysis of SRI performance. Building on the existing literature and attempting to ensure the empirical validity of our empirical assessment of SRI, we have devised a research design that does not rely on the classification of ESG and conventional funds. Unlike previous studies where a group of SR funds are compared to a group of conventional funds (or an index), we examine all mutual funds in the U.S. on a continuous basis, where the Morningstar

portfolio ESG score is utilized to measure the extent to which a portfolio (fund) is ESG-oriented. The second research design specification that is different from previous studies is that we use multiple performance measures to investigate the extent to which different conceptualizations and measures of performance yield different empirical findings. Moreover, our research design allows us to conduct a cross-sectional regression controlling for other factors that may have a significant impact on fund performance. The pairing approach used in most previous studies may not effectively control for various cross-sectional factors, such as fund size, fund age, manager tenure, manager ownership, the fee structure, the type of the fund, and other influential fund characteristics.

The remainder of the paper proceeds as follows. Section 2 discusses the literature and our testing hypotheses. Section 3 describes the data and methodology. Empirical results are presented in Section 4, followed by concluding remarks in Section 5.

2. LITERATURE REVIEW & HYPOTHESES

2.1 Literature Review

Scholars have been searching to discern the impact of applying social criteria to business for several decades. In the financial investments arena, launch of the KLD benchmark held out exciting prospects for systematic study. At its broadest level, traditional financial theory predicts that investors should pay a price for choosing to screen for social criteria because this limits gains from the most complete diversification (Adler & Kritzman, 2008). This argument builds on Rudd (1981) and Grossman and Sharpe (1986), who argue that imposing any constraints on the investment process will result in inferior portfolio performance. The most common finding in studies of the impact for investors across a wide variety of differences, including unit of analysis, geographic domain, data sources used, measures deployed, and model specification, is that despite

this conventional theory, no material difference exists between screened and traditional investment strategies (e.g., Bauer et al., 2005; Diltz, J., 1995; Goldreyer, Ahmed, & Diltz, 1999; Hamilton et al., 1993; Kreander et al., 2005; Managi et al., 2012; Shank et al., 2005; Statman, 2000).This broad characterization of the literature obscures a range of results. A significant number of studies find that applying ESG criteria yields better performance (e.g., Derwall et al., 2005; Kempf & Osthoff, 2007; Henke & Maehlmann, 2015; Nofsinger & Varma, 2014). Other studies confirm the conventional theory and show that applying ESG criteria results in inferior fund performance (e.g., Ferruz et al., 2012; Girard et al., 2007; Jones et al., 2008; Renneboog et al., 2008a).

Despite almost thirty years of research, differences in research design and method still hinder understanding the impact of ESG screening on investment performance. Rather than continuing to test whether ESG screening has negative, neutral, or positive impact, arguably the most needed work in this field would be to help researchers discern the impact of critical differences in research strategy. One promising trend in addressing this concern is an effort to isolate the impact of using different data sources for measuring the ESG content of any given investment portfolio. Several recent studies attempt to test for alignment across the major sources commonly used in the investment world. Unlike credit ratings, which are highly consistent across the main rating firms, ESG rating schemes face various challenges, such as the lack of standardized corporate disclosure and the use of data from third parties. For comprehensive coverage of environmental, economic, social, and governance performance, two other comprehensive and research-ready data repositories emerged to compete with KLD and are now consolidated under ownership of global data services firms. These were Asset4 and Sustainalytics. RiskMetrics bought KLD in 2009 and shortly following that MSCI purchased RiskMetrics. In 2010, Thompson Reuters bought Asset4. In 2017, Morningstar bought and rolled Sustainalytics into Morningstar's

Sustainability Rating. Until more recently the majority of studies used KLD data (van den Heuvel, 2012), although the other two are gaining in popularity in more recent research (Bouten et al., 2016). There is some evidence suggesting acceptable levels of comparability across these three major sources of ESG data is sufficient to assume that the choice of data source is not a critical determinant of divergent research findings (van den Heuvel, 2012).

Another important difference in studies with divergent findings is how one conceptualizes better performance. In the literature, one common practice is to use the strategy of matching funds categorized as "social impact" funds with "regular" funds and to compare the relative performance of these two different groups of investment vehicles. The matching strategy design (e.g., Bauer et al., 2005; Ferruz, et al., 2012; Hamilton et al., 1993; Kreander et al., 2005; Renneboog et al., 2008a; Statman, 2000) poses the challenge of identifying which entities fall into the social impact category. The challenge of classifying SR funds is evident in the inconsistent lists of SR funds used in previous studies. For example, Hamilton et al. (1993) use the Lipper's list, Bello (2005) and Statman (2000) use the Morningstar list, Benson and Humphrey (2008) use the Social Investment Forum (SIF) report, Derwall, Koedijk, and Ter Horst (2011) and Kempf and Osthoff (2007) use the KLD Research & Analytics, and Ferruz et al. (2012) use a combination of socialfunds.com, the SIF report, and the fund's prospectus. Another common method is to use funds or other entities that self-identify as being in the social impact category. However, inconsistent practices across those self-identifying funds would call into question the findings that the performance of such funds is not statistically significantly different from that of conventional funds. One promising strategy to overcome this challenge is to examine all funds on a continuous basis (Kreander et al. 2005). In this study, we have devised a research design that does not rely on the classification of the social funds and conventional funds.

In addition to the inconsistency in data sources and the classification of SR funds, the variation in measures used to capture fund performance may also help explain the mixed conclusions in the empirical investigation of ESG investing.

Prior to Treynor (1965), Sharpe (1966), and Jensen (1968), investment performance had been generally evaluated by comparing the realized portfolio return to the market index return, and little attention had been paid to the level of risk the investor or manager takes in producing that return. The introduction of the capital asset pricing model (CAPM) refined the performance evaluation practices by demonstrating the important role of risk played in driving investment performance. Building on the implications of the CAPM, a series of risk-adjusted measures have been developed. In particular, Treynor (1965) proposes that the performance of a portfolio should be evaluated using the ratio of realized excess return (over the risk-free rate) to its systematic risk (as measured by beta). Building upon his own groundbreaking work as one of the originators of the CAPM, Sharpe (1966) introduces the well-known reward-to-volatility ratio or the Sharpe ratio. Similar to the Treynor ratio, the Sharpe ratio also measures excess return (over the risk-free rate) per unit of risk, while standard deviation is utilized as the measure of risk to account for incomplete diversification. Along the same line, Jensen (1968) devises a method of evaluating portfolio performance using the abnormal return (i.e., alpha) over the required return based on the CAPM. While developed in the 1960s, the Treynor ratio, the Sharpe ratio, and Jensen's alpha are still the most prevalent risk-adjusted measures used by investment practitioners nowadays.

Since the introduction of the CAPM and risk-adjusted performance measures, academic researchers have been in command of a battery of performance measures and an expanding body of literature has accumulated. Most recent expansion is centered on the development of various asset pricing models in deriving alpha, among which the most famous ones are the Fama-French

three factor model (Fama & French, 1993) and the Carhart four factor model (Carhart, 1997). In the Fama-French three factor model, size premium (as measured by market capitalization) and valuation premium (as measured by the book-to-market ratio) are included in addition to the market risk premium used in the CAPM. The Carhart four factor model adds a momentum factor to the Fama-French model.

With the development of various performance measures, the major difficulty encountered by investment practitioners in attempting to evaluate fund performance has been the lack of a thorough understanding of different performance measures developed in the academic literature. To better illustrate whether, and to what extent, different conceptualizations and measures of performance yield different empirical findings, we clarify various performance measures into two distinct groups: 1) model-independent measures, and 2) model-dependent measures. Widely-used model-independent measures include pre-tax return, post-tax return, standard deviation (a measure of total risk), and the Sharpe ratio. Commonly-used model-dependent measures include alpha (may be derived from the CAPM or various multi-factor models), beta (a measure of systematic risk), the Treynor ratio, and the information ratio. The information ratio measures active return (alpha) per unit of tracking error (the standard deviation of alpha). It is important to note that a potential problem associated with model-dependent measures is that they are highly dependent on the model specifications and benchmarks used in the least-squares regressions deriving alpha and beta. In particular, alpha depends on two factors: 1) the accuracy of systematic risk, as measured by beta, and 2) the strength of the linear relationship between the fund and the index, as measured by the R-squared value. Generally speaking, a high R-squared value indicates more reliable alpha and beta figures, while a low R-squared value indicates that the fund's movements are not well

explained by the movements in its benchmark index. Therefore, model-dependent measures are not necessarily reliable if the R-squared value is low.

2.2 Hypothesis Development

Our first alternative hypothesis is developed on the conventional finance framework. As we know, SRI applies various screening processes based on a set of ESG criteria, such as avoiding businesses involved in alcohol, tobacco, gambling, and weapons. From the modern portfolio theory perspective, SRI screening inevitably imposes a constraint on the investment universe available to investors (or fund managers in this case). This may limit the diversification potential and, consequently, shift the efficient frontier towards less favorable risk-return tradeoffs in the mean-variance space. As such, it is extremely difficult, if not impossible, for SR funds to outperform conventional funds.

This reasoning underpins the following hypothesis:

H1a: ESG investing has a negative impact on fund performance, all else being equal.

On the other hand, screening assets based on ESG criteria may generate value-relevant information otherwise not available to investors, resulting in better financial performance. If this is the case, then SRI is not necessarily inconsistent with investors' value-maximization objective. Reasons supporting this alternative hypothesis are as follows. First, increased CSR standard and practice may help cultivate good corporate governance, which ultimately translates into favorable financial performance. Second, as Fombrun and Shanley (1990) point out, firms may use CSR as an information signal to influence stakeholders' assessments and to improve their reputational status. This may also translates into favorable financial performance. Moreover, social and environmental screening may reduce the possibility of incurring high costs that emerge during

corporate social crises or environmental disasters, which financial markets tend to undervalue (e.g., Renneboog et al., 2008a, b).

Based on the discussion above, the following alternative hypothesis is derived:

H1b: ESG investing has a positive impact on fund performance, all else being equal.

As previously discussed, among various potential explanations for the inconsistent evidence in the literature, the variation in measures used to capture fund performance may also account for the mixed conclusions in the empirical investigation of SR funds. Generally speaking, there are two distinct types of performance measures: 1) model-independent measures, such as pre-tax return, post-tax return, standard deviation, and the Sharpe ratio; 2) model-dependent measures, such as alpha, beta, the Treynor ratio, and the information ratio. The major difference is that model-dependent measures are highly dependent on the model specifications and benchmarks used in in the least-squares regressions deriving alpha and beta. In particular, alpha depends on two factors: 1) the accuracy of systematic risk, as measured by beta, and 2) the strength of the linear relationship between the fund and the index, as measured by R-squared. A low R-squared value indicates that the fund's movements are not well explained by the movements in its benchmark index. Therefore, model-dependent measures are not necessarily reliable if the R-squared value is low. Because different quantifications of inputs may lead to different results, close attention has to be paid to the measures used to capture fund performance.

The discussion above motivates the following hypothesis:

H2: The relationship between ESG investing and fund performance is sensitive to the performance measures used. This is especially the case for model-dependent measures.

3. DATA AND METHODOLOGY

3.1 Data & Main Variables

Our empirical investigation is based on a cross-sectional sample of all mutual funds in the United States in 2017. Our data are compiled from Morningstar and Thomson Reuters. This study uses cross-sectional data in 2017 because Morningstar started to report ESG scores for mutual funds in late 2016. U.S.-based mutual funds are chosen as the research focus of this study because the Morningstar ESG scores are not generally available for international funds. After eliminating funds with missing data on ESG scores, we are left with 14,507 mutual funds.

Unlike most previous studies where a group of SR funds are compared to a group of conventional funds or an index (e.g., Bauer et al., 2005; Ferruz, et al., 2012; Hamilton et al., 1993; Kreander et al., 2005; Renneboog et al., 2008a; Statman, 2000), we do not attempt to classify funds into two different groups. As Statman and Glushkov (2016) point out in their study, "it is impossible to draw clear boundaries between socially responsible and conventional funds, because no clear boundaries exist." Instead, funds are arrayed on a scale with different degree of ESG screening. The challenge of classifying SR funds is evident in the inconsistent classification means used in previous studies, including the Lipper's list, the Morningstar list, the SIF report, the KLD Research & Analytics, and fund prospectus (Bello, 2005; Benson & Humphrey, 2008; Derwall et al., 2011; Ferruz et al., 2012; Hamilton et al., 1993; Kempf & Osthoff, 2007; Statman, 2000).

Instead of trying to classify the funds into two different categories, this study examines all mutual funds in the U.S. on a continuous basis, where the Morningstar portfolio ESG score is utilized to measure the extent to which a portfolio (fund) is ESG-oriented. The Morningstar portfolio ESG score is an asset-weighted average of normalized company-level ESG scores from Sustainalytics, a leading provider of firm-level ESG research. In particular, Sustainalytics assesses a firm's performance on ESG issues relative to other firms in the same global industry peer group,

using a 0-100 scale. To make the ESG scores comparable across peer groups, Morningstar normalizes the scores of each peer group using a z-score transformation, i.e., $z_i = \frac{ESG_i - \mu_{PG}}{\sigma_{PG}}$. Here, ESG_i is the ESG score of an individual firm, μ_{PG} is mean ESG score of its peer group, and σ_{PG} is the standard deviation of the ESG scores of its peer group. The z-scores are used to create the normalized ESG scores on a 0-100 scale. Once the company ESG scores are normalized, they are aggregated to a portfolio ESG score using an asset-weighted average of all covered securities. Covered securities include equity and fixed-income securities issued by companies that have ESG scores. To receive a portfolio ESG score, at least 50% of a portfolio's assets under management must have firm-level ESG scores.[1]

Table 1 provides summary statistics of the sample. The fund-related variables are defined as follows: *ESG* is the portfolio ESG score. *SUS* is the portfolio sustainability score. *SIZE* is the size of the fund, calculated as the natural log of the net asset value of the fund. *AGE* is the age of the fund, which is calculated based on the fund's inception date. *TEN* is manager tenure, measured as the number of years that the current manager has been the portfolio manager of the fund. For funds with more than one managers, the average tenure is used. *TURN* is the annual turnover ratio of the fund, which indicates the percentage of the portfolio's holdings that have changed over the year. A low turnover ratio (20% to 30%) would indicate a buy-and-hold strategy, while a high turnover (more than 100%) would indicate an investment strategy involving considerable buying and selling of securities. *EXP* is the net expense ratio, measured as the percentage of fees deducted from total assets each fiscal year, including 12b-1 fees, management fees, administrative fees,

[1] For a more rigorous analysis, the Morningstar portfolio sustainability score is utilized as an alternative measure. The main results are unaffected. The Morningstar portfolio sustainability score is an asset-weighted average of the company ESG scores with deductions made for holdings involved in controversial incidents, as measured by the portfolio controversy score.

operating costs, and all other asset-based costs incurred by the fund. *PC* is the portfolio concentration ratio, measured as the percentage of fund assets in top 10 holdings. The portfolio concentration ratio measures the idiosyncratic (non-market) risk taken on by a fund, where a lower portfolio concentration ratio is better for investors seeking to avoid security-specific or sector-specific risks. *MO* is the percentage of firm assets with manager ownership.

The performance-related variables are defined as follows: *RET1* is the annualized pre-tax return. *RET2* is the annualized post-tax return. *SD* is the annualized standard deviation. *SR* is the Sharpe ratio, calculated as excess return (over the risk-free rate) divided by standard deviation. *Alpha* is the excess return, measured as the difference between a fund's actual return and its expected return based on its level of risk (beta). *Beta* is the sensitivity of the fund return to market movements. The lower the beta is, the lower the fund's systematic risk. Note that both alpha and beta are based on a least-squares regression of the excess return of the fund and the excess returns of the fund's benchmark index (i.e., the CAPM). *RSQ* is the R-squared value from the least-squares regression. It measures the strength of the linear relationship between the fund and the index, where high R-squared values indicate more reliable alpha and beta. *TR* is the Treynor ratio, calculated as excess return (over the risk-free rate) divided by market risk (beta). *IR* is the information ratio, measured as active return (alpha) divided by tracking error (the standard deviation of alpha).

[Insert Table 1 about here]

The correlation matrix is reported in Table 2. Based on the correlation test, both the ESG score and the sustainability score tend to be positively associated with firm size, portfolio

concentration, pre-tax and post-tax returns, the Sharpe ratio, and beta; they are negatively associated with manager tenue, the fund turnover ratio, the expense ratio, manager ownership, and standard deviation. While Table 2 provides some preliminary evidence of the relationships among key variables, such an analysis must be viewed cautiously, given that other cross-sectional factors are not taken into consideration in the correlation matrix.

[Insert Table 2 about here]

3.2 Multivariable Regressions

To better understand the impact of ESG investing on fund performance, a series of multivariable regressions are conducted. The dependent variable in the regressions are various performance measures, *PERM*, including pre-tax return, post-tax return, standard deviation, the Sharpe ratio, alpha, beta, the Treynor ratio, and the information ratio. The key independent variable in the models is the fund ESG score, *ESG*. The control variables include fund size, *SIZE*, fund age, *AGE*, manager tenue, *TEN*, the fund turnover ratio, *TURN*, the net expense ratio, *EXP*, portfolio concentration, *PC*, manager ownership, *MO*, the team management dummy, *TEAM*, and fund type, *TYPE*. Here, *TEAM* is a dummy variable, which takes the value of 1 if the fund is team managed and 0 otherwise. *TYPE* is the type of the fund. Based on Morningstar's global broad category group classification, there are six categories: equity, fixed income, alternative, allocation, convertibles, and tax preferred. All other variables are defined the same way as before. In particular, the following model is estimated:

$$PERM = \lambda_0 + \lambda_1 ESG + \lambda_2 SIZE + \lambda_3 AGE + \lambda_4 TENURE + \lambda_5 TURN + \lambda_6 EXP + \lambda_7 PC + \lambda_8 TEAM + \lambda_9 MO + \lambda_{10} TYPE + \varepsilon$$

To ensure robust model specifications, close attention has been paid to potential endogeneity and multicollinearity problems. First, unlike most of the performance-based studies, endogeneity is not a serious concern in this study, given that the ESG scores are compiled from the firm-level first, and then aggregated to derive portfolio ESG scores. Second, both a correlation test and VIF statistics are used to detect potential multicollinearity problems. As Table 2 indicates, except for one pair of variables (i.e., the ESG score and the sustainability score, our alternative SRI measures), all other explanatory variables have correlation coefficients of less than 0.26 and VIF statistics of less than 2.0. Therefore, the concern about multicollinearity among the independent variables does not appear to be warranted. Moreover, as an additional robustness check, we also repeated the regressions using the Morningstar portfolio sustainability score as an alternative measure. The main results are unaffected.

4. EMPIRICAL RESULTS

Table 3 reports the regression results regarding the impact of ESG investing on fund performance, where the dependent variables are pre-tax return, post-tax return, standard deviation, and the Sharpe ratio. The main finding from Table 3 is that ESG investing has a positive impact on fund performance when model-independent measures are utilized.[2] In particular, we find that funds with higher ESG scores tend to have higher absolute returns (both pre-tax and post-tax), lower risk (as measured by standard deviation), and higher risk-adjusted returns (as measured by the Sharpe ratio). In terms of control variables, we find that fund size tends to have a positive impact on fund performance, while fund age and manager tenue appear to have a negative impact on fund performance. In addition, we find that the turnover ratio of the fund and portfolio

[2] We consider the impact as "positive" if it leads to higher return, lower risk, and higher risk-adjusted return.

concentration tend to result in higher pre-tax and post-tax returns but also higher risk, while manager ownership is associated with lower pre-tax and post-tax returns but also lower risk.

[Insert Table 3 about here]

Table 4 reports the regression results regarding the impact of ESG investing on fund performance, where the dependent variables are alpha, beta, the Treynor ratio, and the information ratio. The main results from Table 4 is that ESG investing tends to result in higher systematic risk (as measured by beta), but its impact on relative performance is unclear, as indicated by the insignificant coefficient estimates on alpha and the Treynor ratio but the significant negative coefficient estimate on the information ratio. Regarding control variables, we find that larger funds are generally associated with higher alpha, higher beta, higher information ratio, and lower Treynor ratio. In addition, we find that older funds tend to have lower information ratio. Both manager tenure and manager ownership are associated with higher alpha and lower systematic risk, as measured by beta. The expense ratio appears to have a negative impact on alpha but a positive impact on the Treynor ratio. Portfolio concentration tends to result in higher information ratio but also higher systematic risk (beta).

[Insert Table 4 about here]

There are two possible explanations for the mixed results in Table 4. First, ESG investing fails to add value on a relative basis; it results in higher systematic risk and lower information ratio. Second, as previously discussed, a potential problem associated with model-dependent measures

is that they are highly reliant on the model specifications and benchmarks used in the least-squares regressions deriving alpha and beta. In particular, alpha depends on two factors: 1) the accuracy of systematic risk, as measured by beta, and 2) the strength of the linear relationship between the fund and the index, as measured by R-squared. A low R-squared value would indicate that the alpha and beta figures are not reliable.

To better understand the methodology used in deriving our alpha and beta inputs, further investigation is conducted on the benchmarks and model specifications used by Morningstar. Using the CAPM, the fund's alpha and beta statistics are derived from the least-squares regressions of the fund's excess returns (over the risk-free rate) on the excess returns of the fund's broad asset class index. In the regressions, the S&P 500 index is utilized as the benchmark for U.S. equity funds, the MSCI ACWI Ex USA index is utilized as the benchmark for international equity funds, the Bloomberg Barclays US Aggregate Bond TR USD index is utilized as the benchmark for fixed income funds, the Credit Suisse Mgd Futures Liquid TR USD index is utilized as the benchmark for alternative funds, the Morningstar Moderate Target Risk TR USD index is utilized as the benchmark for allocation funds, and the Morningstar Mod Tgt Risk TR USD is utilized as the benchmark for convertibles funds. As Table 1 indicates, the R-squared values from these least-squares regressions are generally low (41.5% on average). This suggests that the alpha and beta figures, as well as the Treynor ratio and the information ratio calculated based on alpha and beta, are not necessarily reliable. Therefore, the inconsistent results in Table 4 may be largely attributable to different model specifications and benchmarks applied in the least-squares regression deriving alpha and beta in the first place.

A review of the literature indicates that most previous studies use alpha as the measure of fund performance. The model specifications vary from the traditional CAPM (e.g., Hamilton et

al., 1993; Statman, 2000), to the Fama and French (1993) three-factor model, to the Carhart (1997) four-factor model (e.g., Bauer et al., 2005; Bauer et al., 2006; Ferruz, et al., 2012; Kempf & Osthoff, 2007; Renneboog et al., 2008a), to the recently developed six-factor model (e.g., Statman & Glushkov, 2016). The benchmark indices used in the models include the S&P 500 (e.g., Statman, 2000), the NYSE index (e.g., Hamilton et al., 1993), the CRSP value-weighted index (e.g., Kempf & Osthoff, 2007), the market indices supplied by Worldscope (e.g., Bauer et al., 2005; Bauer, et al., 2006; Renneboog et al., 2008a), and a combination of NYSE, AMEX, and NASDAQ indices (e.g., Ferruz, et al., 2012). As with our empirical results, the variation in model specifications and benchmarks employed by different researchers may, to a large extent, account for the mixed evidence in the literature.

5 CONCLUDING REMARKS

This study examines the impact of ESG investing on mutual fund performance in 2017. Using multiple performance measures, we find that different conceptualizations and measures of performance yield different empirical findings. In particular, ESG investing adds value when model-independent performance measures are utilized (e.g., pre-tax return, post-tax return, standard deviation, and the Sharpe ratio), while the role of ESG investing is not clear when model-dependent performance measures are considered (e.g., alpha, beta, the Treynor ratio, and the information ratio). The results are robust after controlling for other factors that may have a significant impact on fund performance, such as fund size, fund age, manager tenure, manager ownership, the fee structure, and the type of the fund.

This study adds to the literature in several ways. First, it offers a possible explanation for the mixed conclusions in the empirical investigation of SRI. In particular, we find that the

conflicting evidence may be, to a large extent, a result of various performance measures employed by different researchers. Because unwarranted quantifications of inputs may lead to misleading results, more robust and consistent measures of relative performance have to be developed and utilized. Second, unlike most previous studies where a group of SR funds are compared to a group of conventional funds (or an index), we have devised a research design that is not reliant on the classification of SR funds. Using the Morningstar portfolio ESG scores, this study examines all different types of mutual funds in the U.S. on a continuous basis. Moreover, our research design allows us to conduct a cross-sectional regression controlling for other factors that may have a significant impact on fund performance. As previously noted, the pairing approach used in most previous studies may not be able to effectively control for various cross-sectional factors, such as fund size, fund age, manager tenure, manager ownership, the fee structure, and the type of the fund, etc. This study also bridges this gap. In addition to its contributions to the academic literature, this study also offers some practical ideas for investors, managers, and policy makers to consider as they seek to undertake SRI or ESG investing.

As with other exploratory research, the present paper also points out some promising areas for future research. First, in addition to the classification of SR funds and the variation in performance measures, the contextual parameters of the research design, such as the timeframe and the country setting, may also account for the mixed conclusions in the empirical investigation of SR funds. Due to data limitations, the empirical investigation of this study is based on U.S.-based mutual funds in 2017. Future research may benefit from exploring other country settings and a longer time frame to determine the generalizability of our results. Second, as the empirical results of the study indicate, the conflicting evidence on SRI may be, to a large extent, a result of various performance measures employed by different researchers. Because different

quantifications of inputs may lead to different results, this study calls for a better understanding of performance measures used in portfolio evaluation in general and SRI analysis in particular.

REFERENCES

Adler, M., & Kritzman, T. (2008). The Cost of Socially Responsible Investing. *Journal of Portfolio Management*, 35(1), 52-56.

Bauer, R., Koedijk, K., & Otten, R. (2005). International Evidence on Ethical Mutual Fund Performance and Investment Style. *Journal of Banking & Finance*, 29, 1751-1767.

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

Bello, Z. (2005). Socially Responsible Investing and Portfolio Diversification. *Journal of Financial Research*, 27, 41-57.

Benson, K., & Humphrey, J. (2008). Socially Responsible Investment Funds: Investor Reaction to Current and Past Returns. *Journal of Banking and Finance*, 32, 1850-1859.

Bouten, L., Cho, C., Michelon, G., & Robins, R. (2016). Does the choice of CSR performance rating matter? http://thepublicinterest.ca/wp/wp-content/uploads/2017/10/bouten.pdf

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

Derwall, J. N., Guenster, N., Bauer, R., & Koedijk, K. (2005). The eco-efficiency premium puzzle. *Financial Analysts Journal*, 61(2), 53-64.

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

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.

Ferruz, L., Muñoz, F., & Vargas, M. (2012). Managerial Abilities: Evidence from Religious Mutual Fund Managers. *Journal of Business Ethics*, 105, 503-517.

Fombrun, C., & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of Management Journal*, 33 (2), 233-258.

Grossman, B., & Sharpe, W. (1986). Financial Implications of South Africa Divestment. *Financial Analysts Journal*, 42, (4), 15-29.

Goldreyer, E., Ahmed, P., & Diltz, J. (1999). The performance of socially responsible mutual funds: Incorporating sociopolitical information. *Journal of Financial Research*, 28, 41–57.

Hamilton, S., Jo, H., & Statman, M. (1993). Doing Well While Doing Good: The Investment Performance of Socially Responsible Mutual Funds. *Financial Analysts Journal*, 49, 62-66.

Henke, H., & Maehlmann, T. (2015). The Effect of Social Screening on Bond Mutual Fund Performance http://dx.doi.org/10.2139/ssrn.2429221

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), 1465-1493.

Jensen, M.C. (1968). The Performance of Mutual Funds in the Period 1945-1964. *Journal of Finance*, 23, 389-416.

Jones, S., van der Laan, S., Frost, G., & Loftus, J. (2008). The investment performance of socially responsible investment funds in Australia. *Journal of Business Ethics*, 80(2), 181–203.

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

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

Renneboog, L., Ter Horst, J. & Zhang, C. (2008b). Socially Responsible Investments: Institutional Aspects, Performance, and Investor Behavior. *Journal of Banking and Finance*, 32, 1723-1742.

Rudd, A. (1981). Social Responsibility and Portfolio Performance. *California Management Review*, 23(4), 55-69.

Sharpe, W. F. (1966). Mutual Fund Performance. *Journal of Business*, 39 (1), 119-138.

Statman, M. (2000). Socially Responsible Mutual Funds. *Financial Analysts Journal*, 56, 30-39.

Statman, M. & Glushkov, D. (2016). Classifying and Measuring the Performance of Socially Responsible Mutual Funds. *Journal of Portfolio Management*, 42 (2), 140-151.

Treynor, Jack L. (1965). How to Rate Management of Investment Funds. *Harvard Business Review*, 43, 63–75.

Van den Heuvel, R. (2016). How Robust are CSR Benchmarks, MA Thesis, Tilburg University.

Table 1. Summary Statistics

This table reports the summary statistics of the sample. The variables in the table are defined as follows: *ESG* is the portfolio ESG score; *SUST* is the portfolio sustainability score; *SIZE* is the size of the fund, calculated as the natural log of the net asset value of the fund; *AGE* is the age of the fund; *TEN* is the manager tenure; *TURN* is the annual turnover ratio of the fund; *EXP* is the net expense ratio; *PC* is the portfolio concentration ratio, measured as the percentage of fund assets in top 10 holdings; *MO* is the percentage of firm assets with manager ownership; *RET1* is the annualized pre-tax return; *RET2* is the annualized post-tax return; *SD* is the annualized standard deviation; *SR* is the Sharpe ratio, calculated as excess return (over the risk-free rate) divided by standard deviation; *Alpha* is the difference between a fund's actual return and its expected return based on its level of risk (beta); *Beta* is the sensitivity of the fund return to market movements; *RSQ* is the R-squared value from the least-squares regression that derives alpha and beta; *TR* is the Treynor ratio, calculated as excess return (over the risk-free rate) divided by market risk (beta); and *IR* is the information ratio, measured as active return (alpha) divided by tracking error (the standard deviation of alpha).

Variable	# Obs.	Mean	Std Dev	Minimum	Maximum
ESG	14507	50.169	5.365	35.060	65.680
SUST	14507	45.249	3.699	29.220	59.770
SIZE	14500	19.974	2.110	9.266	26.177
AGE	14507	11.741	9.238	1.000	94.000
TEN	14507	6.659	4.546	0.080	82.170
TURN	14347	0.714	1.037	0.000	25.680
EXP	14501	1.204	0.560	0.000	8.360
PC	14480	0.259	0.556	-28.323	1.221
MO	14263	0.740	0.275	0.000	1.000
RET1	13672	0.210	0.111	-0.208	1.204
RET2	13668	0.114	0.063	-0.132	0.598
SD	14269	0.057	0.029	0.003	0.535
SR	14269	3.470	1.781	-5.274	10.654
Alpha	14269	0.671	9.747	-82.846	88.138
Beta	14269	0.853	0.416	-3.701	3.880
RSQ	14269	0.415	0.271	0.000	1.000
TR	14269	0.223	5.585	-419.916	277.140
IR	14269	0.084	2.360	-56.239	6.655

Table 2. Correlation Matrix

This table reports the correlation coefficients of key variables. The variables in the table are defined as follows: The variables in the table are defined as follows: *ESG* is the portfolio ESG score; *SUST* is the portfolio sustainability score; *SIZE* is the size of the fund, calculated as the natural log of the net asset value of the fund; *AGE* is the age of the fund; *TEN* is the manager tenure; *TURN* is the annual turnover ratio of the fund; *EXP* is the net expense ratio; *PC* is the portfolio concentration ratio, measured as the percentage of fund assets in top 10 holdings; *MO* is the percentage of firm assets with manager ownership; *RET1* is the annualized pre-tax return; *RET2* is the annualized post-tax return; *SD* is the annualized standard deviation; *SR* is the Sharpe ratio, calculated as excess return (over the risk-free rate) divided by standard deviation; *Alpha* is the difference between a fund's actual return and its expected return based on its level of risk (beta); *Beta* is the sensitivity of the fund return to market movements; *RSQ* is the R-squared value from the least-squares regression that derives alpha and beta; *TR* is the Treynor ratio, calculated as excess return (over the risk-free rate) divided by market risk (beta); and *IR* is the information ratio, measured as active return (alpha) divided by tracking error (the standard deviation of alpha). Here * indicates $p < 0.10$, ** indicates $p < 0.05$, *** indicates $p < 0.01$.

	ESG	**SUST**	**SIZE**	**AGE**	**TEN**	**TURN**	**EXP**	**PC**	**MO**	**RET1**	**RET2**	**SD**	**SR**	**Alpha**	**Beta**	**RSQ**	**TR**	**IR**
ESG	1																	
SUST	0.89***	1																
SIZE	0.13***	0.08***	1															
AGE	0.04***	0.002	0.26***	1														
TEN	-0.05***	-0.04***	0.31***	0.26***	1													
TURN	-0.06***	-0.07***	-0.21***	-0.03***	-0.03***	1												
EXP	-0.15***	-0.11***	-0.26***	0.05***	0.04***	0.18***	1											
PC	0.03***	0.02**	-0.003	0.03***	0.04***	-0.001	0.05***	1										
MO	-0.02***	-0.02**	0.09***	0.005	0.12***	-0.09***	0.08***	0.007	1									
RET1	0.17***	0.21***	0.10***	0.03***	0.011	-0.06***	-0.04***	0.15***	-0.03***	1								
RET2	0.14***	0.19***	0.13***	0.01*	0.04***	-0.09***	-0.10***	0.13***	-0.03***	0.96***	1							
SD	-0.34***	-0.31***	-0.14***	0.02***	0.08***	0.10***	0.18***	0.33***	-0.07***	0.02*	-0.02*	1						
SR	0.36***	0.37***	0.20***	0.02***	-0.04***	-0.11***	-0.16***	-0.09***	0.03***	0.69***	0.68***	-0.50***	1					
Alpha	-0.013	0.02**	0.04***	-0.003	0.03***	0.001	-0.04***	-0.004	0.03***	0.55***	0.55***	-0.24***	0.54***	1				
Beta	0.19***	0.14***	0.090	0.05***	-0.02**	-0.07***	-0.07***	0.12***	-0.06***	0.22***	0.20***	0.14***	0.08***	-0.60***	1			
RSQ	0.48***	0.39***	0.21***	0.05***	-0.06***	-0.11***	-0.24***	-0.18***	-0.03***	0.22***	0.22***	-0.40***	0.45***	-0.21***	0.52***	1		
TR	-0.004	0.001	0.013	0.003	-0.001	-0.17***	-0.003	-0.006	0.009	-0.02*	-0.007	-0.03***	0.02**	-0.010	0.000	-0.001	1	
IR	-0.09***	-0.04***	-0.003	-0.03***	0.012	0.02*	-0.007	0.02**	0.07***	0.37***	0.37***	-0.13***	0.45***	0.69***	-0.43***	-0.31***	0.013	1

Table 3. Impact of ESG Investing on Fund Performance: Model-Independent Measures

This table reports the regression results regarding the impact of ESG investing on fund performance, as measured by pre-tax return, post-tax return, standard deviation, and the Sharpe ratio. The independent variables in the models are defined as follows: *ESG* is the portfolio ESG score; *SUST* is the portfolio sustainability score; *SIZE* is the size of the fund, calculated as the natural log of the net asset value of the fund; *AGE* is the age of the fund; *TEN* is the manager tenure; *TURN* is the annual turnover ratio of the fund; *EXP* is the net expense ratio; *PC* is the portfolio concentration ratio, measured as the percentage of fund assets in top 10 holdings; *MO* is the percentage of firm assets with manager ownership; *TEAM* is a dummy variable, which takes the value of 1 if the fund is team managed and 0 otherwise; and *TYPE* is the type of the fund. The t-values are in parentheses. Here * indicates $p < 0.10$, ** indicates $p < 0.05$, *** indicates $p < 0.01$.

	Pre-Tax Return	**Post-Tax Return**	**Standard Deviation**	**Sharpe Ratio**
Intercept	-0.2046*** -3.44	-0.1044*** -3.09	0.1601*** 11.62	-6.4601*** -6.95
ESG	0.0041*** 24.05	0.0019*** 19.54	-0.0016*** -41.75	0.1139*** 44.14
SIZE	0.0041*** 8.42	0.0025*** 9.00	-0.0009*** -7.95	0.0962*** 12.98
AGE	-0.0002** -2.03	-0.0003*** -4.45	0.0001*** 3.27	-0.0040*** -2.60
TEN	-0.0006*** -3.09	0.0001 0.79	0.0003*** 7.21	-0.0257*** -8.06
TURN	0.0038*** 4.24	0.0009* 1.80	0.0028*** 13.73	-0.0064 -0.46
EXP	0.0027 1.60	-0.0050*** -5.18	0.0034*** 8.79	-0.1004*** -3.85
PC	0.0327*** 5.92	0.0172*** 5.49	0.0446*** 35.28	-1.3942*** -16.35
TEAM	0.0153 1.32	0.0067 1.01	0.0051* 1.90	0.0515 0.29
MO	-0.0105*** -3.22	-0.0068*** -3.68	-0.0079*** -10.69	0.2317*** 4.64
TYPE	Controlled	Controlled	Controlled	Controlled
Adj R-Sq	0.1859	0.1849	0.3397	0.2195

Table 4. Impact of ESG Investing on Fund Performance: Model-Dependent Measures

This table reports the regression results regarding the impact of ESG investing on fund performance, as measured by alpha, beta, the Treynor ratio, and the information ratio. The independent variables in the models are defined as follows: *ESG* is the portfolio ESG score; *SUST* is the portfolio sustainability score; *SIZE* is the size of the fund, calculated as the natural log of the net asset value of the fund; *AGE* is the age of the fund; *TEN* is the manager tenure; *TURN* is the annual turnover ratio of the fund; *EXP* is the net expense ratio; *PC* is the portfolio concentration ratio, measured as the percentage of fund assets in top 10 holdings; *MO* is the percentage of firm assets with manager ownership; *TEAM* is a dummy variable, which takes the value of 1 if the fund is team managed and 0 otherwise; and *TYPE* is the type of the fund. The t-values are in parentheses. Here * indicates $p < 0.10$, ** indicates $p < 0.05$, *** indicates $p < 0.01$.

	Alpha	**Beta**	**Treynor Ratio**	**Information Ratio**
Intercept	2.7149 0.47	-0.0512 -0.22	2.3204 0.71	1.8787 1.39
ESG	-0.0195 -1.23	0.0125*** 19.68	-0.0092 -1.01	-0.0369*** -9.85
SIZE	0.18731*** 4.10	0.0097*** 5.29	-0.0464* -1.77	0.0374*** 3.47
AGE	-0.0068 -0.72	0.0006 1.53	0.0012 0.23	-0.0074*** -3.31
TEN	0.0521*** 2.65	-0.0047*** -6.02	0.0003 0.03	0.0015 0.31
TURN	0.1140 1.34	0.0055 1.62	-0.9811*** -20.1	0.0716*** 3.57
EXP	-0.6426*** -3.99	-0.0012 -0.19	0.2127** 2.31	-0.0554 -1.46
PC	0.3208 0.61	0.2299*** 10.95	-0.4463 -1.48	0.5493*** 4.43
TEAM	1.5903 1.43	-0.0370 -0.83	0.3051 0.48	0.1025 0.39
MO	1.2891*** 4.19	-0.1096*** -8.91	-0.1627 -0.92	0.6041*** 0.6041
TYPE	Controlled	Controlled	Controlled	Controlled
Adj R-Sq	0.0143	0.1318	0.0287	0.0530