Les Cahiers du GERAD ISSN: 0711–2440

Luck Versus Skill in the Cross-Section of Ethical Mutual Funds

M.A. Ayadi, H. Ben-Ameur
L. Kryzanowski

G–2011–07

February 2011

Luck Versus Skill in the Cross-Section of Ethical Mutual Funds

Mohamed A. Ayadi

Department of Finance, Operations, and Information Systems
Faculty of Business
Brock University
St. Catharines (Ontario) Canada
mayadi@brocku.ca

Hatem Ben-Ameur

GERAD & Department of Management Sciences
HEC Montréal
Montréal (Québec) Canada
hatem.ben-ameur@hec.ca

Lawrence Kryzanowski

Concordia University Research Chair in Finance
Department of Finance
Concordia University
Montréal (Québec) Canada
lawrence.kryzanowski@concordia.ca

February 2011

Les Cahiers du GERAD

G–2011–07

Abstract

The risk-adjusted selection and timing performance (alphas and gammas) of a comprehensive and survivorship-free sample of Canadian equity SRI funds after (before) management-related costs is negative (positive) and is sensitive to the choice of the return-generating process. It is not statistically different from that of non-SRI funds. Examination of funds in the tails of the performance distribution using the block-bootstrap method suggests that "bad luck" causes the underperformance of extreme left-tail funds and almost no fund possesses truly superior management and timing skills.

Résumé

Ce papier étudie la performance liée à la sélection des titres et à la synchronisation des mouvements de marché (alphas et gammas) d'un échantillon de fonds mutuels socialement responsables (SRI) au Canada. Les résultats montrent que la performance ajustée au risque de ce fonds après (avant) les coûts de gestion est négative (positive) et est sensible au choix du processus générateur des rendements. Elle n'est pas statistiquement différente de celle des fonds non-SRI. Examen des fonds dans les queues de la distribution des performances en utilisant la méthode des bootstrap en block suggère que la "malchance" explique la sous-performance extrême des fonds à gauche de la queue. Presque aucun fonds ne possède des qualités et compétences vraiment supérieures de sélection des titres et de synchronisation des mouvements de marché.

1 Introduction

Ethical or socially responsible investment (SRI) mutual funds are investment portfolios governed by ethical rules and social screens to select or exclude assets. Their growth in assets under management (AUM) and the number has been rapid over the past 20 years worldwide. Dollars invested in Canadian SRI funds has increased from 0.102 billion to 5.537 billion CDN from 1989 to 2008.[1] Currently 105 SRI funds account for almost 1.11% of total fund assets.

The rich and growing literature on SR investments and mutual funds has addressed several issues with the ultimate objective of assessing the impact of the social screens and SRI rules on the investment management process and its performance. The most important issue is the assessment of ethical fund performance relative to that of non-SRI funds based on gross and net returns. Studies in this research stream also examine the validity of the benchmark models and the potential role of SRI indices in the underlying return generating process of SRI funds.[2] For instance, Asmundson and Foerster (2001) report no statistically significant performance difference between four SRI and non-SRI Canadian funds. Based on standard and social index benchmarks, Bauer et al. (2007) find that the performance differential between eight SRI and non-SRI funds is neutral (not significant) for the period 1994-2003.

These neutral findings for Canadian SRI funds corroborate the results using samples of SRI funds that are international (Bauer et al., 2005), Spanish (Fernandez-Izquierdo and Matallin-Saez, 2007), German, Swiss and American (Schroder, 2004), and European (Cortez et al., 2009).[3] SRI fund underperformance is found by Bauer et al. (2006), by Jones et al. (2008) for Australian funds, Reeneboog et al. (2008a) for many European and Asia-Pacific funds,[4] and by Gil-Bazo et al. (2010) for both gross and net returns but not when the SRI funds are managed by non-specialized companies. SRI fund outperformance is reported by Luther et al. (1992) and Mallin et al. (1995) for UK ethical unit trusts. Ding and Wermers (2009) argue that manager characteristics (including experience, performance and track records) and governance quality (e.g., positive relationship with board size) are good indicators of stock selection and overall portfolio performance, and the replacement of underperforming managers in the case of board size. Gregory and Whittaker (2007) demonstrate that UK ethical fund performance depends on the presence of conditioning information in the benchmark model and find evidence of performance persistence over longer time horizons.

Most of these studies are essentially related to US, UK, and other small European SRI fund markets. Most fail to provide strong justification for the appropriateness of their benchmark models and conduct their tests using net returns on small samples of active SRI funds, while focusing on the selection abilities of these funds with little attention given to tests of timing performance. More importantly, these studies do not examine the performance inferences based on individual funds given the correlation structure of fund returns. Standard performance tests do not properly account for the presence of funds in the tails of the cross-sectional performance distribution (genuine positive or negative alphas) and hence may be misleading. Kosowski et al. (2006, 2007) use the bootstrap methodology to test the alpha significance of individual funds corresponding to various quantiles of the cross-section of estimated performance.[5]

To further extend the growing literature on SRI funds, this paper has two major objectives. The first is to provide extensive and robust evidence on the performance of Canadian SRI equity funds based on (un)conditional linear single- and multi-factor benchmark models for a relatively large sample consisting of

[1] There is a growth of 24% in total assets between 2006 and 2008, in concordance with the growth in the S&P/TSX index (Socially Responsible Investment Review, 2008). SRI mutual funds grew by 13% from $179 billion in 2005 to $201.8 billion in 2007 in the US (Social Investment Forum).

[2] Another important issue is related to the smart money effect with the relationship between SRI fund performance and money-flows. Reeneboog et al. (2006) find that past returns are a major determinant of ethical fund flows worldwide. Funds with increasing social screens tend to attract money inflows and have better future performance. Similarly, Benson and Humphrey (2008) report that flows to ethical funds are less sensitive to past and current returns than conventional funds.

[3] Derwall and Koedijk (2009) find mixed results. The average SRI bond fund performed similar to non-SRI funds, while the average SRI balanced fund outperformed its non-SRI peers by more than 1.3% per year.

[4] Renneboog et al. (2008b) find that a two standard deviation increase in SRI screening intensity generates 2.6% in abnormal returns per annum.

[5] The same method is used by Ayadi and Kryzanowski (2005, 2010) for samples of Canadian equity and fixed-income funds. Similarly, Cuthbertson et al. (2008) adopt the cross-section bootstrap for a large sample of UK unit trusts. More recently, Fama and French (2010) assess the presence of luck and skill in the performances of US equity funds based on the same methodology.

67 Canadian active and terminated SRI equity funds over the period 1988-2008. As in previous research on SRI funds, the frameworks used herein are suitable to perform evaluations of fixed-weight and dynamic strategies.[6] The models accommodate some unique features of ethical funds such as time-variation in their expected returns and risks and the availability of a Canadian ethical index. We also compare the performance statistics and risk measures of this category of funds to those of conventional funds at the individual fund and portfolios of funds levels. This addresses the following important question: Do Canadian SRI and non-SRI mutual funds emit different performances before and after the considerations of fees? As such, this paper appears to be the first to examine full conditional single- and multi-factor selection and timing benchmark models using a comprehensive sample of both surviving and terminated SRI funds over a long study period. While we refer to equity funds that have not declared that they are SRI as non-SRI, this should not be interpreted as these funds not considering SR factors when making investment decisions.

The second major objective of this paper is to study the performance of extreme SRI funds (i.e., in the tails of the performance distribution) using the cross-section bootstrap method and its variants. This is designed to successfully identify (un)skilled managers by accounting for individual fund cross-dependencies and to disentangle the effects of sampling variation or "luck" and superior/inferior management skills. This method can handle special distributional features of fund returns and risks such us asymmetries and fat tails. Our approach uses both net and gross fund returns (i.e., before management fees and other expenses) to assess the true ability of fund managers to generate minimum performance. In the same vein, we address the issue of whether SRI fund managers have the ability to time the market based on the forecasts of market returns.[7] This extends the findings of the few studies that use standard statistical tests. Various return-based timing measures are developed and estimated with and without adjusting for public information, and statistical inferences rely on the bootstrap methodology.

This paper's first major finding is that the performance of Canadian ethical fund managers after (before) costs or expenses is weak to neutral (positive), and is sensitive to the choice of return-generating process. Performance statistics and inferences improve with conditioning and when the benchmark becomes multifactor. No evidence of market-return timing performance is found using both net and gross returns.

The second major finding is that investment performance (selectivity and timing) does not differ between SRI and conventional funds for both gross and net returns. This implies that SRI funds are a legitimate investment alternative, especially for investors who integrate personal and societal values into their investment decisions.

The third major finding is that sampling variation or "bad" ("good") luck is the major cause of the poor (good) performance of extreme SRI funds based on a block bootstrap analysis of funds in the tails of selection and timing performance distributions. This result is mostly robust to the presence of conditional information, an alternative ranking scheme, and standard bootstrapping.

The remainder of the paper is organized as follows: In Section 2, a background on the socially responsible investment industry, with focus on Canadian SRI mutual funds, is provided. Section 3 presents the various benchmark models for stock picking and market timing and discusses the estimation methodology and the construction of the tests. In Section 4, the bootstrap methodology is fully explained for the two components of performance. In Section 5, the samples of funds and data used in the empirical tests reported herein are discussed. Section 6 presents and analyzes the main empirical findings. Section 7 concludes the paper.

2 Background

Socially responsible investments (SRIs) are based on ethical criteria as defined by an investor's moral code and implemented using investment screens (negative of firms or positive of industries) or a best of sector

[6] Examples include Chen and Knez (1996), Ferson and Schadt (1996), Kryzanowski et al. (1997), Christopherson et al. (1998), Ayadi and Kryzanowski (2005), Bauer et al. (2006), and Renneboog et al. (2007, 2008a). Conditioning is performed using information publicly available to uninformed investors, such as dividend yield, interest rates, and default and term structure variables.

[7] Grinblatt and Titman (1989) point out that the Jensen measure is biased when the fund and benchmark returns are jointly non-normal or are nonlinear. They propose the unbiased and robust Positive Period Weight Measure (PPW) as an alternative performance metric.

approach. The percentages of portfolios using each of the top five screens in parentheses in Canada in 2000 are tobacco (83%), environment (64%), alcohol (63%), military (62%) and employee's rights (50%) (Asmundson and Foerster, 2001).

The best of sector approach, where investments are made in companies considered best at addressing the considered social issues, is currently the most popular technique used in Canada. To illustrate its use, the best of sector approach would search for a mining company with the best social and environmental standards for its class and not avoid the mining sector. SRIs would then attempt to implement change in the investee towards being more socially responsible through shareholder activism. The number of SR resolutions has increased by 68% from 1999 to 2007. For the 990 funds and 57 shareholder proposals examined, SRI (non-SRI) funds gave a combined support of 79% (31%), and SRI funds voted against management 51% more often than did non-SRI funds on the selected proposals (Social Investment Organization, 2007). Bauer et al. (2007) identify the following as being important Canadian CSI issues for Canadians: environmental (e.g., carbon emissions management, biodiversity preservations and reducing atmospheric carbon through the use of the boreal forest), social concerns, fair treatment of employees, giving back to community and trustworthiness and transparency of organizations. As a result, SRI portfolios are tilting towards greater holdings of companies delivering sustainable technologies, such as alternative energy sources.[8] SRI proponents argue that SRI funds are higher quality funds because their investments are titled towards companies that are better managed, with responsible employment practices and safe and useful products. An online consumer survey conducted during July 2010 finds that two-thirds of Canadians report "that corporate reputation has a significant impact on which brands they choose", and that "negative reputations are more damaging than favourable ones are helpful" (Bensimon Byrne, 2010, page 5).

According to Bauer et al. (2007), SRI opponents argue that SRI criteria negatively affect risk-adjusted portfolio performance due to the restriction of the portfolio choice opportunity set, the associated costs embedded in the management expense ratio (MER) due to SRI screening, and that "irresponsible activities are more lucrative and recession proof". The average MER for equity-only Canadian SRI funds examined herein is 1.97%. This is considerably higher than the global benchmark of 1.29% for non-SRI equity funds, and is considerably lower than the MERs of 2.53% and 2.56% reported by Morningstar Canada for all-equity funds in 2004 and 2007, respectively. Other observers (e.g., Carrick, 2006) hold a neutral position where they argue SRI funds have similar performance, fees and investment strategies as non-SRI funds.

3 Benchmark Models and Empirical Methodology

3.1 Empirical Issues

Much of the previous research on performance measurement bases its performance statistics and inferences on individual funds and averages thereof. This approach produces unreliable and biased results since individual estimated alphas are most likely correlated (not independent as assumed) and average significance levels are without meaning. We first address this problem using a portfolio-based approach using equal- and total net asset value or size-weighted (E&SW) portfolios of SRI and non-SRI funds constructed using individual fund returns.[9]

The second approach uses individual funds[10] and the bootstrap sampling methodology. This robust method is an alternative for dealing with possible fund return nonlinearity and spatial correlation which most likely results in performances and test statistics being dependent across individual funds. This method is used for performance evaluation by Ayadi and Kryzanowski (2005, 2010), Kosowski et al. (2006, 2007), Huij and Derwall (2008), and Cuthbertson et al. (2008). Bootstrapped statistics are constructed for all and

[8] Daryl-Lynn, Carlson. Feel good, make money, National Post, February 02, 2008.

[9] Equal- and size-weighted portfolios of funds provide evidence on potential size effects in performance. Such portfolios can be interpreted as funds-of-funds, since they represent diversified investments that do not suffer from the most common criticism of funds-of-funds that they add an extra layer of costs. Other constructions could be based on specific screens or investment themes.

[10] Our analysis of individual funds is conducted on all funds with at least 30 monthly observations given the increasing dimension of some of the conditional versions of the multi-factor models.

specific cross-sections of the individual SRI funds for samples of 2000 where funds are ranked according to their estimated performances and their t-statistics.[11] Two resampling schemes are adopted, residuals only and independently of both the residuals and the factors.[12] A detailed description of the bootstrap approach with all data stages is provided in Section 5.[13]

An alpha t-statistic is closely related to the information (Rosenberg, 1976) or appraisal (Ferguson, 1980) ratio that is commonly used to assess selection returns (also, see Treynor and Black, 1973). Unlike the t-statistics, the information ratio is obtained by dividing the alpha component of total returns by the standard deviation (and not the standard error) of these excess alpha returns.

3.2 Benchmark Models

The performance of our sample of SRI and non-SRI funds is examined using various benchmark models with single- and multi-factor structures that integrate the role of conditioning information. Each model's risk-adjusted performance represents the non-systematic component of fund returns which is not replicated by the appropriate systematic factors or indices.[14] Several performance timing models are also developed and employed.

The performance of SRI funds in several markets and for different time frames and sample compositions are examined using both unconditional (Kreander et al., 2005; Bauer et al., 2005; Geczy et al., 2005; Bollen, 2007; Jones et al., 2008) and conditional models (Schroder, 2004; Bauer et al., 2006, 2007; Renneboog et al., 2007, 2008a; Cortez et al., 2009).

3.2.1 Models for Measuring Selection Performance

The traditional one-factor CAPM is widely used as the benchmark model to measure risk-adjusted portfolio performance (e.g., Jensen, 1968). The assumption that the systematic risk of the portfolio is stationary over the evaluation period is not tenable when the portfolio manager is timing the market by adjusting her exposure to the movements in the market return (Grinblatt and Titman, 1989) or when the portfolio manager uses derivatives securities that alter the characteristics or the return distribution of the portfolio under management (Dybvig and Ross, 1985).

The unconditional setting for performance measurement can be augmented to include time-variation in betas as in Ferson and Schadt (1996) and time-variation in both alphas and betas as in Christopherson et al. (1998). In these conditional benchmark models, we test if private information or signals contain useful information beyond that available publicly and whether or not this information has been used to add value by the fund manager.

Let x_{t-1} be a vector of pre-determined information or conditioning variables with zero means. When the beta(s) of the fund vary over time with a linear relation to the information variables, the conditional single factor model with time-varying alpha and betas is expressed as

$$r_{i,t} = \alpha_{i0} + \alpha_i' x_{t-1} + b_{i0} r_{M,t} + b_{i1}'(x_{t-1} r_{M,t}) + u_{i,t}, \tag{1}$$

where α_{i0} is the conditional risk-adjusted performance, α_i' is the vector of slope coefficients that measures the response of the conditional alpha to movements in the innovations in the conditioning variables, b_{i0} is the unconditional mean of the conditional beta, and b_{i1}' is the vector of slope coefficients that measures the response of the conditional beta to movements in the innovations in the conditioning variables. This model is an extended unconditional multi-factor model where the additional factors are the products of the excess returns on the market portfolio and the lagged information variables.

[11] Results for t-statistics-based bootstrapping are only reported since it adjusts for high-risk taking funds and produces better test sizes than the results that rely on rankings of the estimated coefficients (Hall and LePage, 1996).

[12] Independent resampling is sufficient to ensure the required independence assumption between errors and regressors.

[13] Our analysis is extended to bootstrapping blocks of overlapping observations to allow for some persistence in fund returns (Künsch, 1989).

[14] Some papers develop a style-adjusted performance measures for SRI funds (see for example Fernandez-Izquierdo and Mattalan-Saez, 2007).

We also estimate risk-adjusted performance using the four-factor model of Carhart (1997), which adds a momentum term to the three-factor model of Fama and French (1993), in order to improve the average pricing errors implied by the single factor model. The full conditional specification of the Carhart model is expressed as:

$$\begin{aligned} r_{i,t} =& \alpha_{i0} + \alpha_i' x_{t-1} + b_{i01} r_{M,t} + b_{i02}\mathrm{SMB}_t + b_{i03}\mathrm{HML}_t + b_{i04}\mathrm{MOM}_t + b_{i1}'(x_{t-1} r_{M,t}) + b_{i2}'(x_{t-1}\mathrm{SMB}_t) \\ &+ b_{i3}'(x_{t-1}\mathrm{HML}_t) + b_{i4}'(x_{t-1}\mathrm{MOM}_t) + u_{i,t}, \end{aligned} \quad (2)$$

where $r_{M,t}$ is the excess return on the benchmark portfolio M between t-1 and t, SMB_t (small minus big) is the mimicking portfolio return for the size factor (difference in returns across small and big stock portfolios controlling for the same weighted average book-to-market) and HML_t (high minus low) is the mimicking portfolio return for the book-to-market factor (difference in returns between high and low book-to-market equity portfolios), MOM_t is the mimicking portfolio return for the momentum factor (difference in returns of two equally-weighted portfolios of firms, one in the highest 30% eleven-month return and the other in the lowest 30% eleven-month return, both lagged one month);[15] and b_{i01}, b_{i02}, b_{i03}, and b_{i04} are the sensitivities or betas of the fund's excess returns to the market, size, book-to-market, and momentum factors, respectively, and $u_{i,t}$ is the random error of fund i in month t. In all regression models, a positive (negative) and statistically significant intercept or alpha is interpreted as evidence of superior (inferior) performance.

3.2.2 Market Timing Benchmark Models

The performance of a SRI fund can be decomposed into timing and selectivity ability since Grinblatt and Titman (1989), among others, shows that Jensen's alpha is statistically biased in the presence of timing skill. Treynor and Mazuy (1966) demonstrate that the relation between the excess returns of the portfolio and the market become nonlinear when the portfolio manager times the market. The unconditional specification of their model requires that stock returns not be co-skewed with the benchmark return. Timing benchmark models are used by Schroder (2004) for small samples of US, German, and Swiss SRI funds, and by Renneboog et al. (2008a) for SRI funds from 17 countries.

Our two timing models accommodate multi-factor benchmarks as in Lehman and Modest (1987) and conditioning information as in Ferson and Schadt (1996), and are given by:

$$r_{i,t} = \alpha_{i0} + \alpha_i' x_{t-1} + b_{i0} r_{M,t} + b_i'(x_{t-1} r_{M,t}) + \gamma_i r_{M,t}^2 + u_{i,t}, \quad (3)$$

$$r_{i,t} = \alpha_{i0} + \alpha_i' x_{t-1} + b_{i0} r_{M,t} + b_i'(x_{t-1} r_{M,t}) + \beta_{i2}\mathrm{SMB}_t + \beta_{i3}\mathrm{HML}_t + \beta_{i4}\mathrm{MOM}_t + \gamma_i r_{M.t}^2 + u_{i,t}, \quad (4)$$

where γ_i is the market timing coefficient, and all other terms are as defined earlier. Positive alpha and gamma values indicate that the manager has superior selection and timing skills, respectively.

4 Bootstrap Analysis

4.1 Performance Measures

The bootstrap resampling approach is conducted because it accounts for possible violation of the normality assumption underlying the performance tests and inferences,[16] it accommodates possible nonlinearities in fund returns and does not require the estimation of the complex joint distribution of performance across all mutual funds, and it can deal with time-series dependencies in the data through extensions of the basic setup. The implementation of this nonparametric approach follows the lines of Ayadi and Kryzanowksi (2005, 2010)

[15] See Ayadi et al. (2010) for more details on the construction of the size, book-to-market, and momentum factors in Canada.

[16] Untabulated results on the fund return residuals using all benchmark models show that the null hypothesis of normally distributed residuals is consistently rejected (based on the Jarque-Bera test) for 71% of the SRI funds. Furthermore, additional tests (Breusch-Pagan test for heteroskedasticity and Ljung-Box test for serial correlation) reveal that fund return residuals are often heteroskedastic essentially with unconditional models and that they are serially correlated for more than 55% of all funds across all benchmark models.

and Kosowski et al. (2006, 2007) and is based on the four-factor performance model (2) given above.[17] The steps follow.

First, we run a time-series regression for each fund i, and save all estimated coefficients and the alpha t-statistic (using the Newey and West adjustment for standard errors) $\{\hat{\alpha}_i, \hat{t}_{\alpha_i}, \hat{\beta}_{ik},\ k = 1, ..., 4\}$ as well as the time-series of estimated residuals $\{\hat{u}_{i,t},\ t = 1, ..., T_i\}$.

Second, we independently resample (with replacement) B times ($b = 1,\ 2, ..., B = 2000$ *herein*)the saved fund's residuals from the first step and the four factors. We then generate a time-series of resampled residuals (from a zero-mean noise) $\{\hat{u}^b_{i,t},\ t = \nu^b_1, \nu^b_2, ..., \nu^b_{T_i}\}$ and resampled factors $\{r^b_{M,t},\ \text{SMB}^b_t, \text{HML}^b_t, \text{MOM}^b_t, t = \tau^b_1, \tau^b_2, ..., \tau^b_{T_i}\}$ where $t = \nu^b_1, \nu^b_2, ..., \nu^b_{T_i}$ and $t = \tau^b_1, \tau^b_2, ..., \tau^b_{T_i}$ are the independent time reordering in the bootstrap experiment for the residuals and the factors, respectively. In both cases, we have the same sample size as in the original data for each fund.

Third, we construct time-series of the monthly excess returns for fund i by imposing null true performance ($\alpha_i = 0$) for each bootstrap iteration b:

$$r^b_{i,t} = \hat{\beta}_{i1} r^b_{M,t_I} + \hat{\beta}_{i2}\text{SMB}^b_{t_I} + \hat{\beta}_{i3}\text{HML}^b_{t_I} + \hat{\beta}_{i4}\text{MOM}^b_{t_I} + \hat{u}^b_{i,t_u},\ \ t_I = \tau^b_1, \tau^b_2, ..., \tau^b_{T_i}, \quad t_u = \nu^b_1, \nu^b_2, ..., \nu^b_{T_i}.$$

By construction, the resulting artificial or hypothetical time-series of fund excess returns should produce zero performance (equivalently $t_{\alpha_i} = 0$) using the original benchmark regression model. Any positive or negative estimated alpha is entirely due to sampling variation.

These steps are repeated for all funds $i = 1, ..., N$ and for all bootstrap iterations ($b = 1,\ 2,\ ...,\ 2000$) to obtain cross-sectional distributions of the alpha estimates ($\hat{\alpha}^b_i,\ i = 1, ..., N$) and the corresponding t-statistics ($\hat{t}^b_{\alpha_i},\ i = 1, ..., N$).

For a given bootstrap iteration b, we obtain a cross-sectional distribution of the alpha estimates ($\hat{\alpha}^b_1,\ \hat{\alpha}^b_2, ...,$ $\hat{\alpha}^b_N$) and of the t-statistics of these estimates ($\hat{t}^b_{\alpha_1}, \hat{t}^b_{\alpha_2}, ..., \hat{t}^b_{\alpha_N}$) that can be both ranked from the minimum or worst value ($\hat{\alpha}^b_{\min}; \hat{t}^b_{\min}$) to the maximum or best value ($\hat{\alpha}^b_{\max}; \hat{t}^b_{\max}$). This step is performed for all iterations ($b = 1,\ 2,\ ...,\ 2000$) to obtain cross-sectional distributions of all ranked funds including the best and worst funds as well as those in the 3%, 5%, and 10% percentiles in the left and right tails of the distribution. The alternative ranking of performance rests on the t-statistic. The latter is a pivotal statistic and does not depend on unknown parameters and leads to higher coverage probabilities for confidence intervals and more accurate bootstrap estimates (Horowitz, 2001).

Finally, the bootstrapped p-values are obtained by comparing the originally ranked performance estimates (or the t-statistics) with the corresponding ranked performance estimates (or t-statistics).

All of these steps are easily extended to block bootstrapping by dividing the sample into T_i/ℓ blocks (the smallest integer greater or equal to T_i/ℓ) with $\ell = 3$ overlapping monthly observations. The resampled blocks from the residuals and regressors are used to construct the bootstrapped time-series dynamics $r^{\ell,b}_{i,t}$. This approach helps to preserve the dependence structure within blocks.

4.2 Market Timing Measures

The same bootstrap framework is adopted for the extended timing specifications given earlier by (3) to test the null hypothesis that the SRI fund manager has no timing ability.

First, we run a time-series regression for each fund i, and save all estimated coefficients and the t-statistics of the alpha and gamma estimates (using the Newey and West adjustment for standard errors) $\{\hat{\alpha}_i, \hat{t}_{\alpha_i}, \hat{\gamma}_i, \hat{t}_{\gamma_i}, \hat{\beta}_{ik},\ k = 1, ..., 4\}$ as well as the time-series of estimated residuals $\{\hat{u}_{i,t},\ t = 1, ..., T_i\}$.

Second, we independently resample (with replacement) B times ($b = 1,\ 2, ..., B$) the saved fund's residuals from the first step and the four factors. We then generate a time-series of resampled residuals (from a

[17] The discussion here illustrates the four-factor benchmark-based performance evaluation but the bootstrap method has been applied to all of our unconditional and conditional benchmark models. It also presents the case with independent resampling of residuals and factors.

zero-mean noise) $\{\hat{u}^b_{i,t},\ t = \nu^b_1, \nu^b_2, ..., \nu^b_{T_i}\}$ and resampled factors $\{r^b_{M,t},\ \text{SMB}^b_t, \text{HML}^b_t, \text{MOM}^b_t, (r^2_{M,t})^b,\ t = \tau^b_1, \tau^b_2, ..., \tau^b_{T_i}\}$ where $t = \nu^b_1, \nu^b_2, ..., \nu^b_{T_i}$ and $t = \tau^b_1, \tau^b_2, ..., \tau^b_{T_i}$ are the independent time reordering in the bootstrap experiment for the residuals and the factors, respectively. In both cases, we have the same sample size B = 2000 as in the original data for each fund.

Third, we construct time-series of the monthly excess returns for fund i for each bootstrap iteration b by imposing the hull hypothesis that the fund has neither stock selection nor market timing ability ($\alpha_i = \gamma_i = 0$). By construction, the resulting artificial or hypothetical time-series of fund excess returns should produce zero stock selection and market timing performance (equivalently $t_{\alpha_i} = 0$ and $t_{\gamma_i} = 0$) using the original market timing regression model. Any positive or negative estimated alpha or gamma is entirely due to sampling variation.

These steps are repeated for all funds $i = 1, ..., N$ and for all bootstrap iterations ($b = 1, 2, ..., 2000$) to obtain cross-sectional distributions of the alpha or gamma estimates ($\hat{\alpha}^b_i,\ \hat{\gamma}^b_i, i = 1, ..., N$) and the corresponding t-statistics ($\hat{t}^b_{\alpha_i}, \hat{t}^b_{\gamma_i}\ , i = 1, ..., N$).

For a given bootstrap iteration b, we obtain a cross-sectional distribution of the alpha or gamma estimates ($\hat{\alpha}^b_1,\ \hat{\alpha}^b_2, ..., \hat{\alpha}^b_N$) and ($\hat{\gamma}^b_1,\ \hat{\gamma}^b_2, ..., \hat{\gamma}^b_N$), and of the t-statistics of these estimates ($\hat{t}^b_{\alpha_1}, \hat{t}^b_{\alpha_2}, ..., \hat{t}^b_{\alpha_N}$) and ($\hat{t}^b_{\gamma_1}, \hat{t}^b_{\gamma_2}, ..., \hat{t}^b_{\gamma_N}$) that can be both ranked from the minimum to the maximum value. This step is performed for all iterations ($b = 1,\ 2,\ ...,\ 2000$) to obtain cross-sectional distributions of all ranked funds.

Finally, the bootstrapped p-values are obtained by comparing the originally ranked market timing performance estimates (or the t-statistics) with the corresponding ranked original performance estimates (or t-statistics). The bootstrap p-value for fund i is used when the ranking is based on the t-statistic and given by the same formula as in the previous section. The extension to block bootstrapping is easily conducted as in the previous section.

5 Data Sources and Sample Description

5.1 Mutual Fund Returns

Two different samples of Canadian SRI and non-SRI equity mutual funds are carefully constructed by adjusting for mergers and name changes over the period from February 1988 through April 2008 using information from the *Fundata* database augmented by industry and individual fund reports from the *SEDAR* database, and specific fund news in the financial press. The first sample consists of 67 self-declared SRI funds (57 active and 10 terminated) representing the entire industry. The second sample includes 517 Canadian non-SRI equity funds with 340 active and 177 terminated portfolios. Since non-surviving funds are considered until their termination, survivorship bias has a minimal impact on our analysis. In both samples, the number of observations varies across funds and ranges from one to 243. Monthly fund returns are given by the monthly changes in the net asset values per share (NAVPS), and are adjusted for all distributions. Fund size is proxied by total net asset (TNA) value. Only equity funds are examined given their relatively long return histories and to facilitate comparisons with previous studies.

Panel A of Table 1 reports summary statistics on the cross-sectional distribution for each fund sample with at least 12 monthly observations. The average annual SRI (non-SRI) fund returns over the study period vary from -22.82% (-18.95%) for the Real Assets Social Leaders fund (Keystone Altamira Capital Growth fund) to 21.48% (32.77%) for the Social Housing Canadian equity fund (Capstone Canadian equity fund), and have a cross-sectional mean of 7.39% (7.77%). These figures suggest that non-SRI funds outperform their SRI counterparts. The SRI fund annual volatilities (standard deviations) range from 2.80% for the Ethical Advantage fund to 23.30% for the Real Assets Social Leaders fund. A larger dispersion is observed with non-SRI funds where the lowest and highest volatilities are 0.25% and 34.80%, respectively. The cross-sectional volatility of the SRI fund group is 10.05% which is lower than that of the non-SRI fund group of 13.09%. These statistics indicate that SRI funds are less risky than non-SRI funds. The annual average mean and volatility of the returns for the TSX index are 10.89%% and 14.02%, respectively, over the same time period.

Table 1: Summary statistics for the returns of SRI and conventional equity funds

This table reports summary statistics for the returns (in %) of individual and portfolios of Canadian (surviving and non-surviving) SRI and non-SRI equity funds using monthly data from February 1988 through April 2008. The prefixes EW and SW refer to equal- and size- (or total asset value-) weighted portfolios of funds, respectively. Panel A provides the statistics on the distribution of various return parameter estimates for two cross-sections of SRI and non-SRI equity funds. N is the number of all funds over the study period. Panel A reports various statistics for individual fund returns of surviving and non surviving mutual funds for both groups with at least 12 monthly observations. Panel B reports some descriptive statistics and tests on the returns of EW and SW portfolios of funds for the samples of SRI and non-SRI funds with at least (1) monthly observation. Monthly data are from February 1988 to April 2008, which correspond to a maximum of 243 observations.

Panel A: Individual mutual fund returns

Fund group	Statistics	Mean	Median	Std. Dev.	Minimum	Maximum	Skew.	Kurt.
SRI funds(N=53)	Mean	0.6160	0.8449	2.9011	-7.5684	6.7383	-0.499	0.909
	Std. Dev.	0.5161	0.7856	1.2988	4.9848	4.8704	0.500	2.209
	Minimum	-1.9014	-2.7971	0.8090	-27.4018	1.9312	-2.978	-0.982
	Q1	0.4123	0.3568	1.9190	-8.4393	3.7156	-0.684	-0.208
	Median	0.6467	0.8868	2.9339	-6.2732	5.9536	-0.490	0.246
	Q3	0.9214	1.3803	3.6238	-4.0696	7.3143	-0.226	0.875
	Maximum	1.7903	2.6332	6.7270	-1.4098	29.0682	0.682	12.151
Conventional funds(N=469)	Mean	0.6424	0.9727	3.7790	-11.2650	9.1122	-0.507	1.242
	Std. Dev.	0.5302	0.6289	1.0395	5.8813	4.5482	0.511	1.917
	Minimum	-1.5791	-2.2642	0.0732	-29.2906	0.6221	-1.911	-1.193
	Q1	0.4302	0.6637	3.1035	-16.9780	6.1046	-0.794	-0.153
	Median	0.7240	1.0452	3.6764	-8.4936	8.3024	-0.572	0.405
	Q3	0.9135	1.3248	4.2231	-6.7308	10.6431	-0.283	2.312
	Maximum	2.7312	2.9535	10.0448	0.2397	39.2593	3.089	12.433

Panel B: Portfolios of funds

Portfolios of funds		Mean	Median	Std. Dev.
Ethical funds, N=67	Ret. EW	0.762	0.977	3.009
	Ret. SW	0.724	0.926	3.356
Conventional funds, N=517	Ret. EW	0.741	1.116	3.346
	Ret. SW	0.78	1.115	3.535
Mean equality test EW (SW)		p-val = 0.94 (0.86)		
Median equality nonparametric test EW (SW)		p-val = 0.93 (0.93)		
Levene's test for equality of variances EW (SW)		p-val = 0.10 (0.45)		

Summary statistics for equal- and size-weighted (henceforth E&SW) portfolios of SRI and non-SRI funds are reported in panel B of Table 1. The SW non-SRI and SRI portfolios exhibit the highest and the lowest unconditional mean returns of 9.36% and 8.695% per annum, respectively. The SW non-SRI portfolio and the EW SRI portfolio have the highest and lowest unconditional volatilities of respectively 12.25% and 10.42% per annum. Formal tests on the equality of the mean returns for the paired EW and paired SW portfolios yield p-values of 0.94 and 0.86, respectively. Hence, the null hypothesis, that the average returns for these two sets of same-weighted portfolios are equal, cannot be rejected. This result is robust to a nonparametric test of their medians. A test of the equality of the portfolio variances based on the Levene statistic finds that the null hypothesis can be rejected at the 10% level only when comparing the volatilities of the paired EW portfolios.

5.2 Benchmark Variables, Risk Factors, and Information Variables

The analysis of equity funds requires the use of equity indices and risk factors consistent with the investment strategies of these portfolios. We retain the value-weighted TSX index of all Canadian stocks as the first benchmark variable and proxy for the market portfolio. An alternative reference point for the performance of ethical funds is the Jantzi social index (JSI)[18] since fund managers of these SR investments have a smaller investment opportunity set. The JSI has underperformed the S&P/TSX composite index over the available period of June 2001 to April 2008 (8.44% versus 10.73%) with a slightly higher volatility of 11.91% compared to 11.80% for the aggregate stock index. The JSI was used by Bauer et al. (2007) to analyze the performance of a portfolio of eight Canadian SRI funds. These equity indices are obtained from the CFMRC and *Fundata* databases.

The first factor in our four-factor Carhart (1997) model is the excess return on the value-weighted portfolio of all TSX stocks. The second through fourth factor are the returns on mimicking portfolios for size, book-to-market, and momentum obtained from Ayadi et al. (2010).

Two instrumental variables obtained from the CANSIM database due to their power to predict stock returns are used in all of the conditional models.[19] The variables are the lagged values of DY or the dividend yield of the S&P/TSX index and TB or the one-month Treasury bill rate.[20] Since the two instruments exhibit high degrees of persistence, they are stochastically detrended by subtracting a moving average over a period of two months, as in Campbell (1991) and Ferson et al. (2003).[21] To allow for a simple interpretation of the estimated coefficients, the variables are demeaned in the conditional tests, as in Ferson and Schadt (1996).

To test if the conditional methodology is likely to be worthwhile, a predictability analysis of the excess returns for the E&SW portfolios of SRI and non-SRI funds is conducted by regressing portfolio excess returns on the stochastically detrended instruments. The unreported results based on Wald tests strongly support a conditional performance analysis. The null hypothesis, that all of the slope coefficients associated with the selected instruments are zeros, is largely rejected.

Descriptive statistics, autocorrelations, and the correlation matrix for these variables are provided in panels A and B of Table 2, respectively. The correlations between the equity indices and risk factors range from -0.41 to 0.92.

6 Empirical Performance Results

In this section, we provide extensive and robust evidence on the (stock selection and timing) performance and the sensitivity of performance inferences for various benchmark models using market-wide and SRI-specific indexes and compared to non-SRI funds for our survivorship-free sample of Canadian SRI equity mutual funds. Our analysis controls for differences in fund expenses to examine if managers add value before expenses (gross returns) and if any of this added value flows to investors (net returns). We also study the (net and gross) performance of extreme SRI funds (i.e., in the tails of the performance distribution) using the bootstrap method and its variants. This is done to successfully identify skilled managers by accounting for individual fund cross-dependencies and to disentangle the effects of sampling variation or "luck" from superior/inferior management skills.

[18] Launched in January 2000, the JSI is a market capitalization weighted index consisting of 60 Canadian companies drawn from the S&P/TSX composite index and non-member companies with exceptional social standards. Companies are selected based on a rating framework which incorporates environmental, social and governance practices. Jantzi Research Inc. merged with Sustainalytics in August 2009, and now operates under the name Jantzi-Sustainalytics (see www.jantziresearch.com for more details on this index). The market cap weighted ECI (Ethical Canadian Index) of 218 securities was created by the Ethical Funds Company, launched in January 2001, and maintained by the S&P. (See www.ethicalfunds.com for more details on this ethical index). The ECI companies are from the S&P/TSX composite index with an acceptable score on the Ethical Funds Company's Corporate Sustainability scorecard. Unlike the JSI, the ECI does not actively seek out companies for inclusion in the index.

[19] Similar inferences result from tests using several other conditioning variables, such as RISK or risk premium as measured by the yield spread between the long-term corporate McLeod, Young, Weir bond index and long-term government of Canada bonds, TERM or the slope of the term structure as measured by the yield spread between long-tem government of Canada bonds and the one period lagged three-month Treasury bill rate, and DUMJ or a dummy variable for the month of January.

[20] DY is used by Ferson and Schadt (1996), Kryzanowski et al. (1997), and Ayadi and Kryzanowski (2005). TB is used by Ferson and Schadt (1996) and Ayadi and Kryzanowski (2005). In the SRI fund performance literature, the two instruments were used by Schroder (2004), Bauer et al. (2006, 2007), Renneboog et al. (2007, 2008a), and Cortez et al. (2009).

[21] A transformation of the persistent instruments is highly recommended in tests of stock return predictability to alleviate any spurious regression biases induced by the use of persistent lagged regressors, and when innovations are highly correlated with returns (Elliot and Stock, 1994; Stambaugh, 1999; Amihud and Hurvich, 2004).

Table 2: Summary statistics for the instrumental variables and factors

This table reports the summary statistics for the monthly returns of the instrumental variables, bond indices, and factors. TB is the yield on one-month Treasury bills in % per month. DY is the dividend yield on the S&P/TSX index. The two instruments are stochastically detrended by subtracting a moving average over a period of two months. The equity factors are the TSXVWX, SMB, HML, and UMD. TSXVWX is the excess return of the value-weighted portfolio of all TSX stocks, SMB (small minus big), HML (high minus low), and UMD (up minus down) are portfolios representing size, value, and momentum risk factors, respectively. They are formed along the lines of Fama and French (1993) and Carhart (1997). JSIX is the excess return on the Jantzi social index. The Jantzi social index is a socially screened, market capitalization-weighted common stock index modeled on the S&P/TSX 60. The JSI consists of 60 Canadian companies that pass a set of broadly-based social and environmental screens. Panel A reports various statistics for all variables, including autocorrelation coefficients of order 1, 3, 6, and 12. Panel B presents the correlation matrix of equity factors and indices. The data cover the period from February 1988 to April 2008, for a total of 243 observations.

Panel A: Descriptive statistics and autocorrelations

Variable	Mean	Median	Std. Dev.	Min.	Max.	Skew.	Kurt.	$\rho 1$	$\rho 3$	$\rho 6$	$\rho 12$
TB	-0.003	-0.002	0.040	-0.134	0.234	0.896	9.324	0.521	0.082	0.077	-0.029
DY	-0.001	-0.001	0.008	-0.036	0.028	0.098	6.162	0.451	0.012	-0.068	-0.064
TSXVWX	0.486	0.778	4.051	-19.623	11.397	-0.696	5.136	0.067	-0.012	0.079	-0.102
SMB	-0.331	-0.372	3.405	-9.025	17.607	0.550	5.786	0.087	0.087	-0.037	0.070
HML	0.844	0.710	3.927	-11.762	20.021	0.584	5.901	0.225	0.131	0.000	0.038
UMD	1.937	2.337	5.427	-25.855	29.054	-0.477	8.641	0.051	0.105	0.083	0.129
JSIX	0.455	0.600	3.441	-7.809	8.599	-0.347	2.887	0.150	-0.019	-0.015	0.071

Panel B: Correlation matrix of equity factors and indices

Factor	TSXVWX	SMB	HML	UMD	JSIX
TSXVWX	1.00				
SMB	0.13	1.00			
HML	-0.25	-0.23	1.00		
UMD	-0.08	-0.23	0.18	1.00	
JSIX	0.92	0.19	-0.41	-0.35	1.00

6.1 Performance Results for Portfolios of Funds

Based on the results summarized in Table 3, E&SW portfolios of SRI funds display negative to neutral performance across all benchmark models. For instance, the estimated monthly alphas of the SW portfolios are all insignificant and range from -0.0998% (-1.1976% per year) using the single factor CAPM model to 0.0557% (0.6684% per year) obtained with the full conditional four-factor model. SRI fund performance improves with a conditional benchmark, and the effect of conditioning is more pronounced with the four-factor benchmark and EW portfolios of funds. A comparison of the alphas of E&SW portfolios suggests that smaller funds outperform (underperform) larger funds using the one- (four-) factor benchmark specification.

However, when we confine our SRI sample to those (on average, smaller) funds that specialize in SRI (see panels C and D compared to panels A and B in Table 3, respectively), we find that SRI selection and timing fund performance improves except for the size-weighted statistics for the four-factor models. All of the timing coefficients for the SRI funds are still not statistically significant at conventional levels. Based on the Wald test results reported in panels C and D of Table 3, the not significantly different selection and timing performances between the SRI and the non-SRI funds remain when we restrict the SRI sample to the specialized subset.

The non-SRI fund portfolios exhibit insignificant negative alphas, which are insignificantly lower based on Wald tests (all p-values = 0.17) than those of the SRI funds across all benchmark models except for the SW portfolios using the single-factor models.[22] The performance differential increases with conditioning and

[22] Consistent inferences are obtained for untablulated tests using alternative control portfolios of non-SRI funds matched on total asset value and age to their SRI counterparts.

is somewhat stable across all models and ranges from 0.52% to 1.14% annually. The "superiority" of SRI over non-SRI funds is maximized (minimized) with the full conditional four-factor model and SW portfolios (unconditional four-factor model and EW portfolios). As for SRI funds, the performance of non-SRI funds improves with conditioning and time-variation in alphas and with the four-factor model.

As reported in Table 4, the estimated market risk sensitivities are high in magnitude for various portfolios of funds for all of the SRI and non-SRI portfolios for all the benchmark models. However, none of the market beta (and other factor loading) differences are significant. When fund returns are measured before expenses, all performance metrics are positive and mostly significant. Since benchmark returns do not incorporate expenses, the gross-return alpha evidence suggests that SRI fund managers exhibit some stock selection skills despite their restricted investment opportunity set. Portfolios of SRI funds insignificantly outperform non-SRI funds based on net return alphas, which suggests that the value added does not cover fund fees and expenses. For all regression models, the adjusted R2 is higher than 0.85 and 0.95 for SRI and non-SRI fund portfolios, respectively, which confirms the suitability of the benchmark models. The overall results are in conformity with the literature that finds weak after-fees underperformance of SRI funds and no material performance differences between SRI and non-SRI funds. However, we extend the literature with our evidence of before-fees outperformance of SRI portfolios.

All gamma coefficients are not significant for the market-timing tests for the two types of portfolios.[23] The point estimates are insignificant and mostly positive (negative) for the SW SRI (non-SRI) fund portfolios. Conditioning and using a multifactor benchmark positively impact the timing statistics using the two types of E&SW portfolios. Most findings persist when fund returns are measured using gross returns, which implies that fund managers lack timing ability.

Further tests using the Jantzi social index (JSI) as an alternative market benchmark variable, which are reported in Table 5, show weak to neutral SRI fund performance across all benchmark specifications. Using the JSI index in the benchmark models alters the positive effect of conditioning and time-variation in alphas. For example, the use of multifactor models lowers SRI fund performance with the JSI. The best (worst) SRI fund performance based on net returns is achieved with the SW portfolio of funds using a full conditional one-factor (four-factor) model at 0.1553% monthly or 1.86% annually (-0.2495% monthly or -2.99% annually). Finally, the use of the JSI increases the factor sensitivity of SRI fund portfolios to the SMB factor.

6.2 Performance Results for Individual Funds

The cross-sectional performance (selection and timing) distribution of individual SRI funds is summarized in panels A and B of Table 4. Most of the selection statistics in panel A suggest weak to neutral performance across all benchmark models. The performance statistics have higher variability using the JSI in the single factor model and the gap increases with conditioning information. The highest standard deviation of SRI fund performance occurs for the four-factor model across all benchmark models with and without the use of the SR index. Conditioning has a positive effect on the mean and median performances based on the two benchmark models. The timing statistics in panel B are mostly consistent with the inferences obtained using portfolios of funds.

The distribution of fund selection performance is negatively skewed with fat tails for all retained benchmark models suggesting that the left tail is longer. The cross-sectional analysis shows that the number of funds with positive and significant alphas or gammas is low but higher for the four-factor model at the 5% and 10% levels, with and without conditioning information. None of the Bonferroni p-values are significant for the maximum t-statistics, which cannot reject the joint hypothesis of zero alphas. This result holds for the minimum t-statistics with the exception of the full conditional 4-factor model using the market return.

[23] Similar market timing inferences are obtained using (un)conditional models of Henriksson and Merton (1981).

Table 3: Performance measures for portfolios of funds using the single and four-factor models

This table reports summary statistics on the performance (α in %) and g (timing) measures for equal and size-weighted portfolios of SRI and non-SRI individual equity mutual funds. Unconditional and conditional alpha and beta benchmark models based on single and four-factor specifications are used. The equity factors are the TSXVWX, SMB, HML, and UMD. TSXVWX is the excess return of the value-weighted portfolio of all TSX stocks, SMB (small minus big), HML (high minus low), and UMD (up minus down) are portfolios representing size, value, and momentum risk factors, respectively. They are formed along the lines of Fama and French (1993) and Carhart (1997). The stochastically detrended instrumental variables used in the conditional models, are the lagged values of the yield on one-month T-bills and dividend yield on the S&P/TSX index. Gross (pre-expense) fund returns are net returns plus 1/12th of a fund's expense ratio. The alphas are the estimates of the intercepts in the unconditional and conditional alpha and beta time-series based regressions. The gammas are the estimates of the quadratic term in the unconditional and conditional alpha and beta time-series based regressions. In the conditional alpha and beta models, the alpha and beta(s) coefficients are linear functions of two lagged instruments. The t-statistics are adjusted for serial correlation and heteroskedasticity (Newey and West, 1987a) and reported in parentheses below the parameter estimates. Walda (Waldg) corresponds to the p-value based on the Newey and West (1987b) test for the hypothesis that the performances (alphas or gammas) of the portfolios of SRI and non-SRI funds are equal for each benchmark model. Panels C and D contain the same results for equal- and size-weighted portfolios of SRI funds managed by specialized companies, respectively. The test is conducted in a GMM system that includes the equations of the two portfolios. The portfolios of funds include all funds whose returns are available in a given month. As a result, the number of funds in each portfolio varies across the years depending on the entry and exit of funds. The asterisks are used to denote the significant alphas (*, ** and *** at the 10%, 5%, and 1% levels of significance, respectively). Monthly data used are from February 1988 to April 2008, for 243 observations per portfolio of funds.

Benchmark models/returns	Unconditional single factor CAPM model		Unconditional four-factor model		Conditional alpha and beta single factor model		Conditional alpha and beta four-factor model	
	Net	Gross	Net	Gross	Net	Gross	Net	Gross
Panel A: Equal-weighted portfolios of funds								
α-SRI	-0.0210 (-0.24)	0.1776** (2.04)	-0.0108 (-0.11)	0.1864* (1.87)	-0.0108 (-0.12)	0.1878** (2.20)	0.0279 (0.29)	0.2248** (2.32)
α-Conventional	-0.0983 (-1.48)	0.0997 (1.53)	-0.0540 (-0.72)	0.1431* (1.88)	-0.0893 (-1.40)	0.1088* (1.72)	-0.0368 (-0.47)	0.1604** (2.02)
Wald$^{\alpha}$	0.28	0.27	0.58	0.57	0.26	0.25	0.41	0.40
γ-SRI	-0.189 (-0.51)	-0.175 (-0.47)	-0.078 (-0.22)	-0.064 (-0.18)	-0.132 (-0.39)	-0.117 (-0.34)	0.003 (0.01)	0.018 (0.05)
γ-Conventional	-0.281 (-0.98)	-0.277 (-0.97)	-0.143 (-0.47)	-0.139 (-0.46)	-0.256 (-0.92)	-0.251 (-0.90)	-0.088 (-0.30)	-0.082 (-0.29)
Wald$^{\gamma}$	0.70	0.67	0.78	0.74	0.60	0.57	0.68	0.65
Panel B: Size-weighted portfolios of funds								
α-SRI	-0.0998 (-1.18)	0.1043 (1.25)	0.0317 (0.39)	0.2349*** (2.82)	-0.0946 (-1.12)	0.1095 (1.32)	0.0557 (0.63)	0.2592*** (2.89)
α-Conventional	-0.0832 (-1.29)	0.0973 (1.53)	-0.0425 (-0.56)	0.1378* (1.80)	-0.0721 (-1.15)	0.1084* (1.75)	-0.0278 (-0.34)	0.1524* (1.85)
Wald$^{\alpha}$	0.83	0.93	0.35	0.23	0.77	0.99	0.29	0.18
γ-SRI	0.074 (0.26)	0.075 (0.26)	0.250 (0.72)	0.251 (0.72)	0.097 (0.33)	0.098 (0.34)	0.312 (0.93)	0.313 (0.94)
γ-Conventional	-0.151 (-0.54)	-0.146 (-0.52)	-0.027 (-0.09)	-0.022 (-0.07)	-0.114 (-0.43)	-0.109 (-0.40)	0.043 (0.15)	0.048 (0.17)
Wald$^{\gamma}$	0.39	0.40	0.20	0.21	0.45	0.46	0.22	0.22
Panel C: Equal-weighted portfolios of SRI funds managed by specialized companies								
α^{sp}-SRI	0.0185 (0.18)	0.2152** (2.10)	-0.0028 (-0.02)	0.1931 (1.60)	0.0314 (0.31)	0.2278** (2.28)	0.0443 (0.38)	0.2396** (2.02)
Wald$^{\alpha}$	0.18	0.20	0.60	0.52	0.15	0.22	0.40	0.39
γ^{sp}-SRI	0.166 (0.26)	0.174 (0.78)	0.266 (0.45)	0.274 (0.46)	0.293 (0.55)	0.302 (0.56)	0.359 (0.73)	0.369 (0.75)
Wald$^{\gamma}$	0.32	0.44	0.31	0.41	0.15	0.31	0.15	0.20
Panel D: Size-weighted portfolios of SRI funds managed by specialized companies								
α^{sp}-SRI	-0.0587 (-0.71)	0.1381* (1.68)	-0.0637 (-0.66)	0.1330 (1.38)	-0.0570 (-0.69)	0.1397* (1.70)	-0.0388 (-0.38)	0.1578 (1.55)
Wald$^{\alpha}$	0.73	0.74	0.79	0.66	0.83	0.77	0.89	0.60
γ^{sp}-SRI	0.123 (0.30)	0.120 (0.29)	0.242 (0.63)	0.240 (0.62)	0.188 (0.54)	0.185 (0.53)	0.294 (0.92)	0.291 (0.91)
Wald$^{\gamma}$	0.26	0.30	0.22	0.25	0.16	0.20	0.15	0.20

Table 4: Risk measures for portfolios of funds using the single and four-factor models

This table reports summary statistics on the risk (beta) measures for equal and size-weighted portfolios of SRI and non-SRI individual funds using unconditional and conditional alpha and beta models based on single and four-factor specifications. The equity risk factors are the TSXVWX, SMB, HML, and UMD. TSXVWX is the excess return of the value-weighted portfolio of all TSX stocks, SMB (small minus big), HML (high minus low), and UMD (up minus down) are portfolios representing size, value, and momentum risk factors, respectively. They are formed along the lines of Fama and French (1993) and Carhart (1997). The stochastically detrended instrumental variables used in the conditional models, are the lagged values of the yield on one-month T-bills and dividend yield on the S&P/TSX index. The betas are the estimates of the slope coefficients related to the factors in the time-series regressions. In the conditional alpha and beta models, the alpha and beta(s) coefficients are linear functions of the two lagged instruments. The portfolios of funds include all funds whose returns are available in a given month. As a result, the number of funds in each portfolio varies across the years depending on the entry and exit of funds. Monthly data used are from February 1988 to April 2008, for 243 observations per portfolio of funds.

Benchmark model/portfolios	Equal-weighted portfolios of funds								Size-weighted portfolios of funds							
	SRI				Conventional				SRI				Conventional			
	β_M	β_{SML}	β_{HML}	β_{UMD}	β_M	β_{SML}	β_{HML}	β_{UMD}	β_M	β_{SML}	β_{HML}	β_{UMD}	β_M	β_{SML}	β_{HML}	β_{UMD}
Panel A: Single factor CAPM model																
Unconditional	0.690				0.806				0.774				0.855			
Conditional alpha and beta	0.686				0.804				0.772				0.853			
Panel B: Four-factor model																
Unconditional	0.697	0.043	0.060	-0.026	0.814	0.016	0.058	-0.047	0.763	0.012	-0.014	-0.057	0.862	0.006	0.053	-0.045
Conditional alpha and beta	0.691	0.032	0.057	-0.034	0.814	0.006	0.050	-0.053	0.752	0.000	-0.015	-0.063	0.863	-0.004	0.045	-0.051

Table 5: Performance and risk measures for portfolios of funds using the single and four-factor models using the Jantzi social index

This table reports summary statistics on the performance (α in %) and risk (beta) measures for equal and size-weighted portfolios of individual SRI equity funds using unconditional and conditional alpha and beta models based on single and four-factor specifications. The Jantzi social index, which is used as the market benchmark, is a socially screened, market capitalization-weighted common stock index modeled on the S&P/TSX 60. It also consists of 60 Canadian companies that pass a set of broadly-based social and environmental screens. The other risk factors are SMB (small minus big), HML (high minus low), and UMD (up minus down), which are portfolios representing size, value, and momentum risk factors, respectively. They are formed along the lines of Fama and French (1993) and Carhart (1997). The stochastically detrended instrumental variables used in the conditional models are the lagged values of the yield on one-month T-bills and dividend yield on the S&P/TSX index. The alphas and the betas are the estimates of the intercept and slope coefficients in the unconditional and conditional single and four-factor based time-series regressions. In the conditional alpha and beta models, the alpha and beta(s) coefficients are linear functions of the two lagged instruments. The portfolios of funds include all funds whose returns are available in a given month. As a result, the number of funds in each portfolio varies across the years depending on the entry and exit of funds. Monthly data used are from February 1988 to April 2008, for 243 observations per portfolio of funds.

Benchmark model/Portfolios	Equal-weighted portfolios of SRI funds					Size-weighted portfolios of SRI funds				
	α	β_M	β_{SML}	β_{HML}	β_{UMD}	α	β_M	β_{SML}	β_{HML}	β_{UMD}
Panel A: Single factor model with the Jantzi social index										
Unconditional	0.0073	0.760				-0.1725	0.873			
Conditional alpha and beta	0.1553	0.739				-0.1395	0.882			
Panel C: Four-factor model with the Jantzi social index										
Unconditional	-0.1864	0.791	0.162	0.059	0.063	-0.1085	0.847	0.059	-0.043	0.001
Conditional alpha and beta	-0.0689	0.741	0.160	0.019	0.057	-0.1334	0.828	0.065	-0.106	0.022

6.3 Performance Results using the Bootstrap Methodology

The bootstrap methodology is used as an alternative method of dealing with the problems of possible fund return nonlinearity and spatial correlation that are likely to cause the performances and test statistics not to be independent across individual funds and to aid in distinguishing skill from luck. Bootstrapping is used in the context of performance evaluation by Ayadi and Kryzanowski (2005, 2010), Kosowski et al. (2006, 2007), Huij and Derwall (2008), and Cuthbertson et al. (2008). Samples of 2000 are obtained for each of two variants of the bootstrap method [i.e., standard bootstrap of Efron (1979) and the moving block bootstrap of Künsch (1989)] for each of two resampling schemes (i.e., residuals only and independently of both the residuals and the factors).[24] Bootstrapped statistics are constructed for all and specific cross-sections of the individual funds (namely, funds in the tails of the performance distribution) where funds are ranked by their estimated performances (selection and market timing) and by their t-statistics based on both net and gross returns. As a test of robustness, individual performance statistics and inferences are further assessed using the block bootstrap method with a block of three consecutive monthly observations.

Two block-bootstrapped p-values are obtained when funds are ranked according to the estimated t-statistics with the two resampling schemes. These values are then compared with the standard one-tailed p-values from the original estimation. The analysis is conducted on the performances of the best and worst funds including the 2^{nd}, 3^{rd}, 4^{th}, and 5^{th} ranked funds in the left and right tails of the distribution. The median value is also provided.

Based on panels A to D of Table 7 and Figure 1, all the bootstrapped p-values of the alphas for the five worst performing funds are less than 0.05 using the unconditional setting with net returns. Thus, their negative extreme performance cannot be attributed to sampling variability or "bad luck" and indicates the absence of management skills. Some of these results are slightly altered for the single-factor benchmark or when ranking is based on the estimated alpha in the bootstrap experiments or when standard bootstrapping is used. The inferences are strikingly different when resampling is implemented using residuals where all the p-values are greater than 0.41 for block and standard bootstrapping. The same tests conducted using gross

[24] The results of the standard bootstrap are discussed but not tabulated. They are available upon request from the authors.

returns indicate that sampling variation has a marginal effect on the original neutral performance inferences of these worst performing ethical funds.

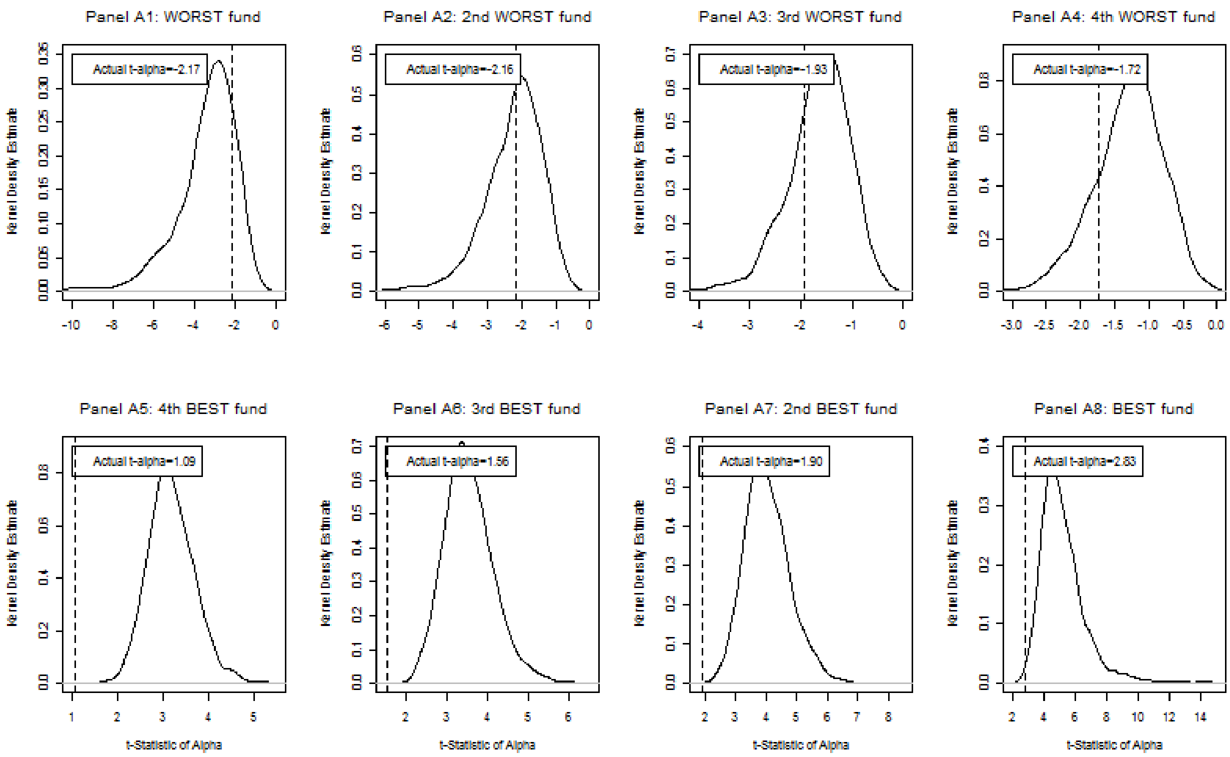

Figure 1: **Estimated alpha t-statistics vs. bootstrapped alpha t-statistic distribution for individual funds at various points in the cross-section using net returns.** The t-statistics are based on the full conditional four-factor model. This figure plots kernel density estimates if the bootstrapped distribution of the t-statistic of alpha for Canadian ethical equity funds with at least 30 monthly observations during the period 1988-2008. The dashed vertical line represents the actual t-statistic of alpha. Panels A1 to A4 present the results for the worst funds (left tail) and panels A5 to A8 presents the results for the best funds (right tail).

When the benchmark model becomes conditional, the corrected bootstrap p-values are consistently greater than 0.17 for the four extreme funds using net returns. This implies that the original negative and significant performance of these funds is essentially related to sampling variation. This result is confirmed using gross returns, a residuals only resampling method, or with the alternative ranking structure. This evidence further justifies the use of bootstrap tests to determine the significance levels of estimated performance in the tails.

In contrast, the block-bootstrapped p-values of the three top performing funds are very different from the original ones using net returns and the full conditional four-factor model. The inference is that sampling variation or "luck" accounts for their good performance. Stronger results are obtained using gross returns and with the standard resampling approach. However, when the bootstrap tests are based on alpha rankings or with a single-factor benchmark specification (with net returns only), there are no differences between the original and corrected performance inferences. Similar adjusted p-values are obtained without conditioning and the true performance of these SRI funds is mainly attributed to sampling variability. All of these results persist when performance is measured using gross returns only with factor returns and residuals independent resampling.

Table 6: Summary statistics for the cross-sections of individual SRI fund performance estimates based on factor models

This table reports the mean, standard deviation, median, minimum, and maximum of individual SRI fund performances (a in % and g) and other cross-sectional statistics. Panel A presents these statistics for the stock selection ability measures using the unconditional and full conditional single and four-factor models. Panel B provide the timing performance statistics. The dividend yield (DY) and the yield on the one-month T-bill (TB) are used as instrumental variables. Time-series regressions are conducted for each individual fund based on simple and extended specifications. Information related to the funds with significant performance at the 5% and 10% levels and with positive significant performance is provided in the table. Only funds with at least 30 observations are considered. The Bonferroni p-values are the minimum and the maximum one-tailed p-values from the t-distribution across all of the funds multiplied by the number of funds. All of the p-values are adjusted for serial correlation and heteroskedasticity (Newey and West, 1987a). Monthly data are used from February 1988 to April 2008 for up to a maximum of 243 observations per fund.

Panel A: Selection performance measures

Benchmark models	**Performance Statistics**											
	Mean	Std. Dev.	Median	Minimum	Maximum	Median Adj. R2	p-max	p-min	% funds with p<5% (10%)	% funds with α >0 & p<5% (10%)	Bonferroni p-value	
											Min. t	Max. t
Unconditional models												
Single factor CAPM model	-0.0921	0.2260	-0.0764	-1.0438	0.3418	0.72	0.95	0.02	2.63 (5.26)	0.00 (0.00)	0.32	1.00
Four-factor model	-0.0863	0.2816	-0.1116	-1.2373	0.4411	0.75	0.93	0.01	5.26 (18.42)	2.63 (10.53)	0.16	0.75
Conditional alpha and beta models												
Single factor CAPM model	0.0249	0.1884	0.0388	-0.3920	0.4015	0.73	0.91	0.06	0.00 (2.63)	0.00 (0.00)	1.00	1.00
Four-factor model	0.0201	0.3497	0.0035	-1.3094	0.7580	0.77	0.99	0.01	7.89 (15.79)	2.63 (5.26)	0.58	0.09

Panel B: Timing performance measures

Benchmark models	**Timing Performance Statistics**										
	Mean	Std. Dev.	Median	Minimum	Maximum	p-max	p-min	% funds with p<5% (10%)	% funds with γ >0 & p<5% (10%)	Bonferroni p-value	
										Min. t	Max. t
Unconditional models											
Single factor CAPM model	0.8922	1.4560	0.9122	-1.2006	4.7594	0.95	0.00	13.16 (15.79)	13.16 (15.79)	1.00	0.00
Four-factor model	1.0076	1.2777	0.7041	-0.5338	5.4735	0.94	0.00	13.16 (15.79)	13.16 (15.79)	1.00	0.00
Conditional alpha and beta models											
Single factor CAPM model	0.9701	1.7797	0.7781	-2.3146	7.7740	0.96	0.00	13.16 (21.05)	13.16 (18.42)	1.00	0.00
Four-factor model	0.5491	1.3726	0.4749	-3.1129	4.6044	0.87	0.00	15.79 (23.68)	13.16 (18.42)	0.87	0.00

Table 7: Bootstrap analysis of the best and worst SRI fund's performances

This table reports least squares monthly estimates of SRI fund performance (α in %) and significance tests using the block bootstrap methodology (with a block of 3 overlapping observations) with independent resampling of the factor returns and the residuals (RF) and with a residuals-only (R) resampling, for all funds. The alpha is estimated using unconditional and conditional alpha and beta models with single and four-factor specifications over the entire period. Conditional alpha and beta refers to models with both time-varying alphas and betas. The dividend yield (DY) and the yield on the one-month T-bill (TB) are used as instrumental variables. Gross (pre-expense) fund returns are net returns plus 1/12th of a fund expense ratio. For each benchmark model, the first, second, third, and fourth columns show the ranked alpha estimate, parametric one-tailed p-value, and bootstrapped p-value under RF and R schemes, respectively. The ranked t-statistic, parametric one-tailed p-value, and t-statistic based bootstrapped p-value under RF and R schemes are reported in columns five, six, seven, and eight, respectively. All t-statistic are based on serial correlation and heteroskedasticity consistent standard errors. All of these p-values concern the distribution of the best (worst) funds in 2000 bootstrapped samples. The first (eleventh) row in each panel reports funds with the lowest (highest) alpha and t-statistic as well as the median. The in-between rows concern the 2^{nd}, 3^{rd}, 4^{th}, and 5^{th} funds in the left (right) tail of the performance distribution as well as the median fund. Only funds with a minimum of 30 observations are included. Monthly data are used from February 1988 to April 2008 for up to 243 observations per fund.

	Unconditional								**Conditional alpha and beta**							
Benchmark model	α	p-val.	boot.RF p-val.	boot.R p-val.	t_α	p-val.	boot.RF p-val.(t)	boot.R p-val.(t)	α	p-val.	boot.RF p-val.	boot.R p-val.	t_α	p-val.	boot.RF p-val.(t)	boot.R p-val.(t)
Panel A: Single market factor model with net returns																
Worst	-1.044	0.15	0.09	0.19	-2.39	0.01	0.10	0.52	-0.392	0.07	0.99	0.87	-1.93	0.03	0.82	0.89
2nd	-0.424	0.04	0.24	0.32	-1.80	0.04	0.04	0.60	-0.289	0.33	0.99	0.90	-1.53	0.07	0.71	0.90
3rd	-0.253	0.13	0.30	0.76	-1.41	0.08	0.02	0.74	-0.288	0.15	0.94	0.75	-1.46	0.07	0.47	0.81
4th	-0.247	0.01	0.12	0.58	-1.15	0.13	0.01	0.83	-0.242	0.03	0.90	0.76	-1.05	0.15	0.55	0.95
5th	-0.245	0.14	0.03	0.36	-1.10	0.14	0.00	0.74	-0.208	0.07	0.81	0.77	-0.83	0.21	0.52	0.97
Median																
5th	0.113	0.34	1.00	0.98	0.91	0.18	1.00	0.88	0.251	0.23	1.00	0.47	0.87	0.19	1.00	0.95
4th	0.162	0.18	1.00	0.93	1.10	0.14	1.00	0.85	0.252	0.19	1.00	0.68	1.05	0.15	1.00	0.92
3rd	0.258	0.14	1.00	0.62	1.11	0.14	1.00	0.93	0.264	0.22	1.00	0.80	1.18	0.12	1.00	0.95
2nd	0.258	0.14	1.00	0.86	1.20	0.12	1.00	0.96	0.295	0.12	1.00	0.87	1.20	0.12	1.00	0.98
Best	0.342	0.12	1.00	0.82	1.50	0.07	1.00	0.95	0.402	0.07	1.00	0.84	1.47	0.07	1.00	0.98
Panel B: Single market factor model with gross returns																
Worst	-0.908	0.17	0.13	0.22	-0.98	0.17	0.71	1.00	-0.200	0.37	1.00	1.00	-0.61	0.27	1.00	1.00
2nd	-0.192	0.21	0.69	0.99	-0.81	0.21	0.46	1.00	-0.159	0.27	1.00	1.00	-0.36	0.36	1.00	1.00
3rd	-0.032	0.44	0.86	1.00	-0.15	0.44	0.82	1.00	-0.070	0.36	1.00	1.00	-0.36	0.36	1.00	1.00
4th	-0.026	0.45	0.70	1.00	-0.12	0.45	0.66	1.00	-0.068	0.40	0.99	1.00	-0.32	0.37	0.98	1.00
5th	-0.004	0.49	0.58	1.00	-0.02	0.49	0.58	1.00	-0.052	0.36	0.98	1.00	-0.25	0.40	0.95	1.00
Median																
5th	0.291	0.02	1.00	0.09	2.05	0.02	1.00	0.01	0.458	0.09	1.00	0.00	2.30	0.01	0.95	0.00
4th	0.362	0.02	1.00	0.05	2.16	0.02	1.00	0.01	0.471	0.09	1.00	0.02	2.73	0.00	0.84	0.00
3rd	0.375	0.06	1.00	0.13	2.20	0.02	1.00	0.05	0.514	0.06	1.00	0.04	2.81	0.00	0.92	0.02
2nd	0.514	0.02	1.00	0.05	2.20	0.02	1.00	0.21	0.551	0.01	1.00	0.11	2.81	0.00	0.98	0.09
Best	0.570	0.02	1.00	0.19	3.92	0.00	0.88	0.03	0.630	0.01	1.00	0.29	3.10	0.00	0.99	0.27

Table 7. Continued.

	Unconditional								Conditional alpha and beta							
Benchmark model	α	p-val.	boot.RF p-val.	boot.R p-val.	t_α	p-val.	boot.RF p-val.(t)	boot.R p-val.(t)	α	p-val.	boot.RF p-val.	boot.R p-val.	t_α	p-val.	boot.RF p-val.(t)	boot.R p-val.(t)
Panel C: Four-factor model with net returns																
Worst	-1.237	0.12	0.08	0.16	-2.62	0.01	0.04	0.46	-1.309	0.05	0.74	0.10	-2.17	0.02	0.80	0.89
*2*nd	-0.334	0.14	0.45	0.73	-1.80	0.04	0.02	0.69	-0.519	0.02	0.96	0.48	-2.16	0.02	0.44	0.65
*3*rd	-0.332	0.14	0.19	0.44	-1.78	0.04	0.00	0.42	-0.487	0.02	0.88	0.24	-1.93	0.03	0.27	0.57
*4*th	-0.330	0.12	0.06	0.21	-1.48	0.07	0.00	0.53	-0.304	0.17	0.91	0.64	-1.72	0.05	0.17	0.53
*5*th	-0.325	0.10	0.02	0.09	-1.30	0.10	0.00	0.53	-0.276	0.03	0.83	0.56	-1.60	0.06	0.09	0.45
Median																
*5*th	0.176	0.16	1.00	0.76	1.37	0.09	1.00	0.40	0.352	0.29	1.00	0.20	1.05	0.15	1.00	0.93
*4*th	0.202	0.02	1.00	0.76	1.68	0.05	1.00	0.27	0.388	0.28	1.00	0.26	1.09	0.14	1.00	0.96
*3*rd	0.225	0.03	1.00	0.83	1.77	0.04	1.00	0.39	0.478	0.20	1.00	0.23	1.56	0.06	1.00	0.83
*2*nd	0.361	0.04	1.00	0.46	1.90	0.03	1.00	0.57	0.679	0.18	1.00	0.14	1.90	0.03	1.00	0.80
Best	0.441	0.09	1.00	0.57	2.06	0.02	1.00	0.77	0.758	0.16	1.00	0.38	2.83	0.00	1.00	0.58
Panel D: Four-factor model with gross returns																
Worst	-1.097	0.13	0.10	0.19	-1.13	0.13	0.57	1.00	-1.200	0.05	0.78	0.13	-1.74	0.05	0.91	0.98
*2*nd	-0.126	0.34	0.86	1.00	-0.69	0.25	0.54	1.00	-0.396	0.05	0.99	0.78	-1.66	0.05	0.72	0.93
*3*rd	-0.125	0.34	0.67	1.00	-0.55	0.29	0.41	1.00	-0.321	0.10	0.98	0.77	-1.34	0.10	0.65	0.94
*4*th	-0.123	0.31	0.41	0.99	-0.49	0.31	0.24	1.00	-0.072	0.41	1.00	1.00	-0.22	0.41	1.00	1.00
*5*th	-0.114	0.25	0.23	0.99	-0.42	0.34	0.14	1.00	-0.026	0.46	1.00	1.00	-0.14	0.45	0.99	1.00
Median																
*5*th	0.368	0.00	1.00	0.01	2.33	0.01	0.98	0.00	0.494	0.02	1.00	0.01	2.07	0.02	0.98	0.07
*4*th	0.384	0.00	1.00	0.04	2.35	0.01	0.99	0.01	0.500	0.19	1.00	0.06	2.14	0.02	0.99	0.15
*3*rd	0.395	0.01	1.00	0.12	3.13	0.00	0.83	0.00	0.589	0.19	1.00	0.07	3.13	0.00	0.74	0.02
*2*nd	0.463	0.08	1.00	0.15	3.14	0.00	0.95	0.03	0.886	0.12	1.00	0.02	3.23	0.00	0.89	0.09
Best	0.478	0.01	1.00	0.47	3.23	0.00	0.99	0.21	0.965	0.10	1.00	0.16	3.44	0.00	0.95	0.33

These results are further supported by Figure 1, which depicts the distribution of the bootstrapped t-statistics for various ranked funds using the full conditional four-factor model with net returns. [25] This figure illustrates cases where bootstrapping and original estimation lead to both similar and divergent conclusions. For example, the worst fund in panel A1 has an actual t-statistic of -2.17 (dashed line) but the null is not rejected using the bootstrap. Panel A5 shows the fourth best fund with an actual t-statistic of 1.09 that rejects the null as does the bootstrap test.

We also assess the market-timing ability of various extreme funds in both tails of the performance distribution using the block bootstrap approach with both methods of resampling for all benchmark specifications. The results presented in Table 8 indicate that the effect of sampling variation or "bad luck" explains the poor conditional timing performance of only the worst and second worst funds where the original and corrected p-values are very different. These finding persist using gross returns, the alternative ranking method, and standard bootstrapping. The impact of luck on these funds mostly vanishes when the benchmark is unconditional or based on single-factor specifications. Furthermore, the five best performing funds have bootstrapped p-values of gamma greater than 0.13, supporting the "good luck" argument for the good originally estimated performances based on net returns. Under the alternative ranking scheme, the original neutral timing performance for the fourth and fifth top SRI funds is confirmed by the bootstrap tests. These conclusions are maintained using gross returns, using an unconditional benchmark, with residuals only resampling, and with standard bootstrapping. As with the bottom performing funds, these findings are reversed using the single-factor timing models.

The market-timing evidence is further supported by Figure 2 which depicts the distribution of the bootstrapped t-statistics for various ranked funds using the conditional multifactor model with net returns. This figure illustrates cases where bootstrapping and the original estimations lead to similar or divergent conclusions. For example, the worst fund in panel A3 has an actual t-statistic of -0.75 (dashed line) and the null is rejected using the bootstrap. Panel A7 shows the second fund with an actual t-statistic of 2.26 that rejects the null but the bootstrap test does not.

These block bootstrap-based inferences differ from those from the Bonferroni tests. They highlight the importance and effects of individual ethical fund cross-correlations. Overall, our findings parallel the recent performance tests of Koswoski et al. (2006) and Kosowski et al. (2007) on U.S. domestic equity mutual funds and on the worst performing hedge funds, respectively. They are also consistent with the evidence on Canadian fixed-income funds by Ayadi and Kryzanowski (2010). All three papers report significant differences in the estimated alphas and bootstrapped alpha distributions. However, they partially differ from the results of Cuthbertson et al. (2008) where all U.K. unit trusts (equity mutual funds) in the left tail have "poor skill".

7 Conclusion

By examining a comprehensive and survivorship-free sample of Canadian SRI equity funds in this paper, we are able to assess the impact of social screens and ethical rules on the investment formation and management process. Our results indicate that performance of SRI funds is weak to neutral. Conditioning information and the multifactor benchmark structure positively impact performance statistics and inferences. Using gross returns, performance becomes positive and significant. This evidence holds for conventional and SRI benchmarks. Market timing tests suggest the absence of such skills for SRI fund managers based on both gross and net returns.

We find no material performance differences between SRI and non-SRI funds using both gross and net returns. This implies that SRI funds are a legitimate investment alternative for investors who integrate personal and societal values (returns in kind) into their investment decisions. Based on the use of the cross-sectional bootstrap, we cannot reject the null hypothesis that most of the top performing SRI funds are simply lucky based on gross- and net-return alphas. Similarly, the evidence is that the worst performing funds suffer from "bad luck".

[25] Kernel density estimators based on Gaussian kernel functions are applied to these distributions.

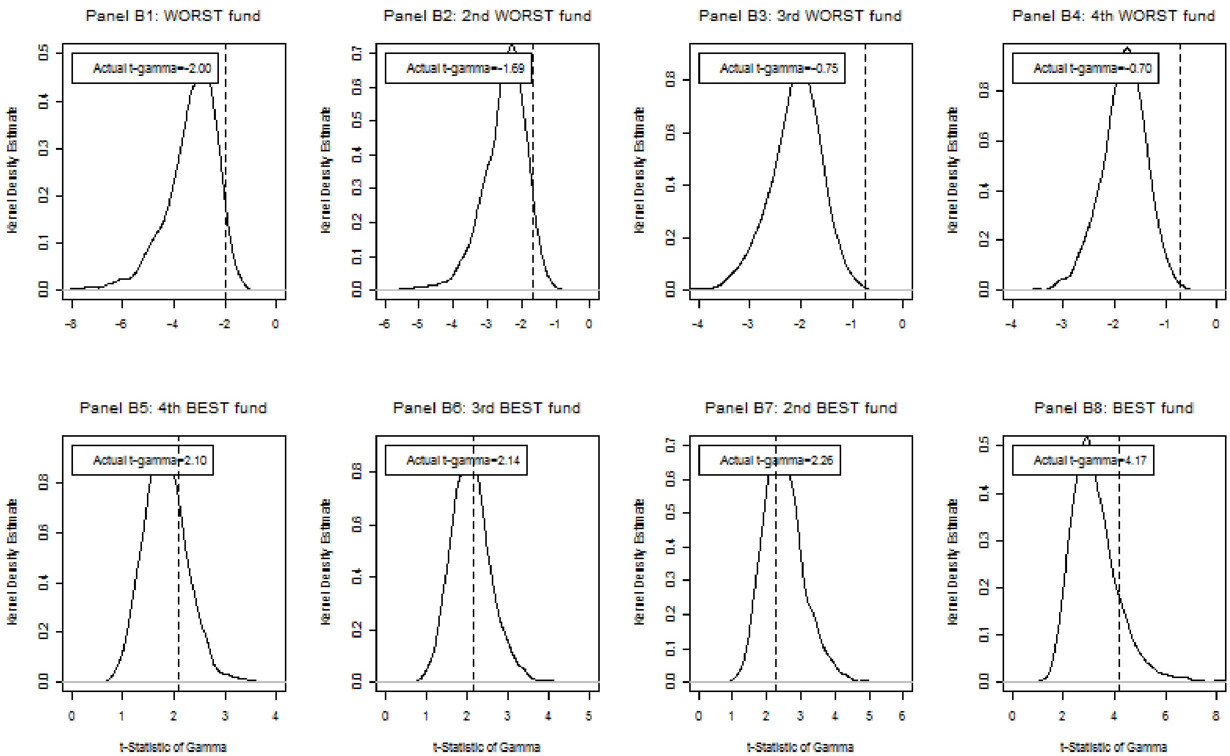

Figure 2: **Estimated gamma t-statistics vs. bootstrapped gamma t-statistic distribution for individual funds at various points in the cross-section using net returns.** The t-statistics are based on the full conditional four-factor model. This figure plots kernel density estimates if the bootstrapped distribution of the t-statistic of gamma for Canadian ethical equity funds with at least 30 monthly observations during the period 1988-2008. The dashed vertical line represents the actual t-statistic of gamma. Panels B1 to B4 present the results for the worst funds (left tail) and panels B5 to B8 presents the results for the best funds (right tail).

The market-timing tests for the extreme funds generally confirm the absence of timing ability based on both gross and net returns among the top five performing Canadian SRI equity mutual funds. Only one fund is identified as possessing market-timing ability, and only when bootstrapping is based on residuals only resampling. Similarly, we cannot reject the hypothesis that the worst two funds with originally significant gammas are merely unlucky, indicating no evidence for market-timing skills.

A number of interesting extensions are left to future work. This includes the use of a benchmark model based on multi-attribute utility functions, identifying the determinants of fund flows based on several fund characteristics, and using the false discovery rate (FDR) framework to deal with the multiple hypothesis testing problems encountered in performance measurement.

Table 8: Bootstrap analysis of the best and worst SRI fund's timing performances

This table reports fund timing performance (g) and significance tests using the block bootstrap methodology (with a block of 3 overlapping observations) with independent resampling of the factor returns and the residuals and with a residuals-only resampling for all funds. The alpha and gamma are estimated using unconditional and conditional alpha and beta models with single and four-factor specifications with a quadratic market variable over the entire period. Conditional alpha and beta refers to models with both time-varying alphas and betas for the market factor. The dividend yield (DY) and the yield on the one-month T-bill (TB) are used as instrumental variables. Gross (pre-expense) fund returns are net returns plus 1/12th of a fund expense ratio. For each benchmark model, the first, second, third, and fourth columns show the ranked gamma estimate, parametric one-tailed p-value, and bootstrapped p-value under RF and R schemes, respectively. The ranked t-statistic, parametric one-tailed p-value, and t-statistic based bootstrapped p-value under RF and R schemes are reported in columns five, six, seven, and eight, respectively. All t-statistic are based on serial correlation and heteroskedasticity consistent standard errors. All of these p-values concern the distribution of the best (worst) funds in 2000 bootstrapped samples. The first (eleventh) row in each panel reports funds with the lowest (highest) alpha and t-statistic as well as the median. The in-between rows concern the 2^{nd}, 3^{rd}, 4^{th}, and 5^{th} funds in the left (right) tail of the timing performance distribution as well as the median fund. Only funds with a minimum of 30 observations are included. Monthly data are used from February 1988 to April 2008, for up to 243 observations per fund.

	Unconditional								**Conditional alpha and beta**							
Benchmark model	γ	p-val.	boot.RF p-val.	boot.R p-val.	t_γ	p-val.	boot.RF p-val.(t)	boot.R p-val.(t)	γ	p-val.	boot.RF p-val.	boot.R p-val.	t_γ	p-val.	boot.RF p-val.(t)	boot.R p-val.(t)
Panel A: Single market factor model with net returns																
Worst	-1.20	0.10	1.00	0.99	-1.30	0.10	1.00	1.00	-2.31	0.04	1.00	0.78	-1.87	0.04	0.94	0.95
2nd	-0.95	0.27	1.00	0.99	-1.03	0.15	0.99	1.00	-1.23	0.12	1.00	0.97	-1.20	0.12	0.99	0.99
3rd	-0.93	0.15	1.00	0.97	-0.88	0.19	0.99	1.00	-0.50	0.29	1.00	1.00	-0.58	0.28	1.00	1.00
4th	-0.67	0.19	1.00	0.99	-0.63	0.27	1.00	1.00	-0.29	0.35	1.00	1.00	-0.55	0.29	1.00	1.00
5th	-0.53	0.28	1.00	0.99	-0.61	0.27	0.99	0.99	-0.26	0.30	1.00	1.00	-0.53	0.30	1.00	1.00
Median																
5th	2.07	0.03	0.89	0.09	2.35	0.01	0.02	0.01	1.77	0.07	0.95	0.20	2.55	0.01	0.01	0.01
4th	2.95	0.00	0.75	0.02	2.72	0.00	0.01	0.01	2.92	0.00	0.77	0.03	2.93	0.00	0.01	0.01
3rd	4.30	0.00	0.46	0.00	3.74	0.00	0.00	0.00	4.71	0.01	0.38	0.00	3.59	0.00	0.00	0.01
2nd	4.75	0.00	0.60	0.04	3.92	0.00	0.01	0.00	5.43	0.00	0.49	0.03	4.38	0.00	0.00	0.00
Best	4.76	0.14	0.87	0.45	4.40	0.00	0.07	0.05	7.77	0.12	0.54	0.23	5.71	0.00	0.01	0.01
Panel B: Single market factor model with gross returns																
Worst	-1.22	0.10	1.00	1.00	-1.32	0.10	0.99	0.99	-2.32	0.04	1.00	0.78	-1.87	0.04	0.94	0.95
2nd	-0.95	0.27	1.00	0.99	-1.05	0.15	0.99	0.99	-1.25	0.12	1.00	0.96	-1.22	0.12	0.99	0.99
3rd	-0.94	0.15	1.00	0.97	-0.87	0.19	0.99	0.99	-0.50	0.30	1.00	1.00	-0.61	0.27	1.00	1.00
4th	-0.66	0.19	1.00	0.99	-0.63	0.27	1.00	1.00	-0.29	0.35	1.00	1.00	-0.54	0.30	1.00	1.00
5th	-0.53	0.28	1.00	0.99	-0.63	0.26	0.99	0.99	-0.27	0.29	1.00	1.00	-0.54	0.29	1.00	1.00
Median																
5th	2.07	0.03	0.89	0.09	2.35	0.01	0.02	0.01	1.77	0.07	0.96	0.21	2.52	0.01	0.01	0.01
4th	2.99	0.00	0.72	0.01	2.71	0.00	0.01	0.01	2.94	0.00	0.76	0.02	2.93	0.00	0.01	0.00
3rd	4.32	0.00	0.47	0.00	3.77	0.00	0.00	0.00	4.68	0.01	0.39	0.00	3.61	0.00	0.00	0.00
2nd	4.72	0.00	0.64	0.04	3.91	0.00	0.00	0.01	5.48	0.00	0.49	0.02	4.37	0.00	0.00	0.00
Best	4.90	0.12	0.88	0.40	4.44	0.00	0.06	0.05	7.98	0.11	0.50	0.19	5.76	0.00	0.02	0.01

Table 8. Continued.

	Unconditional								Conditional alpha and beta							
Benchmark model	γ	p-val.	boot.RF p-val.	boot.R p-val.	t_γ	p-val.	boot.RF p-val.(t)	boot.R p-val.(t)	γ	p-val.	boot.RF p-val.	boot.R p-val.	t_γ	p-val.	boot.RF p-val.(t)	boot.R p-val.(t)
Panel C: Four-factor model with net returns																
Worst	-0.53	0.29	1.00	1.00	-0.73	0.23	1.00	1.00	-3.11	0.05	1.00	0.53	-2.00	0.03	0.95	0.94
*2*nd	-0.39	0.23	1.00	1.00	-0.57	0.29	1.00	1.00	-2.37	0.03	1.00	0.41	-1.69	0.05	0.93	0.92
*3*rd	-0.32	0.36	1.00	1.00	-0.35	0.36	1.00	1.00	-0.58	0.24	1.00	1.00	-0.75	0.23	1.00	1.00
*4*th	-0.31	0.39	1.00	1.00	-0.27	0.39	1.00	1.00	-0.49	0.27	1.00	1.00	-0.70	0.24	1.00	1.00
*5*th	-0.23	0.40	1.00	1.00	-0.25	0.40	1.00	1.00	-0.40	0.35	1.00	1.00	-0.62	0.27	1.00	1.00
Median																
*5*th	2.47	0.01	0.83	0.01	2.25	0.01	0.04	0.02	1.55	0.15	0.98	0.27	2.05	0.02	0.11	0.13
*4*th	2.53	0.01	0.92	0.06	2.29	0.01	0.10	0.06	1.67	0.13	0.99	0.38	2.10	0.02	0.22	0.25
*3*rd	3.06	0.01	0.91	0.07	2.47	0.01	0.14	0.11	2.49	0.02	0.97	0.14	2.14	0.02	0.40	0.44
*2*nd	4.37	0.01	0.79	0.06	2.50	0.01	0.37	0.32	4.04	0.01	0.85	0.04	2.26	0.01	0.59	0.63
Best	5.47	0.11	0.83	0.31	4.56	0.00	0.05	0.04	4.60	0.02	0.94	0.26	4.17	0.00	0.13	0.14
Panel D: Four-factor model with gross returns																
Worst	-0.54	0.28	1.00	1.00	-0.75	0.23	1.00	1.00	-3.11	0.06	1.00	0.52	-1.99	0.03	0.96	0.95
*2*nd	-0.40	0.23	1.00	1.00	-0.58	0.28	1.00	1.00	-2.36	0.03	1.00	0.41	-1.69	0.06	0.93	0.93
*3*rd	-0.34	0.35	1.00	1.00	-0.38	0.35	1.00	1.00	-0.61	0.23	1.00	1.00	-0.77	0.22	1.00	1.00
*4*th	-0.32	0.39	1.00	1.00	-0.28	0.39	1.00	1.00	-0.48	0.28	1.00	1.00	-0.74	0.23	1.00	1.00
*5*th	-0.26	0.39	1.00	1.00	-0.28	0.39	1.00	1.00	-0.41	0.35	1.00	1.00	-0.60	0.28	1.00	1.00
Median																
*5*th	2.50	0.01	0.82	0.02	2.25	0.01	0.03	0.02	1.67	0.14	0.98	0.17	2.05	0.02	0.13	0.12
*4*th	2.53	0.01	0.92	0.07	2.32	0.01	0.08	0.05	1.81	0.30	0.99	0.25	2.12	0.02	0.24	0.22
*3*rd	3.09	0.01	0.91	0.07	2.45	0.01	0.15	0.09	2.52	0.02	0.97	0.13	2.17	0.02	0.41	0.41
*2*nd	4.35	0.01	0.80	0.06	2.53	0.01	0.34	0.28	4.02	0.01	0.88	0.06	2.25	0.01	0.62	0.62
Best	5.59	0.09	0.82	0.27	4.55	0.00	0.05	0.04	4.68	0.02	0.94	0.25	4.17	0.00	0.14	0.14

References

[1] Amihud, Y. and C.M. Hurvich, 2004. Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39, 813–841.

[2] Asmundson, P. and S.R. Foerster, 2001. Socially responsible investing: Better for your soul or your bottom line? *Canadian Investment Review*, 14, Winter.

[3] Ayadi, M.A. and L. Kryzanowski, 2005. Portfolio performance measurement using APM-free kernel models. *Journal of Banking and Finance*, 29, 623–659.

[4] Ayadi, M.A. and L. Kryzanowski, 2010. Fixed-income fund performance: Impact of sponsor ownership and role of luck and ability in tail membership. *Working Paper*, Brock University.

[5] Ayadi, M.A., H. Ben-Ameur, S. Lazrak, and Y. Wang, 2010. Canadian investors and the discount on closed-end funds. *Working Paper*, Brock University.

[6] Bauer, R., J. Derwall, and R. Otten, 2007. The ethical mutual funds performance debate: New evidence from Canada. *Journal of Business Ethics*, 70, 111–124.

[7] Bauer, R., K. Koedijk, and R. Otten, 2005. International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance*, 29, 1751–1767.

[8] Bauer, R., R. Otten, and A. Tourani Dan, 2006. Ethical investing in Australia: Is there a financial penalty? *Pacific-Basin Finance Journal*, 14, 33–48.

[9] Bensimon Byrne, 2010. *Consumerology Report*, August. Available at: http://www.consumerology.ca/pdf/Consumerology_Aug2010.pdf

[10] Benson, L. and J.E. Humphrey, 2008. Socially responsible investment funds: Investor reaction to current and past returns. *Journal of Banking and Finance*, 32, 850–859.

[11] Bollen, N.P., 2007. Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis*, 42, 683–708.

[12] Carhart, M.M., 1997. On persistence in mutual fund performance. *Journal of Finance*, 52, 57–82.

[13] Carrick, R., 2006. At least the name fits: The problem with SRI funds. *Corporate Knights*.

[14] Chen, Z. and P.J. Knez, 1996. Portfolio performance measurement: Theory and applications. *Review of Financial Studies*, 9, 511–555.

[15] Christopherson, J.A., W.E. Ferson, and D.A. Glassman, 1998. Conditioning manager alphas on economic information: Another look at the persistence of performance. *Review of Financial Studies*, 11, 111–142.

[16] Cortez, M.C., F. Silva, and N. Areal, 2009. The performance of European socially responsible funds. *Journal of Business Ethics*, 87, 573–588.

[17] Cuthbertson, K., D. Nitzsche, and N. O'Sullivan, 2008. UK mutual fund performance: Skill or luck? *Journal of Empirical Finance*, 15, 613–634.

[18] Derwall, J. and K. Koedijk, 2009. Socially responsible fixed-income funds. *Journal of Business Finance and Accounting*, 36, 210–229.

[19] Ding, B. and R. Wermers, 2009. Mutual fund performance and governance structure: The role of portfolio managers and boards of directors. *Working Paper*, SUNY.

[20] Dybvig, P.H., and S.A. Ross, 1985. Performance measurement using differential information and a security market line. *Journal of Finance*, 40, 483–496.

[21] Elliot, G. and J.H. Stock, 1994. Inference in time-series regression when the order of integration of a regressor is unknown. *Econometric Theory*, 10, 672–700.

[22] Fama, E.F. and K.R. French, 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3–56.

[23] Fama, E.F. and K.R. French, 2010. Luck versus skill in the cross section of mutual fund alpha estimates. Journal of Finance, *Forthcoming.*

[24] Ferguson, R., 1980. Performance measurement doesn't make sense. *Financial Analysts Journal*, 36, 59–69.

[25] Fernandez-Izquierdo, A. and J.C. Matallin-Saez, 2007. Performance of ethical mutual funds in Spain: Sacrifice or premium. *Journal of Business Ethics*, 81, 247–260.

[26] Ferson, W.E., S. Sarkissian, and T. Simin, 2003. Is stock return predictability spurious? *Journal of Investment Management*, 1, 1–10.

[27] Ferson, W.E. and R. Schadt, 1996. Measuring fund strategy and performance in changing economic conditions. *Journal of Finance*, 51, 425–461.

[28] Geczy, C., R. Stambaugh, and D. Levin, 2005. Investing in socially responsible mutual funds. *Working Paper*, University of Pennsylvania.

[29] Gil-Bazo, G., P. Ruiz-Verdú, and A.A.P. Santos, 2010. The performance of socially responsible mutual funds: The role of fees and management companies. *Journal of Business Ethics*, 94, 243–263.

[30] Gregory, A. and J. Whittaker, 2007. Performance and performance persistence of 'ethical' unit trusts in the UK. *Journal of Business Finance and Accounting*, 34, 1327–1344.

[31] Grinblatt, M. and S. Titman, 1989. Portfolio performance evaluation: Old issues and new insights. *Review of Financial Studies*, 2, 393–422.

[32] Hall, P. and R. LePage, 1996. On bootstrap estimation of the studentized mean. *Annals of the Institute of Statistical Mathematics*, 48, 403–421.

[33] Henriksson, R.D. and R.C. Merton, 1981. On market timing and investment performance II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, 54, 513–533.

[34] Horowitz, J.L., 2001. The bootstrap in econometrics. *Handbook of Econometrics*, 5, 3159–3228.

[35] Huij, J. and J. Derwall, 2008. Hot hands in bond funds. *Journal of Banking and Finance*, 32, 559–572.

[36] Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. *Journal of Finance*, 23, 389–346.

[37] Jones, S., S. van der Laan, G. Frost, and J. Loftus, 2008. The investment performance of socially responsible investment funds in Australia. *Journal of Business Ethics*, 80, 181–203.

[38] Kosowski, R. A. Timmermann, R. Wermers, and H.L. White, 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance*, 61, 2551–2595.

[39] Kosowski, R., N.Y. Naik, and M. Teo, 2007. Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *Journal of Financial Economics*, 84, 229–264.

[40] Kreander, N., G. Gray, D.M. Power, and C.D. Sinclair, 2005. Evaluating the performance of ethical and non-SRI funds: A matched pair analysis. *Journal of Business Finance and Accounting*, 32, 1465–1493.

[41] Kryzanowski, L., S. Lalancette, and M.C. To, 1997. Performance attribution using an APT with prespecified macrofactors and time-varying risk premia and betas. *Journal of Financial and Quantitative Analysis*, 32, 205–224.

[42] Lehmann, B.N. and D.M. Modest, 1987. Mutual funds performance evaluation: A comparison of benchmarks and benchmarks comparisons. *Journal of Finance*, 42, 233–265.

[43] Luther, R. J. Matatko, and D. Corner, 1992. The investment performance of UK ethical unit trusts. *Accounting Auditing and Accountability Journal*, 5, 57–70.

[44] Mallin, C.A., B. Saadouni, and R.J. Briston, 1995. The financial performance of ethical investment funds. *Journal of Business Finance and Accounting*, 22, 483–496.

[45] Newey, W.K. and K.D. West, 1987a. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 703–708.

[46] Newey, W.K. and K.D. West, 1987b. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28, 777–787.

[47] Renneboog, L., J. ter Horst, and C. Zhang, 2006. Is ethical money financially smart? *Working Paper*, Tilburg University and European Corporate Governance Institute.

[48] Renneboog, L., J. ter Horst, and C. Zhang, 2007. Socially responsible investments: Methodology, risk, and performance. *Working Paper*, Tilburg University and European Corporate Governance Institute.

[49] Renneboog, L., J. ter Horst, C. Zhang, 2008a. The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance*, 14, 302–322.

[50] Renneboog, L., J. ter Horst, C. Zhang, 2008b. Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking and Finance*, 32, 1723–1742.

[51] Rosenberg, B., 1976. Security appraisal and unsystematic risk in institutional investment. *Proceedings of the Seminar on the Analysis of Security Prices*. Chicago: University of Chicago Press, November, 171–237.

[52] Schroder, M., 2004. The performance of socially responsible investments: Investment funds and indices. *Financial Markets and Portfolio Management*, 18, 122–142.

[53] Social Investment Organization, 2007. A survey of Canadian mutual funds on proxy voting (Toronto, Ontario). Available at: http://www.socialinvestment.ca/documents/Voting.pdf.

[54] Stambaugh, R.F., 1999. Predictive regressions. *Journal of Financial Economics*, 54, 375–421.

[55] Treynor, J. and F. Black, 1973. How to use security analysis to improve on portfolio selection. *Journal of Business*, 46, 66–86.

[56] Treynor, J. and K. Mazuy, 1966. Can mutual funds outguess the market? *Harvard Business Review*, 44, 131–136.