J Financ Serv Res (2016) 50:57–94
DOI 10.1007/s10693-015-0215-0

Typical and Tail Performance of Canadian Equity SRI Mutual Funds

Mohamed A. Ayadi · Hatem Ben-Ameur · Lawrence Kryzanowski

Received: 25 August 2014 / Revised: 6 February 2015 / Accepted: 18 February 2015 /
Published online: 1 April 2015

Abstract While no evidence for timing ability is identified, selection performance before (after) management-related costs for a comprehensive and survivorship-free sample of Canadian equity SRI funds is significantly positive (insignificant) and not statistically different from that for non-SRI funds. Conditioning and multifactor benchmarking improve selection performance. Based on block-bootstrap tests, luck (and not ability), or the lack thereof, is associated with fund membership in the tails of the cross-sectional selection and timing performance distributions. Accounting for the effects of cross-correlations changes inferences about the interpretation of the significance of traditionally calculated t-values.

Keywords Mutual funds · Socially responsible investment (SRI) · Benchmark models · Block bootstrap · Survivorship bias

JEL classification G15 · G31 · G12

1 Introduction

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 number has been rapid over the past 20 years worldwide. AUMs for Canadian retail mutual funds under SRI guidelines has remained unchanged

M. A. Ayadi (✉)
Department of Finance, Operations, and Information Systems, Goodman School of Business, Brock University, St. Catharines, ON, Canada
e-mail: mayadi@brocku.ca

H. Ben-Ameur
Department of Management Sciences, HEC Montreal, Montreal, QC, Canada
e-mail: hatem.ben-ameur@hec.ca

L. Kryzanowski
Department of Finance, John Molson School of Business, Concordia University, Montreal H3G 1M8 QC, Canada
e-mail: lawrence.kryzanowski@concordia.ca

from 2004 to 2011 at 4.4 billion CDN but is down from 5.5 billion CDN in 2008 (SIO 2013). The corresponding AUMs under SRI guidelines for all Canadian funds of 57.9, 600.9 and 566.7 billion CDN in 2004, 2011 and 2008, respectively, represent 3.2 %, 20.1 % and 20.4 % of total Canadian AUM in each of those years, respectively. As one of the largest retail markets in the world (Social Investment Organization 2000; Bauer et al. 2007) and similar to Germany in AUM at the end of 2010 (Steiauf and Schäfer 2014), the Canadian SRI mutual fund industry has attracted little attention in the literature. As subsequently discussed below, those typically reported in the literature are for non-survivorship-free small samples of SRI funds.

Given its similarities and differences with the U.S. and Europe in terms of legal, institutional and regulatory environments and tolerances for industry concentration and state involvement (Brean et al. 2011), Canada provides an ideal alternative laboratory for examining the effect (if any) of costs, scale, competition and managerial ability on the value of fund provision and management from the perspective of retail investors in SRI and SRI versus non-SRI equity funds. For example, while the U.S. and Canada have similar governance mechanisms and minority shareholder protections, regulatory oversight and enforcement is principle-based in Canada versus rule-based in the U.S. Similarly, the opportunity set of Canadian equity investments also differs from that in the U.S. due to a much greater concentration in a few industrial sectors (e.g., natural resources with their greater possibility of being non-SRI) and to a greater prevalence of small firms and public firms with controlling or concentrated shareholdings, families as the ultimate controlling shareholders, restricted or subordinated voting shares issued, and pyramidal structures (e.g., Gadhoum et al. 2005; King and Santor 2008; Nicholls 2006). Unlike in most countries but like the U.S., fund competition is restricted by not permitting foreign-domiciled funds to register for sale domestically. Like most countries but unlike the U.S., fund management services are subject to domestic consumption taxes in Canada that are paid on most MER components (i.e., management fees and nearly all operating expenses) and the Canadian distribution model uses financial advisors selling and servicing no-load funds (Alpert and Rekenthaler 2011, p. 13). Compared to the U.S., Canadian mutual funds are not required to have a Board of Directors and they realize lower benefits from economies of scale compared to their U.S. counterparts due to both their sponsors and AUM being much smaller. As a result, Canadian funds are ranked as not only continuing to have the highest fees and expenses internationally but as having considerably higher fees compared to the U.S. and much of Europe (Alpert et al. 2013, p. 13). While strong arguments exist in the literature (as reviewed in the next section and in Renneboog et al. 2008b) for both positive and negative impacts on corporate and investor performance of the effect of adhering to socially responsible objectives, we believe that one of the most important issues is whether or not this affects the performance of investor vehicles that implement SRI. If an SRI investment strategy does not negatively impact the risk-adjusted return performance of an investment fund, this should encourage even marginally conscious SR investors to invest in these funds. Unfortunately, the performance findings reported in the literature to date run the gamut from worse to the same to better.[1]

Poorer SRI fund performance is found by Jones et al. (2008) for Australian funds, Renneboog et al. (2008a) for many European and Asia-Pacific funds, 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. Neutral SRI fund performance is reported for funds that are

[1] The smart money effect for the relationship between SRI fund performance and money flows is also important. Major determinants of SRI fund flows are past returns and social screen strength (Renneboog et al. 2008b, 2011), although the fund-flow sensitivity to past and current returns is lower than that for conventional funds (Benson and Humphrey 2008).

Canadian (4 and 8 surviving SRI funds, respectively, in: Asmundson and Foerster 2001; Bauer et al. 2007), international (Bauer et al. 2005), Spanish (Fernandez-Izquierdo and Matallin-Saez 2007), German, Swiss and American (Schroder 2004), and European (Cortez et al. 2009).[2] Superior SRI fund performance is reported by Luther et al. (1992) and Mallin et al. (1995) for UK ethical unit trusts. Ding and Wermers (2012) 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 used for performance assessment.

Most of these studies fail to provide a strong justification for the appropriateness of their benchmark models and conduct their tests using net returns on small samples of active SRI funds using unconditional or *partial* conditional models, while focusing on the selection abilities of these funds with little or no attention given to tests of timing performance or whether winner and loser funds in the tails of the performance distribution are due to managerial ability or lack thereof. 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.[3]

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 SRI equity funds using frameworks suitable to evaluate fixed-weight and dynamic strategies.[4] To the best of our knowledge, we are the first to use *full* conditional selection and timing multifactor benchmark models for this purpose. To this end, we compute performance and risk measures using gross and net returns for a relatively large sample of 67 Canadian active and terminated SRI equity funds over the period 1988–2008, and compare their values to those of conventional funds at the individual fund and portfolio of funds levels. While we refer to equity funds that have not declared that they are SRI as non-SRI or conventional, this should not be interpreted as these funds not considering SR factors when making investment decisions. Our empirical tests address the following important questions: Do Canadian SRI and non-SRI mutual funds exhibit different performances before considerations of fees? Does the comparison change after the consideration of the management fees associated with a typical SRI fund?

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. To the best of our knowledge, this provides the first tests of the managerial skills of specific SRI funds that account for individual fund cross-dependencies, the special

[2] Derwall and Koedijk (2009) find similar and better performance relative to their non-SRI counterparts for SRI bond and balanced funds, respectively.

[3] The same method is used by Ayadi and Kryzanowski (2005, 2011) for samples of Canadian equity and fixed-income funds, by Cuthbertson et al. (2008) for a large sample of UK unit trusts, and by Fama and French (2010) to distinguish between luck and skill for US equity funds.

[4] 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 yields, interest rates, and default and term structure variables.

distributional features of fund returns and risks such as asymmetries and fat tails, and disentangles the effects of sampling variation or "luck" from superior/inferior management skills.[5] In the same vein, we develop various full conditional market-timing models and use the cross-section bootstrap method to address the following question: Do some SRI fund managers have the ability to time the return on the market? This extends the findings of the few studies that use standard statistical tests to examine the market-timing performance of SRI funds that other fund settings find lack robustness (e.g., Ayadi and Kryzanowski 2005, 2011; Kosowski et al. 2006, 2007).

This paper's first major finding is that the selection performance of Canadian equity SRI funds is significantly positive before fees (gross returns), insignificantly weak to neutral after fees (net returns), and improves with conditioning and multifactor benchmarking. No market-return timing performance is identified for Canadian equity SRI funds 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 at least in a Canadian setting are a legitimate investment alternative, especially for retail investors who integrate personal and societal values into their investment decisions.

The third major finding is that sampling variation or "bad" ("good") luck is associated with poor (good) performance of extreme SRI funds based on a block bootstrap analysis of equity funds in the tails of selection and timing performance distributions. This result is mostly robust to the presence of conditioning information, an alternative ranking scheme, and standard bootstrapping. Like all other untabulated and discussed results, the standard bootstrapped results are available if requested.

The remainder of the paper is organized as follows: In section 2, hypotheses based on the conceptual relation between social, environmental and ethical considerations and investment performance are developed. Section 3 presents the various benchmark models for selection 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 Development of the hypotheses based on SR considerations and investment performance

2.1 Conceptual background

Socially responsible investing (SRI) is an investment process that embodies social, environmental and ethical considerations into investment decision making with either passive or active follow-up. SRI appeals to investors that have multi-attribute utility functions based on standard return-risk attributes augmented with personal and societal attributes (Bollen 2007).

Ethical SRI investing has ancient origins in teachings of various religions. Ethical restrictions on loans and investments based on the Old Testament existed in medieval Christian

[5] Kim (2014) discusses the theory of skill versus luck in the assessment of CEO performance. She contends that measures of a CEO's performance can be very noisy even after accounting for systematic economic events over which the CEO has no control because the CEO could have been lucky or unlucky. Hence luck can produce false positive or false negative performance inferences when judging CEO performance.

times. The Catholic Church prohibited usury in 1139 and relaxed the prohibition in the 19th century. Based on the Torah and its interpretations, Judaism has many teachings on how to use money ethically. Similarly, based on the Koran and its interpretations, Islamic investors avoid investments in operations involved in gambling, pornography, pork production and interest-based financial institutions.

In contrast to the religious-motivated ethical investing in the more distant past, modern SRI has a greater grounding in the ethical and social convictions of individual investors. SRI fund development has evolved so that we have four generations of SRI funds in terms of investment screening mechanisms. The first generation is based on negative (or exclusionary) screening of, for example, removing all stocks in the alcohol, tobacco, gambling, nuclear and defense industries from the selection choice set. The second generation relies on positive screens where firms with the "best-of-the-class" CSR (Corporate Social Responsibility) standards in terms of, for example, corporate governance, environment, sustainability, labor relations and cultural diversity, are chosen. The third generation uses both negative and positive screens to determine what is included in their SRI funds. The fourth generation is a third generation SRI fund, which attempts to effect change through moral suasion or exercising its shareholder vote based on a "focus list" which is a set of themes against which companies are measured. In 2009, focus list issues included climate change, biodiversity, safeguarding water, human rights, sweatshops, respecting indigenous rights, engaging communities and building corporate transparency. In terms of 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).

In an experimental setting, Hofmann et al. (2008) test the suitability of three models to explain moral behavior; namely, a multiple attribute utility theory (MAUT), a theory of planned behavior, and an issue-contingent model of ethical decision making in organizations. Controlling for profits, they find that moral considerations influence investment decisions. Save-Soderbergh (2010) finds that human capital, being female and being in an "empathetic" profession are predictors of ethical decision-making while income and wealth are not.

The SRI movement is predicated on the belief that shareholder and societal value maximization can differ. Unlike classical economics (e.g., Adam Smith's invisible hand), modern economic theory finds that profit-maximizing behavior does not imply social-welfare maximization, when (for example) externalities exist. Proponents of CSR argue that CSR plays a role in reducing the costs of societal conflicts (Heal 2005), furthers the completion of competitive product markets with no adverse welfare effects on those not interested in those products, leads to higher firm value by softening competition in product markets (Allen et al. 2009), improves a firm's reputation by signaling its product quality, safety and usefulness (Fombrun and Shanley 1990), reduces monitoring costs by signaling information about the quality of management (Akpinar et al. 2008), acts as a screening device to attract motivated workers due to more responsible employment practices (Brekke and Nyborg 2005), and relieves pressure from lobbyists (Baron 2001). 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 favorable ones are helpful" (Bensimon Byrne 2010, p. 5).

Critics of stakeholder-value maximization point out that under such an objective function managerial performance becomes unaccountable (Jensen 2002; Tirole 2001), the firm by not profit-maximizing may not survive competition and market disciplinary actions for corporate control (e.g., Tirole 2001; not supported by Bagnoli and Watts 2003), unethical corporate behavior may be enhanced (Shleifer 2004), such firms may receive less coverage from financial analysts and have smaller shareholder bases (Hong and Kacperczyk 2009) although Ioannou and Serafeim (2010) find that coverage has improved more recently, and may result in agency problems where insiders over-invest in CSR strategies for their private benefit (Barnea and Rubin 2010).

Whether or not there is a cost to SRI is still a hotly debated topic (e.g., Adler and Kritzman 2008). Richardson and Crag (2010) argue that significant tensions can exist between the goals of being virtuous and prosperous, and that "SRI (and indeed all investment) should not allow the pursuit of maximizing investment returns to prevail over an ethical agenda of promoting social and economic justice and environmental protection". 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". Other observers (e.g., Carrick 2006) hold a neutral position where they argue SRI funds have similar performances, fees and investment strategies as non-SRI funds.

The theory underlying the possible outperformance of SRI funds over non-SRI funds despite being constrained in their portfolio choices includes the arguments proposed by Gil-Bazo et al. (2010). First, the choice set for fund managers typically is much smaller than the investment universe, and it involves a trade-off between breadth (number of firms followed) and depth (how closely they are followed). Breadth reduction is attractive if depth is relatively more profitable than breadth (Van Nieuwerburgh and Veldkamp 2010). Evidence consistent with this conjecture finds that better performance is associated with funds exposed to more idiosyncratic risk (Amihud and Goyenko 2013), with fund families that follow more focused investment strategies (Nanda et al. 2004) and mutual funds holding portfolios concentrated in specific industries (Kacperczyk et al. 2005). Second, SRI constraints could enhance performance if breadth reduction reduces excessive trading and their associated trading costs which directly affect before-MER returns as they are not captured in fund MERs. Third, if SRI funds are associated with better fund governance, and if agency problems between investors (who seek high risk-adjusted returns) and fund managers (who want to maximize fee revenues net of management costs) have a significant effect on performance, then SRI funds could exhibit better performance than conventional funds. Finally, the requirements to be classified as a SRI fund are often not very stringent since it could include a fund that just has a formal policy of excluding companies with interests in the tobacco business.

2.2 Hypotheses

The equilibrium accounting argument of Fama and French (2010) is that funds in the aggregate have neutral performance before management expenses and negative performance after such expenses. Since SRI fund performance can be positively or negatively affected by the constraints that SRI imposes on portfolio choice and is adversely affected by management expenses, the investment return cost of being virtuous through investment in SRI funds needs to be addressed at two levels (e.g., Berk and van Binsbergen 2013). The first level examines the gross return performance of the SRI funds to determine if their average performance based on managerial abilities before reflecting the confounding effects of management expenses differ from that of conventional funds. This first level is examined with the following hypothesis in its alternative form, H_1^A: SRI funds before management expenses underperform their benchmarks and their conventional funds counterparts. The second level tests net return performances to determine if the average performance after management expenses of SRI funds differs from that of conventional funds because any incremental costs associated with SRI fund management (e.g., screening) leads to poorer relative risk-adjusted net returns. The alternative form of the second hypothesis (H_2^A) used to test the incremental impact of higher fees on performance states that the average SRI fund after management expenses underperforms its benchmarks and its conventional fund counterpart.

Various sources (e.g., Nelson 2014; http://www.fundata.com/) periodically provide lists of top and bottom performing funds, fund managers and fund management companies annually receive "best-in-the-class" awards where fund performance and related criteria appear to be the primary determinants of such awards (e.g., Morningstar at http://www.investmentawards.com), and such information is often included in the information disclosures of individual funds (e.g., Mawer 2014). In order to assess the forward-looking value of such information for SRI funds, retail investors need to know if the small group of SRI funds found in either tail of the performance distribution of individual funds is mainly explained by luck (bad or good) or sampling variation. The hypothesis in its null form used to address this issue is H_3^0: managers of funds are in the left (right) tail of the performance distribution due to bad (good) luck or sampling variation and not managerial ability.

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- (EW) and total net asset value or size-weighted (SW) portfolios of SRI and non-SRI funds constructed using individual fund returns. These portfolios of funds provide evidence on potential size effects on 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.

The second approach uses all individual funds with at least 30 monthly observations given the increasing dimension of some of the conditional versions of the multi-factor models used herein. It also uses the bootstrap sampling methodology including bootstrapping blocks of overlapping observations to allow for some persistence in fund returns (Künsch 1989). This robust method is an alternative for dealing with possible fund return non-normality 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, 2011), Kosowski et al. (2006, 2007), Huij and Derwall (2008), and Cuthbertson et al. (2008). Bootstrapped statistics are constructed for all and 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. 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). Two resampling schemes are adopted, residuals only and independently of both the residuals and the factors which is sufficient to ensure the required independence assumption between errors and regressors. A detailed description of the bootstrap approach with all data stages is provided in section 5.

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.[6] 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), partial conditional models (Schroder 2004; Bauer et al. 2006, 2007; Renneboog et al. 2007, 2008a) and full conditional one-factor model (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, for example, 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 Kryzanowski et al. (1994) and 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:

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

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

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

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.

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:

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

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), UMD_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 % 11-month return and the other in the lowest 30 % 11-month return, both lagged 1 month); 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. Partial (not full) conditional 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, including seven from Canada.

Our two timing models accommodate multi-factor benchmarks as in Lehmann and Modest (1987) and conditioning information for alpha and the market beta as in Christopherson et al. (1998), and are given by:

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

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

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 violations of the normality assumption underlying the performance tests and inferences,[7] it accommodates possible nonlinearities in fund returns and does not require the estimation of the complex joint distribution of performances across all mutual funds, and it can deal with time-series dependencies in the data through extensions of the basic setup. All of our experiments explicitly control for luck in performance measurement. In other words, they address the question whether a small group of SRI fund managers (in the tails of the performance distribution) possess genuine stock picking or market timing skills or whether such extreme performance is mainly explained by luck or sampling variation. That is, they provide a test of the third hypothesis presented earlier.

The implementation of the nonparametric bootstrap resampling approach follows the lines of Ayadi and Kryzanowski (2005, 2011) and Kosowski et al. (2006, 2007) and is illustrated here for the case with independent resampling of residuals and factors based on the four-factor performance model (2) given above. Nevertheless, the bootstrap method is applied to all of our unconditional and conditional benchmark models. 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) $\left\{\widehat{\alpha}_i, \widehat{t}_{\alpha_i}, \widehat{\beta}_{ik}, k=1,\ldots,4\right\}$ as well as the time-series of estimated residuals $\{\hat{u}_{i,t}, t=1,\ldots, T_i\}$.

Second, we independently resample (with replacement) B times ($b=1,2,\ldots,B=2000$ *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) $\left\{\widehat{u}^b_{i,t}, t=v^b_1, v^b_2,\ldots,v^b_{T_i}\right\}$ and resampled factors $\left\{r^b_{M,t}, \mathrm{SMB}^b_t, \mathrm{HML}^b_t, \mathrm{UMD}^b_t, t=\tau^b_1,\tau^b_2,\ldots,\tau^b_{T_i}\right\}$ where $t=v^b_1, v^b_2, \ldots, v^b_{T_i}$ and $t=\tau^b_1,\tau^b_2,\ldots,\tau^b_{T_i}$ 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*:

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

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

[7] 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.

benchmark regression model. Any positive or negative estimated alpha is entirely due to sampling variation.

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

For a given bootstrap iteration b, we obtain a cross-sectional distribution of the alpha estimates $\left(\widehat{\alpha}_1^b,\ \ \widehat{\alpha}_2^b,\ \ldots,\widehat{\alpha}_N^b\right)$ and of the t-statistics of these estimates $\left(\widehat{t}_{\alpha_1}^b,\widehat{t}_{\alpha_2}^b,\ \ldots,\widehat{t}_{\alpha_N}^b\right)$ that can be both ranked from the minimum or worst value $\left(\widehat{\alpha}_{\min}^b;\widehat{t}_{\min}^b\right)$ to the maximum or best value $\left(\widehat{\alpha}_{\max}^b;\widehat{t}_{\max}^b\right)$. This step is performed for all iterations ($b=1,\ 2,\ldots,\ 2000$) 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 $^{T_i}/_{\ell}$ blocks (the smallest integer greater or equal to $^{T_i}/_{\ell}$) with $\ell=3$ overlapping monthly observations. The resampled blocks from the residuals and regressors are used to construct the bootstrapped time-series dynamics $r_{i,t}^{\ell,b}$. 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) $\left\{\widehat{\alpha}_i,\widehat{t}_{\alpha_i},\widehat{\gamma}_i,\widehat{t}_{\gamma_i},\widehat{\beta}_{ik},\ \ k=1,\ \ldots,4\right\}$ as well as the time-series of estimated residuals $\{\widehat{u}_{i,t},\ t=1,\ldots,T_i\}$.

Second, we independently resample (with replacement) B times ($b=1,\ 2,\ldots,B$) the saved fund's residuals from the first step and the four factors. We then generate a time-series of resampled residuals (from a zero-mean noise) $\left\{\widehat{u}_{i,t}^b,\ \ t=v_1^b,v_2^b,\ \ldots,v_{T_i}^b\right\}$ and resampled factors $\left\{r_{M,t}^b,\ \ \mathrm{SMB}_t^b,\mathrm{HML}_t^b,\mathrm{UMD}_t^b,\left(r_{M,t}^2\right)^b,\ \ t=\tau_1^b,\tau_2^b,\ldots,\tau_{T_i}^b\right\}$ where $t=v_1^b,v_2^b,\ldots,$ $v_{T_i}^b$ and $t=\tau_1^b,\tau_2^b,\ \ldots,\tau_{T_i}^b$ 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 $\left(\widehat{\alpha}_i^b,\ \widehat{\gamma}_i^b, i=1,\ldots,N\right)$ and the corresponding t-statistics $\left(\widehat{t}_{\alpha_i}^b, \widehat{t}_{\gamma_i}^b\ \ , i=1,\ldots,N\right)$.

For a given bootstrap iteration b, we obtain a cross-sectional distribution of the alpha or gamma estimates $\left(\widehat{\alpha}_1^b,\ \widehat{\alpha}_2^b,\ldots,\widehat{\alpha}_N^b\right)$ and $\left(\widehat{\gamma}_1^b,\ \widehat{\gamma}_2^b,\ldots,\widehat{\gamma}_N^b\right)$, and of the t-statistics of these estimates $\left(\widehat{t}_{\alpha_1}^b, \widehat{t}_{\alpha_2}^b,\ldots,\widehat{t}_{\alpha_N}^b\right)$ and $\left(\widehat{t}_{\gamma_1}^b, \widehat{t}_{\gamma_2}^b,\ldots,\widehat{t}_{\gamma_N}^b\right)$ 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

We only examine funds in the Canadian Equity category given their relatively long return histories and to facilitate comparisons with previous studies. Funds in this category must invest at least 90 % of their equity holdings in securities domiciled in Canada, and their average market capitalization must be greater than the Canadian small/mid cap threshold. To control for selection and survival biases, we include all active and terminated funds in our portfolio tests. To accommodate the increasing dimension of some of the conditional versions of the multi-factor models used herein, we exclude funds without at least 30 monthly observations when examining individual fund performance.

We carefully construct two different samples of SRI and non-SRI Canadian equity mutual funds 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 funds. 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.

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 %). 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 %. 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.

Summary statistics for equal- (EW) and size-weighted (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 8.70 % 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.

Panel C shows descriptive statistics of various fund characteristics such as size (TAV or total net assets of the fund), MER (management expense ratio), management fees (MGF), flow (as suggested by Frazzini and Lamont 2008) and age (since inception) for the ethical and conventional groups. These statistics suggest that, on average, non-SRI funds are larger ($92.63 vs. $86.53 millions), older (6.22 vs. 5.07 years), and with higher MER (2.02 % vs. 1.97 %) and MGF (1.83 % vs. 1.60 %) and less monthly flows ($55.12 vs. $97.16 millions) than SRI funds. However, the only differences that are significant are for MGF and age.

The MER and MGF results are especially interesting because they are counter to the general expectation in the literature that fund expenses are higher for SRI versus non-SRI mutual funds. Such expenses are expected in the literature to be higher for SRI versus non-SRI mutual funds for one or more of the following reasons: (1) SRI funds incur additional monitoring costs of the firms in which they invest to ensure that they maintain socially responsible policies (Gil-Bazo et al. 2010, p. 251); (2) investors in SRI funds are likely to be less performance sensitive (Gil-Bazo et al. 2010, p. 251) and studies find that fees are inversely related with investor performance sensitivity (Christoffersen and Musto 2002; Gil-Bazo and Ruiz-Verdu 2009); and (3) SRI funds may have higher costs since the smaller size of their sponsors and their assets under management (AUM) lead to less economies from scale (as found by Bauer et al. 2005, for German and UK funds, and by Bauer et al. 2006, for Australian funds).

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 benchmark for the performance of SRI funds is the Jantzi social index (JSI)[8] 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

[8] 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).

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. 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. Panel C shows basic statistics on fund characteristics (TAV, MER, MGF, Flow, and Age). TAV is total net assets of the fund in $millions. MER is the fund's annual management expense ratio in %. MGF is the fund's annual management fees in %. Flow is the fund flows as suggested by Frazzini and Lamont (2008) in $millions. Age is the fund's age as given by the fund launch date (years). 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)		

Panel C: Fund characteristics

Fund group	Statistics	Mean	Std. Dev.	Minimum	Maximum
Ethical funds;(N=67)	TAV	86.53	305.91	0.001	1981.66
	MER	1.97	1.32	0.25	8.71
	MGF	1.60	0.67	0.50	3.25
	Flow	97.16	551.31	0.001	3984.21
	Age	5.07	5.02	0.58	22.17
Conventional funds;(N=517)	TAV	92.63	258.81	0.003	2660.70
	MER	2.02	0.99	0.06	11.20
	MGF	1.83	0.91	0.45	3.88
	Flow	55.12	359.66	−14.50	2114.97
	Age	6.22	5.23	0.08	22.17

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. (2013). Two instrumental variables obtained from the CANSIM database due to their power to predict stock returns are used in all of the conditional models for which we report tabulated results. The variables are the lagged values of DY or the dividend yield of the S&P/TSX index and TB or the 1-month Treasury bill rate.[9] Since the two instruments exhibit high degrees of persistence, they are stochastically detrended by subtracting a moving average over a period of 2 months, as in Campbell (1991) and Ferson et al. (2003). 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; Valkanov 2003; Torous et al. 2004). To allow for a simple interpretation of the estimated coefficients, the variables are demeaned in the conditional tests, as in Ferson and Schadt (1996).

Nevertheless, similar inferences are drawn from untabulated test results using several other commonly used conditioning variables. They include the 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, the slope of the term structure as measured by the yield spread between long-term government of Canada bonds and the one period lagged 3-month Treasury bill rate, and a dummy variable for the month of January.

To test if the conditional methodology is likely to be worthwhile, a predictability analysis of the excess returns for the EW and 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. 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 a SRI-specific index and compared to non-SRI funds for our survivorship-free sample of Canadian equity SRI mutual funds. Our analysis first examines if managers add value before management expenses (using gross returns) to test the first alternative hypothesis and then if any of this added value flows to investors after management expenses (using net returns) to test the second alternative hypothesis. 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

[9] 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).

and its variants through the third testable hypothesis. This is done to better identify managerial skill by accounting for individual fund cross-dependencies and to disentangle the effects of sampling variation or "luck" from superior/inferior management abilities.

6.1 Performance results for portfolios of funds

The empirical findings for stock selection (alphas) and timing (gamma) performances for our portfolios of SRI and non-SRI funds are reported in Table 3. Before proceeding to a discussion of the estimated alphas, we observe that none of the estimated gammas are significant at conventional levels.

6.1.1 Security selection performance before management expenses

Before examining the more detailed empirical findings on selection performance using returns before management expenses or gross returns, we observe that these performance results are inconsistent with the first alternative hypothesis. Portfolios of SRI funds outperform and do not underperform their benchmarks, and they outperform (but not significantly) their corresponding portfolios of non-SRI funds. Thus, constraints on the investment choice of SRI funds appear to neither hinder nor aid their relative investment performances.

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 1-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 2 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.	ρ_1	ρ_3	ρ_6	ρ_{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

An examination of the detailed findings reported in Table 3 shows that EW and SW portfolios of SRI funds display positive and significant performances across all benchmark models with the exception of the SW portfolios using the one-factor benchmarks. For example, the estimated monthly alphas of the SW portfolios range from 0.1043 % (1.25 % annually) using the single factor CAPM model to 0.2592 % (3.11 % annually) using the full conditional four-factor benchmark. Since fund returns do not incorporate management expenses, the gross-return alpha evidence suggests that SRI fund managers exhibit some security selection skills despite or because of their restricted investment opportunity set. With conditioning, the increase in the four- and one-factor benchmarks alphas are, respectively, 0.29 and 0.06 % annually for the SW SRI portfolios and 0.46 and 0.12 % annually for the EW SRI portfolios. A comparison of the alphas of these two types of portfolios suggests that smaller funds outperform (underperform) larger funds using the one- (four-) factor benchmark specification.

The non-SRI fund portfolios exhibit significant positive alphas with the exception of the unconditional one-factor model. These metrics are insignificantly lower based on Wald tests (all p-values≥0.18) than those of the SRI funds across all benchmark models. The performance differential increases with conditioning and is somewhat stable across all models and ranges from 0.01 to 1.28 % annually. The “superior” performance of SRI fund portfolios over non-SRI fund portfolios is maximized (minimized) for SW portfolios with the full conditional four- (one-) factor benchmark. As for the SRI fund portfolios, the performances of the non-SRI fund portfolios improve with conditioning and with the four-factor model where the alphas increase by 0.11 to 0.21 % annually. These comparisons are confirmed when the non-SRI fund portfolios are replaced by control portfolios matched on total asset value or age of the fund. For all benchmark models, the adjusted R^2 is higher than 0.85 and 0.95 for SRI and non-SRI fund portfolios, respectively, which confirms the suitability of the benchmark models.

We now examine a portfolio of our SRI sample consisting of those (on average, smaller) funds sponsored by management companies that specialize in SRI. Comparing their alphas reported in panels C and D of Table 3 with those for the portfolio of all SRI funds in panels A and B of Table 3, we observe that the alphas (as well as their significance levels) improve by from 0.08 to 0.48 % annually. The only exception is the lower alpha for the SW portfolio based on the four-factor benchmark which remains positive but is not significant. Based on the Wald test results reported in panels C and D of Table 3, the not significantly different selection performances between the SRI and the non-SRI fund portfolios remain when we confine membership in the SRI portfolio to the SRI funds with sponsors (management companies) specializing in SRI.

6.1.2 Security selection performance after management expenses

The selection performance findings based on returns after deducting management expenses or net returns also are reported in Table 3 and the estimated factor sensitivities using the various benchmark models are reported in Table 4. Before presenting the more detailed results for alpha performance, we observe from Table 3 that the alpha performance results are inconsistent with the second alternative hypothesis. Specifically, the portfolios of SRI funds do not significantly underperform their benchmarks or their corresponding portfolios of non-SRI funds. This is due to similar management expenses for SRI and non-SRI which suggests that the significantly lower management fees (MGF) associated with the SRI funds compared to the non-SRI funds has fully compensated for the remaining component of management expenses (MER). Furthermore, as was the case for the untabulated findings using gross returns, all of the various factor sensitivities reported in Table 4 are not significantly different and all of the market sensitivities are less than one.

Based on an examination of the detailed alpha findings reported in Table 3, we find that the alpha performance of the SRI fund portfolios are negative to neutral across all

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 γ (timing) measures for equal and size-weighted portfolios of SRI and non-SRI individual equity mutual funds. We also include the performance summary statistics (with MT subscript) of non-SRI control portfolios matched based on total net asset value or age of the fund. 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 1-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. $Wald^{\alpha}$ ($Wald^{\gamma}$) 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

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	0.1776**	−0.0108	0.1864*	−0.0108	0.1878**	0.0279	0.2248**
	(−0.24)	(2.04)	(−0.11)	(1.87)	(−0.12)	(2.20)	(0.29)	(2.32)
α-Conventional	−0.0983	0.0997	−0.0540	0.1431*	−0.0893	0.1088*	−0.0368	0.1604**
	(−1.48)	(1.53)	(−0.72)	(1.88)	(−1.40)	(1.72)	(−0.47)	(2.02)
α^{MT}-Conv.	−0.0877	0.0553	−0.0452	0.0985*	−0.0781	0.0881*	−0.0663	0.1154*
	(−1.50)	(1.40)	(−1.01)	(1.70)	(−1.45)	(1.66)	(−0.88)	(1.88)
$Wald^{\alpha}$	0.28	0.27	0.58	0.57	0.26	0.25	0.41	0.40
γ-SRI	−0.189	−0.175	−0.078	−0.064	−0.132	−0.117	0.003	0.018
	(−0.51)	(−0.47)	(−0.22)	(−0.18)	(−0.39)	(−0.34)	(0.01)	(0.05)
γ-Conventional	−0.281	−0.277	−0.143	−0.139	−0.256	−0.251	−0.088	−0.082
	(−0.98)	(−0.97)	(−0.47)	(−0.46)	(−0.92)	(−0.90)	(−0.30)	(−0.29)
γ^{MT}-Conv.	−0.333	−0.217	−0.211	−0.202	−0.389	−0.374	−0.117	−0.111
	(−1.01)	(−1.00)	(−0.95)	(−0.97)	(−1.22)	(−1.18)	(−1.14)	(−1.17)
$Wald^{\gamma}$	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	0.1043	0.0317	0.2349***	−0.0946	0.1095	0.0557	0.2592***
	(−1.18)	(1.25)	(0.39)	(2.82)	(−1.12)	(1.32)	(0.63)	(2.89)
α-Conventional	−0.0832	0.0973	−0.0425	0.1378*	−0.0721	0.1084*	−0.0278	0.1524*
	(−1.29)	(1.53)	(−0.56)	(1.80)	(−1.15)	(1.75)	(−0.34)	(1.85)
α^{MT}-Conv.	−0.1100	0.0478	−0.0333	0.0884*	−0.0621	0.0752*	−0.0551	0.1002*
	(−1.41)	(1.33)	(−1.12)	(1.72)	(−1.33)	(1.74)	(−0.79)	(1.90)
$Wald^{\alpha}$	0.83	0.93	0.35	0.23	0.77	0.99	0.29	0.18
γ-SRI	0.074	0.075	0.250	0.251	0.097	0.098	0.312	0.313
	(0.26)	(0.26)	(0.72)	(0.72)	(0.33)	(0.34)	(0.93)	(0.94)
γ-Conventional	−0.151	−0.146	−0.027	−0.022	−0.114	−0.109	0.043	0.048
	(−0.54)	(−0.52)	(−0.09)	(−0.07)	(−0.43)	(−0.40)	(0.15)	(0.17)

Table 3 (continued)

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
γ^{MT}-Conv.	−0.200	−0.177	−0.092	−0.079	−0.226	−0.208	−0.095	−0.088
	(−0.99)	(−0.78)	(−0.88)	(−0.90)	(−0.87)	(−0.97)	(−1.19)	(−1.22)
$Wald^{\gamma}$	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	0.2152**	−0.0028	0.1931	0.0314	0.2278**	0.0443	0.2396**
	(0.18)	(2.10)	(−0.02)	(1.60)	(0.31)	(2.28)	(0.38)	(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	0.174	0.266	0.274	0.293	0.302	0.359	0.369
	(0.26)	(0.78)	(0.45)	(0.46)	(0.55)	(0.56)	(0.73)	(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	0.1381*	−0.0637	0.1330	−0.0570	0.1397*	−0.0388	0.1578
	(−0.71)	(1.68)	(−0.66)	(1.38)	(−0.69)	(1.70)	(−0.38)	(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	0.120	0.242	0.240	0.188	0.185	0.294	0.291
	(0.30)	(0.29)	(0.63)	(0.62)	(0.54)	(0.53)	(0.92)	(0.91)
$Wald^{\gamma}$	0.26	0.30	0.22	0.25	0.16	0.20	0.15	0.20

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 when measured using fund returns after expenses. For instance, the estimated monthly alphas of the SW SRI portfolios are all insignificant and range from −0.0998 % (−1.20 % annually) using the single factor CAPM to 0.0557 % (0.67 % annually) obtained using the full conditional four-factor model. These findings suggest that whether the value added by fund managers covers fund expenses depends on the choice of performance benchmark. Moreover, the alphas of SRI SW fund portfolios increase by 0.29 and 0.06 % annually with the conditioning of the four- and the one-factor models, respectively. The corresponding respective increases are 0.46 and 0.12 % annually for the alphas of the SRI EW fund portfolios.

The non-SRI fund portfolios exhibit insignificantly negative alphas. As for the SRI fund portfolios, the performances of non-SRI fund portfolios improve from 0.11 to 0.21 % annually with conditioning and with the use of a four-factor benchmark. The alphas of the non-SRI fund portfolios are insignificantly lower based on Wald tests (all *p*-values≥0.28) than those of the SRI portfolios across all benchmark models except for the SW non-SRI fund portfolios using the single-factor benchmarks. These differentials in the performance point estimates increase insignificantly by 0.52 to 1.00 % annually with conditioning and are somewhat stable across all models. Alternative control portfolios of non-SRI funds matched on total asset value or age to their SRI counterparts yield consistent inferences.

The alphas improve from 0.10 to 0.51 % annually but remain statistically insignificant for the SRI fund portfolios consisting of funds managed by specialized SRI management companies. The only exception is for the SW SRI fund portfolios based on the four-factor

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 1-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_{SMB}	β_{HML}	β_{UMD}	β_M	β_{SMB}	β_{HML}	β_{UMD}	β_M	β_{SMB}	β_{HML}	β_{UMD}	β_M	β_{SMB}	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

benchmarks where the alphas become negative. Based on the Wald test results reported in panels C and D of Table 3, the not significantly different selection performances between the SRI and the non-SRI fund portfolios remain when we restrict the SRI portfolio to the SRI funds whose management companies profess such a specialty.

As reported in Table 5, the use of the Jantzi social index (JSI) as an alternative market benchmark variable increases the factor sensitivities of the SRI fund portfolios to the SMB factor but still exhibits similar weak to neutral SRI fund alpha performance after management expenses across all benchmark specifications. The use of JSI lowers SRI fund alpha performance measured against multifactor benchmarks and alters the positive effect of conditioning on the alphas.

6.1.3 Some further observations

Figure 1 provides a visual (and possibly more intuitive) depiction of the portfolio findings presented above. In the interests of compactness, both panels are for the size-weighted (SW) SRI and non-SRI portfolios for the complete time period from February 1988 to April 2008 using both gross and net returns, and the measures of abnormal performance (α), systematic risk (Sys. SD) and residual risk (Res. SD) are based on a full conditional four-factor benchmark specification. The left-most panel of Fig. 1 plots the average monthly returns, and their standard deviations

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 1-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

and systematic standard deviations for both portfolios. The right-most panel plots the average monthly alphas and the average monthly residual standard deviations for both portfolios. We observe from the panels that the SRI portfolio has the lower average monthly total return but a higher average monthly alpha, and lower average monthly total and systematic risks but a higher average monthly residual risk than the non-SRI portfolio. When we capture the trade-off between the mean return measure and its associated risk by dividing the former by the later, we find that only one of six possible ratios favors (marginally) the non-SRI portfolio. This is the ratio of the average monthly total net return to the average monthly standard deviation of net returns where the values are 0.221 and 0.216 for the non-SRI and SRI portfolio, respectively. Furthermore, the only reward-to risk ratio that is materially different favors the SRI portfolio. This is for the information ratio used by practitioners to measure the reward for bearing active investment risk given by the ratio of the average monthly net return alpha to its corresponding average monthly residual risk. The values are −0.037 and 0.047 (or −0.448 and 0.567 annualized) for the non-SRI and SRI portfolio, respectively.

6.2 Performance results for individual funds

The cross-sectional selection (alpha) and timing (gamma) performance distributions for individual SRI funds are summarized in panels A and B of Table 6, respectively. Most of the statistics reported in panel A suggest weak to neutral alpha performance across all benchmark models, and that both mean and median performance improves with conditioning. The highest standard deviation of alphas for individual SRI funds occurs for the four-factor benchmark models. The gamma statistics reported in panel B of Table 6 are generally consistent with the inferences discussed earlier using the portfolios of funds.

The distributions of fund alphas are negatively skewed with fat (longer left) tails for all retained benchmark models. The number of funds with positive and significant alphas or gammas is low cross-sectionally but higher for the four-factor model at the 5 and 10 % levels with and without

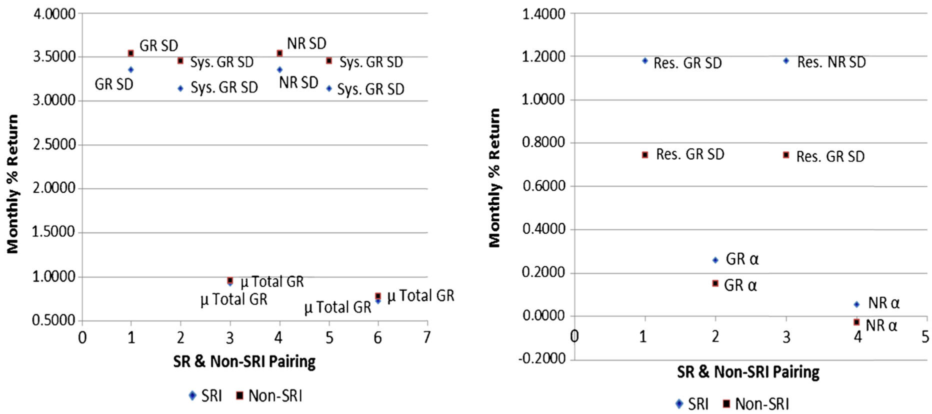

Fig. 1 Plots of various average monthly return and risk measures for size-weighted SR and non-SRI portfolios. Both panels are for the complete time period from February 1988 to April 2008. The systematic (Sys.) and residual (Res.) components of the monthly standard deviations and the alphas for each portfolio are estimated using the full conditional four-factor benchmark model. The left-most panel plots the mean monthly total % gross return (µ Total GR) and net return (µ Total NR), the mean monthly standard deviation of total gross returns (GR SD) and net returns (NR SD), and the mean monthly systematic standard deviation of total gross returns (Sys. GR SD) and net returns (Sys. NR SD). The right-most panel plots the mean monthly % gross return alpha (GR α) and net return alpha (NR α), and the mean monthly residual standard deviation of gross returns (Sys. GR SD) and net returns (Sys. NR SD)

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 (α in % and γ) 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 1-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	Selection Performance Statistics											
	Mean	Std. Dev.	Median	Minimum	Maximum	Median Adj. R^2	p-max	p-min	% funds with $p<5$ % (10 %)	% funds with $\alpha>0$ & $p<5$ % (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 $\gamma>0$ & $p<5$ % (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

conditioning information, respectively. Since none of the Bonferroni *p*-values are significant for the maximum *t*-statistics, we cannot reject the joint hypothesis of zero alphas. This result holds for the minimum *t*-statistics with the exception of the full conditional four-factor benchmark model.

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 non-normality 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 (i.e., to test our third hypothesis). Bootstrapping is used in the context of performance evaluation by Ayadi and Kryzanowski (2005, 2011), 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). 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 (alphas and gammas) for both gross and net returns and by their *t*-statistics. As tests of robustness, individual performance statistics and inferences are further assessed using the block bootstrap method with a block of three consecutive monthly observations. For compactness, we only report the results for the four-factor benchmark models using the residuals and the factors resampling structure, and rely on the *t*-statistics for performance ranking.

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

6.3.1 Bootstrapped selection performance of individual SRI funds

Based on panels A to D of Table 7 and Fig. 2, all the bootstrapped *p*-values with *t*-statistics ranking and with independent resampling of residuals and factors 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. The bootstrapped *p*-values of the worst fund become significant at the 0.10 and not 0.05 level 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. In contrast, the same tests conducted using gross returns indicate that sampling variation has a marginal effect on the original neutral performance inferences of these worst performing SRI funds.

When the benchmark models become conditional, the corrected bootstrap *p*-values are consistently greater than 0.17 for the four extreme worst-performing funds using net returns. This implies that the negative and significant performances of these funds based on the unconditional benchmark models are not an indicator of management skills as they essentially are related to sampling variation (i.e., they are simply unlucky). This result is confirmed using gross returns, a residuals only resampling method, and 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 when distinguishing between ability and luck.

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 1-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 2nd, 3rd, 4th, and 5th 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

Benchmark model	Unconditional benchmark								Conditional alpha and beta benchmark							
	Ranked alpha estimates				Ranked *t*-statistics				Ranked alpha estimates				Ranked *t*-statistics			
	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

Table 7 (continued)

Benchmark model	Unconditional benchmark								Conditional alpha and beta benchmark							
	Ranked alpha estimates				Ranked *t*-statistics				Ranked alpha estimates				Ranked *t*-statistics			
	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)
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
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
2nd	−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
3rd	−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
4th	−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
5th	−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																
5th	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
4th	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
3rd	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
2nd	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

Table 7 (continued)

Benchmark model	Unconditional benchmark								Conditional alpha and beta benchmark							
	Ranked alpha estimates				Ranked *t*-statistics				Ranked alpha estimates				Ranked *t*-statistics			
	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)
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
2nd	−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
3rd	−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
4th	−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
5th	−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																
5th	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
4th	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
3rd	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
2nd	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

In contrast, the block-bootstrapped p-values of the three top performing funds are very different from the non-bootstrapped p-values using net returns and the full conditional four-factor model. The resulting 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.

These results are further supported by Fig. 2, which depicts the distribution of the bootstrapped t-statistics for various ranked funds using the full conditional four-factor model with net returns. Kernel density estimators based on Gaussian kernel functions are applied to these distributions. 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 a traditional t-statistic of −2.17 (dashed line) but the null is not rejected using the bootstrap. Panel A5 shows the fourth best fund with a traditional t-statistic of 1.09 that rejects the null as does the bootstrap test.

Further tests are reported in Table 8 for the two successive subperiods of February 1988 to February 2003 and March 2003 to April 2008. The bootstrap-based inferences obtained for the entire period are confirmed in the second subperiod where the resampled p-values are all greater than 10 % for the top and worst three funds (with one exception). These findings are robust to the use of gross or net returns and to including conditioning information in the four-factor benchmark model. Noticeable differences are evident in the first period, especially using net returns for all extreme funds except the worst one (all bootstrapped p-values less than

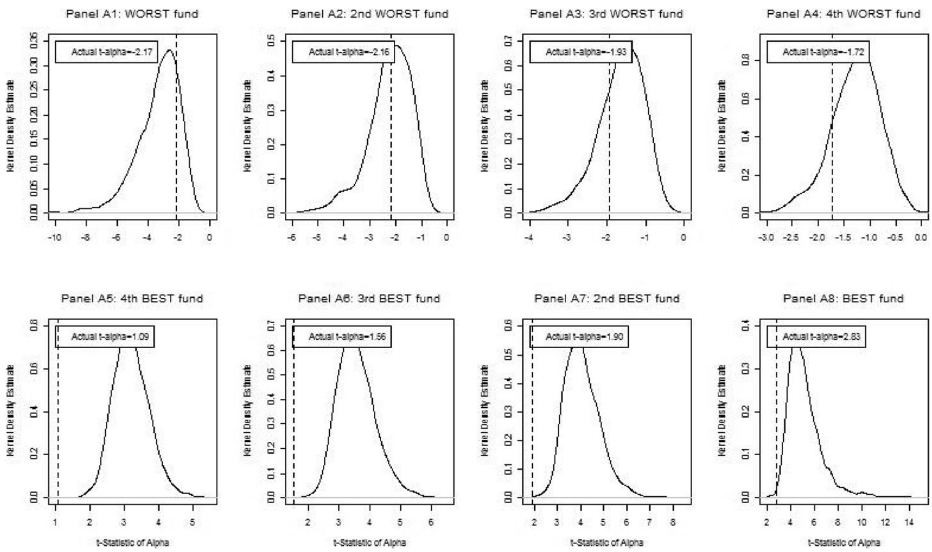

Fig. 2 Estimated alpha t-statistics vs. bootstrapped alpha t-statistic distribution for individual funds at various points in the cross-section using net returns. The t-statistics are based on the full conditional four-factor model. This figure plots kernel density estimates if the bootstrapped distribution 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)

Table 8 Bootstrap analysis of the best and worst SRI fund's performances over subperiods. This table reports least squares monthly estimates of SRI fund performance (α in %) over two successive subperiods 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 for all funds. The alpha is estimated using unconditional and conditional alpha and beta four-factor models over two successive subperiods. The first subperiod is February 1988 to March 2003 and the second subperiod is April 2003 to April 2008. Conditional alpha and beta refers to models with both time-varying alphas and betas. The dividend yield (DY) and the yield on the 1-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 and period 1 (2), the ranked *t*-statistic, parametric one-tailed *p*-value, and *t*-statistic based bootstrapped *p*-value are reported in columns one (four), two (five), and three (six), 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 (seventh) row in each panel reports funds with the lowest (highest) *t*-statistic as well as the median. The in-between rows concern the 2nd and 3rd funds in the left (right) tail of the performance distribution. Only funds with a minimum of 30 observations are included

Benchmark model	Unconditional benchmark						Conditional alpha and beta benchmark					
	Feb. 1988 to Feb. 2003			Mar. 2003 to Apr. 2008			Feb. 1988 to Feb. 2003			Mar. 2003 to Apr. 2008		
	$t\alpha_1$	*p*-val.	boot. *p*-val.(t)	$t\alpha_2$	*p*-val.	boot. *p*-val.(t)	$t\alpha_1$	*p*-val.	boot. *p*-val.(t)	$t\alpha_2$	*p*-val.	boot. *p*-val.(t)
Panel A: Four-factor model with net returns												
Worst	−3.42	0.00	0.81	−1.78	0.04	0.15	−3.35	0.00	0.76	−2.17	0.02	0.53
2nd	−2.54	0.01	0.02	−1.48	0.07	0.13	−2.30	0.02	0.02	−2.16	0.02	0.79
3rd	−1.73	0.05	0.01	−1.30	0.10	0.13	−1.90	0.03	0.03	−1.90	0.03	0.70
Median	−0.65			0.07			−0.49			0.54		
3rd	−0.48	0.31	0.01	1.51	0.07	0.09	−0.15	0.44	0.07	1.05	0.15	0.54
2nd	−0.16	0.44	0.03	2.06	0.02	0.74	−0.10	0.46	0.02	1.09	0.14	0.55
Best	1.45	0.07	0.02	2.25	0.01	0.19	1.69	0.05	0.04	2.33	0.01	0.51
Panel B: Four-factor model with gross returns												
Worst	−2.59	0.01	0.27	−1.07	0.15	0.38	−2.55	0.01	0.32	−1.86	0.04	0.54
2nd	−1.18	0.12	0.76	−0.69	0.25	0.39	−1.33	0.10	0.70	−1.74	0.05	0.32
3rd	−0.90	0.19	0.11	−0.55	0.29	0.41	−1.18	0.12	0.11	−1.34	0.10	0.43
Median	0.11			1.22			0.24			1.14		
3rd	0.88	0.19	0.09	2.61	0.01	0.51	1.02	0.16	0.38	2.07	0.02	0.32
2nd	1.15	0.13	0.81	2.87	0.00	0.54	1.09	0.14	0.72	2.07	0.02	0.29
Best	2.86	0.00	0.86	3.14	0.00	0.28	3.00	0.00	0.81	2.88	0.00	0.30

0.07). However, the bootstrapped *p*-values are mostly greater than 0.10 based on gross returns, which suggests that the best (worst) fund was simply lucky (unlucky).

6.3.2 Bootstrapped timing performance of individual SRI funds

We now assess the market-timing ability of various extreme funds in both tails of the performance distribution using the block bootstrap approach with both resampling methods for all benchmark specifications. The results presented in Table 9 indicate that the effect of sampling variation or "bad luck" explains the poor conditional timing performance of only the worst and second worst performing funds where the traditionally calculated and bootstrapped corrected *p*-values are very different. These findings are robust to the use of gross or net returns, the alternative ranking method, and standard bootstrapping. The five funds with the best timing performance based on

Table 9 Bootstrap analysis of the best and worst SRI fund's timing performances. This table reports fund timing performance (γ) 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 1-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 2nd, 3rd, 4th, and 5th 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

Benchmark model	Unconditional benchmark								Conditional alpha and beta benchmark							
	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

Table 9 (continued)

Benchmark model	Unconditional benchmark								Conditional alpha and beta benchmark							
	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)
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
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
2nd	−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
3rd	−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
4th	−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
5th	−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																
5th	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
4th	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
3rd	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
2nd	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
2nd	−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

Table 9 (continued)

Benchmark model	Unconditional benchmark								Conditional alpha and beta benchmark							
	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)
3rd	−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
4th	−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
5th	−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																
5th	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
4th	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
3rd	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
2nd	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

net returns have bootstrapped p-values greater than 0.13 for gamma, which is consistent with good luck. These conclusions are maintained using gross returns, an unconditional benchmark, with residuals only resampling, and with standard bootstrapping. As expected, extreme timing performance generally cannot be attributed to luck when the benchmark is less robust (i.e., unconditional or based on a single-factor specification).

The market-timing inferences are further supported by Fig. 3 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 a traditionally calculated t-statistic of −0.75 (dashed line) and the null is rejected using the bootstrap. Panel A7 shows the second fund with a traditionally calculated t-statistic of 2.26 that rejects the null but the bootstrap test does not.

We further assess the temporal stability of these timing-performance findings over the same two successive subperiods examined earlier for the alphas. Based on Table 10, there are some timing-performance differences for the extreme funds during the first and second subperiods. For example, the fund with the worst timing performance is revealed to be unlucky only during the second subperiod using all benchmark models with both gross and net returns. The inferences related to the second best timing-performance fund using the full conditional model with net returns differ from those using gross returns for these two subperiods. While this fund has significant bootstrapped and traditionally calculated p-values less than 0.05 in the first subperiod, the bootstrapped p-values are insignificant (>0.40) while the traditionally calculated p-values of 0.01 are significant in the second subperiod. This indicates that this fund's timing performance was simply due to luck in the second subperiod.

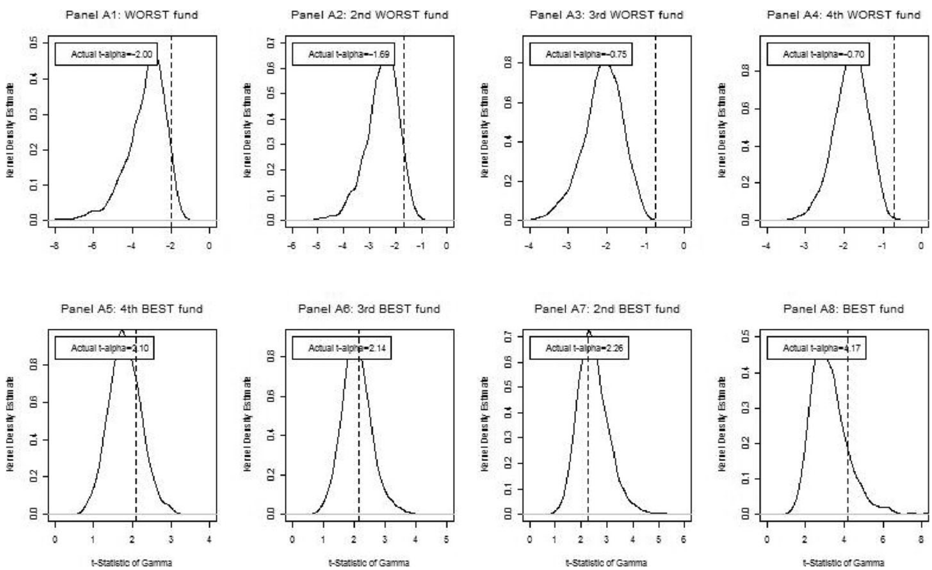

Fig. 3 Estimated gamma t-statistics vs. bootstrapped gamma t-statistic distribution for individual funds at various points in the cross-section using net returns. The t-statistics are based on the full conditional four-factor model. This figure plots kernel density estimates if the bootstrapped distribution 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)

Table 10 Bootstrap analysis of the best and worst SRI fund's timing performances over subperiods. This table reports fund timing performance (γ) and significance tests over two successive subperiods using the block bootstrap methodology (with a block of 3 overlapping observations) with independent resampling of the factor returns and the residuals for all funds. The alpha and gamma are estimated using unconditional and conditional alpha and beta four-factor models with a quadratic market variable. The first subperiod is February 1988 to February 2003 and the second subperiod is March 2003 to April 2008. 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 1-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 and period 1 (2), the ranked *t*-statistic, parametric one-tailed *p*-value, and *t*-statistic based bootstrapped *p*-value are reported in columns one (four), two (five), and three (six), 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 2nd and 3rd funds in the left (right) tail of the timing performance distribution. Only funds with a minimum of 30 observations are included

Benchmark model	Unconditional benchmark						Conditional alpha and beta benchmark					
	Feb. 1988 to Feb. 2003			Mar. 2003 to Apr. 2008			Feb. 1988 to Feb. 2003			Mar. 2003 to Apr. 2008		
	$t\gamma_1$	*p*-val.	boot *p*-val.(t)	$t\gamma_2$	*p*-val.	boot. *p*-val.(t)	$t\gamma_1$	*p*-val.	boot. *p*-val.(t)	$t\gamma_2$	*p*-val.	boot. *p*-val.(t)
Panel A: Four-factor model with net returns												
Worst	−0.82	0.21	0.67	−1.87	0.03	0.24	−0.72	0.24	0.22	−2.01	0.03	0.47
2nd	−0.49	0.31	0.14	−0.35	0.36	0.13	−0.41	0.34	0.15	−2.00	0.03	0.48
3rd	−0.34	0.37	0.21	0.05	0.48	0.29	0.01	0.50	0.21	−1.69	0.06	0.57
Median	1.03			0.54			1.15			0.17		
3rd	1.71	0.05	0.74	2.36	0.01	0.30	3.54	0.00	0.27	v2.14	0.03	0.44
2nd	1.89	0.03	0.22	2.57	0.01	0.34	3.88	0.00	0.03	2.27	0.01	0.43
Best	6.10	0.00	0.65	2.64	0.01	0.26	5.10	0.00	0.25	2.41	0.01	0.61
Panel B: Four-factor model with gross returns												
Worst	−0.83	0.20	0.69	−1.87	0.03	0.24	−0.75	0.23	0.21	−1.99	0.03	0.49
2nd	−0.49	0.31	0.13	−0.38	0.35	0.13	−0.42	0.34	0.15	−1.91	0.03	0.46
3rd	−0.34	0.37	0.20	0.04	0.48	0.29	0.03	0.49	0.20	−1.69	0.06	0.57
Median	1.04			0.54			1.16			0.17		
3rd	1.71	0.05	0.71	2.36	0.01	0.39	3.55	0.00	0.26	2.17	0.02	0.44
2nd	1.89	0.03	0.21	2.57	0.01	0.34	3.87	0.00	0.03	2.28	0.01	0.44
Best	6.09	0.00	0.65	2.64	0.01	0.26	5.11	0.00	0.24	2.42	0.01	0.61

6.3.3 Implications

The above block bootstrap-based inferences differ from those from the Bonferroni tests. They highlight the importance and effects of individual SRI fund cross-correlations. Overall, our findings parallel the recent performance tests of Kosowski 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 (2011). 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 equity SRI 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".

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 equity SRI 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.

Acknowledgments Financial support from Senior Concordia University Research Chair in Finance, Goodman School of Business, IFM2 and SSHRC are gratefully acknowledged. We would like to thank Michael Brennan, Yunieta Nainggolan (discussant), Emilia Peni (discussant) and participants at the 2011 Mid-West meetings (Chicago) and the 2011 EFA meetings (Savannah). We also thank Kaveh Moradi Dezfouli for his excellent research assistance.

References

Adler T, Kritzman M (2008) The cost of socially responsible investing. J Portf Manag 35:52–56

Akpinar A, Jiang Y, Gomez-Mejia LR, Berrone P, Walls JL (2008) Strategic use of CSR as a signal for good management. Working Paper, IE Business School

Allen F, Carletti E, Marquez R (2009) Stakeholder capitalism, corporate governance and firm value. Working Paper, University of Pennsylvania
Alpert B, Rekenthaler J (2011) Global fund investor experience 2011. Morningstar Research Papers. Available at: https://corporate.morningstar.com/us/documents/ResearchPapers/GlobalFundInvestorExperience2011.pdf
Alpert B, Rekenthaler J, Suh S (2013) Global fund investor experience 2013. Morningstar Research Papers. Available at: https://corporate.morningstar.com/us/documents/MethodologyDocuments/FactSheets/Global-Fund-Investor-Experience-Report-2013.pdf
Amihud Y, Goyenko R (2013) Mutual fund's R2 as predictor of performance. Rev Financ Stud 26:667–694
Amihud Y, Hurvich CM (2004) Predictive regressions: a reduced-bias estimation method. J Financ Quant Anal 39:813–841
Asmundson P, Foerster SR (2001) Socially responsible investing: better for your soul or your bottom line? Can Invest Rev 14
Ayadi MA, Kryzanowski L (2005) Portfolio performance measurement using APM-free kernel models. J Bank Financ 29:623–659
Ayadi MA, Kryzanowski L (2011) Fixed-income fund performance: role of luck and ability in tail membership. J Empir Financ 18:379–392
Ayadi MA, Ben-Ameur H, Lazrak S, Wang Y (2013) Canadian investors and the discount on closed-end funds. J Financ Serv Res 43:69–98
Bagnoli M, Watts S (2003) Selling to socially responsible consumers: competition and the private provision of public goods. J Econ Manag Strateg 12:419–445
Barnea A, Rubin A (2010) Corporate social responsibility as a conflict between shareholders. J Bus Ethics 97:71–86
Baron D (2001) Private politics, corporate social responsibility, and integrated strategy. J Econ Manag Strateg 10:7–45
Bauer R, Derwall J, Otten R (2007) The ethical mutual funds performance debate: new evidence from Canada. J Bus Ethics 70:111–124
Bauer R, Koedijk K, Otten R (2005) International evidence on ethical mutual fund performance and investment style. J Bank Financ 29:1751–1767
Bauer R, Otten R, Tourani Rad A (2006) Ethical investing in Australia: is there a financial penalty? Pac Basin Financ J 14:33–48
Bensimon B (2010) Consumerology Report, August. Available at: http://www.consumerology.ca/pdf/Consumerology_Aug2010.pdf
Benson L, Humphrey JE (2008) Socially responsible investment funds: investor reaction to current and past returns. J Bank Financ 32:850–859
Berk JB, van Binsbergen JH (2013) Measuring skill in the mutual fund industry. Working Paper, Stanford University and NBER
Bollen NP (2007) Mutual fund attributes and investor behavior. J Financ Quant Anal 42:683–708
Brean DJS, Kryzanowski L, Roberts GS (2011) Canada and the United States: different roots, different routes to financial sector regulation. Bus Hist 53:249–269
Brekke K, Nyborg K (2005) Moral hazard and moral motivation: Corporate social responsibility as labour market screening. Working Paper, The Ragner Frisch Centre for Economic Research, Oslo
Campbell J (1991) A variance decomposition for stock returns. Econ J 101:157–179
Carhart MM (1997) On persistence in mutual fund performance. J Financ 52:57–82
Carrick R (2006) At least the name fits: the problem with SRI funds. Corporate Knights
Chen Z, Knez PJ (1996) Portfolio performance measurement: theory and applications. Rev Financ Stud 9:511–555
Christoffersen S, Musto D (2002) Demand curves and the pricing of money management. Rev Financ Stud 15:1499–1524
Christopherson JA, Ferson WE, Glassman DA (1998) Conditioning manager alphas on economic information: another look at the persistence of performance. Rev Financ Stud 11:111–142
Cortez MC, Silva F, Areal N (2009) The performance of European socially responsible funds. J Bus Ethics 87:573–588
Cuthbertson K, Nitzsche D, O'Sullivan N (2008) UK mutual fund performance: skill or luck? J Empir Financ 15:613–634
Derwall J, Koedijk K (2009) Socially responsible fixed-income funds. J Bus Financ Account 36:210–229
Ding B, Wermers R (2012) Mutual fund performance and governance structure: the role of portfolio managers and boards of directors. Working Paper, SUNY
Dybvig PH, Ross SA (1985) Performance measurement using differential information and a security market line. J Financ 40:483–496
Efron B (1979) Bootstrap methods: another look at the Jacknife. Ann Stat 7:1–26
Elliot G, Stock JH (1994) Inference in time-series regression when the order of integration of a regressor is unknown. Economet Theor 10:672–700
Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. J Financ Econ 33:3–56

Fama EF, French KR (2010) Luck versus skill in the cross section of mutual fund alpha estimates. J Financ 65: 1915–1947
Ferguson R (1980) Performance measurement doesn't make sense. Financ Anal J 36:59–69
Fernandez-Izquierdo A, Matallin-Saez JC (2007) Performance of ethical mutual funds in Spain: sacrifice or premium. J Bus Ethics 81:247–260
Ferson WE, Sarkissian S, Simin T (2003) Is stock return predictability spurious? J Invest Manag 1:1–10
Ferson WE, Schadt R (1996) Measuring fund strategy and performance in changing economic conditions. J Financ 51:425–461
Fombrun C, Shanley M (1990) What's in a name? Reputation building and corporate strategy. Acad Manag J 33: 233–258
Frazzini A, Lamont OA (2008) Dumb money: mutual fund flows and the cross-section of stock returns. J Financ Econ 88:299–322
Gadhoum Y, Lang LHP, Young L (2005) Who controls us? Eur Financ Manag 11:339–363
Geczy C, Stambaugh R, Levin D (2005) Investing in socially responsible mutual funds. Working Paper, University of Pennsylvania
Gil-Bazo J, Ruiz-Verdu P (2009) The relation between price and performance in the mutual fund industry. J Financ 64:2153–2183
Gil-Bazo G, Ruiz-Verdú P, Santos AAP (2010) The performance of socially responsible mutual funds: the role of fees and management companies. J Bus Ethics 94:243–263
Gregory A, Whittaker J (2007) Performance and performance persistence of 'ethical' unit trusts in the UK. J Bus Financ Account 34:1327–1344
Grinblatt M, Titman S (1989) Portfolio performance evaluation: old issues and new insights. Rev Financ Stud 2: 393–422
Hall P, LePage R (1996) On bootstrap estimation of the studentized mean. Ann Inst Stat Math 48:403–421
Heal G (2005) Corporate social responsibility: an economic and financial framework. Geneva Pap Risk Insur–Issues Pract 30:387–409
Hofmann E, Hoelzl E, Kirchler E (2008) A comparison of models describing the impact of moral decision making on investment decisions. J Bus Ethics 82:171–187
Hong H, Kacperczyk M (2009) The price of sin: the effects of social norms on markets. J Financ Econ 93:15–36
Horowitz JL (2001) The bootstrap in econometrics. Handb Econ 5:3159–3228
Huij J, Derwall J (2008) Hot hands in bond funds. J Bank Financ 32:559–572
Ioannou I, Serafeim G (2010) The impact of corporate social responsibility on investment recommendations. Working Paper, Harvard Business School
Jensen MC (1968) The performance of mutual funds in the period 1945–1964. J Financ 23:389–416
Jensen M (2002) Value maximization, stakeholder theory, and the corporate objective function. Bus Ethics Q 12: 235–256
Jones S, van der Laan S, Frost G, Loftus J (2008) The investment performance of socially responsible investment funds in Australia. J Bus Ethics 80:181–203
Kacperczyk M, Sialm C, Zheng L (2005) On the industry concentration of actively managed equity mutual funds. J Financ 60:1983–2011
Kim S (2014) Measuring luck in CEO outperformance. Working Paper, Santa Clara University
King MR, Santor E (2008) Family values: ownership structure, performance and capital structure of Canadian firms. J Bank Financ 32:2423–2432
Kosowski R, Timmermann A, Wermers R, White H (2006) Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. J Financ 61:2551–2595
Kosowski R, Naik NY, Teo M (2007) Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. J Financ Econ 84:229–264
Kreander N, Gray G, Power DM, Sinclair CD (2005) Evaluating the performance of ethical and non-SRI funds: a matched pair analysis. J Bus Financ Account 32:1465–1493
Kryzanowski L, Lalancette S, To MC (1994) Performance attribution using a multivariate intertemporal asset pricing model with one state variable. Can J Adm Sci 11:75–85
Kryzanowski L, Lalancette S, To MC (1997) Performance attribution using an APT with prespecified macrofactors and time-varying risk premia and betas. J Financ Quant Anal 32:205–224
Künsch HR (1989) The jacknife and the bootstrap for general stationary observations. Ann Stat 17:1217–1241
Lehmann BN, Modest DM (1987) Mutual funds performance evaluation: a comparison of benchmarks and benchmarks comparisons. J Financ 42:233–265
Luther R, Matatko J, Corner D (1992) The investment performance of UK ethical unit trusts. Account Audit Account J 5:57–70
Mallin CA, Saadouni B, Briston RJ (1995) The financial performance of ethical investment funds. J Bus Financ Account 22:483–496

Mawer (2014) Mawer named analysts' choice fund company of the year at 2014 Morningstar Awards for the second year in a row. Press Release, Mawer Investment Management Ltd., November 27. Available at: http://www.mawer.com/assets/Press-Releases/Morningstar-Press-Release-Mawer-2014.pdf

Nanda V, Wang ZJ, Zheng L (2004) Family values and the star phenomenon: strategies of mutual fund families. Rev Financ Stud 17:667–698

Nelson J (2014) Fifteen top Canadian equity funds, The Globe and Mail, April 14. Available at: http://www.theglobeandmail.com/globe-investor/investment-ideas/number-cruncher/canadian-energy-sector-fuelled-fund-gains/article17997910/

Nicholls C (2006) The characteristic of Canada's capital markets and the illustrative case of Canada's legislative regulatory response to Sarbanes–Oxley. Research study commissioned by the Task Force to modernize securities legislation in Canada, pp 129–203

Newey WK, West KD (1987a) A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55:703–708

Newey WK, West KD (1987b) Hypothesis testing with efficient method of moments estimation. Int Econ Rev 28:777–787

Renneboog L, ter Horst J, Zhang C (2011) Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds. J Financ Intermed 20:562–588

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

Renneboog L, ter Horst J, Zhang C (2008a) The price of ethics and stakeholder governance: the performance of socially responsible mutual funds. J Corp Financ 14:302–322

Renneboog L, ter Horst J, Zhang C (2008b) Socially responsible investments: institutional aspects, performance, and investor behavior. J Bank Financ 32:1723–1742

Richardson BJ, Crag W (2010) Being virtuous and prosperous: SRI's conflicting goals. J Bus Ethics 92:21–39

Rosenberg B (1976) Security appraisal and unsystematic risk in institutional investment. Proceedings of the Seminar on the Analysis of Security Prices. University of Chicago Press, Chicago, pp 171–237

Save-Soderbergh J (2010) Who lets ethics guide his economic decision-making? An empirical analysis of individual investments in ethical funds. Econ Lett 94:270–272

Schroder M (2004) The performance of socially responsible investments: investment funds and indices. Fin Mkts Portfolio Mgmt 18:122–142

Shleifer A (2004) Does competition destroy ethical behavior? Am Econ Assoc Pap Proc 94:414–418

Social Investment Organization: 2000. Canadian social investment review 2000: A comprehensive survey of socially responsible investment in Canada

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

SIO (2013) Canadian socially responsible investment review 2012, Social Investment Organization, Toronto. Available at: http://riacanada.ca/sri-review/

Stambaugh RF (1999) Predictive regressions. J Financ Econ 54:375–421

Steiauf T, Schäfer H (2014) From integration to impact—a new investment climate for Germany's SRI landscape. J Sustain Financ Invest 4:38–60

Tirole J (2001) Corporate governance. Econometrica 69:1–35

Torous W, Valkanov R, Yan S (2004) On predicting stock returns with nearly integrated explanatory variables. J Bus 77:937–966

Treynor J, Black F (1973) How to use security analysis to improve on portfolio selection. J Bus 46:66–86

Treynor J, Mazuy K (1966) Can mutual funds outguess the market? Harv Bus Rev 44:131–136

Valkanov R (2003) Long-horizon regressions: theoretical results and applications. J Financ Econ 68:201–232

Van Nieuwerburgh S, Veldkamp L (2010) Information acquisition and under-diversification. Rev Econ Stud 77:779–805