Tri-criterion inverse portfolio optimization with application to socially responsible mutual funds

Sebastian Utz\textsuperscript{a}, Maximilian Wimmer\textsuperscript{a}, Markus Hirschberger\textsuperscript{b}, Ralph E. Steuer\textsuperscript{c,}\textsuperscript{*}

\textsuperscript{a}Department of Finance, University of Regensburg, 93040 Regensburg, Germany
\textsuperscript{b}Munich Re, 80802 Munich, Germany
\textsuperscript{c}Department of Finance, University of Georgia, Athens, Georgia 30602-6253, USA

Abstract

We present a framework for inverse optimization in a Markowitz portfolio model that is extended to include a third criterion. The third criterion causes the traditional nondominated frontier to become a surface. Until recently, it had not been possible to compute such a surface. But by using a new method that is able to generate the nondominated surfaces of tri-criterion portfolio selection problems, we are able to compute via inverse optimization the implied risk tolerances of given funds that pursue an additional objective beyond risk and return. In applying this capability to a broad sample of conventional and socially responsible (SR) mutual funds, we find that after the screening process there appears to be no significant difference between how assets are allocated in socially responsible and conventional mutual funds, which is likely to be different from what most SR investors would expect.

Keywords: Socially responsible investing, Inverse optimization, Portfolio selection, Multiple criteria optimization, Nondominated surfaces, Multiple criteria decision making

1. Introduction

In the seminal work of Markowitz (1952, 1956) and later in books (1959, 1987, 2000) only the expected returns and the covariances of the returns of all considered assets are taken into account when attempting to locate optimal portfolios. However,
recent studies suggest that in many situations a more complex decision model may be at work (Abdelaziz et al., 2007; Ballestero et al., 2012; Bollen, 2007; Dorfleitner et al., 2012; Dorfleitner & Utz, 2012; Hallerbach et al., 2004; Steuer et al., 2007; Xidonas et al., 2012). Toward that end, in Hirschberger et al. (2013), a methodology is developed for extending the portfolio selection model of Markowitz to include a third criterion. Whether the third criterion is a financial or, as in this paper, a non-financial one, this causes the nondominated frontier\(^1\) in two-dimensional space to become a nondominated surface in three-dimensional space, and that paper describes an algorithm for computing tri-criterion nondominated surfaces exactly. In this paper, going one step beyond, we utilize information generated by the algorithm for tri-criterion nondominated surfaces in an inverse portfolio optimization context, and the contributions of this paper are twofold.

One is that we show how inverse portfolio optimization can be used in a tri-criterion model. In bi-criterion mean-variance portfolio selection, inverse portfolio optimization is described by Zagst & Pöschik (2008). In their description, one first computes the entire nondominated frontier, for instance with the critical line method of Markowitz (1956), and then notes that each portfolio along the nondominated frontier has its own risk tolerance (or risk aversion) parameter. The goal is to compute the implied risk tolerances of given portfolios (which are likely not to be on the nondominated frontier) by matching them with close-by portfolios that are on the nondominated frontier and then considering those portfolios’ risk tolerances as proxies for the risk tolerances of the given portfolios.

A fundamental input for the inverse portfolio optimizations conducted in this paper is a nondominated surface. Here, our paper relates to Xidonas & Mavrotas (2012), who use an \(\varepsilon\)-constraint method (see Haimes et al., 1971) to generate a discretized representation of a nondominated surface. Because accurate knowledge of the nondominated surface is important in the inverse optimization, we use the approach of Hirschberger et al. (2013) which enables us to compute the entire nondominated surface analytically, rather than being consigned to working only with a collection of dispersed dots.

The second contribution consists of an empirical part, where we apply inverse portfolio optimization in a tri-criterion framework to socially responsible (SR) mutual funds. Here, our aim is to understand whether the term “socially responsible” is merely a sales pitch or whether fund managers in fact really take social responsibility into account throughout the whole asset allocation process.

\(^1\)Markowitz (1952) calls this the efficient frontier, but we prefer the term nondominated frontier.
A socially responsible investing setting is appropriate for studying, in addition to financial criteria, a non-financial (third) criterion for two reasons. Firstly, the amounts already invested in SR mutual funds point to the demand for such products. Thus, there are clearly investors with further preferences besides financial ones. Secondly, with there being agencies that rate the socially responsible efforts of firms, studies can be conducted to investigate the ways SR mutual funds do or do not actually incorporate social responsibility into their operations.

We examine a broad sample of conventional and SR mutual funds and use ESG-scores from the Thomson Reuters ASSET4 database. These ESG-scores (where ESG stands for “environment, social and governance”) are assessments of firms’ efforts to satisfy standards with respect to social responsibility as in the AA1000 AccountAbility Principles. A typical ESG-score is computed as follows. A firm is rated on a number of issues such as employment quality, health and safety, human rights, product responsibility, emissions, board composition, and so forth. Finally, the ratings are aggregated in such a way that the best firm in the rating universe gets a score of 100 percent, and the worst gets a score of 0.

The results of the empirical part in this paper relate to the recent literature comparing the performance of conventional (or unscreened) mutual funds to SR mutual funds or screened portfolios in general (Bello, 2005; Guerard, 1997; Hamilton et al., 1993). In particular, we cannot confirm that conventional mutual funds exhibit superior financial performance. Moreover, in contrast to Hamilton et al. (1993), we find that conventional mutual funds tend of have, if anything, a higher portfolio return volatility. Although the screening process leads to fewer opportunities for diversification and hence a smaller feasible region in decision space, we find that SR investors do not have to accept significantly higher risk. This indicates that socially responsible firms can be less prone to earnings shocks.

However, comparing only financial performance ignores that investors may gain additional utility by specifically investing in socially responsible companies. In general, this additional utility stems from higher ESG fund scores. Commonly, SR funds follow a two-stage process. In the first stage, they filter out firms that do not meet their specific requirements regarding social responsibility by a screening process. In the second stage, the fund’s total wealth is allocated across the remaining assets. Somewhat surprisingly, we find that SR mutual funds do not exhibit significantly higher ESG-scores than their conventional counterparts. Yet, on average, SR mutual funds show higher minimum ESG-scores than conventional funds. This confirms the practicability of screening procedures. Using inverse portfolio optimization, we also find that SR mutual funds’ managers are not too anxious to give up financial performance in favor of higher ESG-scores in the
second stage.

The paper is organized as follows. The tri-criterion model, which produces a surface as opposed to a frontier, is theoretically introduced in Section 2. This is followed by an explanation of our inverse portfolio optimization process in Section 3. We explain our data in Section 4. Section 5 enumerates the hypotheses tested. Results are discussed in Section 6, and with final remarks, the paper concludes in Section 7.

2. Model

We now describe the model used in our study. Starting with a general von Neumann & Morgenstern (1947) utility function \( u \), the expected utility, as shown by Pratt (1964), of a portfolio with random portfolio return \( R_P \) can be computed as

\[
E[u(R_P)] = u(\mu_P) + \frac{1}{2} u''(\mu_P) \sigma_P^2 + o(\sigma_P^2) \tag{1}
\]

where \( \mu_P \) denotes expected portfolio return and \( \sigma_P^2 \) denotes the variance of portfolio return. The residual term \( o(\sigma_P^2) \) is of smaller order than \( \sigma_P^2 \). Following common practice, we go along with Feldstein (1969); Tobin (1969); Tsiang (1972); Bierwag (1974); Levy (1974); Chamberlain (1983) who show that under the assumption that the investor’s utility function is quadratic or that security returns follow a multinormal distribution, maximizing (1) is *equivalent* to maximizing

\[
\hat{\Psi}(\mu_P, \sigma_P^2, A) = -\frac{1}{2} A \sigma_P^2 + \mu_P, \tag{2}
\]

where \( A = -u''(\mu_P)/u'(\mu_P) \) is Arrow’s absolute risk aversion (Arrow, 1965). By “equivalent” we mean that maximizing (1) and (2) yield the same solutions. Substituting \( \lambda_\mu = 2/A \) and multiplying (2) by \( \lambda_\mu \), we have\(^2\)

\[
\Psi(\mu_P, \sigma_P^2, \lambda_\mu) = -\sigma_P^2 + \lambda_\mu \mu_P. \tag{3}
\]

In (3), \( \lambda_\mu \) represents the risk tolerance of the investor regarding expected return. Varying the risk tolerance parameter \( \lambda_\mu \) over the nonnegative portion of the real line and maximizing \( \Psi \) causes expected utility to yield the Markowitz nondominated frontier in variance-expected return space as in Figure 1. It is to be noted that

\(^2\)As a technical note, this transformation requires \( u'' \neq 0. \) However, for risk-averse investors, \( u'' < 0 \) always holds due to the concavity of \( u. \)
the nondominated frontier is segment-wise parabolic. That is, it is made up of a connected collection of parabolic segments. The dots in Figure 1 along the nondominated frontier are where the different parabolic segments connect with one another.

Since expected utility $\Psi$ cannot explain why certain investors would specifically choose SR mutual funds, it is suggested by Bollen (2007) and Derwall et al. (2011, and references therein) that these investors also obtain utility from the social component of their investments. Furthermore, Ballestero et al. (2012) develop a financial-ethical model to select optimal portfolios for SR funds based on the three criteria of expected return, variance of return, and risk aversion. They use the concept of stochastic goal programming (Ballestero, 2001) to incorporate an ethical goal into the framework of classical utility theory under uncertainty.

As discussed earlier, we confine the general Bernoulli decision theory setting (1) to a mean-variance setup (3). Here, expected portfolio return is calculated as

$$\mu_p = \mu^T x$$

and portfolio variance is calculated as

$$\sigma_p^2 = x^T \Sigma x,$$

where $x$ is a portfolio composition vector (elements sum to one), $\mu$ is a vector of individual security expected returns, and $\Sigma$ is an $n \times n$ covariance matrix. As usual, $x^T$ symbolizes the transpose vectors of $x$. 
We now temporarily consider the approach of Bollen (2007) who proposes adding \( \nu_P \) to the expected utility expression of (3) where \( \nu_P \) is a simple indicator function equaling one if a portfolio satisfies the investor’s requirement for social responsibility. However, this makes \( \nu_P \) a subjective quantity depending on each investor’s perception. Therefore, we strive for a more objective measure of \( \nu_P \) and feel that the social component is to be taken into account more appropriately in an investor’s expected utility function by the continuous quantity

\[
\nu_P = \nu^T x
\]

where \( \nu \) is a vector of individual security ESG-scores. While the future ESG-score of a portfolio could be interpreted as a stochastic quantity, we consider in this paper the ESG-score to be deterministic. Hence, in updating (3) to account for social responsibility, we now have for expected utility

\[
\Phi(x, \Sigma, \mu, \nu, \lambda, \Lambda) = -\sigma_P^2 + \lambda_\mu \mu_P + \lambda_\nu \nu_P = -x^T \Sigma x + \lambda_\mu \mu^T x + \lambda_\nu \nu^T x
\]

(4)

where \( \lambda_\nu \) is the risk tolerance of the investor regarding social responsibility.

Varying the risk tolerance pair \((\lambda_\mu, \lambda_\nu)\) over the nonnegative portion of \( \mathbb{R}^2 \) (which we call parameter space) and maximizing \( \Phi \) causes expected utility to yield, in contrast to the nondominated frontier, a nondominated surface in variance-expected return-ESG space as in Figure 2. In particular, maximizing \( \Phi \) for \( \lambda_\mu \to \infty \) (while keeping \( \lambda_\nu \) fixed) yields a maximum expected return portfolio, and maximizing \( \Phi \) for \( \lambda_\nu \to \infty \) (while keeping \( \lambda_\mu \) fixed) yields a maximum ESG portfolio. It is to be noted that the nondominated surface is platelet-wise paraboloidic. That is, it is made up of a connected collection of (curved) paraboloidic platelets. The lines on the surface show where the different paraboloidic platelets abut one another.

Whereas, since Markowitz (1956), it has been possible to compute a nondominated frontier as in Figure 1, it has not been until recently, namely by Hirschberger et al. (2013), that it has been possible to compute a tri-criterion nondominated surface as in Figure 2. Using the CIOS (Custom Investment Objective Solver) code from Hirschberger et al. (2013) to solve for the tri-criterion nondominated surface, we are able to obtain from its output, for each paraboloidic platelet, (a) the platelet’s 2-dimensional polyhedron of \((\lambda_\mu, \lambda_\nu)\) vectors in parameter space and (b) the platelet’s polyhedron of efficient portfolio composition vectors in decision space. Recall that a point is efficient if and only if its image in criterion space is nondominated (where a point is nondominated if and only if it is impossible to move from it to another without at least deteriorating one criterion).
To conclude the section, we wish to discuss on a few general issues that one might come up with when considering expected utility $\Phi$ from equation (4).

1. Notice that the concept of the rate of substitution for two objectives (see Krugman & Wells, 2009; Pindyck & Rubinfeld, 2005) is not general enough to consider indifference curves of constant utility depending on variance, expected return and level of social responsibility. As both expected return and social responsibility are linear objectives, their marginal rate of substitution appears to be linear, but this is only true if variance is zero. If variance is positive the indifference curves follow the exact differential equation

$$0 = \frac{\partial \Phi}{\partial \mu_p} d\mu_p + \frac{\partial \Phi}{\partial \nu_p} d\nu_p + \frac{\partial \Phi}{\partial \sigma_p^2} d\sigma_p^2.$$ 

Here, the left hand side is zero since the marginal utility of the entire utility function does not change on a fixed indifference curve. Clearly, the rate of substitution in the values of expected return and social responsibility is linear, since both functions are linear in the portfolio weights. However, the change of portfolio weights also affects the variance of the portfolio, which is quadratic in the portfolio weights. Hence, given a fixed level of variance, the expected return/ESG-score indifference curve is not linear.
2. Our setting allows that risk aversion depends on initial wealth. More precisely, the level of risk tolerance depends on the level of the initial wealth due to the fact that each decision maker chooses an efficient portfolio which is appropriate to her risk appetite according to both linear objectives as well as her level of initial wealth.

3. When considering equation (3) in the case of a risk-free investment $R_f$, the expected utility $\Psi = \lambda \mu R_f$ depends on the risk tolerance $\lambda \mu$. On first glance, this might seem counter-intuitive, since in equation (2) the equivalent expected utility $\hat{\Psi} = R_f$ was fixed for risk-free investments. However, notice that equation (2) does not denote the investor’s expected utility directly, but is an equivalent utility function when it comes to maximizing the expected utility (1). In fact, according to equation (1), the expected utility of a risk-free investment is $E[u(R_f)] = u(R_f)$ (see for instance Copeland et al., 2005, Chapter 3), which also depends on the investor’s individual utility function.

3. Inverse Optimization Process

For the experiments of this paper, inverse optimizations are needed to compute the values of the risk tolerance parameters $\lambda \mu$ and $\lambda \nu$ that are implied by a given fund’s portfolio composition vector on a given date (subsequently called a reporting date). In regular optimization, there is an objective function whose parameters are all fixed. Then we find the point in decision space that optimizes it. In inverse optimization, we start with a point in decision space. Usually, this point, which we designate $w$, is not efficient. Then the endeavor is to identify the efficient point that is closest to $w$ as a proxy point. Continuing, we then try to find what the values of certain parameters in the objective function would have to be for the proxy point to optimize the objective function. Strictly speaking, these are the parameters of the closest efficient point only, but they are then considered to be the implied parameters for $w$. The parameter values we are looking for in this paper are the $\lambda \mu$ and $\lambda \nu$ in $\Phi$. Because the $\lambda \mu$ and $\lambda \nu$ of a portfolio cannot be computed directly, inverse portfolio optimization is used, which here involves the following steps:

1. Start with a given fund’s portfolio composition vector $w$ on a given reporting date.

2. Compute for the given fund on the given date its tri-criterion (financial variance-reward-ESG) nondominated surface, designated $\hat{E}$.

Here, we make two assumptions before calculating the nondominated surface. One is that the assets a fund is invested in on a given reporting date comprise
all of the assets that the fund can actually invest in at that time, for the asset allocation. The other is that due to risk control, the fund enforces a minimal and maximal investing rule. That is, for the calculation of the nondominated surface there is a minimum and maximum amount that can be invested in each asset. These amounts are given by the actual individual minimum and maximum investments of the fund, i.e., by \( \min_j \{w_j\} \) and \( \max_j \{w_j\} \).

Therefore, the nondominated surface for a given vector \( w \) consists of the images of all efficient portfolios that are achievable with the given assets present in \( w \) and under the aforementioned minimum and maximum constraints.

We use the CIOS algorithm to compute the nondominated surface. As an output from CIOS, we obtain information about all of the efficient polyhedra in decision space whose images form the platelets which together form the nondominated surface in criterion space.

3. Now, from the set of all efficient portfolio composition vectors \( E \), we need to find in decision space the \( x \in E \) closest (in Euclidean norm) to \( w \). We start by first comparing all of the vertices of all of the efficient polyhedra in decision space with \( w \). We only work with vertices at this point because in the output of CIOS, the polyhedra of efficient portfolio composition vectors are only characterized by their vertices.

4. After finding the closest vertex, we determine all of the polyhedra of efficient portfolio composition vectors (ascertainable from the output of CIOS) that are incident to this vertex (i.e., all polyhedra that have the “closest vertex” as one of its vertices).

5. Pixelate the polyhedron in \((\lambda_{\mu}, \lambda_{\nu})\) parameter space associated with each polyhedron of efficient portfolio composition vectors identified in the above step. Utilizing the relationship in Hirschberger et al. (2013) that enables us to specify the \( x \in E \) for each point in \((\lambda_{\mu}, \lambda_{\nu})\) parameter space, we are able to pixelate all incident polyhedra in \( E \).

6. Compare each pixelated point in \( x \)-space with \( w \).

7. Then the point in parameter space associated with the pixelated \( x \)-vector that is closest (in Euclidean norm) to \( w \) yields the desired \( \lambda_{\mu} \) and \( \lambda_{\nu} \) implied risk tolerance values.

Given a fund containing \( n \) securities, the inverse optimization process is expressed mathematically as
\[
\min_{\lambda, \mu, \nu} \|x - w\| \\
\text{s.t. } x \in \mathcal{E}(\lambda, \mu, \nu)
\]  

where \( \mathcal{E}(\lambda, \mu, \nu) \) is the solution of

\[
\max_x \Phi(x, \Sigma, \mu, \nu, \lambda, \mu, \lambda, \nu) \\
\text{s.t. } \sum_{i=1}^{n} x_i = 1 \\
x_i \geq \min_j \{w_j\} \text{ for all } i \\
x_i \leq \max_j \{w_j\} \text{ for all } i
\]

With regard to the outer minimization problem (5), note that the inner maximization problem (6a)–(6d) is a scalarized version of the actual tri-criterion portfolio selection model

\[
\min x^T \Sigma x \\
\max x^T \mu \\
\max x^T \nu \\
\text{s.t. } \sum_{i=1}^{n} x_i = 1 \\
x_i \geq \min_j \{w_j\} \text{ for all } i \\
x_i \leq \max_j \{w_j\} \text{ for all } i
\]

that we have been alluding to all along. The solution \( x \) of the outer minimization problem (5) is the efficient portfolio corresponding to portfolio \( w \). Figure 3 displays the procedure of inverse optimization graphically.

Notice that according to equation (4), the nondominated surface \( \mathcal{E}(\lambda, \mu, \nu) \) could also be computed by solving a family of quadratic programs. One possibility would be the weighting method, where (6a)–(6d) is solved for many \( (\lambda, \mu, \nu) \) pairs. Another would be to use the \( \varepsilon \)-constrained method, where the portfolio variance is minimized for a given level of expected return and a given level of ESG-score. While there has been some progress on improving the performance of the \( \varepsilon \)-constrained method (Hamacher et al., 2007), the accuracy of an inverse portfolio optimization still critically depends on the accuracy of the generated nondominated surface. Yet, in order to obtain a reasonable degree of accuracy for
Figure 3: A picture illustrating the inverse portfolio optimization process where $\tilde{w}$ and $\tilde{x}$ are the criterion vectors of $w$ and $x$, and $x$ is the efficient portfolio closest to $w$ in decision space. The shaded area is the tri-criterion nondominated surface $\tilde{E}$. The white lines across the surface are iso-quants of the variance objective.

The nondominated surface, it would have to be pixelated to the extent of at least several thousand points thus requiring that many quadratic programs to be solved to generate a single nondominated surface. Since we calculate more than 23,000 nondominated surfaces in total (see Section 4), such a direct approach we require the solution of upwards of 50 million quadratic programming problems.

4. Data and Summary Statistics

In this study, after computing returns from stock prices in Thomson Reuters Datastream, we use a rolling window approach where each window has a length of 120 months and ends on a date that a fund reports its portfolio composition. In particular, we perform an inverse portfolio optimization for each reported fund composition and compute the one-month out-of-sample returns, i.e., the returns one month after the reported fund composition, both for the fund’s portfolio and for its respective efficient portfolio.

As for ESG-scores, they are based on actions already implemented in an observed company as well as on planned or started programs regarding the socially responsible performance of a company. Consequently, an ESG-score obtained from existing data at the end of month $t$ is an appropriate proxy for the socially responsible performance in month $t + 1$. 

11
The mutual fund compositions, which constitute the main information needed for the inverse optimizations are derived from the CRSP Survivor-Bias-Free US Mutual Fund database over the period December 31, 2001 to February 27, 2010. In contrast to returns, data on mutual fund compositions are typically not available more frequently than quarterly, with the dates on which composition information is reported called, as referred to earlier, *reporting dates*. Because of incomplete ESG data over the period of the study, certain funds are dropped. If it is not possible to obtain ESG-scores covering over 70% of a fund’s portfolio, the fund is dropped. Finally, we filter out all funds with \( n < 10 \) or \( n > 150 \) securities. For identifying which are SR mutual funds, we utilize the classifications of the U.S. Social Investment Forum (SIF)\(^3\). This then leaves us with 27 SR mutual funds and 2,346 conventional mutual funds. For the 27 SR mutual funds, we have a total of 276 different portfolio compositions over the different reporting dates. For the 2,346 conventional mutual funds, we have a total of 22,746 different portfolio compositions over the different reporting dates.

In Table 1 we list, by year, the number of different portfolio compositions, the mean portfolio ESG-score, median portfolio ESG-score, minimum portfolio ESG-score, maximum portfolio ESG-score, and the standard deviation of the different portfolio composition ESG-scores. A fund’s composition is included in a given year depending upon the date a change in composition is reported. Note that the average mean and average maximum ESG-scores of the SR mutual funds do not exceed those of the conventional mutual funds for all years. Nevertheless, the average minimum ESG-scores of SR mutual funds – that an SR investor would suppose to be higher – are indeed higher than those of conventional mutual funds. This indicates that the screening approaches applied by SR mutual funds are capable of building portfolios that exceed minimal requirements in social responsibility.

As the totals of SR mutual funds in 2002 and 2010 are 0 and 2, respectively, we do not consider these two years furthermore.

## 5. Hypothesis Development

In this section, we develop testable hypotheses regarding the investment policies of SR mutual funds. While the first two hypotheses consider the actual level of social responsibility of a fund, the next four hypotheses regard solely financial performance.

\(^3\)SIF provides a list with all known, socially responsible screened mutual funds online at the US SIF website at [http://ussif.org/resources/mfpc/screening.cfm](http://ussif.org/resources/mfpc/screening.cfm).
As already hinted above, asset allocation in a SR mutual fund is typically conducted in a two-stage approach. In the first stage, a set of suitable assets is selected by some kind of screening of all available assets, which is a binary selection process of the assets an investor is willing to buy. Among the criteria for the screening process can be requirements about the size and liquidity of a stock, or certain predefined standards regarding social responsibility (see Renneboog et al., 2008). In the second stage, the fund manager then allocates the fund’s total wealth to the selected assets. We wish to analyze whether this asset allocation ignores social responsibility, i.e., is influenced only by the expected returns and the covariances of the returns, or whether social responsibility is taken into account in the second stage, too.

**Hypothesis 1a:** Asset allocation after screening in SR mutual funds depends upon the social responsibility of the individual assets.
Hypothesis 1a is about whether ESG-scores continuing to play a role in the asset allocation of the second stage, i.e., after the screening process has been conducted. Notice that the importance of the ESG-score in the second-stage asset allocation is measured by the implied risk tolerance with respect to the ESG-score $\lambda_v$. For instance, small values of $\lambda_v$ indicate that fund managers weight the ESG-scores only marginally, whereas a high $\lambda_v$ indicates a high appreciation of the ESG-scores.

Consequently, it can be asked (below) whether the ESG-scores of SR mutual fund portfolios exceed those of their conventional peers.

**Hypothesis 1b:** SR mutual funds show higher weighted ESG-scores than conventional mutual funds.

A high ESG-score can be explained either by the screening process or by giving assets with high ESG-scores more weight in a fund’s portfolio, i.e., by operating with a higher risk tolerance $\lambda_v$.

Having challenged funds’ abilities to incorporate ESG-scores into their asset allocations, we now continue with hypotheses about their skills in generating financial performance. The financial performance of SR mutual funds is a heavily discussed area in literature (see Bauer et al., 2005; Bello, 2005; Guerard, 1997; Hamilton et al., 1993; Kreander et al., 2005; Mallin et al., 1995; Statman, 2000). We compare the implied financial risk tolerances, the overall returns, the overall risks (which we measure in terms of standard deviation), and the out-of-sample returns of the conventional and SR mutual funds.

**Hypothesis 2a:** SR mutual funds differ from conventional mutual funds in terms of financial risk tolerance.

**Hypothesis 2b:** SR mutual funds show lower returns than conventional mutual funds.

**Hypothesis 2c:** SR mutual funds differ from conventional mutual funds in terms of risk.

**Hypothesis 2d:** SR mutual funds differ from conventional mutual funds in terms of out-of-sample return.

Table 2 formally summarizes the hypotheses from above. As for notation, in line H1a, the Alternative Hypothesis is that the risk tolerance parameter regarding social responsibility is higher in SR mutual funds than in conventional mutual
Table 2: Test hypotheses.

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>((\lambda_v)_C \geq (\lambda_v)_S)</td>
<td>((\lambda_v)_C &lt; (\lambda_v)_S)</td>
</tr>
<tr>
<td>((w^T\nu)_C \geq (w^T\nu)_S)</td>
<td>((w^T\nu)_C &lt; (w^T\nu)_S)</td>
</tr>
<tr>
<td>((\lambda_p)_C = (\lambda_p)_S)</td>
<td>((\lambda_p)_C \neq (\lambda_p)_S)</td>
</tr>
<tr>
<td>((w^T\mu)_C \leq (w^T\mu)_S)</td>
<td>((w^T\mu)_C &gt; (w^T\mu)_S)</td>
</tr>
<tr>
<td>((\sqrt{w^T\Sigma w})_C = (\sqrt{w^T\Sigma w})_S)</td>
<td>((\sqrt{w^T\Sigma w})_C \neq (\sqrt{w^T\Sigma w})_S)</td>
</tr>
<tr>
<td>((w^T\Omega)OS)_C = (w^T\Omega)OS)_S)</td>
<td>((w^T\Omega)OS)_C \neq (w^T\Omega)OS)_S)</td>
</tr>
</tbody>
</table>

funds. As another example, in line H2d, the Null Hypothesis is that the out-of-sample returns of conventional mutual funds are equal to the out-of-sample returns of the actual portfolios of SR mutual funds.

6. Results

Using a rolling window approach, the latest available ESG-scores before a reporting date, the inverse portfolio optimization process, and out-of-sample returns, we now have, for our hypothesis testing, the descriptive statistics of Table 3. This table and the following tests are based on the computation of an inverse optimization for each of the 23,022 different fund compositions reported above.

Continuing with the notation that \(x\) is an efficient portfolio and \(w\) is an already existing mutual fund portfolio, we now have by year, in Table 3, median values for portfolio in-sample volatility\(^4\), implied financial risk tolerance, monthly portfolio in-sample returns, and monthly portfolio out-of-sample returns. Also, we have median values for implied socially responsible risk tolerance, and in-sample portfolio ESG-scores. In the last column of the table we have median values for the distances between the \(x\)’s and their \(w\)’s in our experiments.

Especially for the implied \(\lambda_v\) and \(\lambda_p\), the median is a more appropriate statistic than the mean to depict the data in a single quantity because of skewness and outliers. Even so, the mean values (not reported) of the empirical distributions of all columns, apart from those of the two implied risk tolerances, do not differ much from the reported medians anyway. The implied risk tolerance parameters \(\lambda_p\) and \(\lambda_v\) range between 0 and 13,112 \((\lambda_p)\), and 0 and 2,215 \((\lambda_v)\), respectively. However,

\(^4\)We report volatility in opposite to variance in the remainder of this section since volatility is given in the same unit (percent per month) as return.
the informative value of these intervals is rather limited due to the fact that the level of the risk tolerance parameters depends on the scale of $\mu$ and $\nu$ as well as on the specification of the utility function (see Kallberg & Ziemba, 1983, for an extensive discussion).

As mean values of the parameters could be biased by outliers, we test our hypotheses using the non-parametric Mann-Whitney $U$-Test. We report the test statistics and the corresponding $p$-values of each test in Table 4.

The parameters $\lambda_\mu$ and $\lambda_\nu$ display the implied risk tolerances, which we compute via inverse portfolio optimization as described earlier. Since the two risk tolerance parameters are computed from the efficient portfolios $x$ rather than from the funds’ real portfolios $w$, one could be fearful that the implied parameters do not reflect the funds’ actual parameters reasonably well. However, when considering the distances $\|x - w\|$ reported in Table 3, we see that the two fund compositions do not differ a lot.

<table>
<thead>
<tr>
<th>Year</th>
<th>$\sqrt{w^T \Sigma w}$</th>
<th>$\lambda_\mu$</th>
<th>$w^T \mu$</th>
<th>$w^T \mu^{OS}$</th>
<th>$\lambda_\nu$</th>
<th>$w^T \nu$</th>
<th>$|x - w|$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>0.0482</td>
<td>0.22</td>
<td>0.0145</td>
<td>0.0279</td>
<td>0.0084</td>
<td>0.64</td>
<td>0.15</td>
</tr>
<tr>
<td>2004</td>
<td>0.0484</td>
<td>0.16</td>
<td>0.0156</td>
<td>-0.0080</td>
<td>0.0080</td>
<td>0.67</td>
<td>0.15</td>
</tr>
<tr>
<td>2005</td>
<td>0.0467</td>
<td>0.16</td>
<td>0.0139</td>
<td>0.0165</td>
<td>0.0088</td>
<td>0.69</td>
<td>0.14</td>
</tr>
<tr>
<td>2006</td>
<td>0.0470</td>
<td>0.14</td>
<td>0.0125</td>
<td>0.0151</td>
<td>0.0108</td>
<td>0.69</td>
<td>0.14</td>
</tr>
<tr>
<td>2007</td>
<td>0.0462</td>
<td>0.13</td>
<td>0.0110</td>
<td>-0.0154</td>
<td>0.0122</td>
<td>0.70</td>
<td>0.14</td>
</tr>
<tr>
<td>2008</td>
<td>0.0478</td>
<td>0.09</td>
<td>0.0089</td>
<td>-0.0033</td>
<td>0.0061</td>
<td>0.71</td>
<td>0.15</td>
</tr>
<tr>
<td>2009</td>
<td>0.0608</td>
<td>0.17</td>
<td>0.0082</td>
<td>0.0360</td>
<td>0.0146</td>
<td>0.71</td>
<td>0.14</td>
</tr>
</tbody>
</table>

**Panel (S): SR Mutual Funds**

<table>
<thead>
<tr>
<th>Year</th>
<th>$\sqrt{w^T \Sigma w}$</th>
<th>$\lambda_\mu$</th>
<th>$w^T \mu$</th>
<th>$w^T \mu^{OS}$</th>
<th>$\lambda_\nu$</th>
<th>$w^T \nu$</th>
<th>$|x - w|$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>0.0466</td>
<td>0.23</td>
<td>0.0144</td>
<td>0.0276</td>
<td>0.0094</td>
<td>0.61</td>
<td>0.17</td>
</tr>
<tr>
<td>2004</td>
<td>0.0452</td>
<td>0.06</td>
<td>0.0157</td>
<td>-0.0061</td>
<td>0.0038</td>
<td>0.64</td>
<td>0.12</td>
</tr>
<tr>
<td>2005</td>
<td>0.0447</td>
<td>0.11</td>
<td>0.0124</td>
<td>0.0054</td>
<td>0.0037</td>
<td>0.68</td>
<td>0.11</td>
</tr>
<tr>
<td>2006</td>
<td>0.0463</td>
<td>0.35</td>
<td>0.0122</td>
<td>0.0117</td>
<td>0.0107</td>
<td>0.69</td>
<td>0.13</td>
</tr>
<tr>
<td>2007</td>
<td>0.0443</td>
<td>0.14</td>
<td>0.0104</td>
<td>-0.0126</td>
<td>0.0093</td>
<td>0.70</td>
<td>0.13</td>
</tr>
<tr>
<td>2008</td>
<td>0.0479</td>
<td>0.14</td>
<td>0.0087</td>
<td>-0.0386</td>
<td>0.0074</td>
<td>0.68</td>
<td>0.15</td>
</tr>
<tr>
<td>2009</td>
<td>0.0576</td>
<td>0.20</td>
<td>0.0073</td>
<td>0.0436</td>
<td>0.0096</td>
<td>0.69</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Table 4: Test statistics and p-values for the hypotheses H1a–H2d of Table 2 where *, **, *** denote significant parameters at a 10%, 5%, and 1% level, respectively.

<table>
<thead>
<tr>
<th>Corresp. Hyp.</th>
<th>λ_ν</th>
<th>w^T_ν</th>
<th>λ_μ</th>
<th>w^T_μ</th>
<th>√w^TΣw</th>
<th>w^Tμ^OS</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003 Test stat.</td>
<td>3959</td>
<td>6302</td>
<td>4240</td>
<td>4421</td>
<td>5165</td>
<td>4596</td>
<td>(0.2828)</td>
</tr>
<tr>
<td>2004 Test stat.</td>
<td>15963</td>
<td>15920</td>
<td>17782</td>
<td>14380</td>
<td>19448</td>
<td>12766</td>
<td>(0.8957)</td>
</tr>
<tr>
<td>2005 Test stat.</td>
<td>30354</td>
<td>30366</td>
<td>27142</td>
<td>29950</td>
<td>32411</td>
<td>27359</td>
<td>(0.9551)</td>
</tr>
<tr>
<td>2006 Test stat.</td>
<td>45245</td>
<td>43022</td>
<td>34188</td>
<td>44738</td>
<td>42450</td>
<td>44320</td>
<td>(0.7091)</td>
</tr>
<tr>
<td>2007 Test stat.</td>
<td>44807</td>
<td>47262</td>
<td>40688</td>
<td>40845</td>
<td>39783</td>
<td>40055</td>
<td>(0.7832)</td>
</tr>
<tr>
<td>2008 Test stat.</td>
<td>132479</td>
<td>147577</td>
<td>105374</td>
<td>129152</td>
<td>118989</td>
<td>139180</td>
<td>(0.7528)</td>
</tr>
<tr>
<td>2009 Test stat.</td>
<td>410838</td>
<td>414671</td>
<td>361563</td>
<td>404457</td>
<td>461939</td>
<td>362837</td>
<td>(0.8786)</td>
</tr>
</tbody>
</table>

Generally, zero values for λ_μ or λ_ν imply infinite risk aversion. By this we mean that no efforts will be made to increase expected return (in the case of λ_μ = 0) or ESG-score (in the case of λ_ν = 0) if it comes at the expense of increasing risk above its minimal value. While a few funds exhibit a zero λ_μ or a zero λ_ν, we see that most fund managers are willing to take on risk above the minimum to achieve additional expected return and/or higher ESG-scores, as the medians of both λ’s are positive.

To determine whether screening is the only approach used to build SR mutual funds, we test to see if the implied tolerances λ_ν are greater in SR mutual funds than in conventional mutual funds. In principle, conventional mutual funds are to maximize expected utility containing only financial terms implying a λ_ν of zero. However, conventional mutual funds may exhibit non-zero λ_ν’s due to random noise stemming from the inverse optimizations. If SR mutual funds value assets with high ESG-scores more, then we would expect their λ_ν’s to exceed the random λ_ν’s of their conventional counterparts. Yet, we find no statistically significant evidence to reject the null hypothesis of H1a in any year. Therefore, we cannot confirm that managers of SR mutual funds pursue any socially responsible strategies in their
second-stage asset allocations that conventional fund managers do not. This is in clear contrast to the common investor opinion that SR mutual funds would not be disregarding social responsibility issues after their initial screenings have been made. Next, we test whether there are any differences in the socially responsible performance of the two types of funds at all. Since we have already found that SR mutual funds do not show higher risk tolerances with respect to social responsibility, higher ESG performances had to be attributed to the screening processes. However, we cannot reject the null of H1b at any arbitrary confidence level. Therefore, our data suggest that SR mutual funds do not exhibit higher ESG-scores than conventional mutual funds.

With regard to the commonly used term of financial risk tolerance displayed as $\lambda_{\mu}$, we investigate H2a, but we find no strong evidence with the Mann-Whitney U-Test in any year that there is a difference between the $\lambda_{\mu}$’s of the SR mutual funds and the conventional funds. Nevertheless, we find some evidence to reject the null in the years 2004, 2006 and 2008, which indicates that the financial risk tolerance of conventional and SR mutual funds differ in those years.

Adler & Kritzman (2008) and Dorfleitner & Utz (2012) show that building portfolios in which ESG-score is an objective yields a decrease in returns relative to portfolios where social responsibility is of only minor importance. Following this result, we test in H2b whether the returns of SR mutual funds are significantly lower than the returns of conventional funds. Apart from weak empirical evidence to reject the null in 2005, we do not find any statistical evidence in other years. This indicates that we cannot reject the null hypothesis, which states that the distribution function of the portfolio return of the conventional mutual funds is not significantly higher than that of the SR mutual funds, at any arbitrary significance level. Therefore, we are in line with several former studies (Bauer et al., 2005; Bello, 2005; Guerard, 1997; Hamilton et al., 1993; Kreander et al., 2005; Mallin et al., 1995; Statman, 2000) that find no significant differences in fund performance between conventional and SR mutual funds.

We also test the standard deviations of the mutual funds and find that the null of H2c can be rejected at the 1% significance level in 2004 and 2009 and at the 5% significance level in 2005. Thus, in our sample, conventional mutual funds differ with some significance from SR mutual funds in terms of standard deviation. Moreover, the data in Table 3 show that the standard deviations of conventional mutual all years but 2008.

The out-of-sample fund returns, which describe real fund performance in the month following the date of portfolio composition, are tested in H2d. We
check whether the distribution of one-month out-of-sample returns of conventional mutual funds differs from the distribution of one-month out-of-sample returns of SR mutual funds. We do not find any statistical evidence to reject the null. Thus, we cannot conclude that there are statistically significant differences between the two respective distributions.

Summarizing the results of Table 4, we cannot find any statistical evidence that implies profound differences between conventional and SR mutual funds—just that SR mutual funds have, as seen in Table 1, higher minimum ESG-scores than conventional funds showing that screening methods successfully keep firms with poor socially responsible records out of the portfolios.

Furthermore, we check the robustness of our results. With regard to supposed sensitivity of the results of a Markowitz model to its input parameters (see for example Fabozzi et al., 2002), we apply different methods for estimating (i.e., exponentially weighted moving average) the financial parameters and we also vary the time period upon which the estimation is based. In summary, we do not find any profound differences in the tests’ results.

7. Summary and Conclusion

In this article, we present a framework for inverse portfolio optimization in a Markowitz model, which is extended to include a third criterion. Using the prescribed inverse optimization process we are able to compute the implied risk tolerance parameters of mutual funds that pursue a third, non-financial criterion.

In applying this capability to a broad sample of conventional and socially responsible (SR) mutual funds, we find that after the screening process there appears to be no significant difference in how assets are allocated in socially responsible and conventional mutual funds. By considering the implied risk tolerance with respect to the ESG-score, we are unable to find evidence that social responsibility is taken into account in the asset allocation stage at all. With regard to key indicators such as volatility, expected return, and ESG-scores (used to measure social responsibility), we find a slightly lower volatility for the SR mutual funds’ returns, but we find no significant differences for the expected return and the mean ESG-scores. Merely, on average, the minimum ESG-score in an SR mutual fund is markedly higher than the minimum ESG-score in a conventional mutual fund.

The methodological ideas developed in this paper pave the ground to various new research. For instance, the approach is not bound to the application of SR mutual funds. Applying a tri-criterial model to other real-world problems might be of interest both for practitioners and the scientific community. Moreover, while the
decision model in the inverse optimization in this paper was tri-criterial, further research could broaden the model to include more than three criteria.

References


