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

Sebastian Utz [a], Maximilian Wimmer [a], Markus Hirschberger [b], Ralph E. Steuer [c,*]

[a] Department of Finance, University of Regensburg, 93040 Regensburg, Germany
[b] Munich Re, Koeniginstrasse 107, 80802 Munich, Germany
[c] Department of Finance, University of Georgia, Athens, GA 30602-6253, USA

ARTICLE INFO

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 there appears to be no significant evidence that social responsibility issues, after the screening stage, are further taken into account in the asset allocation process, which is a result that is likely to be different from what many SR investors would expect.

© 2013 Elsevier B.V. All rights reserved.

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, Aouni, & Fayedh, 2007; Ballestero, Bravo, Pérez-Gladish, Arenas-Parra, & Plà-Santamaria, 2012; Bollen, 2007; Dorfleitner, Leidl, & Reeder, 2012; Dorfleitner & Utz, 2012; Hallerbach, Ning, Soppe, & Spronk, 2004; Steuer, Qi, & Hirschberger, 2007; Xidonas, Mavrotas, Krintas, Psarras, & Zopounidis, 2012). Toward that end, in Hirschberger, Steuer, Utz, Wimmer, and Qi (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 nonfinancial 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 and 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 and Mavrotas (2013), who use an ε-constraint method (see Haimes, Lasdon, & Wismer, 1971) to generate a discretized representation of a nondominated surface. Because accurate knowledge of the nondominated surface is important in inverse optimization, we use the approach of Hirschberger et al. (2013) which enables us to compute the entire nondominated surface exactly, 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

---

* Corresponding author. Tel.: +1 706 542 3782; fax: +1 706 542 9434.
  E-mail addresses: sebastian.utz@ur.de (S. Utz), maximilian.wimmer@ur.de (M. Wimmer), rsteuer@uga.edu (R.E. Steuer).

[1] Markowitz (1952) calls this the *efficient frontier*, but we prefer the term *nondominated frontier*.

pitch or whether fund managers in fact really take social responsibility into account throughout the whole investment process.

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%, 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, Jo, & Statman, 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 (screening process). In the second (asset allocation) 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. 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 and 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

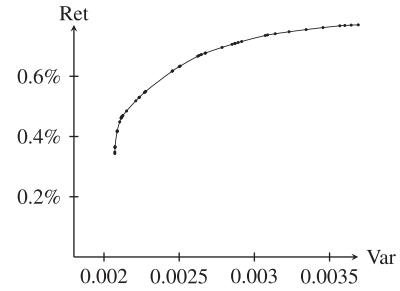$$E[u(R_P)] = u(\mu_P) + \frac{1}{2}u''(\mu_P)\sigma_P^2 + o(\sigma_P^2) \qquad (1)$$

Fig. 1. A bi-criterion nondominated frontier. It is composed of a connected collection of parabolic segments. The frontier shown is from a bi-criterion problem with $n = 50$ securities and has 46 segments.

where μ_P denotes expected portfolio return and σ_P^2 denotes the variance of portfolio return. The residual term $o(\sigma_P^2)$ is of smaller order than σ_P^2. Following common practice, we go along with Feldstein (1969), Tobin (1969), Tsiang (1972), Bierwag (1974), Levy (1974), and 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

$$\widehat{\Psi}(\mu_P, \sigma_P^2, A) = -\frac{1}{2}A\sigma_P^2 + \mu_P, \qquad (2)$$

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 λ_μ, we have[2]

$$\Psi(\mu_P, \sigma_P^2, \lambda_\mu) = -\sigma_P^2 + \lambda_\mu\mu_P. \qquad (3)$$

In (3), λ_μ represents the risk tolerance of the investor regarding expected return. Varying the risk tolerance parameter λ_μ over the nonnegative portion of the real line and maximizing Ψ causes expected utility to yield the Markowitz nondominated frontier in variance-expected return space as in Fig. 1. It is to be noted that the nondominated frontier is segment-wise parabolic. That is, it is made up of a connected collection of parabolic segments. The dots in Fig. 1 along the nondominated frontier are where the different parabolic segments connect with one another.

Since expected utility Ψ cannot explain why certain investors would specifically choose SR mutual funds, it is suggested by Bollen (2007) and Derwall, Koedijk, and Horst (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

$$\mu_P = \boldsymbol{\mu}^T\boldsymbol{x}$$

and portfolio variance is calculated as

$$\sigma_P^2 = \boldsymbol{x}^T\Sigma\boldsymbol{x},$$

---

[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.

where $\boldsymbol{x}$ is a portfolio composition vector (elements sum to one), $\boldsymbol{\mu}$ is a vector of individual security expected returns, and $\boldsymbol{\Sigma}$ is an $n \times n$ covariance matrix. As usual, $\boldsymbol{x}^T$ symbolizes the transpose of $\boldsymbol{x}$.

We now temporarily consider the approach of Bollen (2007) who proposes adding $\lambda_v v_P$ to the expected utility expression of (3) where v_P is a simple indicator function equaling one if a portfolio satisfies the investor's requirement for social responsibility. However, this makes v_P a subjective quantity depending on each investor's perception. Therefore, we strive for a more objective measure of v_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

$$v_P = \boldsymbol{v}^T \boldsymbol{x}$$

where $\boldsymbol{v}$ 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

$$\begin{aligned} \Phi(\boldsymbol{x}, \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{v}, \lambda_\mu, \lambda_v) &= -\sigma_P^2 + \lambda_\mu \mu_P + \lambda_v v_P \\ &= -\boldsymbol{x}^T \boldsymbol{\Sigma} \boldsymbol{x} + \lambda_\mu \ \boldsymbol{\mu}^T \boldsymbol{x} + \lambda_v \ \boldsymbol{v}^T \boldsymbol{x} \end{aligned} \quad (4)$$

where λ_v is the risk tolerance of the investor regarding social responsibility.

Varying the risk tolerance pair (λ_μ, λ_v) over the nonnegative portion of $\mathbb{R}^2$ (which we call *parameter space*) and maximizing Φ causes expected utility to yield, in contrast to the nondominated frontier, a nondominated surface in variance-expected return-ESG space as in Fig. 2. In particular, maximizing Φ for $\lambda_\mu \to \infty$ (while keeping λ_v fixed) yields a maximum expected return portfolio, and maximizing Φ for $\lambda_v \to \infty$ (while keeping λ_μ 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 different paraboloidic platelets abut one another.

Whereas, since Markowitz (1956), it has been possible to compute a nondominated frontier as in Fig. 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 Fig. 2. Using the CIOS (Custom Investment Objective Solver) code from Hirschberger et al. (2013) to solve for a tri-criterion nondominated
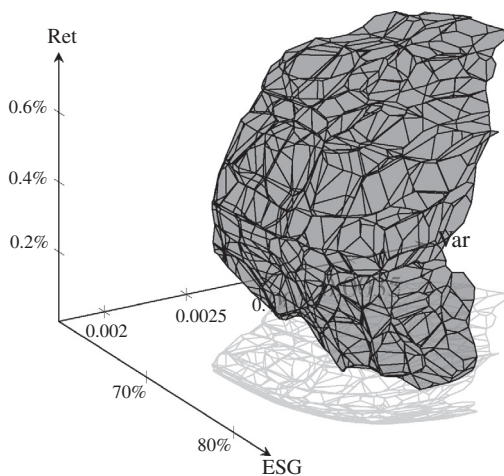
surface, we are able to obtain from its output, for each paraboloidic platelet, (a) the platelet's 2-dimensional polyhedron of (λ_μ, λ_v) vectors in parameter space and (b) the platelet's polyhedron of efficient portfolio composition vectors in decision space. Recall that a point is *efficient* if and 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 discuss a few general issues that one might encounter when considering expected utility Φ from Eq. (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

$$0 = \frac{\partial \Phi}{\partial \mu_P} \mathrm{d}\mu_P + \frac{\partial \Phi}{\partial v_P} \mathrm{d}v_P + \frac{\partial \Phi}{\partial \sigma_P^2} \mathrm{d}\sigma_P^2.$$

   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 Eq. (3) in the case of a risk-free investment R_f, the expected utility $\Psi = \lambda_\mu R_f$ depends on the risk tolerance λ_μ. On first glance, this might seem counter-intuitive, since in Eq. (2) the equivalent expected utility $\widetilde{\Psi} = R_f$ was fixed for risk-free investments. However, notice that Eq. (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 Eq. (1), the expected utility of a risk-free investment is $\boldsymbol{E}[u(R_f)] = u(R_f)$ (see for instance Copeland, Weston, & Shastri (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 λ_μ and λ_v 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 $\boldsymbol{w}$, is not efficient. Then the endeavor is to identify the efficient point that is closest to $\boldsymbol{w}$ 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

Fig. 2. A tri-criterion nondominated surface. Note that while a nondominated frontier is composed of a connected collection of parabolic segments, a nondominated surface is composed of a connected collection of paraboloidic "platelets." The surface shown is from a tri-criterion problem with $n = 50$ securities and has 1460 platelets.

only, but they are then considered to be the *implied* parameters for $\boldsymbol{w}$. The parameter values we are looking for in this paper are the λ_μ and λ_v in Φ. Because the λ_μ and λ_v 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 $\boldsymbol{w}$ 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 $\widetilde{\mathcal{E}}$.

   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. 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 $\boldsymbol{w}$ consists of the images of all efficient portfolios that are achievable with the given assets present in $\boldsymbol{w}$ 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 $\mathcal{E}$, we need to find in decision space the $\boldsymbol{x} \in \mathcal{E}$ closest (in Euclidean norm) to $\boldsymbol{w}$. We start by first comparing all of the vertices of all of the efficient polyhedra in decision space with $\boldsymbol{w}$. 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 if its vertices).
5. Pixelate the polyhedron in (λ_μ, λ_v) 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 $\boldsymbol{x} \in \mathcal{E}$ for each point in (λ_μ, λ_v) parameter space, we are able to pixelate all incident polyhedra in $\mathcal{E}$.
6. Compare each pixelated point in $\boldsymbol{x}$-space with $\boldsymbol{w}$.
7. Then the point in parameter space associated with the pixelated $\boldsymbol{x}$-vector that is closest (in Euclidean norm) to $\boldsymbol{w}$ yields the desired λ_μ and λ_v implied risk tolerance values.

Given a fund containing n securities, the inverse optimization process is expressed mathematically as

$$\min_{\lambda_\mu, \lambda_v} \quad \|\boldsymbol{x} - \boldsymbol{w}\| \tag{5}$$
$$\text{s.t.} \quad \boldsymbol{x} \in \mathcal{E}(\lambda_\mu, \lambda_v)$$

where $\mathcal{E}(\lambda_\mu, \lambda_v)$ is the solution of

$$\max_{\boldsymbol{x}} \quad \Phi(\boldsymbol{x}, \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{v}, \lambda_\mu, \lambda_v) \tag{6a}$$
$$\text{s.t.} \quad \sum_{i=1}^n x_i = 1 \tag{6b}$$
$$x_i \geqslant \min_j\{w_j\} \quad \text{for all } i \tag{6c}$$
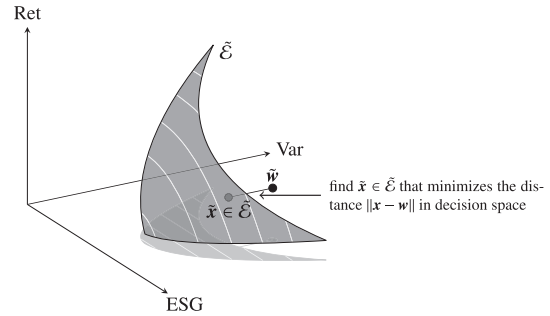$$x_i \leqslant \max_j\{w_j\} \quad \text{for all } i \tag{6d}$$

Fig. 3. A picture illustrating the inverse portfolio optimization process where $\bar{\boldsymbol{w}}$ and $\bar{\boldsymbol{x}}$ are the criterion vectors of $\boldsymbol{w}$ and $\boldsymbol{x}$, and $\boldsymbol{x}$ is the efficient portfolio closest to $\boldsymbol{w}$ in decision space. The shaded area is the tri-criterion nondominated surface $\widetilde{\mathcal{E}}$. The white lines across the surface are iso-quants of the variance objective.

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

$$\min \quad \boldsymbol{x}^T \boldsymbol{\Sigma} \boldsymbol{x}$$
$$\max \quad \boldsymbol{x}^T \boldsymbol{\mu}$$
$$\max \quad \boldsymbol{x}^T \boldsymbol{v}$$
$$\text{s.t.} \quad \sum_{i=1}^n x_i = 1$$
$$x_i \geqslant \min_j\{w_j\} \quad \text{for all } i$$
$$x_i \leqslant \max_j\{w_j\} \quad \text{for all } i$$

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

Notice that according to Eq. (4), the nondominated surface $\widetilde{\mathcal{E}}$ 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 (λ_μ, λ_v) pairs. Another would be to use the ε-constraint method, where 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 ε-constrained method (Hamacher, Pedersen, & Ruzika, 2007), the accuracy of an inverse portfolio optimization still depends on the accuracy of the generated nondominated surface. Yet, in order to obtain a reasonable degree of accuracy for the nondominated surface, it would have to be pixelated to the extent of at least several thousand points thus requiring many quadratic programs to be solved to generate a single nondominated surface. Since we calculate more than 22,000 nondominated surfaces in total (see Section 4), such a direct approach would require the solution of tens of millions of 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 com-

Table 1
Summary ESG statistics for conventional and socially responsible mutual funds.

Year	Compositions	Mean ESG	Median ESG	Min ESG	Max ESG	Std ESG
Panel (C): Conventional mutual funds						
2002	152	0.621	0.626	0.408	0.817	0.060
2003	1119	0.641	0.642	0.402	0.799	0.054
2004	1372	0.666	0.674	0.437	0.854	0.064
2005	1663	0.683	0.693	0.417	0.888	0.077
2006	2522	0.683	0.692	0.416	0.882	0.075
2007	2594	0.696	0.705	0.365	0.874	0.073
2008	4930	0.688	0.709	0.342	0.886	0.099
2009	8003	0.678	0.707	0.311	0.908	0.115
2010	391	0.681	0.710	0.346	0.877	0.111
Panel (S): SR mutual funds						
2002	0					
2003	8	0.611	0.614	0.568	0.664	0.031
2004	20	0.653	0.642	0.584	0.732	0.051
2005	32	0.656	0.676	0.470	0.758	0.083
2006	34	0.668	0.693	0.474	0.796	0.099
2007	33	0.677	0.697	0.528	0.758	0.069
2008	51	0.672	0.683	0.518	0.842	0.073
2009	96	0.674	0.690	0.430	0.822	0.091
2010	2	0.608	0.608	0.524	0.693	0.119

pany. 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 $t + 1$.

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 2346 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 2346 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. 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.

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, Horst, and Zhang, 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 continue 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 λ_v with respect to the ESG-score. For instance, small values of indicate that fund managers weight the ESG-scores only marginally, whereas a high λ_v indicates a high appreciation of 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 λ_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 the literature (see Bauer, Koedijk, & Otten, 2005; Bello, 2005; Guerard, 1997; Hamilton et al., 1993; Kreander, Gray, Power, & Sinclair, 2005; Mallin, Saadouni, & Briston, 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

Table 2
Test hypotheses.

	Null hypothesis	Alternative hypothesis
H1a	$(\lambda_v)_C \geqslant (\lambda_v)_S$	$(\lambda_v)_C < (\lambda_v)_S$
H1b	$(\mathbf{w}^T\mathbf{v})_C \geqslant (\mathbf{w}^T\mathbf{v})_S$	$(\mathbf{w}^T\mathbf{v})_C < (\mathbf{w}^T\mathbf{v})_S$
H2a	$(\lambda_\mu)_C = (\lambda_\mu)_S$	$(\lambda_\mu)_C \neq (\lambda_\mu)_S$
H2b	$(\mathbf{w}^T\boldsymbol{\mu})_C \leqslant (\mathbf{w}^T\boldsymbol{\mu})_S$	$(\mathbf{w}^T\boldsymbol{\mu})_C > (\mathbf{w}^T\boldsymbol{\mu})_S$
H2c	$\left(\sqrt{\mathbf{w}^T\Sigma\mathbf{w}}\right)_C = \left(\sqrt{\mathbf{w}^T\Sigma\mathbf{w}}\right)_S$	$\left(\sqrt{\mathbf{w}^T\Sigma\mathbf{w}}\right)_C \neq \left(\sqrt{\mathbf{w}^T\Sigma\mathbf{w}}\right)_S$
H2d	$(\mathbf{w}^T\mathbf{r}^{OS})_C = (\mathbf{w}^T\mathbf{r}^{OS})_S$	$(\mathbf{w}^T\mathbf{r}^{OS})_C \neq (\mathbf{w}^T\mathbf{r}^{OS})_S$

Table 3
Descriptive statistics for use in hypothesis testing.

Year	Financial				ESG		$\|\mathbf{x} - \mathbf{w}\|$
	$\sqrt{\mathbf{w}^T\sum\mathbf{w}}$	λ_μ	$\mathbf{w}^T\boldsymbol{\mu}$	$\mathbf{w}^T\mathbf{r}^{OS}$	λ_v	$\mathbf{w}^T\mathbf{v}$	
Panel (C): Conventional mutual funds							
2003	0.0482	0.22	0.0145	0.0279	0.0084	0.64	0.15
2004	0.0484	0.16	0.0156	−0.0080	0.0080	0.67	0.15
2005	0.0467	0.16	0.0139	0.0165	0.0088	0.69	0.14
2006	0.0470	0.14	0.0125	0.0151	0.0108	0.69	0.14
2007	0.0462	0.13	0.0110	−0.0154	0.0122	0.70	0.14
2008	0.0478	0.09	0.0089	−0.0033	0.0061	0.71	0.15
2009	0.0608	0.17	0.0082	0.0360	0.0146	0.71	0.14
Panel (S): SR Mutual Funds							
2003	0.0466	0.23	0.0144	0.0276	0.0094	0.61	0.17
2004	0.0452	0.06	0.0157	−0.0061	0.0038	0.64	0.12
2005	0.0447	0.11	0.0124	0.0054	0.0037	0.68	0.11
2006	0.0463	0.35	0.0122	0.0117	0.0107	0.69	0.13
2007	0.0443	0.14	0.0104	−0.0126	0.0093	0.70	0.13
2008	0.0479	0.14	0.0087	−0.0386	0.0074	0.68	0.15
2009	0.0576	0.20	0.0073	0.0436	0.0096	0.69	0.13

mutual funds than in conventional mutual 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 22,477 different fund compositions reported in Table 1.

Continuing with the notation that $\mathbf{x}$ is an efficient portfolio and $\mathbf{w}$ 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 $\mathbf{x}$'s and their $\mathbf{w}$'s in our experiments.

Especially for the implied λ_v and λ_μ, 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 range between 0 and 13,112 (λ_μ), and 0 and 2215 (λ_v), respectively. However, 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 μ and v 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 λ_μ and λ_v 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 $\mathbf{x}$ rather than from the funds' real portfolios $\mathbf{w}$, one could be fearful that the implied parameters do not reflect the funds' actual parameters reasonably well. However, when considering the distances $\|\mathbf{x} - \mathbf{w}\|$ reported in Table 3, we see that the two fund compositions do not differ a lot.

Generally, zero values for λ_μ or λ_v imply infinite risk aversion. By this we mean that no efforts will be made to increase expected return (in the case of $\lambda_\mu = 0$) or ESG-score (in the case of $\lambda_v = 0$) 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_v, we see that most fund managers are willing to take on risk above the minimum.

To determine whether screening is the only approach used to build SR mutual funds, we test to see if the implied tolerances λ_v 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_v of zero. However, conventional mutual funds may exhibit non-zero λ_v'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_v's to exceed the random λ_v'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 have 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 λ_μ, 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 λ_μ'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 and Kritzman (2008) and Dorfleitner and 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

---

[4] We report volatility in opposition to variance in the remainder of this section since volatility is given in the same unit (percent per month) as return.

Table 4
Test statistics and *p*-values for hypotheses H1a–H2d of Table 2.

Corresp. Hyp.		Mann–Whitney U-test					
		λ_v H1a	$\boldsymbol{w}^T\boldsymbol{v}$ H1b	λ_μ H2a	$\boldsymbol{w}^T\boldsymbol{\mu}$ H2b	$\sqrt{\boldsymbol{w}^T\Sigma\boldsymbol{w}}$ H2c	$\boldsymbol{w}^T\boldsymbol{r}^{OS}$ H2d
2003	Test stat.	3959	6302	4240	4421	5165	4596
	p-value	(0.2828)	(0.9757)	(0.7878)	(0.5293)	(0.4620)	(0.4535)
2004	Test stat.	15,963	15,920	17,782	14,380	19,448	12,766
	p-value	(0.8957)	(0.8912)	(0.0228)**	(0.3559)	(0.0013)***	(0.7037)
2005	Test stat.	30,354	30,366	27,142	29,950	32,411	27,359
	p-value	(0.9551)	(0.9555)	(0.6128)	(0.0610)*	(0.0140)**	(0.2789)
2006	Test stat.	45,245	43,022	34,188	44,738	42,450	44,320
	p-value	(0.7091)	(0.5123)	(0.0417)**	(0.3329)	(0.9180)	(0.3692)
2007	Test stat.	44,807	47,262	40,688	40,845	39,783	40,055
	p-value	(0.7832)	(0.9130)	(0.8538)	(0.5586)	(0.6917)	(0.6303)
2008	Test stat.	132,479	147,577	105,374	129,152	118,989	139,180
	p-value	(0.7528)	(0.9847)	(0.0480)**	(0.3606)	(0.5226)	(0.0901)*
2009	Test stat.	410,838	414,671	361,563	404,457	461,939	362,837
	p-value	(0.8786)	(0.9092)	(0.3185)	(0.1874)	(0.0006)***	(0.8263)

* Significant parameter at 10%.
** Significant parameter at 5%.
*** Significant parameter at 1%.

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 funds are at least as high as the standard deviations of SR mutual funds in 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. Furthermore, we checked 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, Gupta, & Markowitz, 2002), we applied different methods for estimating (i.e., exponentially weighted moving average) the financial parameters and we also varied the time period upon which the estimation is based. In summary, we did 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 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.

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.

Acknowledgement

References

Abdelaziz, F. B., Aouni, B., & Fayedh, R. E. (2007). Multi-objective stochastic programming for portfolio selection. *European Journal of Operational Research, 177,* 1811–1823.

Adler, T., & Kritzman, M. (2008). The cost of socially responsible investing. *Journal of Portfolio Management, 35,* 52–56.

Arrow, K. (1965). *Aspects of the theory of risk bearing.* Helsinki: Academic Book Store.

Ballestero, E. (2001). Stochastic goal programming: A mean–variance approach. *European Journal of Operational Research, 131,* 476–481.

Ballestero, E., Bravo, M., Pérez-Gladish, B., Arenas-Parra, M., & Plà-Santamaria, D. (2012). Socially responsible investment: A multicriteria approach to portfolio selection combining ethical and financial objectives. *European Journal of Operational Research, 216,* 487–494.

Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance, 29,* 1761–1767.

Bello, Z. Y. (2005). Socially responsible investing and portfolio diversification. *The Journal of Financial Research, 28,* 41–57.

Bierwag, G. O. (1974). The rationale of the mean-standard deviation analysis: Comment. *American Economic Review, 64*, 431–433.

Bollen, N. P. B. (2007). Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis, 42*, 683–708.

Chamberlain, G. (1983). A characterization of the distributions that imply mean–variance utility functions. *Journal of Economic Theory, 29*, 185–201.

Copeland, T. E., Weston, J. F., & Shastri, K. (2005). *Financial theory and corporate policy*. Boston, Massachusetts: Pearson.

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

Dorfleitner, G., Leidl, M., & Reeder, J. (2012). Theory of social returns in portfolio choice with application to microfinance. *Journal of Asset Management, 13*, 384–400.

Dorfleitner, G., & Utz, S. (2012). Safety first portfolio choice based on financial and sustainability returns. *European Journal of Operational Research, 221*, 155–164.

Fabozzi, F. J., Gupta, F., & Markowitz, H. M. (2002). The legacy of modern portfolio theory. *Journal of Investing, 11*, 7–22.

Feldstein, M. S. (1969). Mean–variance analysis in the theory of liquidity preference and portfolio selection. *Review of Economic Studies, 36*, 5–12.

Guerard, J. B. (1997). Additional evidence on the cost of being socially responsible in investing. *Journal of Investing, 6*, 31–36.

Haimes, Y. Y., Lasdon, L. S., & Wismer, D. A. (1971). On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man and Cybernetics, SMC-1*, 296–297.

Hallerbach, W., Ning, H., Soppe, A., & Spronk, J. (2004). A framework for managing a portfolio of socially responsible investments. *European Journal of Operational Research, 153*, 517–529.

Hamacher, H. W., Pedersen, C. R., & Ruzika, S. (2007). Finding representative systems for discrete bicriterion optimization problems. *Operations Research Letters, 35*, 336–344.

Hamilton, S., Jo, H., & Statman, M. (1993). Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal, 49*, 62–66.

Hirschberger, M., Steuer, R. E., Utz, S., Wimmer, M., & Qi, Y. (2013). Computing the nondominated surface in tri-criterion portfolio selection. *Operations Research, 61*, 169–183.

Kallberg, J. G., & Ziemba, W. T. (1983). Comparison of alternative utility functions in portfolio selection problems. *Management Science, 29*, 1257–1276.

Kreander, N., Gray, R., Power, D., & Sinclair, C. (2005). Evaluating the performance of ethical and non-ethical funds: A matched pair analysis. *Journal of Business Finance and Accounting, 32*, 1465–1493.

Krugman, P., & Wells, R. (2009). *Microeconomics*. New York: Worth Publishers.

Levy, H. (1974). The rationale of the mean-standard deviation analysis: Comment. *American Economic Review, 64*, 434–441.

Mallin, C., Saadouni, B., & Briston, R. (1995). The financial performance of ethical investment funds. *Journal of Business Finance and Accounting, 22*, 483–496.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance, 7*, 77–91.

Markowitz, H. M. (1956). The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly, 3*, 111–133.

Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments*. New York: John Wiley and Sons.

Markowitz, H. M. (1987). *Mean–variance analysis in portfolio choice and capital markets*. Oxford: Basil Blackwell.

Markowitz, H. M., & Todd, P. (2000). *Mean–variance analysis in portfolio choice and capital markets*. New Hope, Pennsylvania: Frank J. Fabozzi Associates.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behaviour* (2nd ed.). Princeton, New Jersey: Princeton University Press.

Pindyck, R., & Rubinfeld, D. (2005). *Microeconomics*. Upper Saddle River, New Jersey: Pearson Prentice Hall.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica, 32*, 122–136.

Renneboog, L., Horst, J. T., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking and Finance, 32*, 1723–1742.

Statman, M. (2000). Socially responsible mutual funds. *Financial Analysts Journal, 56*, 30–38.

Steuer, R. E., Qi, Y., & Hirschberger, M. (2007). Suitable-portfolio investors, nondominated frontier sensitivity, and the effect of multiple objectives on standard portfolio selection. *Annals of Operations Research, 152*, 297–317.

Tobin, J. (1969). Comment on Borch and Feldstein. *Review of Economic Studies, 36*, 13–14.

Tsiang, S. C. (1972). The rationale of the mean-standard deviation analysis, skewness preference, and the demand for money. *American Economic Review, 62*, 354–371.

Xidonas, P., & Mavrotas, G. (2013). Multiobjective portfolio optimization with non-convex policy constraints: Evidence from the Eurostoxx 50. European Journal of Finance (in press).

Xidonas, P., Mavrotas, G., Krintas, T., Psarras, J., & Zopounidis, C. (Eds.). (2012). *Multicriteria Portfolio Management*. New York: Springer.

Zagst, R., & Pöschik, M. (2008). Inverse portfolio optimisation under constraints. *Journal of Asset Management, 9*, 239–253.