Performance Measurement and Best-Practice Benchmarking of Mutual Funds:

Combining Stochastic Dominance criteria with Data Envelopment Analysis

Timo Kuosmanen
Environmental Economics and Natural Resources Group
Wageningen University
Hollandseweg 1
6706 KN Wageningen
The Netherlands
Tel. +31 317 484738
Fax. +31 317 484933
E-mail. Timo.Kuosmanen@wur.nl

Abstract

We propose a method for mutual fund performance measurement and best-practice benchmarking, which endogenously identifies a dominating benchmark portfolio for each evaluated mutual fund. Dominating benchmarks provide information about efficiency improvement potential as well as portfolio strategies for achieving them. Portfolio diversification possibilities are accounts for by using Data Envelopment Analysis (DEA). Portfolio risk is accounted for in terms of the full return distribution by utilizing Stochastic Dominance (SD) criteria. The approach is illustrated by an application to US based environmentally responsible mutual funds.

Key Words: *activity analysis, diversification, fund management, portfolio choice, risk management.*

JEL classification: G11, D81, C61.

1. Introduction

Mutual fund performance measurement and benchmark selection are closely related topics. The traditional measure of mutual fund performance is the Jensen's alpha (Jensen, 1967), derived from a linear beta pricing model such as the Capital Asset Pricing Model (CAPM). The alpha coefficient is the intercept from the regression of the excess returns of the mutual fund on the excess returns of a pre-selected benchmark portfolio. The alpha represents how much the fund manager's forecasting ability contributes to the fund's returns.[1] The modern approach to estimating Jensen's alpha augments the regression function with market capitalization and book-to-market ratio, following the three factor model by Fama and French (1992, 1993).[2] This approach takes into account the size risk and the book-to-market effect, in addition to the market risk captured by the beta. Despite these extensions, the alpha measure remains sensitive to the choice of the benchmark portfolio (see e.g. Roll, 1978, for discussion)

The benchmark portfolio can be defined as a passive version of the fund manager's investment portfolio. The traditional approach of using a risk-adjusted market index as a common benchmark for all funds has been challenged by several authors. For example, Clarke and Ryan (1994) argue that the diversity of investment strategies implies that no single benchmark is suitable for all funds. Rather, the benchmark should reflect the investment objectives, philosophy, and the fund manager's attitude to risk. Brown et al. (1992) suggest that the median benchmark gives fund managers incentives to adopt short term and possibly high risk investment strategies that are not in the long term interest of their trustees. To quote Bailey (1992): "a good benchmark will increase the proficiency of performance evaluation, highlighting the management contribution of the managers." By contrast, poor benchmarks

[1] An parallel interpretation of the alpha intercept of the regression as a managerial performance measure is also well known in the context of production frontier estimation with panel data; see e.g. Schmidt and Sickles (1984).
[2] Fama and French showed empirically that these three factors (beta, market capitalization, and book-to-market ratio) explain about 95% of the variability of stock market returns.

"obscure the managers skills' and promote inefficient managers' allocations and increase the likelihood of unpleasant surprises in total fund performance."

Ansell et al. (2003) point out an interesting difference between the "business" versus "finance" applications of benchmarking. The business definition of benchmarking encompasses identification of industry best-practices as well as "systematic comparison of elements of performance of an organization against those of other organizations, usually with the aim of mutual improvement" (Thor, 1996). By contrast, "finance" interpretation of benchmarking is based on average practices, and is typically geared towards ranking and rating the fund managers. In words of Ansell et al. (2003): "Benchmarking of investment funds is rarely used to select the best product or organization. Instead it is employed to ensure that the product or organization meets a performance standard comparable to the rest of the population. It could be argued that the preferred strategy is to achieve close to the average performance, rather than to outperform the average."

This paper presents a new approach for mutual fund performance measurement that is based on best-practice benchmarking. More specifically, we compare the mutual fund's performance with an endogenously selected benchmark portfolio that optimally tracks the evaluated fund's risk profile. Our main objective is to provide the fund managers and investors with information about efficiency improvement potential and identify portfolio strategies that can achieve such improvements. To this end, we model the investment universe and its portfolio diversification possibilities by means of Data Envelopment Analysis (DEA: Charnes et al., 1978), which is a widely-used performance assessment and benchmarking tool in business, public sector, and non-profit organizations. DEA has already been extensively applied to efficiency analysis of financial institutions such as bank branches; see Berger and Humphrey (1996) for a review.[3]

The main departure from the traditional DEA literature concerns the characterization of the mutual fund's risk profile of a fund by means of stochastic dominance (SD) criteria,[4] building on the results of Kuosmanen (2001, 2004). The SD approach does not restrict to a specific functional form of risk preferences but apply to a general class of monotonic increasing and concave Von Neumann – Morgenstern utility functions. SD does not require any assumptions about the functional form of the return distribution, but applies to any function capable of representing a cumulative probability distribution. These features of the SD are in harmony with the nonparametric orientation of DEA.

The novel contributions of the paper can be summarized as follows:

1) We establish intimate links between the absolute dominance criteria used in DEA and the state-space representations of stochastic dominance criteria derived by Kuosmanen (2001, 2004). We also phrase the portfolio efficiency problem in the standard terminology of DEA literature, to make it more transparent to the readers familiar with DEA.

2) We consider two alternative efficiency metrics: the slack-based Pareto-Koopmans measure by Charnes et al. (1985) and the directional distance function by Chambers et al. (1998). Some key properties of these two measures are discussed and compared.

3) We derive dual expressions for the efficiency measures, and discuss their utility theoretic interpretations.

4) We discuss some practical issues involved in implementation of the SD based DEA approach, including the specification of benchmark units, preprocessing of data, statistical

---

[3] DEA has also been applied to mutual fund performance assessment by Murthi et al. (1997), Morey and Morey (1999), and Basso and Funari (2001, 2003), among others. The present approach differs substantially from these earlier weighting approaches by presenting a systematic efficiency measurement framework with a sound theoretical foundation.

[4] We refer to Bawa et al. (1979) and Levy (1992, 1998) for insightful surveys of the SD literature.

testing of normality hypothesis, specification of weight restrictions, and the interpretation and illustration of the results. To this end, an application to the performance evaluation of eight U.S.-based environmentally responsible mutual funds is presented.

The remainder of the paper is organized as follows. Section 2 introduces the notion of return possibilities set and shows how one can operationalize it by means of DEA. Section 3 introduces the dominance concepts, and re-interprets them in the multi-dimensional state-space where the return possibility set is defined. Section 4 combines these insights and introduces six different additive efficiency measures, derives their dual expressions, and discusses their properties. Section 5 presents an illustrative application to environmentally responsible mutual funds. Section 6 discusses the merits of the proposed measures and points out some challenges for future research.

2. Return possibilities set

Prior DEA approaches to mutual fund performance assessment have presented DEA as a tool for solving a weighting problem. This paper adopts a more rigorous approach inspired by the axiomatic production theory. This section starts by formally characterizing the feasible set of portfolios by introducing the notion of *return possibilities set*.

The general aim of the paper is to assess performance of a single mutual fund, whose returns are observed in S states of nature.[5] The states are randomly drawn without replacement from a pool of possible states, such that each state occurs with the same probability. The returns of the evaluated fund are represented by the S dimensional column vector $\mathbf{r}_0$. The fund can invest in N alternative assets available in the market. Return of asset n in state s is denoted by R_{sn}, and the $S \times N$ matrix of fund returns is denoted by $\mathbf{R}$.

[5] In empirical studies, states of nature are usually interpreted as certain observed time-periods such as years or months, but they could equally well represent some hypothetical condition (e.g. bull vs. bear market).

In terminology of DEA, returns $\mathbf{r}_0$ and $\mathbf{R}$ can be interpreted as input-output variables (or briefly netputs); note that return observations can take positive or negative values.[6] The evaluated mutual fund can be seen as the evaluated Decision Making Unit (DMU 0). The N assets the fund may hold form the set of reference DMUs, from which we construct a benchmark portfolio; note that a mutual fund is simply a convex combinations of assets.

According to the minimum extrapolation principle (Banker et al., 1984), the DEA production possibility set is the smallest set that contains input-output vectors of all observed DMUs, and satisfies the maintained production assumptions. In the similar vein, we can characterize the *return possibilities set* $\mathfrak{R}$ as the minimal set that contains return vectors of all observed assets, and satisfies the following two properties:

I) $\mathfrak{R}$ is negative monotonic ($\mathfrak{R} - \mathbb{R}_+^S = \mathfrak{R}$).

II) $\mathfrak{R}$ is convex.

In DEA terminology, the first property is interpreted as free disposability of inputs and outputs. In the present context, the first property implies that wasting money is possible. The second convexity property also has a compelling interpretation: it means that portfolio diversification is possible. That is, the mutual fund can distribute proportions of its wealth over several assets. Combining properties I) and II), we can characterize $\mathfrak{R}$ as

(1) $$\mathfrak{R} \equiv \left\{ \mathbf{r} \in \mathbb{R}^S \,\middle|\, \mathbf{r} \leq \mathbf{R}\boldsymbol{\lambda}; \boldsymbol{\lambda} \in \Lambda \right\},$$

where $\boldsymbol{\lambda} \in \mathbb{R}^N$ is a vector of portfolio weights, and $\Lambda \subset \mathbb{R}^N$ is their feasible domain. In the DEA literature, variables $\boldsymbol{\lambda}$ are called intensity weights. In DEA, intensity weights are usually constrained to be non-negative. In the portfolio analysis, the negative intensity/portfolio weight can be interpreted as a short position in the corresponding asset.

[6] The model could be enriched by additional features such as transaction costs, minimum investment requirement, or fund manager's experience (compare with Murthi et al., 1997; and Basso and Funari, 2001, 2003). These variables could be modeled as inputs or environmental factors in the spirit of traditional DEA.

Therefore, we may or may not allow for negative weights in the portfolio analysis, depending on the application.

If we do allow for short sales, then it is important to impose some additional constraints on portfolio weights $\boldsymbol{\lambda}$; otherwise the return possibilities set $\mathfrak{P}$ can be unbounded and the efficiency measures (to be introduced below) become infeasible. The additional constraints could be bounds on portfolio weights of a single asset or a group of assets. For example, the rules of most mutual funds dictate very specifically the maximum portfolio weight of any single asset, as well as the maximum proportion of holdings that can be invested in certain group of assets (e.g., domestic stocks, foreign stocks, bonds) as a whole. These types of constraints can be modeled as linear inequalities, which characterize the set $\Lambda \subset \mathbb{R}^N$ as a polyhedral cone. According to Bailey (1992), a good benchmark is *unambiguous* (i.e., the components of the benchmark are clearly defined), *investable* (i.e., managers may forego active management and simply hold the benchmark portfolio), and *appropriate* (the benchmark matches with the manager's investment style and objectives). These criteria are worth keeping in mind in the specification of Λ.

It is worth emphasizing that our approach to construct the benchmark portfolio directly from stocks and other assets forms an important point of departure from earlier DEA approaches. The earlier studies have compared mutual fund performance against the performance of a portfolio of other mutual funds (the "fund of funds"). This limits the domain of the benchmark portfolios to the observed portfolios of the evaluated mutual funds, and the convex combinations thereof. By contrast, our domain of benchmark portfolios covers the entire investment universe of the fund managers. Our approach has at least three important advantages. First, modeling the full investment universe improves the discriminatory power of the model. In terms of the usual DEA setting, our approach would correspond to a possibility of constructing virtual benchmark firms by mixing labor inputs of one firm with capital inputs of another firm.[7] Second, we are not dependent on a large, homogenous sample of reference funds; we may assess efficiency of a single fund relative to the optimal benchmark portfolio constructed from stocks and other assets. Third, we can ensure the benchmark portfolio matches the profile and characteristics of the evaluated fund (e.g., its industrial diversification, market capitalizations, or international spread) by imposing explicit constraints on portfolio weights of stocks.

One could also form benchmarks from groups of stocks instead of individual stocks; this has been the standard approach in the empirical testing of the capital asset pricing model (CAPM) since Black et al. (1972). Groups are usually formed by some prior criteria related to stock returns (e.g., percentiles of the book-to-market distribution), and are periodically rebalanced to account for changes in stocks' risk profile and to avoid discontinuities due to new stock entries and exits. However, grouping of stocks necessarily involves a loss of information, and decreases the discriminating power of the model (see e.g. Berk, 2000, for a critical discussion in the context of CAPM). In the present context, a prior grouping of stocks will fix the relative portfolio weights of assets within each group. Grouping of stocks could be used for the purposes of strategic asset allocation, to determine which types of assets to hold. However, the fund managers usually take the strategy of the fund as given, and are more concerned about the tactical asset allocation. For that purpose, constructing benchmarks directly from individual stocks and bonds seems a preferred option.

3. Absolute and stochastic dominance criteria

Both the DEA and portfolio choice literatures build their efficiency tests and measures on dominance criteria. The purpose of this section is to provide explicit links between the absolute dominance criteria of DEA and the SD criteria utilized in the portfolio theory.

[7] While conventional DEA models always diagnose some DMUs as efficient by default, this is not the case in the present setting.

We start by adapting the usual dominance criterion of DEA (which is closely related to the Pareto-Koopmans notion of technical efficiency) to the present setting as follows:[8]

Definition: absolute dominance (AD). Portfolio $\boldsymbol{\lambda}$ dominates mutual fund 0 in the sense of absolute dominance iff portfolio $\boldsymbol{\lambda}$ yields a return greater than or equal to that of mutual fund 0 in all states, and a strictly greater return in some state, formally:

(2) $\mathbf{R}\boldsymbol{\lambda} \geq \mathbf{r}_0$ and $\mathbf{R}\boldsymbol{\lambda} \neq \mathbf{r}_0$.

AD represents a very strong preference relation: if mutual fund 0 is AD dominated by portfolio $\boldsymbol{\lambda}$, then all non-satiated decision makers, whose preferences can be represented by an increasing state-contingent utility function $u^{sc} \in U^{SC}$, will unanimously prefer portfolio $\boldsymbol{\lambda}$ to mutual fund 0. Obviously, such strong dominance condition is not easily satisfied in practice. This motivates one to consider somewhat weaker dominance requirements.

The SD criteria assume that preferences are *state-independent*: the decision maker's utility only depends on return of the portfolio; occurrence of a specific state does not influence the utility as such. The SD literature has developed a progressive series of alternative dominance criteria which impose further assumptions about risk preferences: risky option A is said to dominate option B by *first-order* SD if every non-satiated decision maker (weakly) prefers option A to B. Option A dominates option B by *second-order* SD if every non-satiated and risk averse decision maker prefers option A to B. Third and higher order SD notions are also known (e.g. Levy, 1992, 1998), but those have proved to be more difficult to operationalize in the portfolio analysis.

The SD criteria can be formally stated in three equivalent ways. First, we may phrase them in terms of probability distribution functions. The distribution function of mutual fund 0

[8] Absolute dominance is particularly prominent in the Free Disposable Hull (FDH) model (e.g. Tulkens, 1993).

is defined as a step-function $F_0(r) \equiv card\{s \in \sigma | r \geq r_{0s}\}/S$, where $\sigma \equiv \{1,2,...,S\}$. Similarly, the distribution function of portfolio $\boldsymbol{\lambda}$ is $F_\lambda(r) \equiv card\left\{s \in \sigma \middle| r \geq \sum_{n=1}^{N} R_{sn}\lambda_n\right\}/T$. Second, we may phrase SD in terms of von Neumann – Morgenstern expected utility functions Let u represent the von Neumann –Morgenstern utility function: a decision maker with $u'(.) \geq 0$ is said to be non-satiated, and decision maker with $u''(.) \leq 0$ is said to be risk-averse. Third, we may express SD using permutation and doubly stochastic matrices. Denote the set of permutation matrices by $\Pi \equiv \left\{\left[P_{ij}\right]_{S\times S} \middle| P_{ij} \in \{0,1\}; \sum_{i=1}^{S} P_{ij} = \sum_{j=1}^{S} P_{ij} = 1 \ \forall i,j \in \sigma\right\}$, and the set of doubly stochastic matrices by $\Xi \equiv \left\{\left[W_{ij}\right]_{S\times S} \middle| 0 \leq W_{ij} \leq 1; \sum_{i=1}^{S} W_{ij} = \sum_{j=1}^{S} W_{ij} = 1 \ \forall i,j \in \sigma\right\}$. The following two theorems state the links between these three perspectives:

FSD Theorem: The following conditions are equivalent:

1) Portfolio $\boldsymbol{\lambda}$ dominates mutual fund 0 by FSD.

2) $F_0(r) - F_\lambda(r) \geq 0 \ \ \forall r \in \mathbb{R}$ with strict inequality for some $r \in \mathbb{R}$.

3) Every non-satiated decision maker prefers portfolio $\boldsymbol{\lambda}$ over mutual fund 0, with a strict preference for at least one such decision maker.

4) There exists a permutation matrix $\mathbf{P} \in \Pi$ such that $\mathbf{R}\boldsymbol{\lambda} \geq \mathbf{P}\mathbf{r}_0$ and $\mathbf{R}\boldsymbol{\lambda} \neq \mathbf{P}\mathbf{r}_0$.

Proof. The equivalence of conditions 1)-3) was first established by Quirk and Saposnik (1962). The equivalence of 1)-3) and 4) follows from Theorem 2 of Kuosmanen (2004), where it was proved that the FSD dominating set of vector $\mathbf{r}_0$ is a positive monotonic hull of all permutations of vector $\mathbf{r}_0$. Condition 4) states that vector $\mathbf{R}\boldsymbol{\lambda}$ belongs to this dominating set. □

SSD Theorem: The following conditions are equivalent:

1) Portfolio $\boldsymbol{\lambda}$ dominates mutual fund 0 by SSD.

2) $\int_{-\infty}^{r} \left[F_0(s) - F_\lambda(s) \right] ds \geq 0 \;\; \forall r \in \mathbb{R}$ with strict inequality for some $r \in \mathbb{R}$.

3) Every non-satiated, risk-averse decision maker prefers portfolio $\boldsymbol{\lambda}$ over mutual fund 0, with a strict preference for at least one such decision maker.

4) There exists a doubly stochastic matrix $\mathbf{W} \in \Xi$ such that $\mathbf{R}\boldsymbol{\lambda} \geq \mathbf{W}\mathbf{r}_0$ and $\mathbf{R}\boldsymbol{\lambda} \neq \mathbf{W}\mathbf{r}_0$.

Proof. The equivalence of conditions 1)-3) was first proved by Fishburn (1964); see also Rothschild and Stiglitz (1970). The equivalence of 1)-3) and 4) follows from Theorem 3 in Kuosmanen (2004). □

In these two theorems, equivalence conditions 1)-3) are well established (see e.g. Levy 1992, 1998). For our purposes, the fourth conditions are the most remarkable ones, as they indicate a direct relationship between the AD, FSD, and SSD notions. For example, the FSD criterion can be seen as the absolute dominance of some permuted return vector. Clearly, AD implies FSD, which in turn implies SSD. It is also easy to see that the converse does not hold (e.g., FSD does not imply AD).

Besides the conceptual link, the fourth conditions are also of considerable practical importance. Note that the return distribution of a portfolio is a rather complicated function of the portfolio weights. It is generally impossible to construct the distribution function of a diversified portfolio from the distribution functions of the individual assets; in particular, $F_\lambda(r) \neq \sum_{n=1}^{N} \lambda_n F_n(r)$ (see Kuosmanen, 2004, Appendix 2 for a numerical example). This is the main reason for the inability of the traditional, distribution function based SD algorithms to deal with diversification. Introduction of the permutation and doubly-stochastic matrices allows us to re-express the FSD and SSD conditions in the S-dimensional state space where the AD condition as well as the return possibilities set have been defined.

4. Efficiency measures

Having established the return possibilities set and the dominance concepts in the same S-dimensional state space, we are equipped to measure efficiency of mutual fund 0. The fund is said to be efficient, in the sense of AD, FSD, or SSD, if the portfolio possibility set $\mathfrak{P}$ does not contain any feasible portfolio that dominates the fund by AD, FSD, or SSD, respectively. Besides testing for efficiency, our purpose is to measure the degree of inefficiency by some suitable efficiency metric. In the DEA literature, many alternative economically meaningful efficiency measures have been considered, including the multiplicative and radial Farrell (1957) measures, the additive slack-based measures, and the directional distance functions. In principle, any of these metrics could be adapted to the present context as well. However, any multiplicative measures will prove difficult to operationalize in the present framework, since multiplication of the permutation matrix or the doubly stochastic matrix by another variable would render the measurement problem highly non-radial. Therefore, in this section we consider two additive measures: 1) the classic Pareto-Koopmans type additive measure introduced by Charnes et al. (1985), and 2) the directional distance function by Chambers et al. (1998).[9]

4.1 Additive Pareto-Koopmans efficiency measures

[9] Briec et al. (2004) introduced the directional distance functions to the financial performance assessment in the classic Mean-Variance framework.

Let us first introduce the additive Pareto-Koopmans (PK) efficiency measure for all three dominance concepts, and then interpret them. The measure can be formally presented in terms of AD, FSD, and SSD as follows:

(3) $$PK^{AD}(\mathbf{r}_0) \equiv \max_{\mathbf{s}\geq\mathbf{0}} \left\{ \mathbf{1}'\mathbf{s} \middle| \mathbf{r}_0 + \mathbf{s} \in \mathcal{P} \right\} = \max_{\mathbf{s},\boldsymbol{\lambda}} \left\{ \mathbf{1}'\mathbf{s} \middle| \mathbf{r}_0 + \mathbf{s} = \mathbf{R}\boldsymbol{\lambda}; \mathbf{s} \geq \mathbf{0}; \boldsymbol{\lambda} \in \Lambda \right\}.$$

(4) $$PK^{FSD}(\mathbf{r}_0) \equiv \max_{\mathbf{s}\geq\mathbf{0},\mathbf{P}\in\Pi} \left\{ \mathbf{1}'\mathbf{s} \middle| \mathbf{P}\mathbf{r}_0 + \mathbf{s} \in \mathcal{P} \right\} = \max_{\mathbf{s},\boldsymbol{\lambda},\mathbf{P}} \left\{ \mathbf{1}'\mathbf{s} \middle| \mathbf{P}\mathbf{r}_0 + \mathbf{s} = \mathbf{R}\boldsymbol{\lambda}; \mathbf{s} \geq \mathbf{0}; \boldsymbol{\lambda} \in \Lambda; \mathbf{P} \in \Pi \right\}.$$

(5) $$PK^{SSD}(\mathbf{r}_0) \equiv \max_{\mathbf{s}\geq\mathbf{0},\mathbf{W}\in\Xi} \left\{ \mathbf{1}'\mathbf{s} \middle| \mathbf{W}\mathbf{r}_0 + \mathbf{s} \in \mathcal{P} \right\} = \max_{\mathbf{s},\boldsymbol{\lambda},\mathbf{W}} \left\{ \mathbf{1}'\mathbf{s} \middle| \mathbf{W}\mathbf{r}_0 + \mathbf{s} = \mathbf{R}\boldsymbol{\lambda}; \mathbf{s} \geq \mathbf{0}; \boldsymbol{\lambda} \in \Lambda; \mathbf{W} \in \Xi \right\}.$$

Vector **s** represents the so-called 'slack variables' that here identify inefficiencies in each state. The weights of the reference portfolio are chosen to maximize the sum of slacks.

The AD measure (3) can be interpreted as a variant of the additive DEA model by Charnes et al. (1985). The FSD and SSD measures were introduced in Kuosmanen (2001, 2004). The only difference between these formulations is that the evaluated return vector $\mathbf{r}_0$ is multiplied by a permutation matrix **P** in (4), and by a doubly stochastic matrix **W** in (5), which do not appear in (3); compare with conditions 4) of the FSD and SSD Theorems. Introduction of these matrices makes the efficiency criterion more stringent in the sense that

(6) $$PK^{SSD}(\mathbf{r}) \geq PK^{FSD}(\mathbf{r}) \geq PK^{AD}(\mathbf{r}) \ \forall \mathbf{r} \in \mathbb{R}^S.$$

In the present context, the additive measure has a compelling economic interpretation: $PK(\mathbf{r}_0)/S$ can be seen as the potential increase in the mean return that can be achieved without worsening the risk exposure of the fund; if $PK(\mathbf{r}_0) > 0$, the reference portfolio $\boldsymbol{\lambda}^*$ dominates the mutual fund 0 in the sense of AD, FSD, or SSD. Thus, $PK(\mathbf{r}_0) = 0$ is a necessary condition for portfolio efficiency. In the cases of AD and FSD, $PK(\mathbf{r}_0) = 0$ is also a sufficient condition. In the SSD case, however, $PK^{SSD}(\mathbf{r}_0) = 0$ does not yet guarantee that no SSD dominating portfolio exists: even if the mean return of the evaluated fund cannot be increased, it may be possible to find a portfolio that yields the same mean return with a lower risk. However, this paper focuses on the measurement of the degree of inefficiency, and thus we abstract from the sufficiency tests (see Kuosmanen, 2004, for further details).

The PK measures (3)-(5) can never take negative values: slacks **s** are restricted to the non-negative orthant to guarantee that the reference portfolio $\boldsymbol{\lambda}^*$ dominates the evaluated mutual fund. If the mutual fund 0 is not contained in the return possibilities set $\mathfrak{P}$, problems (3)-(5) will turn out as infeasible. Note that the benchmark portfolio represents a "passive" strategy (i.e., the portfolio weights are fixed), while the fund manager be "active" in adjusting portfolio weights. Thus, an active manager may be able to beat any passive benchmark portfolio. The simplest way to avoid the infeasibility problem is to insert the evaluated fund as a new column in the return matrix **R**.

The PK measures (3)-(5) are monotonic in the sense that for funds *A* and *B*: $\mathbf{r}_A \geq \mathbf{r}_B \Rightarrow PK(\mathbf{r}_A) \leq PK(\mathbf{r}_B) = 0$. Monotonicity is important for the logical consistency of the efficiency measures. If $PK(\mathbf{r}_A) > PK(\mathbf{r}_B)$, then monotonicity implies that fund *A* cannot dominate fund *B*. This is a particularly attractive property if one ranks funds based on their efficiency. However, condition $PK(\mathbf{r}_A) > PK(\mathbf{r}_B)$ does not imply that fund *B* dominates *A*. Thus, some investors with well behaved preferences will prefer fund *A* over fund *B*, even if *B* is rated as more efficient. In this sense, ranking of funds based on efficiency measures can be misleading, especially if funds differ in terms of their investment strategy and risk profile.

In terminology of Kuosmanen (2004), the set of return vectors that dominate $\mathbf{r}_0$ (by AD, FSD, or SSD) is called the (AD, FSD, SSD) dominating set of $\mathbf{r}_0$. It is easy to verify that the AD dominating set is always convex. Also the SSD dominating sets are convex, as shown by Kuosmanen (2004). Convexity implies that the AD and SSD measures (3) and (5) are continuous functions of return vectors. Continuity is an attractive property for the stability of efficiency indices. Unfortunately, the FSD measure (4) is not continuous; an infinitesimal change in the return vector $\mathbf{r}_0$ can cause a significant jump in the efficiency index.[10] The reason for the possible instability of the FSD efficiency measure lies in the non-convex nature of the FSD dominating set (see Kuosmanen, 2004). In traditional DEA settings, similar discontinuities will arise if one uses the additive PK efficiency measure together with the non-convex free disposable hull (FDH) reference technology.

The additive efficiency measures are not unit invariant. While this is a rather unpleasant feature in the usual DEA where inputs and outputs are usually measured in incommensurable units, it is not a problem in the present setting where the returns (netputs) are measured in the same units of measurement (usually percents). Thus, the efficiency measure is also expressed in the same units.

When the domain of portfolio weights Λ takes a form of a convex polyhedron, the AD and the SSD problems (3) and (5) can be expressed as linear programming problems. Thus, they have equivalent dual formulations. For simplicity, we here focus on the case where the set of portfolio weights takes the form of a simplex $\Lambda = \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^N \middle| \mathbf{1}'\boldsymbol{\lambda} = 1 \right\}$. In this case, the dual problem of (3) can be stated as

(7) $$PK^{AD}(\mathbf{r}_0) = \min_{\phi,\mathbf{v}} \left\{ \phi \middle| \phi\mathbf{1} \geq \mathbf{v}'(\mathbf{R} - \mathbf{r}_0\mathbf{1}'); \mathbf{v} \geq \mathbf{1} \right\},$$

where the S-dimensional vector $\mathbf{v}$ represents a tangent hyperplane to the return possibilities set $\mathcal{P}$, analogous to the DEA multiplier weights of inputs and outputs. In the present context, weights $\mathbf{v}$ can be interpreted as average state-contingent utilities, normalized such that $\mathbf{v} \geq \mathbf{1}$.

---

[10] The discontinuity of the FSD measure can be illustrated by the following numerical example. Let N=3, S=2, $\mathbf{R} = \begin{pmatrix} 1 & 1.5 & 3 \\ 4 & 1 & 2.5 \end{pmatrix}$, and $\Lambda = \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^N \middle| \mathbf{1}'\boldsymbol{\lambda} = 1 \right\}$. Consider mutual fund $\mathbf{r}_0 = \begin{pmatrix} 2 \\ 3+\varepsilon \end{pmatrix}$. Starting with the value $\varepsilon = 0$, we obtain the efficiency score $PK^{FSD}(\mathbf{r}_0) = 0.5$ with $\boldsymbol{\lambda}^* = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}'$. However, an infinitesimal increase in the value of ε will change the efficiency score discontinuously to $PK^{FSD}(\mathbf{r}_0) = \frac{1}{3} - \frac{4}{3}\varepsilon$ with $\boldsymbol{\lambda}^* = \begin{pmatrix} \frac{1+2\varepsilon}{3} & 0 & \frac{2-2\varepsilon}{3} \end{pmatrix}'$. [The author thanks an anonymous referee for this example.]

If $PK^{AD}(\mathbf{r}_0) = 0$, there exists a non-satiated decision maker with an increasing state-contingent utility function who prefers mutual fund 0 over any alternative feasible portfolio. If $PK^{AD}(\mathbf{r}_0) > 0$, such decision maker does not exist. The objective function ϕ represents the difference in normalized, state-contingent utility levels between the most preferred reference asset (i.e., the maximum element of vector $\mathbf{v}'\mathbf{R}$) and the mutual fund 0 (i.e., $\mathbf{v}'\mathbf{r}_0$).

In the FSD case, we cannot directly derive a dual problem for (4) which involves binary integer variables in its permutation matrix **P**. Still, it is possible to exploit the embedded linear structure of problem (4) to recover a utility function for a representative decision maker who (most) prefers the evaluated portfolio. Note first that the FSD efficiency measure can found by solving the AD efficiency measure for all possible permutations of vector $\mathbf{r}_0$, and taking the maximum of thus obtained solutions, i.e., $PK^{FSD}(\mathbf{r}_0) = \max_{\mathbf{P}\in\Pi} PK^{AD}(\mathbf{P}\mathbf{r}_0)$.[11] Thus, we may first identify the optimal permutation matrix $\mathbf{P}^*$ that maximizes problem (4), permute the returns of fund 0 using $\mathbf{P}^*$, and subsequently solve the dual AD problem (7) for the permuted values $\mathbf{P}^*\mathbf{r}_0$. This provides us with vector **v**. If elements of **v** are interpreted as average utilities $v_s = u(r_{0s})/r_{0s}$, then the expected utility of mutual fund 0 can be expressed as $\mathbf{v}'\mathbf{r}_0 / S$. Note that by introducing the permutations of $\mathbf{r}_0$, we essentially require that the preferences must be state-independent: each permutation of $\mathbf{r}_0$ is equally preferred. Utilizing vector **v**, it is possible to derive bounds for the class of utility functions that can rationalize an efficient portfolio choice, following Varian (1983).

The SSD problem (5) amends itself to the linear programming format, so it is possible to derive the dual problem analytically. The dual problem can be expressed as follows.

[11] In fact, by enumerating all possible permutation matrices, we could express the FSD efficiency measure (4) as an enormously large linear programming problem. However, the "brute-force" strategy of considering all possible permutations becomes highly expensive for almost any non-trivial number of states (e.g., if S=100, the number of permutations is $100! = 100\cdot 99\cdot \ldots \cdot 2\cdot 1 \approx 9.33\cdot 10^{157}$). It is therefore advisable to use modern integer programming algorithms (such as branch-and-bound) for solving problem (4).

(8) $$PK^{SSD}(\mathbf{r}_0) = \min_{\beta,\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{v}} \left\{ \beta - (\mathbf{1}'\boldsymbol{\theta} + \mathbf{1}'\boldsymbol{\tau}) \middle| \beta\mathbf{1} \geq \mathbf{v}'\mathbf{R}; v_s r_{0t} \geq \theta_t + \tau_s \ \forall t, s \in \sigma; \mathbf{v} \geq \mathbf{1} \right\}.$$

The interpretation of vector $\mathbf{v}$ as average utilities is the same as in (7). Difference $\beta - (\mathbf{1}'\boldsymbol{\theta} + \mathbf{1}'\boldsymbol{\tau})$ in the objective function of (8) is analogous to variable ϕ in (7): here β represents the expected utility of the most preferred reference asset (i.e., the largest element of vector $\mathbf{v}'\mathbf{R}$), while sum of variables θ, τ represents the expected utility of mutual fund 0. Variables θ, τ represent the dual variables of the doubly stochastic matrix $\mathbf{W}$ in (5), and their role is to facilitate transfer of utility from one state to another, and thus guarantee the state-independence property of the expected utility theory.

4.2 Directional distance function

We next turn to the directional distance function (DD) measures. These efficiency indices require a prior specification of the direction vector, denoted by *S*-dimensional column vector $\mathbf{g} : \mathbf{g} \geq \mathbf{0}, \mathbf{g} \neq \mathbf{0}$. Again, we first present the measures and then interpret them.

(9) $$DD^{AD}(\mathbf{r}_0) \equiv \max_{\delta} \left\{ \delta \middle| \mathbf{r}_0 + \delta\mathbf{g} \in \mathcal{P} \right\} = \max_{\delta,\boldsymbol{\lambda}} \left\{ \delta \middle| \mathbf{r}_0 + \delta\mathbf{g} = \mathbf{R}\boldsymbol{\lambda}; \boldsymbol{\lambda} \in \Lambda \right\}.$$

(10) $$DD^{FSD}(\mathbf{r}_0) \equiv \max_{\delta,\mathbf{P}} \left\{ \delta \middle| \mathbf{P}\mathbf{r}_0 + \delta\mathbf{g} \in \mathcal{P}; \mathbf{P} \in \Pi \right\} = \max_{\delta,\boldsymbol{\lambda},\mathbf{P}} \left\{ \delta \middle| \mathbf{P}\mathbf{r}_0 + \delta\mathbf{g} = \mathbf{R}\boldsymbol{\lambda}; \boldsymbol{\lambda} \in \Lambda; \mathbf{P} \in \Pi \right\}.$$

(11) $$DD^{SSD}(\mathbf{r}_0) \equiv \max_{\delta,\mathbf{W}\in\Xi} \left\{ \delta \middle| \mathbf{W}\mathbf{r}_0 + \delta\mathbf{g} \in \mathcal{P} \right\} = \max_{\delta,\boldsymbol{\lambda},\mathbf{W}} \left\{ \delta \middle| \mathbf{W}\mathbf{r}_0 + \delta\mathbf{g} = \mathbf{R}\boldsymbol{\lambda}; \boldsymbol{\lambda} \in \Lambda; \mathbf{W} \in \Xi \right\}.$$

Analogous to the inequalities (6), these three efficiency measures satisfy

(12) $$DD^{SSD}(\mathbf{r}) \geq DD^{FSD}(\mathbf{r}) \geq DD^{AD}(\mathbf{r}) \ \forall \mathbf{r} \in \mathbb{R}^S.$$

The interpretation of this efficiency measure depends on the specification of the direction vector, which also determines the units of measurement. Since all states are assumed to be equally likely, a natural specification is to set $\mathbf{g} = \mathbf{1}$. As noted by Kuosmanen (2004, p. 1395), this specification allows us to interpret $DD(\mathbf{r}_0)$ as the minimum risk free premium that must be added to $\mathbf{r}_0$ to make the fund efficient (in the sense of AD, FSD, or SSD).

Since we do not require a priori that the mutual fund 0 (or any of its permutations) is contained in the return possibilities set $\mathfrak{P}$, the DD measure may yield negative values. This is analogous to the super-efficiency DEA model (Andersen and Petersen, 1993). This super-efficiency feature is important because an active manager may beat any passive benchmark portfolio, as noted above. This feature also enables one to assess efficiency of hypothetical, non-existing portfolios, which might include, for example, derivative instruments or new securities entering the market. On the other hand, these DD measures fall subject to similar infeasibility problems that frequently arise in the DEA.

If $DD(\mathbf{r}_0)>0$, the reference portfolio the reference portfolio $\boldsymbol{\lambda}^*$ dominates the mutual fund 0 in the sense of AD, FSD, or SSD. Thus, $DD(\mathbf{r}_0) \le 0$ is a necessary condition for portfolio efficiency for all dominance criteria. If $DD(\mathbf{r}_0) < 0$, the evaluated fund lies outside the boundaries of the return possibilities set $\mathfrak{P}$, and is classified as super-efficient. Clearly, super-efficient funds cannot be dominated by any feasible portfolio, so $DD(\mathbf{r}_0) < 0$ is a sufficient efficiency condition for all dominance criteria considered. In the limiting case of $DD(\mathbf{r}_0) = 0$, however, we cannot determine the efficiency status based on the optimal solutions to (9)-(11). In the AD and FSD cases, return vector $\mathbf{r}_0$ can be located in the "weakly efficient" boundary of the return possibilities set $\mathfrak{P}$ where it is possible to increase return in at least one state, but not simultaneously in all states. In the SSD case, the directional distance measure is incapable of detecting a dominating portfolio that yields an equally high return as mutual fund 0 with a lower risk (compare with Kuosmanen, 2004, Appendix 2, Ex. 3).

The directional distance functions are monotonic in the sense that $\mathbf{r}_A \ge \mathbf{r}_B \Rightarrow DD(\mathbf{r}_A) \le DD(\mathbf{r}_B) = 0$.[12] In the AD and SSD cases, the DD measures are

[12] Properties of the directional distance function in the AD case have been examined in more detail by Chambers et al. (1998).

continuous functions of return vectors, but in the FSD case, the DD measure can change in a discontinuous manner when return vector $\mathbf{r}_0$ is perturbed. As in the case of the PK measure, the instability of this DD measure is caused by non-convexity of the FSD dominating set.

When the set of portfolio weights Λ takes a form of a simplex $\Lambda = \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^N \middle| \mathbf{1}'\boldsymbol{\lambda} = 1 \right\}$, we can derive the following dual problem for the AD distance function (9):

$$(13) \qquad DD^{AD}(\mathbf{r}_0) = \min_{\delta, \mathbf{v}} \left\{ \delta \middle| \delta \mathbf{1} \geq \mathbf{v}'(\mathbf{R} - \mathbf{r}_0 \mathbf{1}'); \mathbf{v}'\mathbf{g} = 1 \right\}.$$

The interpretation of this dual problem is directly analogous to that of problem (7). The only difference between problems (7) and (13) concerns the normalization of the average utilities $\mathbf{v}$: problem (7) imposes a constraint $\mathbf{v} \geq \mathbf{1}$, whereas problem (13) uses a normalization $\mathbf{v}'\mathbf{g} = 1$.

In the FSD case, we may exploit the property $DD^{FSD}(\mathbf{r}_0) = \max_{\mathbf{P} \in \Pi} DD^{AD}(\mathbf{P}\mathbf{r}_0)$. Thus, we can first identify the optimal permutation matrix $\mathbf{P}^*$ as the optimal solution to (10), and subsequently derive average utilities $\mathbf{v}$ using (13).

In the SSD case, the dual problem of (11) can be stated as

$$(14) \qquad DD^{SSD}(\mathbf{r}_0) = \min_{\beta, \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{v}} \left\{ \beta - (\mathbf{1}'\boldsymbol{\theta} + \mathbf{1}'\boldsymbol{\tau}) \middle| \mathbf{v}'\mathbf{g} = 1; \beta \mathbf{1} \geq \mathbf{v}'\mathbf{R}; v_s r_{0t} \geq \theta_t + \tau_s \ \forall t, s \in \sigma; \mathbf{v} \geq \mathbf{1} \right\}.$$

The interpretation of this dual is directly analogous to that of problem (8). The only difference between dual problems (8) and (14) lies in the normalization of the average utilities $\mathbf{v}$.

4.3 Comparison

We have discussed six different portfolio efficiency measures pertaining to three different dominance concepts and two alternative efficiency metrics. All these measures can be expressed in additive form. We found that all these efficiency measures have a meaningful economic interpretation, and they all are monotonic decreasing functions of the return vector of the evaluated mutual fund. Both the PK and DD measures are continuous functions of the evaluated return vector in the AD and SSD cases, and both efficiency metrics fail the continuity property in the FSD case. The FSD measures also lack explicit dual formulations, but their embedded linear structure can be exploited to derive preference information in the form of average utilities. The additive PK^{AD} and PK^{FSD} measures offer a clear-cut necessary and sufficient efficiency condition. By contrast, PK^{SSD} and the three DD measures may fail to identify a dominating portfolio, and thus may fail to classify a mutual fund as efficient or inefficient. One advantage of the DD measures is their ability to handle mutual funds that are not contained in the return possibility set, in the spirit of the super-efficiency DEA model.

The dual formulations were derived in the case where the domain of portfolio weights takes the form of a simplex, i.e., $\Lambda = \left\{ \boldsymbol{\lambda} \in \mathbb{R}_{+}^{N} \middle| \mathbf{1}'\boldsymbol{\lambda} = 1 \right\}$. These dual formulations provide additional insight, but need to be appropriately adjusted if the domain Λ changes. While the primal problems can be directly enriched with additional constraints on portfolio weights $\boldsymbol{\lambda}$, in the dual problems we would need to adjust the set of benchmark assets, which proves a more complicated task. This is the key advantage of the primal formulations.

Performance measures are frequently used for ranking funds. Attractively, all six efficiency measures are monotonic increasing, which is important for logical consistency of efficiency rankings. However, monotonicity does not guarantee that all investors agree with the ranking. Moreover, the efficiency ranking can be sensitive to the choice of the efficiency metric (or the specification of the direction vector $\mathbf{g}$ of the directional distance function). Furthermore, we noted above that the FSD efficiency measures exhibit discontinuities, which can make the FSD efficiency ranking rather unstable. On the other hand, we are not aware of any superior method for performance ranking of mutual fund. Apart from the discontinuity of the FSD measures, the usual DEA rankings are subject to similar difficulties.

We conclude this section by briefly comparing the proposed approach with the linear programming SD efficiency tests proposed by Post (2003). Post's approach is geared towards testing the SSD efficiency of the market portfolio, not measuring the degree of inefficiency.

Indeed, Post himself emphasized that "*...the test statistics do not represent meaningful performance measures that can be used for ranking portfolios based on the "degree of efficiency"."* (Post, 2003, p.1912-1913). The main problem in applying Post's method for performance measurement lies in the fact that Post test procedure does not identify a dominating benchmark portfolio. Even though his dual test statistic (the ordered mean difference measure) does identify a reference portfolio (denoted by λ), this reference portfolio does not necessarily dominate the evaluated portfolio by SSD.[13] This feature is problematic for the benchmarking purposes, but it also makes it difficult to build meaningful performance measures on Post's approach. Another limitation of Post's framework is that it does not apply to the FSD case. Although Post (2003, p. 1926) suggested to test for FSD efficiency by relaxing the concavity restriction of his SSD test, he failed to account for the permutation possibilities which would guarantee the state independence. As a result, Post's relaxed test does not capture FSD efficiency, but rather boils down to the AD efficiency. This can be verified by comparing problem (7) with the FSD test suggested by Post (2003, eq. 19).

5. Application to Environmentally Responsible Mutual Funds

The purpose of the following application is to illustrate the practical application of the proposed approach. We draw attention to some critical issues in data collection and processing, propose some prior diagnostic tests that may be useful for validating the approach, and suggest ways for interpreting and illustrating the results.

---

[13] For example, let N=2, S=2, $\mathbf{R} = \begin{pmatrix} 2 & 7 \\ 5 & 1 \end{pmatrix}$, and $\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}_+^N | \mathbf{1}'\boldsymbol{\lambda} = 1\}$. Consider mutual fund $\mathbf{r}_0 = (2 \quad 5)'$. Post's dual measure (12) will identify vector $(7 \quad 1)'$ as the reference portfolio, i.e., $\boldsymbol{\lambda}^* = \arg\max \psi(\tau) = (0 \quad 1)'$; note that 7>2, and 7+1>2+5. However, it is easy to verify that vector $(7 \quad 1)'$ does not dominate the evaluated mutual fund by SSD since 2>1.

5.1 Background

Socially responsive investing (SRI) has emerged in recent years as a dynamic, quickly growing segment of financial services industry. In the United States, professionally managed SRI holdings amounted to $2.34 trillion in 2001 (see e.g. Schueth, 2003, for a review of SRI in the U.S.). The next application presents a new approach to measuring performance of a sample of SRI funds with emphasis on the environment. Instead of focusing on mean return or some narrowly defined indicators of risk, we take the shape of the entire return distribution into account by using the SD enhanced DEA tools discussed in the previous sections. In contrast to Basso and Funari (2003), we do not include an ethical index as an evaluation criterion. Rather, we assess performance of SRI funds purely on financial grounds (portfolio efficiency), to see whether any ”ethical premium” can be identified in the empirical data at all.

5.2 Data and benchmarks

A sample of US-based SRI funds was selected from the homepage of SRI World Group, Inc.; a financial information services and consulting company (see www.ishareowner.com). As the selection criteria, we required that

1) The fund applies a *positive screen* to environmental criteria (i.e., the fund seeks companies with a positive record or achievement in terms of the environment).
2) The fund applies an *exclusionary screen* to environmental criteria (the fund avoids companies with a poor record or achievement in terms of the environment).
3) The shares of the fund have been traded in the NYSE since January 1, 2000 or earlier.
4) The fund is characterized as a large-blend equity fund following growth strategy.

These criteria resulted with a sample of eight mutual funds with similar enough characteristics to allow for comparison.

To specify the feasible domain for benchmark portfolios, we considered the following criteria suggested by Bailey (1992). According to Bailey, a good benchmark is

1) *unambiguous* (i.e., the components of the benchmark are clearly defined),

2) *investable* (i.e., it should be open to managers to forego active management and simply hold the benchmark portfolio),

3) *measurable* (i.e., benchmark's return can be calculated on frequent basis) ,

4) *appropriate* (the benchmark matches with the manager's investment style and objectives),

5) *reflective of current investment opinions*, and

6) *specified in advance*.

Our approach satisfies requirements 1) and 3) by default, since portfolio weights $\boldsymbol{\lambda}$ and return matrix $\mathbf{R}$ are explicitly represented in the framework. On the other hand, any ex post assessment with endogenously selected benchmarks will fail condition 6). Conditions 2), 4), and 5) call for appropriate specification of domain Λ for portfolio weights of the benchmark.

The investment universe of the fund managers mainly consists of stocks traded in the New York Stock Exchange (NYSE). However, not all stocks are eligible to the benchmark portfolio. Each fund has a unique investment strategy described (in qualitative terms) in the prospectus of the fund. The US based growth funds generally favor high capitalized, US-based firms included in the Standard and Poor's S&P500 index. Most funds allow only a limited investment in foreign companies, or small emerging firms involving high risk. To make the benchmark portfolios as realistic as possible from the point of view of SRI funds, we limit the investment universe to stocks included in the Dow Jones STOXX Sustainability Index (henceforth DJSI),[14] which were traded in the NYSE throughout the entire study period from 26 November 2001 to 26 November 2002. As a low risk investment alternative we

[14] DJSI is one of the most standard benchmarks for the green mutual funds. The components of this index represent the top 20 per cent of the leading sustainability companies in each industry group within the DJSI STOXX investable universe. For further details, see http://www.sustainability-index.com/.

included the standard one-month treasury bill. This gives the total number of 175 assets: 49 stocks of U.S.-based companies, 125 stocks of international companies, and a single bond.

We further imposed the following constraints on the portfolio weights of the benchmark:

1) Portfolio weight of any individual asset is less than 5.8 per cent of the total holdings of the fund: $\lambda_i \leq 0.058 \ \forall i$.

2) Not more than 5.8 per cent of the total holdings can be invested in bonds: $\sum_{i \in \{bonds\}} \lambda_i \leq 0.058 \ \forall i$

3) At least 65 per cent of the total holdings should be invested in Standard and Poor's rated U.S. stocks: $\sum_{i \in \{\text{domestic stocks}\}} \lambda_i \geq 0.65 \ \forall i$

4) Short sales are not allowed: $\lambda_i \geq 0 \ \forall i; \sum_{i=1}^{N} \lambda_i = 1$.

These constraints do not take into account all explicit and implicit bounds faced by every fund manager. Still, constraints 1) – 4) should suffice to guarantee that the benchmarks will be well diversified and reflect the overall investment strategy of these mutual funds by giving emphasis on the stocks of large U.S. firms. These weight bounds can be directly incorporated in the primal formulations, so we therefore focus exclusively on the primal problems in this section.

Weekly data of closing prices (adjusted to splits and dividends) of all stocks and mutual funds for the one-year evaluation period November 26, 2001 – November 26, 2002 were obtained from the website http://finance.yahoo.com. For simplicity, each week in this evaluation period represents one state of nature. Denoting the weekly closing prices of asset n in weeks t and t-1 by $P_{t,n}$ and $P_{t-1,n}$, respectively, the annual rate of return for asset n in week t was calculated as $R_{t,n} = \frac{P_{t,n}}{P_{t-1,n}} \cdot 52$.

Table 1 presents some descriptive statistics regarding the samples of mutual fund and stock returns. During the evaluation period, the stock values were in decline as seen from the negative signs of the mean returns. The stocks provided a somewhat higher mean return, as expected. The stock returns were also more volatile, as can be seen from the standard deviation, and the wide spread between the minimum and maximum values.

Table 1: Descriptive statistics of mutual fund and stock returns

	Mean	St.dev.	Min	Max
Mutual funds	-0.179	1.449	-6.494	3.289
Stocks	-0.138	3.154	-33.481	35.360

5.3 Normality test

Before proceeding to the efficiency analysis, we tested for normality of the return distributions. If returns are normally distributed, the SSD criterion reduces to the classic Mean-Variance rule. As we only have the sample estimates for the mean and variance, the standard test procedure is to use the Shapiro-Wilks test. We ran this test for all eight mutual funds and the subset of 49 S&P500 stocks in the DJSI index. Table 2 summarizes the results.

Table 2: Shapiro-Wilks normality test for return distributions: the frequency of rejections at different significance levels

Significance level	1%	5%	10%
SRI funds (N=8)	0	1	1
Stocks (N=49)	13	17	22

Return distributions of the eight mutual funds could be relatively well described by normal distributions. Only one fund showed statistically significant violation of normality. By contrast, almost half of the S&P500 stocks in the DJSI violated normality at significance level of ten percent or lower. This gives a motivation for using the SD approach that does not depend on the normality assumption. By visual inspection of empirical distribution functions, the patterns that lead to rejecting normality were rich in number. In many cases, the high kurtosis of the distribution was at odds with normality. Also many stocks had fat tails, either left or right, but a few stocks had thinner tails than normality would imply. For some stocks, the return distributions appeared almost uniform, except for their extreme tails.

5.4. Results

We calculated the PK and the DD measures (with direction vector **g = 1**) for all eight mutual funds using AD, FSD, and SSD dominance criteria. Table 3 summarizes the results. The PK measures are reported in normalized form as *PK/S* to facilitate the interpretation of efficiency score as potential increase in mean return without worsening the risk profile.

Table 3: Summary of results; the eight SRI funds (% points p.a.)

Mutual fund	*PK* / *S* measure			*DD* (**g** = **1**) measure		
(ticker symbol)	AD	FSD	SSD	AD	FSD	SSD
CSIEX	0.000	0.350	0.350	0.000	0.349	0.349
CSECX	0.000	0.358	0.358	0.000	0.357	0.357
DESRX	0.237	0.433	0.433	0.014	0.432	0.432
DSEFX	0.000	0.507	0.507	0.000	0.506	0.506
GCEQX	0.000	0.485	0.485	0.000	0.484	0.484
NBSRX	0.000	0.432	0.432	0.000	0.432	0.432
FEMMX	0.000	0.364	0.364	0.000	0.363	0.363
ADVOX	0.368	0.454	0.454	0.102	0.454	0.454
Average	0.076	0.423	0.423	0.015	0.422	0.422
St. Dev.	0.144	0.060	0.060	0.036	0.060	0.060

Only two funds were diagnosed as inefficient in terms of the AD criterion. The low discriminatory power of the AD criterion is explained by the curse of dimensionality; the AD case corresponds to a DEA model with 51 outputs and only 175 reference DMUs. It is quite remarkable that we could construct passive benchmark portfolios that beat these two funds in every week during the study period. The fact that all eight funds were found inefficient in the SD cases illustrates the power of the state independence property. Note that the two AD inefficient funds were not the least efficient funds according to the FSD and SSD criteria.

The PK and DD measures yield very similar results in the FSD and SSD cases; except for the last fund, the DD measures were slightly smaller than the PK measures. Interestingly, the FSD and SSD criteria gave almost identical results (with accuracy of three decimal points) for all eight funds in terms of both PK and DD measures.[15] The equality of the FSD and SSD measures was confirmed by the doubly stochastic matrices **W***, obtained as the optimal solution to problems (5) and (11): almost all elements of these matrices were binary integers. A few non-integer solutions occurred when the corresponding elements in the vector $\mathbf{r}_0$ had equal values.

The economic interpretation of the equality of the FSD and SSD measures is that the assumption of risk aversion (of SSD) does not play a role in these measurements. In general, risk aversion becomes critical when the evaluated fund has volatile returns, and thus the fund appeals to risk-seeking investors but not to risk-averse ones. However, mutual funds (as any well diversified portfolios) usually provide relatively low mean return combined with low risk. These characteristics make the mutual funds attractive investment alternatives only for risk-averse investors; hence imposing risk-aversion by assumption does not change the result.[16] This argument is further elaborated and illustrated in the Appendix.

Besides the efficiency scores, it is interesting to examine the composition of the benchmark portfolios. Table 4 reports the averages and standard deviations of the portfolio weights of the stocks included in the benchmark portfolio in the FSD and SSD cases (including both PK and DD measures). Total of twenty stocks were assigned a positive weight in one of the sixteen benchmark portfolios that exhibited reasonable spread over different industries and well-known US and international companies. The 5.8 percent upper bound for the portfolio weights turned out as a binding constraint for most stocks in the benchmark: fifteen stocks appeared with the full 5.8 percent weight in all benchmarks, accounting for 87 percent of the total portfolio. The remaining thirteen percent was composed of five different

[15] Similarly, Kuosmanen (2004) found no difference between the FSD and SSD efficiency measures calculated for the marker portfolio.
[16] A parallel example from the usual DEA settings: the assumption of constant returns to scale does not influence the efficiency measures when the evaluated firm operates on the most productive scale size.

stocks with varying weights. The reason for the similarity of the benchmarks is that all eight mutual funds had very similar return distributions and risk profiles (see Figure 1 below).

Table 4: Portfolio weights of the benchmark portfolios in the FSD and SSD cases

Company name	Industry	average	st.dev
Alcoa Inc.	Aluminum	0.058	0
Amcor Ltd	Packaging & Containers	0.058	0
Dell Inc.	Personal Computers	0.058	0
Eastman Kodak	Photographic Equipment & Supplies	0.058	0
ENSCO International Inc	Oil & Gas Drilling & Exploration	0.058	0
Entergy Corp.	Electric Utilities	0.058	0
Johnson & Johnson	Drug Manufacturing	0.058	0
Mattel, Inc.	Toys & Games	0.058	0
Mentor Corp.	Medical Appliances & Equipment	0.058	0
Royal Caribbean Cruises Ltd.	Entertainment	0.058	0
Safeway Inc.	Food retail	0.058	0
SKF	Industrial Goods & Services	0.058	0
Boeing	Aerospace & Defense	0.058	0
United Health Group Inc	Health Care Plans	0.058	0
Visteon Corp.	Auto Parts	0.058	0
Canon Inc.	Photographic Equipment & Supplies	0.036	0.016
Intel Corp.	Semiconductors	0.034	0.016
ANZ Banking Group Ltd	Banking	0.033	0.023
Cognos Inc	Application Software	0.031	0.022
BHP Billiton Ltd	Industrial Metals & Minerals	0.001	0.001

Obviously, the specification of the weight restrictions directly influences the weights and efficiency scores, and thus limits the freedom of the data to speak for themselves. Without the weight restrictions, however, the benchmark portfolio would be composed of relatively few stocks, and would thus fail to represent acceptable passive diversification strategy for the mutual funds. Moreover, it is not obvious from the outset which stocks will be eventually selected to the benchmark portfolios.

To illustrate the efficiency improvement potential, we may plot the cumulative distributions of the returns. Figure 1 shows the cumulative return distributions of the eight evaluated mutual funds, represented by the black and grey step functions. We see that the return distributions are almost indistinguishable; all eight return distributions fall within 0.5 percent point interval in almost all percentiles. Only in the left tail of the distribution there is more variation across funds. Return distributions are approximately normal; recall that only one fund violated the Shapiro-Wilks test above.

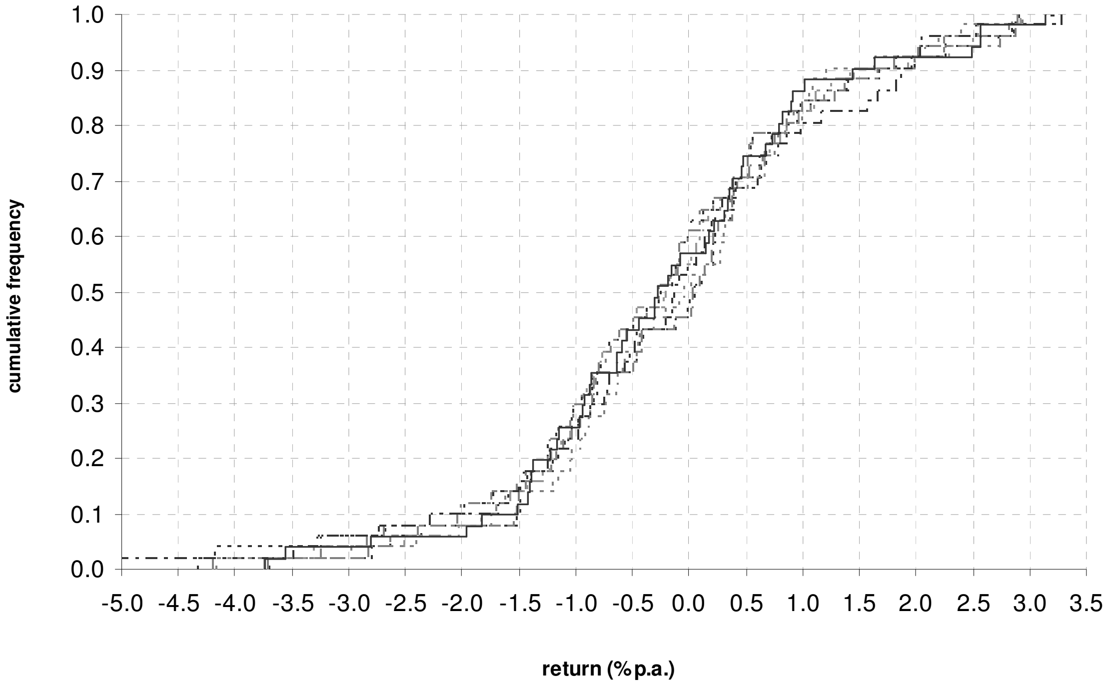

Figure 1: Return distributions of the eight mutual funds.

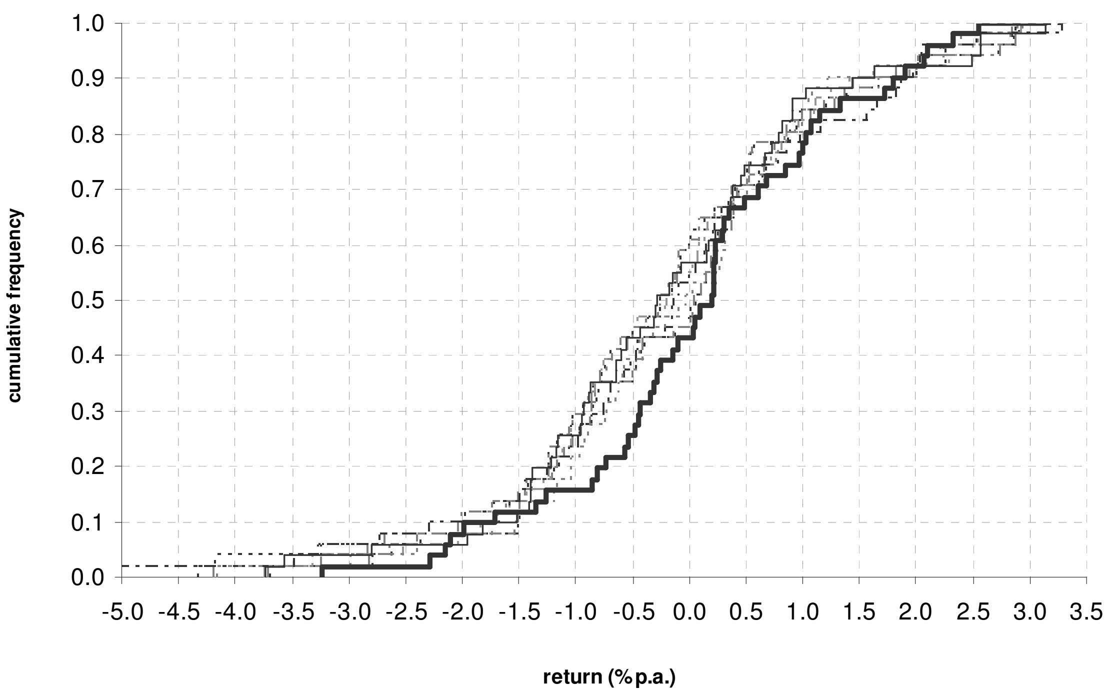

Figure 2: Return distributions of the mutual funds (thin lines) and a SSD dominating benchmark portfolio (thick line)

In Figure 2 we added a cumulative return distribution of a benchmark portfolio, composed of the stocks reported in Table 4, using the average portfolio weights. The benchmark portfolio incurs smaller losses in the left tail of the distribution, and performs notably better within the range from the 15th to the 60th percentile. However, evaluated funds tend to yield higher maximum returns. For all eight funds, the return distribution curves of the fund and the average benchmark cross each other at some point. Thus, the average benchmark does not dominate any fund by FSD. Still, this average benchmark dominates all eight funds by SSD; for any ("negative") area enclosed between the distribution curves of a fund and the benchmark such that the fund's curve is to the right from the benchmark curve, there is a larger ("positive") area enclosed between the distribution curves such that the fund's curve is to the left from the benchmark curve.

The efficiency measures can be interpreted in terms of the return distribution functions as follows. For any given benchmark portfolio, the DD measure (with direction vector specified as $\mathbf{g} = \mathbf{1}$) shifts the return distribution of an inefficient mutual fund to the right, preserving its original shape, until the dominance of the benchmark distribution is broken up. In problems (10) and (11) the benchmark portfolio is not fixed a priori, but is endogenously selected to maximize the right-ward shift. The PK measure similarly shifts the return distribution of an inefficient fund to the right until the dominance of the benchmark is broken up. Whereas the DD measure shifts the entire distribution in its original shape, the PK measure typically adjusts only one of the percentiles.

5.5 Comparative discussions

For sake of comparison, we also calculated the PK and DD performance measures relative to the classic mean-variance (MV) frontier. Interestingly, the resulting efficiency measures were almost equal to the FSD and SSD efficiency scores. This means that imposing the assumption of normally distributed returns does not influence the efficiency indices for these eight funds; recall that the return distributions of the eight mutual funds were approximately normal (see also Figure 1). This example demonstrates that, when we account for the portfolio diversification possibilities and the evaluated return distribution is approximately normal, the general SD criteria can be as powerful as the MV criterion (which imposes more restrictive assumptions of normally distributed returns or quadratic utility functions). This suggests that the low power of the traditional SD crossing algorithms (see e.g. Levy, 1992, 1998) is due to their failure to account for portfolio diversification. Of course, for non-normally distributed returns, the SD and MV measures can differ considerably (see e.g. Heikkinen and Kuosmanen, 2003)

To put the results of the eight SRI funds to proper perspective, we also considered a randomly drawn sample of 35 non-SRI funds from the population of 632 U.S.-based large blend funds with similar characteristics (net assets less than 250 Million US$ and minimum investment less than or equal to $1000). Nine funds were subsequently excluded because of missing data (those were new funds emerging during the evaluation period), which reduced the sample size of the control group to 26.

Table 5: SSD inefficiency of the control group (% points p.a.)

Fund (ticker symbol)	*PK / S* measure	Fund (ticker symbol)	*PK / S* measure
NPPAX	0.000	EVSBX	0.448
ASECX	0.279	HFFYX	0.449
SSLGX	0.323	HIGCX	0.449
WFDMX	0.389	HGRZX	0.452
MMLAX	0.394	FGIBX	0.456
MDLRX	0.396	FBLVX	0.463
OTRYX	0.404	PWSPX	0.468
STVDX	0.417	FLCIX	0.485
PRFMX	0.426	WCEBX	0.495
PRACX	0.430	FRMVX	0.501
GESPX	0.432	IGSCX	0.508
ACQAX	0.433	EGRCX	0.513
IBCCX	0.435	**Average**	0.419
AFEAX	0.444	**Std. Deviation**	0.100

Table 5 presents the normalized PK/S efficiency measures for the control group in the case of SSD. We observe very similar efficiency levels in the control group and the eight SRI funds. The average efficiency is slightly higher in the control group (0.419 versus 0.423), but the control group also exhibits greater dispersion, as measured by the standard deviation (0.100 versus 0.060). Note that the non-SRI funds were compared against reference portfolios formed from the DJSI stocks, which was considered a relevant benchmark for SRI funds. Interestingly, almost all non-SRI stocks were found to be inefficient relative to ”ethical” reference portfolios formed from the DJSI stocks. This suggests that efficient SRI funds do have potential to outperform almost any non-SRI fund in terms of the financial criteria, even if we ignore the ethical criteria. Overall, these efficiency measures do not reveal any systematic difference in the efficiency of the SRI funds and the control funds.

The absolute levels of the inefficiency premia were relatively small, considering the fact that our model completely ignores transaction costs. The portfolio inefficiencies we identified under “ideal” conditions of no transaction costs are so small that taking transaction costs into account would render all funds efficient from the perspective of private investors. Recall that our objective was to assess performance of the fund managers, who operate in nearly ideal conditions. Evaluating efficiency of the mutual funds versus stock portfolios from the perspective of private investors is a slightly different problem, which would require taking into account the transaction costs and possibly also some ethical criteria.

6. Discussion

We have presented a method for mutual fund performance measurement and best-practice benchmarking, which compares the mutual fund performance to an endogenously selected benchmark portfolio that tracks the evaluated fund’s risk profile. Our main objective is to provide the fund managers and investors with information about efficiency improvement potential and identify portfolio strategies that can achieve such improvements. To this end, we model the investment universe and its portfolio diversification possibilities by means of DEA. We departed from the traditional DEA approaches by characterizing the mutual fund's risk profile by means of the SD criteria, building on Kuosmanen (2001, 2004). We examined two alternative efficiency metrics, the slack-based Pareto-Koopmans measure and the directional distance function, and discussed some key properties of these two measures. We also derived dual expressions for the efficiency measures, and discussed their utility theoretic interpretations. Some practical issues involved in implementation of the SD based DEA approach were elaborated by means of an illustrative application.

The proposed approach can provide valuable information for the fund managers, as well as their trustees and clients. From the perspective of fund managers, the portfolio weights of the dominating benchmark portfolio, and their changes over time, are probably the most interesting information. Obviously, the past performance of the benchmark portfolio does not guarantee a good performance in the future. Still, a closer inspection of the industrial diversification and capitalization of the benchmark portfolio can reveal some useful hints for tactical asset allocation.

The best practice performance measures can provide useful incentive instruments for the trustees of the fund. One can see the performance measurement problem as a principal-agent game played between the trustees and the fund managers. In this respect, Bogetoft (1997) has shown how one can build second-best optimal incentive schemes based on DEA-type best practice benchmarks that will induce the agents to minimize their information rents. Similar incentive schemes might be applied in the present context to reward fund managers based on their efficiency.

The proposed efficiency measures could also provide useful information for the individuals who invest in mutual funds. The proposed performance measures can support the portfolio choice by providing information about the performance of the fund managers. Yet, we should emphasize that portfolio choice always involves a subjective element: the most efficient fund is not necessarily the most preferred one for all investors.

The sensitivity of the deterministic DEA to measurement errors, outliers, sampling errors, and missing variables is an ongoing concern. In this respect, it is worth to note that the return data from financial markers are usually much more reliable and accurate than empirical production data usually studied with DEA. Thus, the problem of measurement error seems a less serious concern in the present context. The problem of outliers can occur in the present setting, if the return possibilities set includes such assets that, for any reason, are infeasible investment alternatives for the fund manager. By careful modeling of the investment alternatives as well as the investment criteria and constraints faced the fund managers, the problem of outliers can be remedied. Since we construct the benchmark portfolios directly from stocks and other assets, heterogeneity of the evaluated funds does not obscure the efficiency measures (although it can affect their ranking). The sampling error seems not a major problem either: return data for stocks, bonds, and other investment alternatives are available, and modeling the fund manager's entire investment universe is technically feasible. Moreover, the sampling theory of the DEA efficiency estimators is nowadays well understood (see e.g. Simar and Wilson, 1998, 2000), and the insights from that literature can be directly applied in the present context as well. Therefore, we conclude that the problem of missing variables forms the most important challenge for the presented framework.

In empirical applications, we typically observe a sample of possible states of nature, not all possible states (this is equivalent to missing input or output variables in terms of standard DEA). Thus, the empirical return distribution may differ from the true but unknown distribution of returns. Some authors (e.g. Nelson and Pope 1991; Post, 2003) see this problem similar to the sampling error, and thus suggest to use the standard bootstrapping techniques as a remedy. However, the approaches presented thus far are rather ad hoc: it seems questionable to infer probability of returns in some unobserved states based on the observed ones. Such approach will fail to account for rarely occurring but critical events such as catastrophic risks. Further research is needed to treat the unobserved states in more satisfactory fashion.

Acknowledgements

This paper has benefited of helpful comments from Thierry Post, two anonymous reviewers, and participants in the 4^{th} International DEA Symposium, 5-6 September 2004, Birmingham, UK; the VIII European Workshop of Efficiency and Productivity Analysis, 25-27 September 2003, Oviedo, Spain; and the EURO/INFORMS joint international meeting, 6-10 July 2003, Istanbul, Turkey. The usual disclaimer applies.

This study forms a part of the *NOMEPRE*-research program. For further information, see: http://www.socialsciences.wur.nl/enr/staff/kuosmanen/program1/

Financial support from the Emil Aaltonen Foundation, Finland, for this program is gratefully acknowledged.

References

Andersen, P., and N.C. Petersen (1993): A Procedure for Ranking Efficient Units in Data Envelopment Analysis, *Management Science* 39, 1261-1264.

Ansell, J., P. Moles, and A. Smart (2003): Does benchmarking help? *International Transactions in Operational Research* 10, 339-350.

Bailey, J.V. (1992): Are Manager Universes Acceptable Performance Benchmarks? *Journal of Portfolio Management* 8(3), 9 13.

Banker, R.D., A. Charnes, W. W. Cooper (1984): Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* 30(9), 1078 – 1092.

Basso, A., and S. Funari (2001): A Data Envelopment Analysis Approach to Measure the Mutual Fund Performance, *European Journal of Operational Research* 135, 477-492.

Basso, A., and S. Funari (2003): Measuring the Performance of Ethical Mutual Funds: A DEA Approach, *Journal of the Operational Research Society* 54(5), 521-531.

Bawa, V.S., E. Lindenberg, and L. Rafsky (1979): An Algorithm to Determine Stochastic Dominance Admissible Sets, *Management Science* 25(7), 609-622.

Berger, A. N. and D. B. Humphrey (1997): Efficiency of Financial Institutions: International Survey and Directions for Future Research, *European Journal of Operational Research* 98, 175-212.

Berk, J.B. (2000): Sorting Out Sorts, *Journal of Finance* 55(1), 407-427.

Black, F., M.C. Jensen, and M. Scholes (1972): The Capital Asset Pricing Model: Some Empirical Tests, in M.C. Jensen (Ed.): *Studies in the Theory of Capital Markets*, Praeger Publishers, New York.

Bogetoft, P. (1997): DEA-based Yardstick Competition: The Optimality of Best Practice Regulation, *Annals of Operations Research* 73, 277–298.

Briec, W., K. Kerstens, and J.-B. Lesourd (2004): Nonparametric Tests of Portfolio Investment Efficiency: A Shortage Function Generalization, *Journal of Optimization Theory and Applications* 120(1), 1-27.

Brown, G., D. Davies, and P. Draper (1992): Pension Fund Trustees and Performance Measurement, *Management Accounting* 4, 38-44.

Chambers, R.G., Y. Chung and R. Färe (1998): Profit, Directional Distance Functions, and Nerlovian Efficiency, *Journal of Optimization Theory and Applications* 98(2), 351-364.

Charnes, A., W.W. Cooper, B. Golany, L. Seiford, and J. Stutz (1985): Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Frontiers, *Journal of Econometrics* 30(1-2), 91-107.

Charnes, A., W. Cooper, and E. Rhodes (1978): Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research* 2(6), 429 – 444.

Clarke, B.E., and A.W. Ryan (1994): Proper Overseas Benchmark: A Critical Choice, *Pensions & Investments* 22, 28-29.

Fama, E.F., and K.R. French (1992): The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427-465.

Fama, E.F., and K.R. French (1993): Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3-56.

Fishburn, P.C. (1964): *Decision and Value Theory*, John Wiley & Sons, New York.

Heikkinen, V.-P., and T. Kuosmanen (2003): Stochastic Dominance Portfolio Analysis of Forestry Assets, chapter 14 in J. Wesseler, H.-P. Weikard, and R. Weaver (Eds.): *Risk and Uncertainty in Environmental and Natural Resource Economics*, Edward Elgar, Cheltenham.

Jensen, M.C. (1967): The Performance of Mutual Funds in the Period 1945-1964, *Journal of Finance* 23(2), 389-416.

Kuosmanen, T. (2001): *Stochastic Dominance Efficiency Tests under Diversification*, Working Paper W-283, Helsinki School of Economics and Business Administration, Helsinki.

Kuosmanen, T. (2004): Efficient Diversification According to Stochastic Dominance Criteria, *Management Science* 50(10), 1390-1406.

Levy, H. (1992): Stochastic Dominance and Expected Utility: Survey and Analysis, *Management Science* 38(4), 555-593.

Levy, H. (1998): *Stochastic Dominance: Investment Decision Making Under Uncertainty*, Kluwer Academic Publishers, Boston.

Morey, M.R., and R.C. Morey (1999): Mutual Fund Performance Appraisals: A Multi-Horizon Perspective with Endogenous Benchmarking, *Omega* 27(2), 241-258.

Murthi, B.P.S., Y.K. Choi, and P. Desai (1997): Efficiency of Mutual Funds and Portfolio Performance Measurement, *European Journal of Operational Research* 98, 408-418.

Nelson, R.D., and R. Pope (1991): Bootstrapping Insights into Empirical Applications of Stochastic Dominance, *Management Science* 37, 1182-1194.

Post, G.T. (2003): Empirical Tests for Stochastic Dominance Efficiency, *Journal of Finance* 58(5), 1905-1931.

Quirk, J., and R. Saposnik (1962): Admissibility and Measurable Utility Functions, *Review of Economic Studies* 29, 140-146.

Roll, R. (1978): Ambiguity When Performance is Measured by the Security Market Line, *Journal of Finance* 33, 1051-1069.

Rothschild, M., and J.E. Stiglitz (1970): Increasing Risk I: A Definition, *Journal of Economic Theory* 2, 225-243.

Schmidt, P., and R.C. Sickles (1984): Production Frontiers and Panel Data, *Journal of Business and Economic Statistics* 4, 367-374.

Schueth, S. (2003): Socially Responsible Investing in the United States, *Journal of Business Ethics* 43, 189-194.

Simar, L. and P.W. Wilson (1998): Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models, *Management Science* 44(11), 49–61.

Simar, L. and P.W. Wilson (2000): Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis* 13, 49–78.

Thor, C.G. (1996): Let's Clear Up Some Confusion about Benchmarking, *Journal for Quality &* Participation 19(4), 88-90.

Tulkens, H. (1993): On FDH Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit, *Journal of Productivity Analysis* 4(1-2), 183-210.

Varian, H. (1983): Non-parametric Tests of Models of Investor Behavior, *Journal of Financial and Quantitative Analysis* 18, 269-278.

APPENDIX: Why FSD and SSD efficiency measures yield equal results for mutual funds?

In Section 4 we found that all eight environmentally responsible mutual funds scored equally well in terms of FSD and SSD efficiency. A similar finding was made by Kuosmanen (2004) regarding the marker portfolio. The purpose of this appendix is to rationalize this finding by means of a stylized graphical example.

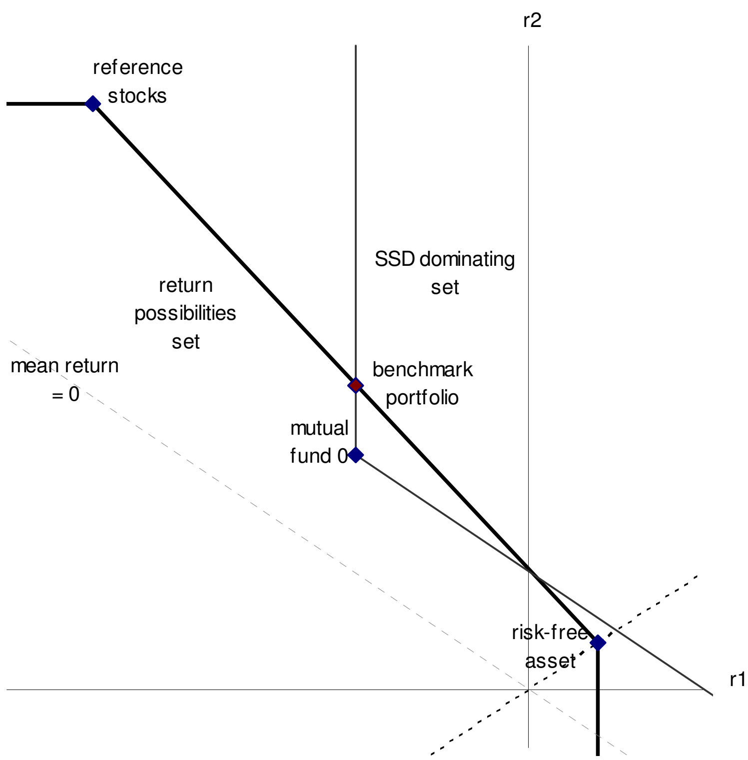

Figure A: Illustration of why FSD and SSD tests yield the same results.

Figure A presents a two-dimensional case where state 1 represents a bear market and state 2 a bull market. A risk-free asset is displayed in the bottom-right corner of the diagram; the broken diagonal line that runs through the risk free asset indicates vectors that yield equal return in both states. Volatile reference stocks are typically found in the top-left corner of the diagram, where return is negative in state 1, and highly positive in state 2. The evaluated mutual fund lies somewhere between the benchmark stock and the risk-free asset, within the return possibility set; this set is indicated by the thick solid piece-wise linear frontier with vertices in the risk-free asset and the reference stocks. The set of return vectors that dominate the evaluated fund by SSD (i.e., "the SSD dominating set", see Kuosmanen, 2004) is indicated by the thin piece-wise linear isoquant that runs through the evaluated fund. This dominating set overlaps with the return possibilities set, and thus fund 0 is SSD inefficient.

Note that the mean return of the reference stocks must typically be higher than the return of the risk-free asset, to compensate for the higher risk. Thus, the slope of the return possibilities frontier must generally be steeper than that of diagonal line-segment of the SSD dominating set, like in Figure 1. Recall that the PK measure selects the benchmark portfolio from the intersection of the return possibility set and the SSD dominating set by maximizing the difference in mean return between the benchmark portfolio and the evaluated fund. In this example, the benchmark portfolio will be found directly above the point representing the return vector of the mutual fund 0, as indicated in Figure A. Note that this benchmark is also included in the FSD dominating set. Thus, exactly the same benchmark is obtained by using the FSD criterion. (A similar argument holds for the DD measure.)

Although this stylized example involves only two states, it does describe the essential features of the phenomenon at hand. Also in the general setting with S states of nature, the maximum mean return over the intersection of the return possibility set and the SSD dominating set is usually found in the corner point where the boundaries of the return possibility set and the SSD dominating set intersect. The FSD and SSD measures will differ when there exists an asset that offers a high mean return with a low risk, or if the evaluated mutual fund itself is highly risky. Given the usual geometry of the return possibilities sets, the FSD and SSD measures are likely to yield the same results in the efficiency assessment of the mutual funds and other well-diversified portfolios such as the market portfolio.