Shortfall as a risk measure: properties, optimization and applications

Dimitris Bertsimas *   Geoffrey J. Lauprete †   Alexander Samarov ‡

October 2000, revised February 2001, January 2002

Abstract

Motivated from second-order stochastic dominance, we introduce a risk measure that we call shortfall. We examine shortfall's properties and discuss its relation to such commonly used risk measures as standard deviation, VaR, lower partial moments, and coherent risk measures. We show that the mean-shortfall optimization problem, unlike mean-VaR, can be solved efficiently as a convex optimization problem, while the sample mean-shortfall portfolio optimization problem can be solved very efficiently as a linear optimization problem. We provide empirical evidence (a) in asset allocation, and (b) in a problem of tracking an index using only a limited number of assets that the mean-shortfall approach might have advantages over mean-variance.

*Boeing Professor of Operations Research, Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, E53-363, Cambridge, MA 02139, dbertsim@mit.edu; corresponding author.

†Operations Research Center, Massachusetts Institute of Technology, lauprete@mit.edu.

‡Sloan School of Management, Massachusetts Institute of Technology and Department of Mathematics, University of Massachusetts, Lowell, samarov@mit.edu.

1 Introduction

The standard deviation of the return of a portfolio is the predominant measure of risk in finance. Indeed mean-variance portfolio selection using quadratic optimization, introduced by Markowitz [25], is the industry standard. It is well known (see Huang and Litzenberger [17] or Ingersoll [18]) that the mean-variance portfolio selection paradigm maximizes the expected utility of an investor if the utility is quadratic or if returns are jointly normal, or more generally, obey an elliptically symmetric distribution [1]. It has long been recognized, however, that there are several conceptual difficulties with using standard deviation as a measure of risk:

(a) Quadratic utility displays the undesirable properties of satiation (increase in wealth beyond a certain point decreases utility) and of increasing absolute risk aversion (the demand for a risky asset decreases as the wealth increases), see e.g., Huang and Litzenberger [17].

(b) The assumption of elliptically symmetric return distributions is problematic because it rules out possible asymmetry in the return distribution of assets, which commonly occurs in practice, e.g., due to the presence of options (see, e.g., Bookstaber and Clarke [6]). More generally, departures from ellipticity occur due to the greater contagion and spillover of volatility effects between assets and markets in down rather than up market movements, see, e.g. King and Wadhwani [20], Hamao, Masulis, and Ng [15], Neelakandan [27], and Embrechts, McNeil, and Straumann [11]. Asymmetric return distributions make standard deviation an intuitively inadequate risk measure because it equally penalizes desirable upside and undesirable downside returns. In fact, Chamberlain [7] has shown that elliptically symmetric are the only distributions for which investor's utility is a function only of the portfolio's mean and standard deviation.

Motivated by the above difficulties, alternative downside-risk measures have been proposed and analyzed in the literature (see the discussion in Section 3.1.4). Though intuitively appealing, such risk measures are not widely used for portfolio selection because of computational difficulties and problems with extending standard portfolio theory results, see, e.g., a recent review by Grootveld and Hallerbach [14].

In recent years the financial industry has extensively used quantile-based downside risk measures. Indeed one such measure, Value-at-Risk, or VaR, has been increasingly used as a risk management tool

[1] In particular, any multivariate random variable $\boldsymbol{R}$ with probability density $f(\boldsymbol{r}) = \frac{1}{\sqrt{\det(\boldsymbol{\Sigma})}} g((\boldsymbol{r}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}-\boldsymbol{\mu}))$, obeys an elliptically symmetric distribution (where the function $g : R^+ \rightarrow R^+$. While there exist elliptic distributions without finite moments, we consider here only elliptic distributions with finite variances.

(see e.g. Jorion [19], Dowd [9] and Duffie and Pan [10]). While VaR measures the worst losses which can be expected with certain probability, it does not address how large these losses can be expected when the "bad", small probability events occur. To address this issue, the "mean excess function", from extreme value theory, can be used (see Embrechts, McNeil, and Straumann [11] for applications in insurance). More generally, Artzner, et al. [1] propose axioms that risk measures (they call them coherent risk measures) should satisfy and show that VaR is not a coherent risk measure because it may discourage diversification and thus violates one of their axioms. They also show that, under certain assumptions, a version of the mean excess function, which they call tail conditional expectation (TailVaR), is a coherent measure (see Section 3.1.5 below).

Our goal in this paper is to propose an alternative methodology for defining, measuring, analyzing, and optimizing risk that addresses some of the conceptual difficulties of the mean-variance framework, to show that it is computationally tractable and has, we believe, interesting and potentially practical implications.

The key in our proposed methodology is a risk measure called **shortfall**, which we argue has conceptual, computational and practical advantages over other commonly used risk measures. It is a variation of the mean excess function and TailVaR mentioned earlier (see also Uryasev and Rockafellar [32]). Some mathematical properties of the shortfall and its variations have been discussed in Uryasev and Rockafellar [32], and Tasche [31].

Our contributions and the structure in this paper are:

1. We define shortfall as a measure of risk in Section 2, and motivate it by examining its natural connections with expected utility theory and stochastic dominance.

2. We discuss structural and mathematical properties of shortfall, including its relations to other risk measures in Section 3. Shortfall generalizes standard deviation as a risk measure, in the sense that it reduces to standard deviation, up to a scalar factor depending only on the risk level, when the joint distribution of returns is elliptically symmetric, while measuring only downside risk for asymmetric distributions. We point out a simple explicit relation of shortfall to VaR and show that it is in general greater than VaR at the same risk level. Moreover, we provide exact theoretical bounds that relate VaR, shortfall, and standard deviation, which indicate how big the error in evaluating VaR and shortfall can possibly be. We obtain closed form expressions for the gradient of shortfall with respect to portfolio weights as well as an alternative representation of shortfall, which shows that shortfall is a convex function of the portfolio weights, giving it an important advantage over VaR, which in general is not a convex function. This alternative expression also leads to an efficient sample mean-shortfall optimization algorithm in Section 5. We propose a natural nonparametric estimator of shortfall, which does not rely on any assumptions about the asset's distribution and is based only on historical data.

3. In Section 4, we formulate the portfolio optimization problem based on mean-shortfall optimization and show that because of its convexity it is efficiently solvable. We characterize the mean-shortfall efficient frontiers and, in the case when a riskless asset is present, prove a two-fund separation theorem. We also define and illustrate a new, risk-level-specific beta coefficient of an asset relative to a portfolio, which represents the relative contribution of the asset to the portfolio shortfall risk. When the riskless asset is present, the optimal mean-shortfall weights are characterized by the equations having the CAPM form involving this risk-level-specific beta. For the elliptically symmetric, and in particular normal, distributions, the optimal mean-shortfall weights, the efficient frontiers, and the generalized beta for any risk level all reduce to the corresponding standard mean-variance portfolio theory objects. However, for more general multivariate distributions, the mean-shortfall optimization may lead to portfolio weights qualitatively different from the standard theory and, in particular, may considerably vary with the chosen risk level of the shortfall, see simulated and real data examples in Section 6.

4. In Section 5, we show that the sample version of the population mean-shortfall portfolio problem, which is a convex optimization problem, can be formulated as a linear optimization problem involving a small number of constraints (twice the sample size plus two) and variables (the number of assets plus one plus the sample size). Uryasev and Rockafellar [32] have independently made the same observation in the context of optimizing the conditional value-at-risk. This implies that the sample mean-shortfall portfolio optimization is computationally feasible for very large number of assets. Together with observations in item **3.** above, this also importantly implies that the mean-shortfall optimization may be preferable to the standard mean-variance optimization, even if the distribution of the assets is in fact normal or elliptic, because in this case it leads to the efficient and stable computation of the same optimal weights and does not require the often problematic estimation of large covariance matrices necessary under the mean-variance approach.

5. In Section 6, we present computational suggesting that the efficient frontier in mean-standard deviation space constructed via mean-shortfall optimization outperforms the frontier constructed via mean-variance optimization. We also numerically demonstrate the ability of the mean-shortfall approach to handle cardinality constraints in the optimization process using standard linear mixed integer programming methods. In contrast, the mean-variance approach to the problem leads to a quadratic integer programming problem for which there is no available commercial software.

2 Definition and Motivation of Shortfall

In this section, we adopt the expected utility paradigm and theorems of stochastic dominance to motivate our definition of shortfall. Consider an investment choice based on expected utility maximization with an investor-specific utility function $u(\cdot)$. This means that an investment with random return X is preferred to the one with random return Y if $E[u(X)] \geq E[u(Y)]$, where the expectations are taken with respect to the corresponding distributions of X and Y respectively.

As it is very difficult to articulate a particular investor's utility, the literature usually considers large classes of utility functions satisfying very general properties. Most commonly considered classes are the class U_1 of increasing utility functions (more is preferred to less) and the class U_2 of increasing and concave utility functions (risk averse investors), see Ingersoll [18] and Huang and Litzenberger [17].

When the joint distribution of returns $\boldsymbol{R}$ is multivariate normal, mean-variance portfolio selection is consistent with expected utility maximization in the sense that given a fixed expected return, any investor with utility function in U_2 will prefer the portfolio with the smallest standard deviation. This means that in this case standard deviation is the appropriate measure of risk simultaneously for all utility functions in U_2. The same is true for elliptically symmetric distributions. However, for more general distributions, variance looses this property: even if $E[X] < E[Y]$ and $\sigma_X > \sigma_Y$, there exist $u \in U_2$, such that $E[u(X)] \geq E[u(Y)]$, (see Ingersoll [18]).

There exists an extensive theory connecting preferences over various utility classes to stochastic dominance relations between the distributions of the investment alternatives, see, e.g., a survey by Levy [23]. We will use the following theorem of Levy and Kroll [24] which characterizes preferences of investors with utilities in U_i, $i = 1, 2$, in terms of the quantile functions of their investments. We define the α-quantile of a random variable X as $q_\alpha(X) := \inf\{x | P(X \leq x) \geq \alpha\}, \quad \alpha \in (0, 1)$.

Theorem 1 (Levy and Kroll [24] and Levy [23]) *Let X and Y be random variables with continuous densities.*

(a) *$E[u(X)] \geq E[u(Y)]$ for all $u \in U_1$ if and only if $q_\alpha(X) \geq q_\alpha(Y), \quad \forall \alpha \in (0, 1)$ and we have strict inequality for some α.*

(b) *$E[u(X)] \geq E[u(Y)]$ for all $u \in U_2$ if and only if $E[X|X \leq q_\alpha(X)] \geq E[Y|Y \leq q_\alpha(Y)]$, $\quad \forall \alpha \in (0,1)$ and we have strict inequality for some α.*

Let us now consider a portfolio of n assets whose random returns are described by the random vector $\boldsymbol{R} = (R_1, \ldots, R_n)'$ having a joint density with the finite mean $\boldsymbol{\mu} = E[\boldsymbol{R}]$. We will assume throughout the paper that the joint distribution of $\boldsymbol{R}$ is continuous. Let $\boldsymbol{x} = (x_1, \ldots, x_n)'$ be portfolio weights, so that the total random return of the portfolio is $X = \boldsymbol{R'x}$. As we would like to concentrate on risk measures, let us fix the mean portfolio return to $E[\boldsymbol{R'x}] = r_p$. Investors are interested in maximizing their expected utility $E[u(X)]$.

According to Theorem 1(a), if $u \in U_1$ this is equivalent to minimizing over $\boldsymbol{x}$ the $(1-\alpha)$-confidence level Value-at-Risk $VaR_\alpha(\boldsymbol{x}) := \boldsymbol{\mu' x} - q_\alpha(\boldsymbol{R'x})$, $\quad \forall \alpha \in (0,1)$, where $q_\alpha(\boldsymbol{R'x})$ denotes the α-quantile of the distribution of the portfolio return $\boldsymbol{R'x}$. (It is a common practice in risk management to center VaR at the expected value, see for example Jorion [19], so that for the normal distribution it is equal to the standard deviation times a factor depending only on α.) Note that $VaR_\alpha(\boldsymbol{x})$ gives the size of the losses below the expected return, which may occur with probability no greater than α. Of course, the minimization of $VaR_\alpha(\boldsymbol{x})$ may not be achieved with a single portfolio simultaneously for all $\alpha \in (0,1)$. But Theorem 1(a) establishes that a portfolio $\boldsymbol{x}$ chosen to minimize $VaR_\alpha(\boldsymbol{x})$ for a fixed α and a given mean is non-dominated, i.e., there is no other portfolio with the same mean which would be preferred to $\boldsymbol{x}$ by all investors with utilities $u \in U_1$. Thus, Theorem 1(a) naturally leads to minimizing $VaR_\alpha(\boldsymbol{x})$ of a portfolio with weights $\boldsymbol{x}$ for some α. However, the fact that $VaR_\alpha(\boldsymbol{x})$ is a non-convex function of $\boldsymbol{x}$ causes theoretical and computational difficulties (see Lemus, et al. [22]).

Let us now consider investors with utility $u \in U_2$, i.e., it is not only increasing but also concave. Again fixing the mean portfolio return to $E[\boldsymbol{R'x}] = r_p$, according to Theorem 1(b), this is equivalent to minimizing the shortfall at the risk level α:

$$s_\alpha(\boldsymbol{x}) := \boldsymbol{\mu' x} - E[\boldsymbol{R'x}|\boldsymbol{R'x} \leq q_\alpha(\boldsymbol{R'x})], \quad \forall \alpha \in (0,1). \tag{1}$$

This definition means that $s_\alpha(\boldsymbol{x})$ measures how large losses, below the expected return, can be expected if the return of the portfolio drops below its α-quantile. As with $VaR_\alpha(\boldsymbol{x})$, the minimization of $s_\alpha(\boldsymbol{x})$ may not be achieved simultaneously for all $\alpha \in (0,1)$ with a single portfolio. But Theorem 1(b) establishes that a portfolio $\mathbf{x}$ chosen to minimize $s_\alpha(\boldsymbol{x})$ for a fixed α and a given mean is non-dominated, i.e., there is no other portfolio with the same mean which would be preferred to $\boldsymbol{x}$ by *all* investors with utilities $u \in U_2$. Thus, one is naturally led to minimize the quantity $s_\alpha(\boldsymbol{x})$ for some α which we call **shortfall** at level α.

3 Properties of Shortfall

The purpose of this section is to deepen our understanding of shortfall. We discuss its relation to other risk measures and outline various properties of shortfall.

3.1 Relation to Other Risk Measures

In this section, we explore the relations of shortfall to other measures of risk.

3.1.1 Relation to Standard Deviation

In this section, we show that for elliptically symmetric distributions the shortfall is proportional to the standard deviation, and thus comparing portfolio risks using shortfall **with any** α is equivalent to using the standard deviation. In this sense, shortfall generalizes standard deviation as a risk measure.

Proposition 1 **(a)** *If the vector of returns $\boldsymbol{R}$ obeys a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, then*

$$s_\alpha(\boldsymbol{x}) = \frac{\phi(z_\alpha)}{\alpha}(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x})^{1/2}, \tag{2}$$

where $\phi(z)$ is the density of the standard normal and z_α is its upper α-percentile, that is $P\{Z > z_\alpha\} = \alpha$ and Z is a standard normal.

(b) *If the vector of $\boldsymbol{R}$ has an elliptically symmetric distribution with a mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, then*

$$s_\alpha(\boldsymbol{x}) = p(\alpha)(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x})^{1/2}, \tag{3}$$

where the factor $p(\alpha)$ depends on the specific form of the elliptical distribution.

Proof **(a)** If $\boldsymbol{R}$ obeys a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, then $X = \boldsymbol{R}'\boldsymbol{x}$ obeys a normal distribution with mean $\mu = \boldsymbol{\mu}'\boldsymbol{x}$ and variance $\sigma^2 = \boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}$. Thus,

$$\begin{aligned} s_\alpha(\boldsymbol{x}) &= \mu - E[X|\ X \le q_\alpha(X)] = \mu - \frac{1}{\alpha\sigma\sqrt{2\pi}}\int_{-\infty}^{q_\alpha(X)} x\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\ &= -\frac{1}{\alpha\sigma\sqrt{2\pi}}\int_{-\infty}^{q_\alpha(X)}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx = -\frac{\sigma}{\alpha\sqrt{2\pi}}\int_{-\infty}^{z_\alpha} y\exp\left(-\frac{y^2}{2}\right)dy = \frac{\phi(z_\alpha)}{\alpha}\sigma. \end{aligned}$$

The proof of part **(b)** is analogous. □

Remark: If the vector of returns $\boldsymbol{R}$ obeys a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, then it is easy to see that $VaR_\alpha(\boldsymbol{x}) = z_\alpha(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x})^{1/2}$, that is in this case, standard deviation, VaR and shortfall are essentially equivalent, see Embrechts et al. [11].

3.1.2 Relation to Value-at-Risk

Recall that $VaR_\alpha(\boldsymbol{x}) := \boldsymbol{\mu}'\boldsymbol{x} - q_\alpha(\boldsymbol{R}'\boldsymbol{x})$.

Proposition 2 *The following relations between VaR and shortfall hold:*

(a) *Shortfall at level α is the average of VaR's for all levels below α, i.e., $s_\alpha(\boldsymbol{x}) = \frac{1}{\alpha}\int_0^\alpha VaR_u(\boldsymbol{x})du$,*

(b) $s_\alpha(\boldsymbol{x}) \geq VaR_\alpha(\boldsymbol{x})$.

(c) *Both $s_\alpha(\boldsymbol{x})$ and $VaR_\alpha(\boldsymbol{x})$ are decreasing functions of α.*

Proof **(a)** The expression follows from $E[X|X \leq q_\alpha(X)] = \dfrac{1}{\alpha}\displaystyle\int_0^\alpha q_u(X)du$.
(b) It is clear from the definition of the quantile that $VaR_\alpha(\boldsymbol{x})$ is a decreasing function of α, which together with part (a) implies that $s_\alpha(\boldsymbol{x}) \geq VaR_\alpha(\boldsymbol{x})$.
(c) Let $0 \leq \alpha_1 \leq \alpha_2 \leq 1$. Since $s_\alpha(\boldsymbol{x}) = \frac{1}{\alpha}\int_0^\alpha VaR_u(\boldsymbol{x})du$, we have

$$\begin{aligned}
s_{\alpha_2}(\boldsymbol{x}) &= \frac{\alpha_1}{\alpha_2}s_{\alpha_1}(\boldsymbol{x}) + \frac{1}{\alpha_2}\int_{\alpha_1}^{\alpha_2} VaR_u(\boldsymbol{x})du \\
&\leq \frac{\alpha_1}{\alpha_2}s_{\alpha_1}(\boldsymbol{x}) + \frac{1}{\alpha_2}VaR_{\alpha_1}(\boldsymbol{x})\int_{\alpha_1}^{\alpha_2} du \qquad (VaR_u(\boldsymbol{x}) \text{ decreasing}) \\
&\leq \frac{\alpha_1}{\alpha_2}s_{\alpha_1}(\boldsymbol{x}) + \frac{\alpha_2-\alpha_1}{\alpha_2}s_{\alpha_1}(\boldsymbol{x}) \qquad (\text{part (b)}) \\
&= s_{\alpha_1}(\boldsymbol{x}).
\end{aligned}$$

□

3.1.3 Optimal Bounds on Shortfall

In this section, for given values of the mean and standard deviation, we obtain universal bounds on VaR and shortfall that are best possible in the sense that there exist probability distributions that attain them. This allows us to compute bounds on VaR and shortfall even if the distribution of returns are unknown. For ease of notation, we write in this subsection $s_\alpha(\boldsymbol{x}) = s_\alpha$ and $q_\alpha(\boldsymbol{R}'\boldsymbol{x}) = q_\alpha$. We use the techniques from Bertsimas and Popescu [5] to derive these bounds.

Theorem 2 *The inequalities shown in Table 1 are valid and best possible.*

Table 2 compares the value of shortfall for the normal case: $s_\alpha = \frac{\phi(z_\alpha)}{\alpha}\sigma$, where, as in Eq. (2), $\phi(z)$ is the density of the standard normal and z_α is its upper α-percentile, with the largest value $\sqrt{(1-\alpha)/\alpha}\sigma$, which s_α may possible achieve for any distribution. Table 2 shows that while for a risk level $\alpha = 10\%$ the maximum possible "model risk" of shortfall estimation is moderate, for the commonly considered smaller risk level $\alpha = 1\%$, the normal model may underestimate shortfall by a factor of up to $9.9499/2.6652 = 3.7$.

(a) Optimal Bounds on $VaR_\alpha := \mu - q_\alpha$ given μ and σ	$-\sigma\sqrt{\alpha/(1-\alpha)} \leqslant VaR_\alpha \leqslant \sigma\sqrt{(1-\alpha)/\alpha}$,
(b) Optimal Bounds on s_α given $\mu,\ \sigma,$ and VaR_α	$\left\{\begin{array}{ll} VaR_\alpha & \text{if } VaR_\alpha \geq 0 \\ -VaR_\alpha(1-\alpha)/\alpha & \text{if } VaR_\alpha < 0 \end{array}\right\} \leqslant s_\alpha \leqslant \sigma\sqrt{(1-\alpha)/\alpha}$,
(c) Optimal Bounds on s_α given μ and σ	$0 \leqslant s_\alpha \leqslant \sigma\sqrt{(1-\alpha)/\alpha}$.

Table 1: Optimal bounds on quantiles and shortfall.

α	$\frac{\varphi(z_\alpha)}{\alpha}$	$\sqrt{\frac{1-\alpha}{\alpha}}$
0.1	1.7550	3.0000
0.05	2.0627	4.3589
0.01	2.6652	9.9499
0.005	2.8919	14.1067

Table 2: Comparison of shortfall under a normal distribution and under the worst case distribution.

3.1.4 Relation to Lower Partial Moments

The general idea of downside risk measures has been extensively discussed in the financial economics literature starting with the safety first approach of Roy [29] and Markowitz's [25] discussion of semi-variance. A more general notion of lower partial moment (LPM) as a measure of risk was introduced and analyzed by Bawa [2], [3], Fishburn [12], and Bawa and Lindenberg [4], see also more recent papers by Harlow and Rao [16] and Grootveld and Hallerbach [14]. The LPM of order a with the threshold τ of a portfolio return $X = \boldsymbol{R'x}$ is defined as follows:

$$LPM_a(\tau; X) := \int_{-\infty}^{\tau} (\tau - t)^a dF_X(t), \quad a \geq 0. \tag{4}$$

Notice that the lower partial moment for $a = 0$ ($LPM_0(\tau; X) = F_X(\tau)$) corresponds to Roy's safety and $a = 2$ corresponds to semi-variance. The threshold parameter τ is usually chosen as a short term interest rate, or expected return, or as a minimal acceptable return. Though intuitively appealing, the LPM risks are not widely used for portfolio selection because of computational difficulties and the fact that standard portfolio theory results, like linear two-fund separation, extend to the LPM risks only for some special values of $\boldsymbol{\tau}$ or for some special families of distributions, see Harlow and Rao [16] and Grootveld and Hallerbach [14]. We have $s_\alpha(\boldsymbol{x}) = \boldsymbol{\mu'x} - q_\alpha(\boldsymbol{R'x}) + \frac{1}{\alpha} LPM_1(q_\alpha(\boldsymbol{R'x}); \boldsymbol{R'x})$. In contrast to LPM measures of risk, we show in Section 5 that shortfall minimization is efficiently solvable.

3.1.5 Relation to Coherent Risk Measures

Artzner et al. [1] propose four axioms which, they argue, every measure of risk should satisfy. They call measures of risk that satisfy those four axioms coherent. While Artzner et al. [1] considered only discrete probability spaces, Delbaen [8] extends their definitions to the case of arbitrary probability spaces. A coherent risk measure $\rho(X)$ of an investment with the random return X is a real-valued function defined on the space of real-valued random variables which satisfies the following axioms:

(i) (Translation invariance). For any $a \in R$, $\rho(X + a) = \rho(X) - a$.

(ii) (Subadditivity). For any random variables X and Y, $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

iii) (Positive homogeneity). For all $t \geq 0$, $\rho(tX) = t\rho(X)$.

(iv) (Positivity). If $X \geq 0$, $\rho(X) \leq 0$.

This definition rules out as incoherent, under general distributional assumptions, risk measures based on standard deviation (violates axiom (iv)), on value-at-risk (violates axiom (ii)), and on semi-variance (violates axiom (iv)). However, when the assets' returns have elliptically symmetric distributions, all these risk measures are coherent.

Recall that we have $X = \boldsymbol{R'x}$. Artzner et. al. [1] introduce the risk measure called tail conditional expectation (TailVaR) $TCE_\alpha(\boldsymbol{x}) = -E[\boldsymbol{R'x}|\boldsymbol{R'x} \leq q_\alpha(\boldsymbol{R'x})]$. It is easy to verify that $TCE_\alpha(\boldsymbol{x})$ is a coherent risk measure when the underlying random variables $\boldsymbol{R}$ have a joint density.

Note that the Artzner's et. al. [1] definition of risk does not consider assets' expected returns separately, while we follow the standard practice and consider the reward, measured by the expected return, separately from risk. So, $s_\alpha(\boldsymbol{x}) = \boldsymbol{\mu'x} + TCE_\alpha(\boldsymbol{x})$, and because of this mean adjustment it violates axiom (i): $s_\alpha(X + a) = s_\alpha(X)$, and axiom (iv): we have $s_\alpha(X) \geq 0$. It satisfies, however, the remaining two axioms (ii) and (iii), see Proposition 3. Note also that if we fix the expected return of the portfolio, then shortfall minimization results in the same allocation $\boldsymbol{x}$ as in the problem of minimizing the tail conditional expectation.

3.2 An Alternative Expression for Shortfall

In this section, we show that the shortfall $s_\alpha(\boldsymbol{x})$ can be expressed in terms of the so-called "check" function $\rho_\alpha(\cdot)$ sometimes used in defining quantiles, see Koenker and Bassett [21]. This representation, which is interesting in its own right, gives an alternative proof of shortfall's convexity and also leads to an efficient sample mean-shortfall optimization algorithm in Section 5.

Let $z \in R$ and $\alpha \in (0,1)$. We define the function

$$\rho_\alpha(z) = \alpha z - z 1_{\{z<0\}} = \begin{cases} \alpha z & \text{if } z \geq 0, \\ (\alpha - 1)z & \text{if } z < 0. \end{cases}$$

It is straightforward to verify that

$$\arg\min_{c \in R^1} E[\rho_\alpha(\boldsymbol{R'x} - c)] = q_\alpha(\boldsymbol{R'x}). \tag{5}$$

Now we can write the expression for the shortfall of portfolio $\boldsymbol{x}$ as

$$\begin{aligned} s_\alpha(\boldsymbol{x}) &= \boldsymbol{\mu}'\boldsymbol{x} - E[\boldsymbol{R'x} | \boldsymbol{R'x} \leq q_\alpha(\boldsymbol{R'x})] = E\left[(\boldsymbol{R'x} - q_\alpha(\boldsymbol{R'x})) - \frac{1}{\alpha}(\boldsymbol{R'x} - q_\alpha(\boldsymbol{R'x}))1_{\{\boldsymbol{R'x} \leq q_\alpha(\boldsymbol{R'x})\}}\right] \\ &= \frac{1}{\alpha} E[\rho_\alpha(\boldsymbol{R'x} - q_\alpha(\boldsymbol{R'x}))] \qquad \text{(from Eq. (5))} = \frac{1}{\alpha} \min_c E[\rho_\alpha(\boldsymbol{R'x} - c)]. \end{aligned} \tag{6}$$

Note that the function $\rho_\alpha(z)$ is convex, and thus it follows that shortfall is a convex function of $\boldsymbol{x}$, see Rockafellar [28]. Note that in contrast to shortfall $s_\alpha(\boldsymbol{x})$, $VaR(\boldsymbol{x})$ is not in general a convex function of $\boldsymbol{x}$ as illustrated in Artzner et. al. [1] [2].

3.3 Mathematical Properties of Shortfall

The following general properties of shortfall are used in various parts of the paper. Formula (7) for the gradient of $s_\alpha(\boldsymbol{x})$ has been also independently obtained by Tasche [31] and Scaillet [30].

Proposition 3 *Shortfall satisfies the following properties:*

(a) $s_\alpha(\boldsymbol{x}) \geq 0$ *for all* $\boldsymbol{x}$ *and* $\alpha \in (0,1)$. *Moreover, the shortfall* $s_\alpha(\boldsymbol{x})$ *is equal to zero for some* $\boldsymbol{x}$ *and* α *if and only if* $\boldsymbol{R'x}$ *is constant with probability* 1.

(b) *The shortfall is positively homogeneous, i.e.,* $s_\alpha(t\boldsymbol{x}) = t s_\alpha(\boldsymbol{x})$, *for all* $t \geq 0$.

(c) *If the distribution of returns has a continuous positive density, then*

$$\nabla_{\boldsymbol{x}} s_\alpha(\boldsymbol{x}) = \boldsymbol{\mu} - E[\boldsymbol{R} | \boldsymbol{R'x} \leq q_\alpha(\boldsymbol{R'x})]. \tag{7}$$

Proof **(a)** Let $X = \boldsymbol{R'x}$. Conditioning on $X \leq q_\alpha(X)$ and its complement, we obtain

$$s_\alpha(\boldsymbol{x}) = E[X] - E[X | X \leq q_\alpha(X)] = (1-\alpha)\left\{E[X | X > q_\alpha(X)] - E[X | X \leq q_\alpha(X)]\right\} \geq 0,$$

[2] Artzner et. al. [1] show that there are distributions for which

$$VaR\left(\frac{1}{2}\boldsymbol{x_1} + \frac{1}{2}\boldsymbol{x_2}\right) > VaR\left(\frac{1}{2}\boldsymbol{x_1}\right) + VaR\left(\frac{1}{2}\boldsymbol{x_2}\right) = \frac{1}{2}VaR(\boldsymbol{x_1}) + \frac{1}{2}VaR(\boldsymbol{x_2}).$$

and $s_\alpha(\boldsymbol{x}) = 0$ for all $\alpha \in (0,1)$ if and only if $P(\boldsymbol{R'x} = c) = 1$.

(b) Clearly, $q_\alpha(t\boldsymbol{R'x}) = tq_\alpha(\boldsymbol{R'x})$ for $t \geq 0$, and thus homogeneity follows.

(c) (see also Scaillet [30]) We will prove (7) for one component of $\nabla_{\boldsymbol{x}} s_\alpha(\boldsymbol{x})$. We have

$$\frac{\partial s_\alpha(\boldsymbol{x})}{\partial x_k} = \mu_k - \frac{1}{\alpha}\frac{\partial}{\partial x_k} E\left[(\boldsymbol{R'x})\ 1\{\boldsymbol{R'x} \leq q_\alpha(\boldsymbol{R'x})\}\right).$$

Writing the last expectation as a bivariate integral in the variables $u = \sum_{j \neq k} x_j r_j$ and $v = r_k$ and differentiating with respect to x_k, we obtain denoting $f(u,v)$ the corresponding bivariate density:

$$\begin{aligned}
\frac{\partial s_\alpha(\boldsymbol{x})}{\partial x_k} &= \mu_k - \frac{1}{\alpha}\frac{\partial}{\partial x_k}\int\int_{\mathcal{R}^2} (u + x_k v) 1\{u + x_k v \leq q_\alpha(\boldsymbol{R'x})\} f(u,v) du dv \\
&= \mu_k - \frac{1}{\alpha}\frac{\partial}{\partial x_k}\int_{-\infty}^{\infty}\int_{-\infty}^{q_\alpha(\boldsymbol{R'x}) - x_k v} (u + x_k v) f(u,v) du dv \\
&= \mu_k - \frac{1}{\alpha}\int_{-\infty}^{\infty}\int_{-\infty}^{q_\alpha(\boldsymbol{R'x}) - x_k v} v f(u,v) du dv \\
&\quad - \frac{1}{\alpha}\int_{-\infty}^{\infty}\left(\frac{\partial q_\alpha(\boldsymbol{R'x})}{\partial x_k} - v\right) q_\alpha(\boldsymbol{R'x}) f(q_\alpha(\boldsymbol{R'x}) - x_k v, v) dv. \qquad (8)
\end{aligned}$$

By the definition of the quantile $q_\alpha(\boldsymbol{R'x})$

$$\alpha = \int\int_{\{(u,v): u + x_k v \leq q_\alpha(\boldsymbol{R'x})\}} f(u,v) du dv = \int_{-\infty}^{\infty}\int_{-\infty}^{q_\alpha(\boldsymbol{R'x}) - x_k v} f(u,v) du dv.$$

Differentiating this equation with respect to x_k we obtain that the last integral in (8) is equal to 0, and thus (7) follows.

$\square$

Remark. Under appropriate conditions on the distribution of $\boldsymbol{R}$, it can be also shown that the Hessian of $s_\alpha(\boldsymbol{x})$ has the form

$$\nabla^2_{\boldsymbol{x}} s_\alpha(\boldsymbol{x}) = \frac{f_{\boldsymbol{R'x}}(q_\alpha(\boldsymbol{R'x}))}{\alpha} Cov(\boldsymbol{R} | \boldsymbol{R'x} = q_\alpha(\boldsymbol{R'x})), \qquad (9)$$

where $f_{\boldsymbol{R'x}}(.)$ is the probability density of $\boldsymbol{R'x}$ and $Cov(\boldsymbol{R}|.)$ is the conditional covariance matrix of $\boldsymbol{R}$. Note that (9) also implies the convexity of $s_\alpha(\boldsymbol{x})$.

3.4 Estimation of Shortfall

In this section, we discuss the estimation of $s_\alpha(\boldsymbol{x})$ given a sample of T returns on the n assets $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T$. Let $r_t(\boldsymbol{x}) = \boldsymbol{r}_t'\boldsymbol{x}$ be the portfolio return in period t given a portfolio allocation $\boldsymbol{x}$. We propose the following natural estimator of shortfall. We sort $r_t(\boldsymbol{x})$ in the increasing order:

$$r_{(1)}(\boldsymbol{x}) \leq r_{(2)}(\boldsymbol{x}) \leq \ldots \leq r_{(T)}(\boldsymbol{x}).$$

Let $\overline{\boldsymbol{r}}$ denote the sample mean of $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T$ and $K = \lfloor \alpha T \rfloor$. Using these definitions we obtain the nonparametric estimator of $s_\alpha(\boldsymbol{x})$:

$$\hat{s}_\alpha(\boldsymbol{x}) = \boldsymbol{x}'\overline{\boldsymbol{r}} - \frac{1}{K}\sum_{j=1}^{K} r_{(j)}(\boldsymbol{x}), \tag{10}$$

which does not rely on any distributional assumptions.

In case one needs to estimate $s_\alpha(\boldsymbol{x})$ for α so small that $\alpha T < 1$, extreme value theory can be used to extrapolate outside the observed sample, i.e., to estimate the expected size of "a yet unseen disaster", see, e.g. Embrechts et al. [11].

4 The Efficient Shortfall Frontier

In this section, we consider the mean-shortfall portfolio optimization problem

$$\begin{aligned} \text{minimize} \quad & s_\alpha(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x}'\boldsymbol{\mu} = r_p, \quad \boldsymbol{e}'\boldsymbol{x} = 1, \end{aligned} \tag{11}$$

where $\boldsymbol{e}$ is the column vector of 1's. Problem (11) is defined analogously to the mean-variance portfolio optimization:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x}'\boldsymbol{\mu} = r_p, \quad \boldsymbol{e}'\boldsymbol{x} = 1. \end{aligned} \tag{12}$$

Because of the convexity of $s_\alpha(\boldsymbol{x})$ a solution $\boldsymbol{x}_\alpha(r_p)$ of Problem (11) exists, and the graph of $s_\alpha(r_p) = s_\alpha(\boldsymbol{x}_\alpha(r_p))$ as a function of r_p gives the minimum α-shortfall frontier. We next show that this frontier is a convex curve in the (r_p, s_α) plane.

Proposition 4 *The frontier curve $s_\alpha(r_p)$ is convex.*

Proof Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be any two frontier portfolios with distinct means $r_{p_1} = \boldsymbol{x}_1'\boldsymbol{\mu}$ and $r_{p_2} = \boldsymbol{x}_2'\boldsymbol{\mu}$. Let $\lambda \in (0,1)$. Then $r_p = \lambda r_{p_1} + (1-\lambda) r_{p_2}$ is the mean of the portfolio $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$. Now since the portfolio $\boldsymbol{x}_\alpha(r_p)$ is a minimizer of $s_\alpha(\boldsymbol{x})$ with portfolio mean r_p and since $s_\alpha(\boldsymbol{x})$ is convex in $\boldsymbol{x}$,

$$s_\alpha(r_p) = s_\alpha(\boldsymbol{x}_\alpha(r_p)) \le s_\alpha(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \le \lambda s_\alpha(\boldsymbol{x}_1) + (1-\lambda) s_\alpha(\boldsymbol{x}_2) = \lambda s_\alpha(r_{p_1}) + (1-\lambda) s_\alpha(r_{p_2}). \quad \square$$

4.1 Minimum shortfall frontier in the presence of a riskless asset

We next consider the mean-shortfall portfolio optimization problem in the presence of a riskless asset with rate of return r_f. Recall first that the minimum variance frontier found by solving the problem

$$\begin{aligned} \sigma(r_p) = \text{minimize} \quad & \boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x}'\boldsymbol{\mu} + (1 - \boldsymbol{e}'\boldsymbol{x})r_f = r_p, \end{aligned} \tag{13}$$

can be generated as a linear combination of the riskless asset and a single risky portfolio obtained by solving (13) for a single value of r_p. In the usual case when $r_p \geq r_f$, the frontier consists of a positively sloped ray in the $(r_p, \sigma(r_p))$ plane:

$$\sigma(r_p) = A(r_p - r_f), \quad \text{where} \quad A = ((\boldsymbol{\mu} - \boldsymbol{e}r_f)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{e}r_f))^{-1/2},$$

see Ingersoll [18] or Huang and Litzenberger [17] for details.

We next show that a similar situation takes place for the minimum shortfall frontier when a riskless asset is present. The minimum shortfall frontier in the presence of a riskless asset is defined as

$$\begin{aligned} s_\alpha(r_p) = \text{minimize} \quad & s_\alpha(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x}'\boldsymbol{\mu} + (1 - \boldsymbol{e}'\boldsymbol{x})r_f = r_p. \end{aligned} \tag{14}$$

Proposition 5 *(Tasche [31]) The minimum shortfall frontier in the $(r_p, s_\alpha(r_p))$ space, with $r_p \geq r_f$, is a ray starting from the point $(r_f, 0)$ and passing through a particular point $(r_p^*, s_\alpha(r_p^*))$ with $r_p^* > r_f$.*

Proof We consider the Lagrangean for the problem (14) $L(\boldsymbol{x}, \gamma) = s_\alpha(\boldsymbol{x}) - \gamma(\boldsymbol{x}'\boldsymbol{\mu} + (1 - \boldsymbol{x}'\boldsymbol{e})r_f - r_p)$, and set its derivatives to zero:

$$\frac{\partial L}{\partial \boldsymbol{x}} = \boldsymbol{\mu} - E(\boldsymbol{R}|\boldsymbol{x}'\boldsymbol{R} \leq q_\alpha(\boldsymbol{R}'\boldsymbol{x})) - \gamma(\boldsymbol{\mu} - \boldsymbol{e}r_f) = \boldsymbol{0}, \quad \frac{\partial L}{\partial \gamma} = r_p - (\boldsymbol{x}'\boldsymbol{\mu} + (1 - \boldsymbol{x}'\boldsymbol{e})r_f) = 0, \tag{15}$$

where we used formula (7) in the first of Eqs. (15). Let $r_p^* > r_f$ be a particular target mean value. Let us denote by $\boldsymbol{x}^*$ and γ^* the solutions of Eqs. (15) for $r_p = r_p^*$.

Consider an arbitrary $r_p \geq r_f$ and choose a scalar λ such that $r_p = \lambda r_f + (1 - \lambda)r_p^*$. Since $q_\alpha((1 - \lambda)\boldsymbol{R}'\boldsymbol{x}^*) = (1 - \lambda)q_\alpha(\boldsymbol{R}'\boldsymbol{x}^*)$ we observe that the vector $(1 - \lambda)\boldsymbol{x}^*$ and the scalar γ^* solve Eqs. (15) for $r_p = \lambda r_f + (1 - \lambda)r_p^*$. Therefore, the entire minimum shortfall frontier can be generated by solving Eqs. (15) for a single r_p^* and then multiplying the solution by a scalar factor. The minimum shortfall corresponding to r_p is $s_\alpha(r_p) = s_\alpha((1 - \lambda)\boldsymbol{x}^*) = (1 - \lambda)s_\alpha(r_p^*)$. $\square$

Note that, in view of (2) and (3), when the joint distribution of returns is normal or, more generally, elliptical with finite variance, the shortfall optimization problems (14) and (11) reduce, for all $\alpha \in (0, 1)$, to the mean-variance problems (12) and (13), respectively. For more general joint distributions of $\boldsymbol{R}$, however, the solutions of the problems (14) and (11) will depend on α, see the examples in Section 6.

4.2 Shortfall beta

In this section, we show that, generalizing the standard beta coefficient in a natural way, we can define a risk-level-specific shortfall beta and interpret it analogously to the CAPM model.

Proposition 6 *(Tasche [31]) The optimal solution of Problem (14) satisfies:*

$$\mu_j - r_f = \beta_{j,\alpha}(\boldsymbol{x}_\alpha)(r_p - r_f), \ \ j = 1, ..., n, \tag{16}$$

$$\beta_{j,\alpha}(\boldsymbol{x}) = \frac{1}{s_\alpha(\boldsymbol{x})} \frac{\partial s_\alpha(\boldsymbol{x})}{\partial x_j} = \frac{\mu_j - E(R_j | \boldsymbol{x}'\boldsymbol{R} \le q_\alpha(\boldsymbol{R}'\boldsymbol{x}))}{\boldsymbol{x}'\boldsymbol{\mu} - E(\boldsymbol{x}'\boldsymbol{R} | \boldsymbol{x}'\boldsymbol{R} \le q_\alpha(\boldsymbol{R}'\boldsymbol{x}))}, \quad j = 1, \dots, n. \tag{17}$$

Proof Multiplying Eq. (15) by $\boldsymbol{x}$, using the second of Eqs. (15), and solving for γ we obtain

$$\gamma = \frac{\boldsymbol{\mu}'\boldsymbol{x} - E(\boldsymbol{x}'\boldsymbol{R} | \boldsymbol{x}'\boldsymbol{R} \le q_\alpha(\boldsymbol{R}'\boldsymbol{x}))}{\boldsymbol{x}'\boldsymbol{\mu} - r_f \boldsymbol{e}'\boldsymbol{x}} = \frac{\boldsymbol{\mu}'\boldsymbol{x} - E(\boldsymbol{x}'\boldsymbol{R} | \boldsymbol{x}'\boldsymbol{R} \le q_\alpha(\boldsymbol{R}'\boldsymbol{x}))}{r_p - r_f}.$$

Substituting the value of γ in Eqs. (15) we obtain (16). □

The quantity $\beta_{j,\alpha}(\boldsymbol{x})$, which we call shortfall beta, can be interpreted as the relative change in shortfall when varying the weight of asset j. Note that, as with the standard beta, $\sum_{j=1}^n x_j \boldsymbol{\beta}_{j\alpha}(\boldsymbol{x}) = 1$, which effectively gives a decomposition of the portfolio shortfall into the individual assets' contributions, see Tasche [31], and also Garman [13], Dowd [9], and Lemus, et al. [22] for a similar decomposition of VaR. For the multivariate normal and for elliptically symmetric distributions of returns, (17) reduces to the standard definition of beta, $\beta_j(\boldsymbol{x}) = Cov(R_j, \boldsymbol{R}'\boldsymbol{x})/Var(\boldsymbol{R}'\boldsymbol{x})$, for all values of $\alpha \in (0, 1)$; this can be verified by direct calculation using the fact that for elliptic distributions the conditional expectation of any linear combination of $\boldsymbol{R}$ given its another linear combination is linear in the conditioning variable, see, e.g., Muirhead [26]. It is also easy to show that the shortfall beta $\beta_{j,\alpha}(\boldsymbol{x})$ is an example of the generalized measure of systematic risk of an asset relative to a portfolio discussed in Ingersoll [18].

The fact that $\beta_{j\alpha}(\boldsymbol{x})$ depends on α can be used to quantify the empirically observed fact that components of the market (portfolio) become more dependent on the market when the latter is more volatile, i.e., far out in the tails, and are less dependent on the market in more quiet periods. For asymmetric distributions, $\beta_{j\alpha}(\boldsymbol{x})$ depends on α, where as for elliptic distributions $\beta_{j,\alpha}(\boldsymbol{x})$ is constant over α (see also Section 6). Thus, elliptic distributions cannot capture this phenomenon. Writing the vector $\boldsymbol{\beta}_\alpha(\boldsymbol{x})$ as $\nabla_{\boldsymbol{x}} \log s_\alpha(\boldsymbol{x})$, it is easy to verify that $\boldsymbol{\beta}_\alpha(\boldsymbol{x})$ is constant over α if and only if $s_\alpha(\boldsymbol{x}) = g(\alpha)b(\boldsymbol{x})$ with some functions $g(\alpha)$ and $b(\boldsymbol{x})$ which depend on the joint distribution of returns.

Note that the validity of Eq. (16) for all $\alpha \in (0, 1)$ implies that $\beta_{j,\alpha}(\boldsymbol{x}_\alpha)$ is constant over α and, in fact, equal to the standard beta $\beta_j(\boldsymbol{x}_{\text{mv}})$ evaluated at the mean-variance optimal portfolio $\boldsymbol{x}_{\text{mv}}$ (assuming that second moments exist).

5 Sample Mean-Shortfall Optimization

In this section, we outline efficient algorithms for the sample mean-shortfall optimization problem. The advantage of our approach is that we do not make any assumptions on the distribution of returns $\boldsymbol{R}$, but rather work directly with the historical data. We show that the sample mean-shortfall optimization can be formulated as a linear optimization problem involving a small number of constraints (twice the sample size plus two) and variables (the number of assets plus one plus the sample size). This formulation replaces sorting, usually required for computing quantile-based quantities, with an appropriately chosen linear program. In fact, more generally, this approach may be computationally competitive for sorting large arrays and it may be of independent interest.

Let $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T$ be the vectors of realized (historical) returns in periods $t = 1, \ldots, T$. Assume that our forecast for the vector of returns for the next period $T+1$ is the historical mean: $\overline{\boldsymbol{r}} = \sum_{i=1}^{T} \boldsymbol{r}_i / T$.

In Eq. (10) we considered the following nonparametric estimator of shortfall: $\hat{s}_\alpha(\boldsymbol{x}) = \boldsymbol{x}'\overline{\boldsymbol{r}} - \sum_{i=1}^{K} r_{(i)}(\boldsymbol{x})/K$, where $\boldsymbol{r}_{(t)}(\boldsymbol{x})$, $t = 1, ..., T$, are the order statistics of the portfolio returns $\boldsymbol{r}_t'\boldsymbol{x}$ and $K = \lfloor \alpha T \rfloor$. Given a fixed $\alpha \in (0,1)$ and a target portfolio return r_p, the sample mean-shortfall optimization problem can be stated as follows:

$$\begin{aligned} Z_{sample} = \text{minimize}_{\boldsymbol{x}} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} - \frac{1}{K}\sum_{i=1}^{K} r_{(i)}(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} = r_p, \quad \boldsymbol{x}'\boldsymbol{e} = 1. \end{aligned} \tag{18}$$

It is possible that there might be additional linear constraints of the form $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ present in the problem, for example, nonnegativity constraints $\boldsymbol{x} \geq \boldsymbol{0}$. It is not obvious how Problem (18) might be solved because of the presence of the order statistics $r_{(i)}(\boldsymbol{x})$.

Note, however that $\displaystyle\sum_{i=1}^{t} r_{(i)}(\boldsymbol{x}) \leq \sum_{i \in S} \boldsymbol{r}_i'\boldsymbol{x}, \qquad S: \; |S| = t, \; t = 1, \ldots, T$. Therefore, Problem (18) can be reformulated as follows:

$$\begin{aligned} Z_{sample} = \text{minimize}_{\boldsymbol{x}} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} - \frac{1}{K}\sum_{i=1}^{K} z_i \\ \text{subject to} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} = r_p, \quad \boldsymbol{x}'\boldsymbol{e} = 1 \\ & \sum_{i=1}^{t} z_i \leq \sum_{i \in S} \boldsymbol{r}_i'\boldsymbol{x}, \qquad S: \; |S| = t, \; t = 1, \ldots, T. \end{aligned} \tag{19}$$

The linear optimization problem (19) has T new variables z_i, $i = 1, \ldots, T$, but 2^T constraints. We will reformulate Problem (19) with only a linear number of variables and constraints. Uryasev and

Rockafellar [32], in the context of the conditional-value-at-risk, have independently derived the linear optimization problem outlined in Theorem 3, using different methods. Let $K = \lfloor \alpha T \rfloor$.

Theorem 3 *Problem (19) is equivalent to the linear optimization problem*

$$\begin{aligned} Z_{sample} = \text{minimize}_{\boldsymbol{x},t,\boldsymbol{z}} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} - t + \frac{1}{K}\sum_{i=1}^{T} z_i \\ \text{subject to} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} = r_p, \quad \boldsymbol{x}'\boldsymbol{e} = 1 \\ & z_i \geq t - \boldsymbol{x}'\boldsymbol{r}_i, \quad z_i \geq 0, \quad i = 1, \dots, T. \end{aligned} \tag{20}$$

Proof Given a vector $\boldsymbol{v}$ we first observe that the value of the linear optimization problem

$$\begin{aligned} \text{minimize}_{\boldsymbol{z}} \quad & \sum_{i=1}^{T} v_i z_i \\ \text{subject to} \quad & \sum_{i=1}^{T} z_i = K, \quad 0 \leq z_i \leq 1, \qquad i = 1, \dots, T \end{aligned} \tag{21}$$

is equal to the sum of the K smallest components of the vector $\boldsymbol{v}$, i.e., it is equal to $\sum_{i=1}^{K} v_{(i)}$. By strong duality, the optimal solution values of Problems (21) and (22) below are equal:

$$\begin{aligned} \text{maximize}_{t,\boldsymbol{y}} \quad & Kt + \sum_{i=1}^{T} y_i \\ \text{subject to} \quad & t + y_i \leq v_i, \quad y_i \leq 0, \quad i = 1, \dots, T. \end{aligned} \tag{22}$$

Therefore, Problem (19) can be formulated as follows:

$$\begin{aligned} Z_{sample} = \text{minimize}_{\boldsymbol{x}} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} - \frac{1}{K}\max_{t,\boldsymbol{y}}(Kt + \sum_{i=1}^{T} y_i) \\ \text{subject to} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} = r_p, \quad \boldsymbol{x}'\boldsymbol{e} = 1 \\ & t + y_i \leq \boldsymbol{x}'\boldsymbol{r}_i, \quad y_i \leq 0, \quad i = 1, \dots, T. \end{aligned}$$

Using the fact that $\max(\theta) = -\min(-\theta)$, we obtain:

$$\begin{aligned} Z_{sample} = \text{minimize}_{\boldsymbol{x},t,\boldsymbol{y}} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} - t - \frac{1}{K}\sum_{i=1}^{T} y_i \\ \text{subject to} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} = r_p, \quad \boldsymbol{x}'\boldsymbol{e} = 1 \\ & t + y_i \leq \boldsymbol{x}'\boldsymbol{r}_i, \quad y_i \leq 0,, \quad i = 1, \dots, T. \end{aligned}$$

Letting $z_i = -y_i$, we obtain Problem (20). $\square$

We next observe, as also observed independently by Uryasev and Rockafellar [32], that the representation (6) of shortfall also leads to the same reformulation (20).

From (6) and the Definition (3.2) of the function $\rho_\alpha(\cdot)$, we obtain that the sample mean-shortfall optimization problem (18) is equivalent to

$$\begin{aligned} Z_{sample} = \text{minimize}_{\boldsymbol{x},t} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} - t + \frac{1}{\alpha T}\sum_{i=1}^{T}(t - \boldsymbol{x}'\boldsymbol{r}_i)^+ \\ \text{subject to} \quad & \boldsymbol{x}'\overline{\boldsymbol{r}} = r_p, \quad \boldsymbol{x}'\boldsymbol{e} = 1, \end{aligned}$$

where $(z)^+ := \max(0, z)$. This can be rewritten as linear optimization problem Formulation (20), which is a linear optimization problem with only $n+T+1$ variables and $2T+2$ constraints and can be solved by classical linear optimization approaches (the simplex method and interior point methods) very efficiently both theoretically (in polynomial time) as well as practically for large number of variables.

5.1 Solving mean-variance portfolio selection as a linear optimization problem

In this section, we explore the idea of solving a mean-standard deviation optimization problem as a linear optimization problem.

Consider the mean-standard deviation optimization problem

$$\begin{aligned} Z_{QP} = \text{minimize} \quad & (\boldsymbol{x'\Sigma x})^{1/2} \\ \text{subject to} \quad & \boldsymbol{Ax} \leq \boldsymbol{b}, \end{aligned} \tag{23}$$

where the vectors $\boldsymbol{c}$, $\boldsymbol{b}$ and the matrices $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma}$ is positive semi-definite) are given.

In Eq. (2) we have established that when returns are normally distributed, then

$$(\boldsymbol{x'\Sigma x})^{1/2} = \frac{\alpha}{\phi(z_\alpha)} s_\alpha(\boldsymbol{x}).$$

Suppose we generate T vectors $\boldsymbol{r}_i$, $i = 1, \ldots, T$ each from a multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$. A nonparametric estimator of $s_\alpha(\boldsymbol{x})$ [see Eq. (10)] is

$$\hat{s}_\alpha(\boldsymbol{x}) = \boldsymbol{x}'\overline{\boldsymbol{r}} - \frac{1}{K}\sum_{j=1}^{K} r_{(j)}(\boldsymbol{x})$$

where $K = \lfloor \alpha T \rfloor$. Combining the previous two equations, we obtain that an estimator of standard deviation is given by:

$$(\boldsymbol{x}'\hat{\boldsymbol{\Sigma}}\boldsymbol{x})^{1/2} = \frac{\alpha}{\phi(z_\alpha)} \left(\boldsymbol{x}'\overline{\boldsymbol{r}} - \frac{1}{K}\sum_{j=1}^{K} r_{(j)}(\boldsymbol{x}) \right). \tag{24}$$

In Eq. (20) we expressed shortfall as a linear optimization problem. Translating the analysis to the expression (24), we obtain that Problem (23) can be solved as

$$
\begin{aligned}
Z_{QP} = \lim_{T\to\infty} \text{minimize}_{\boldsymbol{x},t,\boldsymbol{z}} \quad & \frac{\alpha}{\phi(z_\alpha)} \left(\boldsymbol{x}'\overline{\boldsymbol{r}} - t + \frac{1}{K}\sum_{i=1}^{T} z_i \right) \\
\text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \quad z_i \ge t - \boldsymbol{x}'\boldsymbol{r}_i, \quad z_i \ge 0, \quad i = 1, \ldots, T.
\end{aligned} \tag{25}
$$

Problem (25) is a linear optimization problem. Thus, we are able to solve the mean-standard deviation nonlinear optimization problem by obtaining T samples (each of dimension N) from a multivariate normal distribution $N(0, \boldsymbol{\Sigma})$, and then solving the linear optimization problem (25). For finite T this is of course only an approximation. We perceive, however, some practical advantages in solving a linear optimization problem rather than a nonlinear one:
(a) The matrix $\boldsymbol{\Sigma}$ is estimated from data, which is a nontrivial problem that has an large literature. The proposed approach eliminates the need to estimate the matrix $\boldsymbol{\Sigma}$ altogether as it works directly with the data $\boldsymbol{r}_i$, and unifies the estimation and the optimization problem in a single problem; **(b)** Given that there are decades of experience in solving linear optimization problems, the numerical stability of linear optimization codes is definitely stronger compared to quadratic ones; **(c)** Mean-variance optimization problems in the presence of cardinality constraints, like for example the problem of tracking an index using only a limited number of assets, can be solved using standard linear mixed integer programming (MIP) methods (see Section 6.4) as opposed to quadratic mixed integer programming methods. While there are no commercially available quadratic mixed integer programming codes, there are many excellent linear MIP codes. Thus, mean-variance optimization problems in the presence of cardinality constraints can be solved using a well established methodology.

6 Numerical Examples

Our goal in this section is to shed light to the questions: **(a)** *How different are the allocations produced by the mean-variance and mean-shortfall optimization under varying degree of asymmetry of the distribution of returns?* and **(b)** *How viable and effective is the idea that we can solve mean-variance quadratic optimization problems as linear optimization problems and in particular, in the presence of cardinality constraints?*

We address the first question for symmetric return distributions and simulated data in Section 6.1, asymmetric return distributions and simulated data in Section 6.2, and real data in an asset allocation context in Section 6.3. We address the second question in Section 6.4.

6.1 Comparing mean-variance and shortfall optimization under symmetric distributions

We consider in this section three assets A, B, and C with a multivariate normal distribution having mean vector $\boldsymbol{\mu} = [8\%, 9\%, 12\%]'$, standard deviations $\boldsymbol{\sigma} = [15\%, 20\%, 22\%]'$ and correlation matrix

$$\boldsymbol{Cor} = \begin{bmatrix} 1 & 0.5 & 0.7 \\ 0.5 & 1 & -0.2 \\ 0.7 & -0.2 & 1 \end{bmatrix}.$$

We repeat the following experiment 100 times: **(a)** We generate samples with $T = 50, 500$, and 2000 observations from the multivariate normal distribution described above; **(b)** We apply mean-variance and mean-shortfall optimization on each sample for values of α varying between 2% and 50%, with a target rate of return $r_p = 10\%$, in the presence of a riskless asset with rate of return $r_f = 2.5\%$, and with no further constraints on the weights (note that the population optimal portfolio, for both mean-variance and mean-shortfall, by virtue of the multivariate normality of the assets, is $\boldsymbol{x}^* = (-1.41, 0.88, 1.00)'$.

For each value of T, we plot the median weight, over 100 experiments, assigned to each asset. We also plot the 10% and 90% percentile weights to give intuition about the variability of the weight estimates in Figure 1 for $T = 50$, 500 and 2000. The following insights emerge from this experiment: **(a)** As predicted by theory mean-variance and mean-shortfall optimization yield portfolio weights that are almost indistinguishable as soon as α is above say 15%. For lower values of α, the mean-shortfall estimates appear to be more volatile than the mean-variance estimates, a reflection of the fact that the mean-shortfall portfolio estimator uses only a small fraction of the data, the $\lfloor \alpha T \rfloor$ lowest order statistics of the sample; **(b)** As T increases, the variability of both the mean-variance and mean-shortfall estimates decreases.

6.2 Comparing mean-variance and shortfall optimization under asymmetric distributions

In this section, we compare the allocations and the the risk-sensitivity of portfolios given by mean-variance and mean-shortfall optimization when return distributions are asymmetric.

Data

We generate return data for following three assets. Asset A has a lognormal return distribution. Asset B consists of a stock with a lognormal distribution, combined with a call on 75% of the value of the

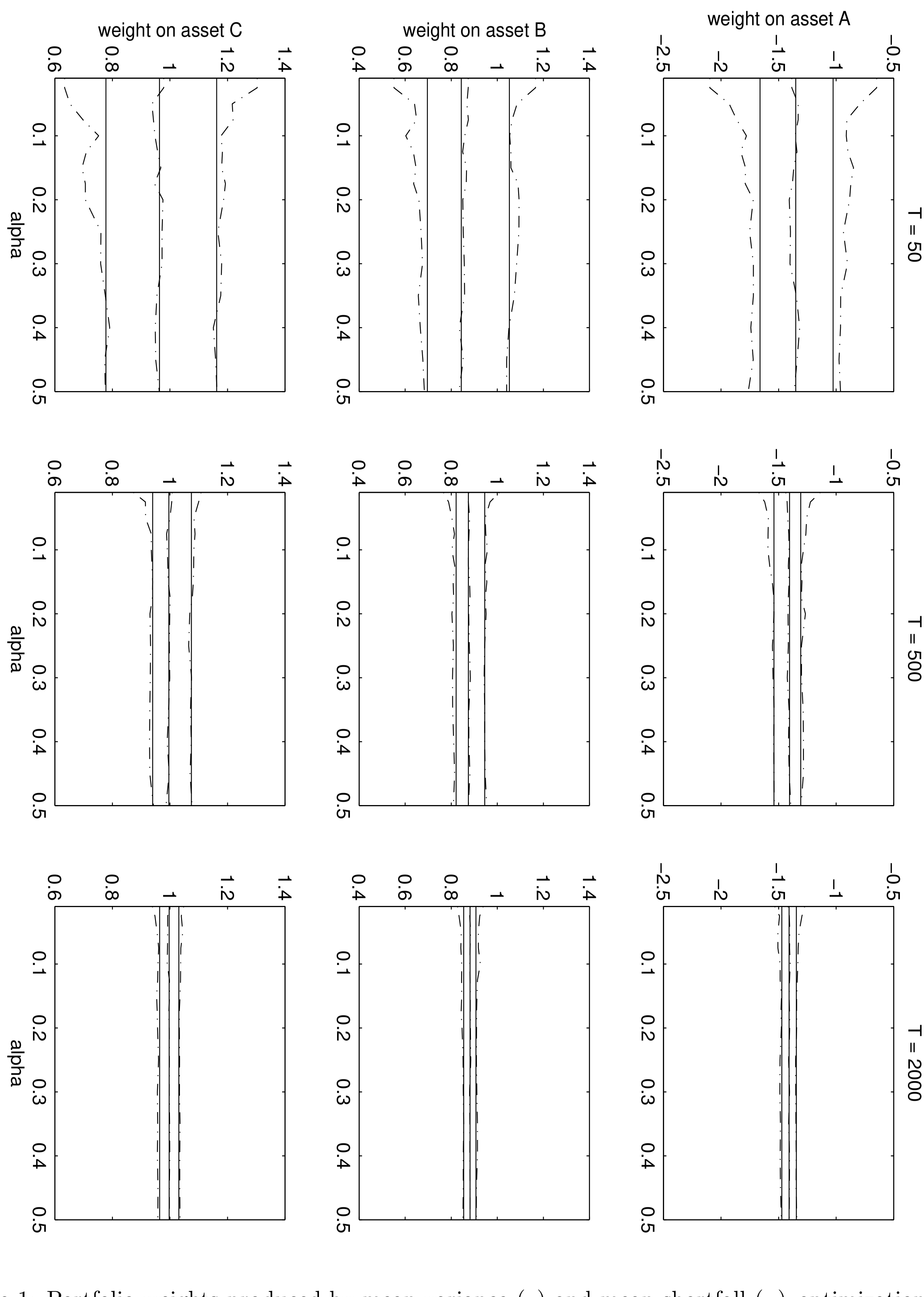

Figure 1: Portfolio weights produced by mean-variance (–) and mean-shortfall (-.) optimization. The median, 10% and 90% percentiles over 100 samples are plotted using the same symbol.

stock, financed by borrowing at a riskless rate $r_f = 2.5\%$. Thus, Asset B has a return distribution that is skewed to the left. Asset C consists of a stock with a lognormal distribution, combined with a put on 75% of the value of the stock, financed by borrowing at a riskless rate $r_f = 2.5\%$. Thus, Asset C has a return distribution that is skewed to the right. The assets are designed to have the same mean and standard deviation, and to be uncorrelated with each other, which will make mean-variance optimization blind to their differences, so that it leads to an equiweighted portfolio.

The price of the call and put options, used to calculate the returns of those options, were determined using the classical Black-Scholes formula, assuming a maturity of one period, and a strike price equal to the price of the asset. In each sample that we use in our experiments, the mean and standard deviation of each asset are standardized to be 8% and 20% respectively.

Shortfall and Shortfall Beta of Fixed Weight Portfolios

In order to obtain some insight on how shortfall can be influenced by different portfolios, we examine the shortfall of three different fixed weight portfolios: $\boldsymbol{x_1} = (1/3, 1/3, 1/3)'$, $\boldsymbol{x_2} = (0.1, 0.8, 0.1)'$ and $\boldsymbol{x_3} = (0.1, 0.1, 0.8)'$ for different values of α. We repeat the following experiment 100 times: **(a)** We generate a sample of $T = 2000$ observations from the asymmetric multivariate distribution described above; **(b)** We calculate the shortfall of each portfolio $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, and $\boldsymbol{x_3}$, and the shortfall beta coefficient of each asset with respect to each portfolio.

In Figure 2, we plot the median shortfall, over the 100 experiments, as a function of α between 2% and 50%, for each of the three portfolios. We also plot the 10% and 90% quantiles, over the 100 experiments, to give an idea about the variability of the shortfall estimates. As expected, portfolio $\boldsymbol{x_2}$ has a higher shortfall than portfolio $\boldsymbol{x_3}$ for values of α below 40%, reflecting that fact that portfolio $\boldsymbol{x_2}$ is highly loaded on the negatively skewed asset B, whereas portfolio $\boldsymbol{x_3}$ is heavily loaded on the positively skewed asset C. Note that portfolio $\boldsymbol{x_1}$, the equally weighted portfolio, has the lowest shortfall of all portfolios, at every value of α, a clear reflection of the power of diversification.

In Figure 3, we report the shortfall beta of each asset with respect to portfolio $\boldsymbol{x_1}$. For portfolio $\boldsymbol{x_1}$ and low values of α, Asset B has the highest shortfall beta, indicating asset B is responsible for most of the shortfall of the portfolio. Asset C has the smallest shortfall beta. For portfolio $\boldsymbol{x_1}$ and high values of α, all assets have shortfall beta about one, indicating comparable contributions to the portfolio's shortfall. The message is that contrary to the standard beta, the shortfall beta can vary with α, indicating that an asset can have different contributions to the risk of a portfolios at different values of α.

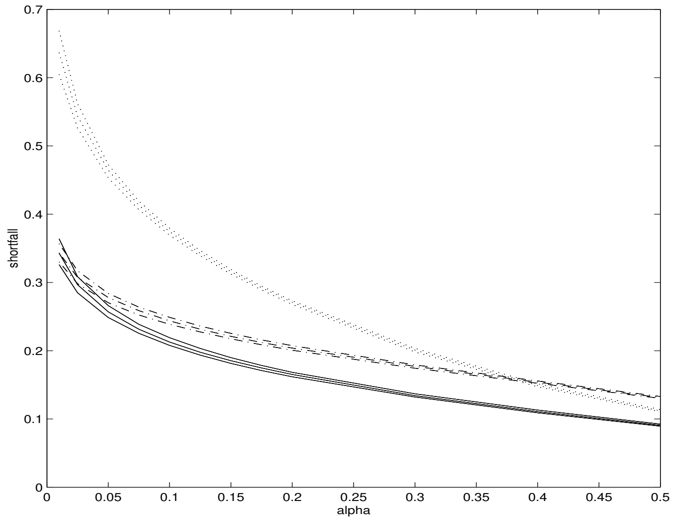

Figure 2: Shortfall of portfolios x_1 (—), x_2 ($\cdots$), and x_3 (-.-.). For each portfolio and α-level combination, the 10%, 50%, and 90% quantiles over 100 samples are represented.

Weights given by Mean-Variance and Mean-Shortfall Optimization

We next compare the portfolio weights obtained via mean-variance and mean-shortfall optimization on samples from an asymmetric distribution. We repeat the following experiment 100 times: **(a)** We generate a sample of $T = 2000$ observations from the asymmetric multivariate distribution described above; **(b)** We apply mean-variance optimization and mean-shortfall optimization for values of α between 2% and 50%. We use a target rate of return of $r_p = 8\%$, and constrain the weights to be non-negative.

Figure 4 shows the cumulative distribution function of returns for the optimal mean-variance (MV) portfolio, the optimal mean-shortfall ($s_{0.01}$) at $\alpha = 1\%$, and the optimal mean-shortfall ($s_{0.10}$) at $\alpha = 10\%$ on one sample with 2000 observations. It is clear that the shortfall portfolios dominate in the tails, as expected, but the MV portfolio dominate in the mid-range (+/- 10%).

Figure 5 gives the weights assigned to each asset, for α ranging from 2% to 50%. We see that mean-variance optimization (which is independent of α) gives equal weight to each asset, as expected. Mean-shortfall optimization, especially for low levels of α, puts less weight on asset B, and extra weight on asset C, also as expected. The weight assigned to asset A seems to be roughly the same for both mean-variance and mean-shortfall optimization.

In summary, the example in this section clearly indicates that allocations based on mean-shortfall optimization differ and often significantly from those based on the mean-variance paradigm and may depend on the risk level α. Furthermore, the risk-sensitivity of any given portfolio may significantly

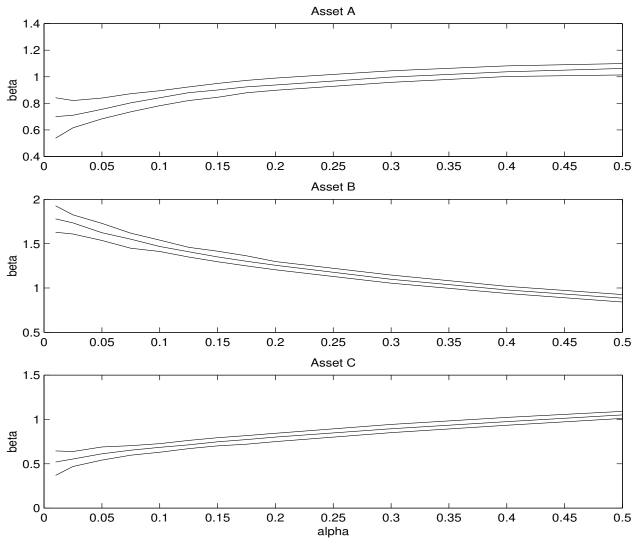

Figure 3: Shortfall beta of each asset with respect to portfolio x_1. For each asset, and α-level combination, the 10%, 50%, and 90% quantiles over 100 samples are represented.

vary with the risk level α.

6.3 Comparing mean-variance and shortfall optimization in asset allocation

In this section, we compare asset allocations computed using mean-variance and mean-shortfall optimization. The data consist of monthly returns on seventeen asset classes: six indices involving US equities (large cap, large cap value companies, large cap growth companies, small cap, small cap value companies, small cap growth companies), the corresponding six indices involving international equities in developed markets, emerging equities, US goverment bonds, international goverment bonds, US treasury, and the US real estate index trust (REIT). The time period under consideration is January, 1994 to September, 2001.

We estimate the historical covariance matrix and use it to solve Problem (12) with nonnegativity constraints on the allocations in order to find the efficient frontier of the mean-variance portfolio. We record the shortfall that this portfolio produces as well. We solve Problem (11) with nonnegativity constraints on the allocations in order to find the efficient frontier of the mean-shortfall portfolio. We record the standard deviation of this portfolio as well.

In Figures 6, and 7 we present the efficient frontiers in mean annual return-annual standard deviation and mean annual return-annual shortfall space for both methods for $\alpha = 16.7\%$ and $\alpha = 2.5\%$. It is clear that for $\alpha = 16.7\%$, the two methods provide almost identical efficient frontiers. As expected, for

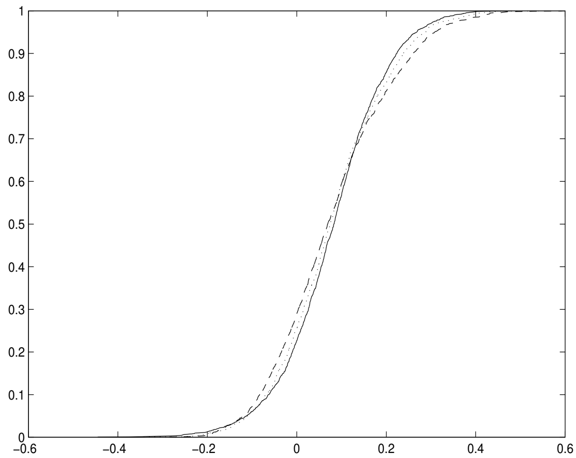

Figure 4: Cumulative distribution of MV (—), $s_{0.10}$ ($\cdots$), and $s_{0.01}$ (–.), sample of 2000 observations.

$\alpha = 2.5\%$, minimizing variance produces a slightly better frontier in mean-variance space and slightly worse in mean-shortfall space than minimizing shortfall. Moreover, it turns out that the allocations of the two portfolios are quite similar. We feel that the closeness of the two optimal portfolios is an indication that the joint return of these asset classes obeys a joint multivariate normal distribution.

6.4 Solving mean-variance portfolio selection with cardinality constraints

In this section, we consider a universe of n different assets. We would like to construct a portfolio of at most $M < n$ assets that "is close to" a given benchmark. Let $\mathbf{R}$ be the vector of returns, which has a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. We consider an equally weighted benchmark, i.e., the allocation $\boldsymbol{x_B} = \boldsymbol{e}/n$, where $\boldsymbol{e}$ is the vector of all ones.

The tracking error of a portfolio $\boldsymbol{x}$ is given by $E\left[(\boldsymbol{x'R} - \boldsymbol{x'_B R})^2\right] = (\boldsymbol{x} - \boldsymbol{x_B})'\boldsymbol{\Sigma}^*(\boldsymbol{x} - \boldsymbol{x_B})$, where $\boldsymbol{\Sigma}^* = \boldsymbol{\mu\mu'} + \boldsymbol{\Sigma}$. We are interested in selecting a portfolio $\boldsymbol{x}$ of at most M assets that minimizes the tracking error. Moreover, the nonzero weights of all assets in the portfolio $\boldsymbol{x}$ are constrained to be in the interval $[a, b]$.

We introduce decision variables y_i, which is equal to one, if asset i is in the portfolio, and zero, otherwise. The problem of minimizing the tracking error under the cardinality constraint is:

$$\begin{array}{ll} \text{minimize} & (\boldsymbol{x} - \boldsymbol{x_B})'\boldsymbol{\Sigma}^*(\boldsymbol{x} - \boldsymbol{x_B}) \\ \text{subject to} & \mathbf{e}'\mathbf{x} = 1, \quad \mathbf{e}'\mathbf{y} \le M, \quad \mathbf{x} \ge \mathbf{0} \\ & a y_i \le x_i \le b y_i, \quad y_i \in \{0,1\}, \quad i = 1, \ldots, n. \end{array} \tag{26}$$

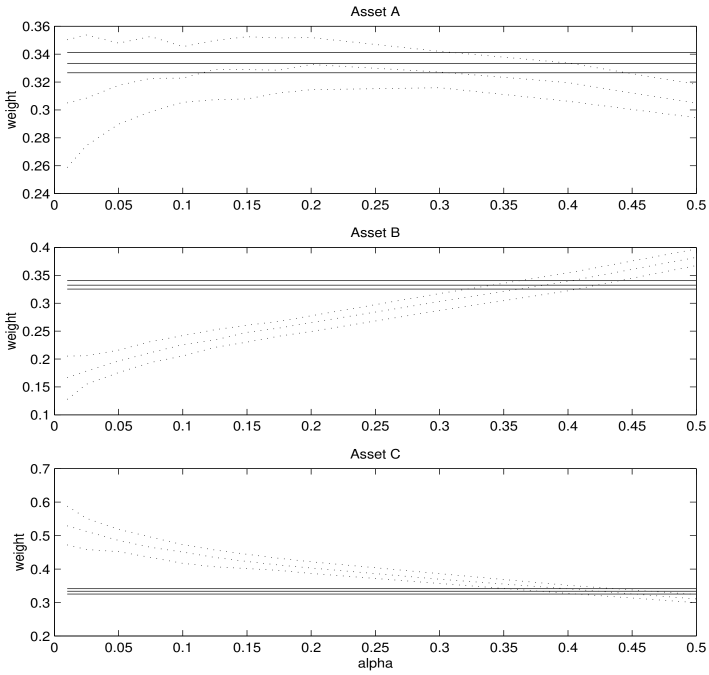

Figure 5: Weights for each asset: MV (—), s_α ($\cdots$). For each portfolio, asset, and α-level combination, the 10%, 50%, and 90% quantiles over 100 samples are plotted.

Problem (26) is a mixed integer quadratic program, for which there is no commercially available software package. Note, however, that by using the equivalence between shortfall and standard deviation when the underlying return distributions are multivariate normal, we can rewrite the tracking error (or rather its square root) as $\sqrt{(\boldsymbol{x}-\boldsymbol{x_B})'\boldsymbol{\Sigma}^*(\boldsymbol{x}-\boldsymbol{x_B})} = \frac{\alpha}{\phi(z_\alpha)} s_\alpha((\boldsymbol{x}-\boldsymbol{x_B})'\boldsymbol{R})$. This last fact suggests a new approach to solving Problem (26): generate samples $\boldsymbol{r}_j$, $j = 1, \ldots, T$ from the distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and solve Problem (26) (see also Eq. (25)) as the following linear mixed integer programming problem:

$$
\begin{aligned}
\text{minimize} \quad & \frac{\alpha}{\phi(z_\alpha)} \left(-t + \frac{1}{K} \sum_{j=1}^{T} z_j \right) \\
\text{subject to} \quad & \mathbf{e}'\mathbf{x} = 1, \quad \mathbf{e}'\mathbf{y} \le M, \quad \mathbf{x}, \boldsymbol{z} \ge \mathbf{0} \\
& a y_i \le x_i \le b y_i, \quad y_i \in \{0, 1\}, \quad i = 1, \ldots, n, \\
& z_j \ge t - (\boldsymbol{x} - \boldsymbol{x_B})'\boldsymbol{r}_j, \quad j = 1, \ldots, T.
\end{aligned}
\tag{27}
$$

Problem (27) is a mixed linear integer programming problem for which there is a large literature as well as several commercially available codes. In order to illustrate this method we used an example of tracking the equally weighted SP100 index with M stocks for $M = 96$, 90, $80, \ldots, 40$.

M	96	90	80	70	60	50	40
$T = 100$	0.0%	0.46%	0.54%	0.66%	0.81%	0.90%	1.1%
$T = 200$	0.0%	0.38%	0.48%	0.63%	0.72%	0.93%	1.1%
$T = 500$	0.0%	0.29%	0.46%	0.59%	0.73%	0.79%	1.0%
$T = 1000$	0.0%	0.27%	0.37%	0.54%	0.60%	0.78%	0.90%

Table 3: Tracking Error (per month) as a function of M and T.

Data

For our experiment, we calculated a mean vector $\overline{\boldsymbol{r}}$ and a covariance matrix $\boldsymbol{\Sigma}$ by using the daily return data on 96 stocks, for the period January 02, 1996, to December 31, 1999. The 96 stocks were selected using the following criteria: they were in the SP100 on December 31, 1999, and had daily return data for the whole four year period under consideration. Out of the 100 stocks in the SP100 on December 31, 1999, the following four stocks did not have daily return data for the entire four year period under consideration: Honeywell Inc. (CRSP Permanent Number 18374), Lucent Technologies Inc. (83332), Rockwell International Corp. (84381), and Raytheon Corp. (85658). The data come from the CRSP database, and were obtained using the Wharton Research Database Service (WRDS). The daily mean vector and covariance matrix calculated from the daily data were each multiplied by 21 to obtain an estimate $\overline{\boldsymbol{r}}$ of the monthly mean vector and an estimate $\boldsymbol{\Sigma}$ of the covariance matrix of the 96 stocks.

Results

We generated a random sample of size T for $T = 100$, 200, 500, 1000 using the historical mean vector $\overline{\boldsymbol{r}}$ and covariance matrix $\boldsymbol{\Sigma}$ described previously. We chose $a = 0.3\%$ and $b = 10\%$. We solved Problem (27) using a state of the art optimization software CPLEX, and we report in Table 3 the tracking error and in Table 4 the running time as a function of M for T. We run the integer programming algorithm until five feasible solutions have been generated. The reason for this is that running times for exact optimality can be excessive. Moreover, the best solution found among the first five solutions found is either the best or very close to the best solution.

The following insights emerge from this experiment: **(a)** The proposed approach successfully solves the problem of minimizing the tracking error with cardinality constraints within reasonable computational times. As expected the tracking error increases as M decreases, since it becomes increasingly more difficult to track the index with a smaller number of stocks; **(b)** The tracking error converges to the

M	96	90	80	70	60	50	40
$T = 100$	0.7	1.7	3.2	5.0	7.4	9.3	12.5
$T = 200$	0.8	7.0	11.7	18.0	21.8	27.1	32.4
$T = 500$	0.9	32.8	59.5	81.6	102.7	121.9	135.8
$T = 1000$	1.2	128.2	206.6	313.3	372.9	436.4	519.3

Table 4: Running time (in seconds) as a function of M and T.

solution of the mean-variance optimization with cardinality constraints as T increases; **(c)** The running times are monotonically increasing as M decreases and T increases. This is expected as problems with smaller M are harder, while problems with larger T have simply more variables.

7 Conclusions

We have shown that shortfall naturally arises as a measure of risk by considering distributional conditions for second-order stochastic dominance. We examined its properties and its connections with other risk measures. We showed that optimization of shortfall leads to a tractable convex optimization problem and to a linear optimization problem in its sample version. Interestingly, portfolio separation theorems as well natural definitions of beta can be derived in direct analogy to standard mean-variance portfolio optimization theory. We showed computationally that the mean-shortfall approach generates portfolios that can outperform those generated by the mean-variance approach. Finally, we showed that the mean-shortfall approach can readily address portfolio optimization problems with cardinality constraints. All these considerations convince us that we should consider more closely the notion of shortfall in real world environments.

Acknowledgements

We thank Chris Darnell for interesting discussions and providing us data for the asset allocation experiment reported in Section 6.3 and Stu Rosenthal for assistance with the computations in this section. We thank Roy Welsch for useful discussions and support. We thank the reviewers of the paper for several insightful comments. This research was partially supported by NSF grants DMS-9626348, DMS-9971579, DMI-9610486, and grants from Merill Lynch and General Motors.

References

[1] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[2] V. S. Bawa. Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2:95–121, 1975.

[3] V. S. Bawa. Safety-first, stochastic dominance, and optimal portfolio choice. *Journal of Financial and Quantitative Analysis*, 13:255–271, 1978.

[4] V. S. Bawa and E. B. Lindenberg. Capital market equilibrium in a mean-lower partial moment framework. *Journal of Financial Economics*, 5:189–200, 1977.

[5] D. Bertsimas and I. Popescu. Optimal inequalities in probability: A convex programming approach. Working paper, Operations Research Center, MIT, 1999.

[6] R. Bookstaber and R. Clarke. Option portfolio strategies: Measurement and evaluation. *Journal of Business*, 57(4):469–492, 1984.

[7] G. Chamberlain. A characterization of the distribution that imply mean-variance utility functions. *Journal of Economic Theory*, 29:185–201, 1983.

[8] F. Delbaen. Coherent risk measures on general probabbility spaces. Technical Report, ETH Zurich., 2000.

[9] K. Dowd. *Beyond Value at Risk*. John Wiley and Sons, Ltd., 1998.

[10] D. Duffie and J. Pan. An overview of value at risk. *The Journal of Derivatives*, pages 7–49, 1997.

[11] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependency in risk management: Properties and pitfalls. Working paper, Risklab, ETH, available at http://www.gloriamundi.org., 1999.

[12] P. C. Fishburn. Mean-risk analysis with risk associated with below-target returns. *The American Economic Review*, 67(2):116–126, 1977.

[13] M. Garman. Taking var to pieces. *Risk*, (10):70–71, 1997.

[14] H. Grootveld and Hallerbach W. Variance vs. downside risk: Is there really that much difference? *European Journal of Operations Research*, 114:304–319, 1999.

[15] Y. Hamao, R. Masulis, and V. Ng. Correlations in price changes and volatility across international stock markets. *The Review of Financial Studies*, 3:282–307, 1990.

[16] W. V. Harlow and R. Rao. Asset pricing in a generalized mean-lower partial moment framework: Theory and evidence. *The Journal of Financial and Quantitative Analysis*, 24:285–311, 1999.

[17] C.F. Huang and R. H. Litzenberger. *Foundations for Financial Economics*. Prentice Hall, 1988.

[18] J. Ingersoll, Jr. *Theory of Financial Decision Making*. Rowman and Littlefield Publishers, Inc., 1987.

[19] P. Jorion. *Value at Risk*. McGraw-Hill, 1997.

[20] M. King and S. Wadhwani. Transmission of volatility between markets. *The Reveiw of Financial Studies*, 3:5–35, 1990.

[21] R. W. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.

[22] G. Lemus, A. Samarov, and R. Welsch. Portfolio analysis based on value-at-risk. In *Proceedings of the 52nd Session of ISI*, volume 3, pages 221–222, 1999.

[23] H. Levy. Stochastic dominance and expected utility: Survey and analysis. *Management Science*, 38(4):555–593, 1992.

[24] H. Levy and Y.. Kroll. Ordering uncertain options with borrowing and lending. *The Journal of Finance*, 31(2), 1978.

[25] H. M. Markowitz. *Portfolio Selection*. Wiley, 1959.

[26] R. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, 1982.

[27] H. Neelakandan. Volatility-correlation dynamics in financial markets. Master's thesis, Sloan School of Management, MIT, 1994.

[28] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[29] A.D. Roy. Safety first and the holding of assets. *Econometrica*, 20:431–449, 1952.

[30] O. Scaillet. Nonparametric estimation and sensitivity analysis of expected shortfall. available at http://www.gloriamundi.org., 2000.

[31] D. Tasche. Risk contributions and performance measurement. Working paper, TU Munich, available at http://www.gloriamundi.org., 2000.

[32] S. Uryasev and R. T. Rockafellar. Optimization of conditional value-at-risk. Report 99-4, available at http://www.gloriamundi.org., 1999.

Figure 6: Efficient frontiers for optimal portfolios obtained by minimizing variance and minimizing shortfall for $\alpha = 16.7\%$.

Figure 7: Efficient frontiers for optimal portfolios obtained by minimizing variance and minimizing shortfall for $\alpha = 2.5\%$.