

# Frontiers of Stochastically Nondominated Portfolios\*

Andrzej Ruszczyński<sup>†</sup> and Robert J. Vanderbei<sup>‡</sup>

April 17, 2002

## Abstract

We consider the problem of constructing a portfolio of finitely many assets whose returns are described by a discrete joint distribution. We propose mean–risk models which are solvable by linear programming and generate portfolios whose returns are nondominated in the sense of second-order stochastic dominance. Next, we develop a specialized parametric method for recovering the entire mean–risk efficient frontiers of these models and we illustrate its operation on a large data set involving thousands of assets and realizations.

*Keywords:* Portfolio optimization, stochastic dominance, mean–risk analysis, least absolute deviations, Fenchel duality, linear programming, parametric simplex method, robust statistics.

## 1 Introduction

The problem of optimizing a portfolio of finitely many assets is a classical problem in theoretical and computational finance. Since the seminal work of Markowitz [17, 19] it is generally agreed that portfolio performance should be measured in two distinct dimensions: the *mean* describing the expected return, and the *risk* which measures the uncertainty of the return. In the mean–risk approach, we select from the universe of all possible portfolios those that are *efficient*: for a given value of the mean they minimize the risk or, equivalently, for a given value of risk they maximize the mean. Such an approach has many advantages: it allows one to formulate the problem as a parametric optimization problem, and it facilitates the trade-off analysis between mean and risk.

---

\*submitted for publication.

<sup>†</sup>Rutgers University, Department of Management Science and Information Systems and RUTCOR, Piscataway, NJ 08854, USA ([rusz@rutcor.rutgers.edu](mailto:rusz@rutcor.rutgers.edu)).

<sup>‡</sup>Princeton University, Department of Operations Research and Financial Engineering, and Bendheim Center for Finance, Princeton, NJ 08544, USA ([rvdb@princeton.edu](mailto:rvdb@princeton.edu)).

Markowitz used the variance of the return as the measure of the risk. It is easy to compute, and it reduces the portfolio selection problem to a parametric quadratic programming problem. One can, however, construct simple counterexamples that show the imperfection of the variance as the risk measure: it treats over-performance equally as under-performance, and more importantly its use may suggest a portfolio which is always outperformed by another portfolio. The use of the semivariance rather than the variance was already recommended by Markowitz himself [18]. But even in this case significant deficiencies remain (see [22, 23]) as we shall explain.

Another theoretical approach to the portfolio selection problem is that of *stochastic dominance* (see [30, 15]). The usual (first order) definition of stochastic dominance gives a partial order in the space of real random variables. More important from the portfolio point of view is the notion of second-order dominance which is also defined as a partial order but which is equivalent to this statement: a random variable  $Y$  dominates the random variable  $Z$  if  $\mathbb{E}[U(Y)] \geq \mathbb{E}[U(Z)]$  for all nondecreasing concave functions  $U(\cdot)$  for which these expected values are finite. Thus, no risk-averse decision maker will prefer a portfolio with return  $Z$  over a portfolio with return  $Y$ . While theoretically attractive, stochastic dominance order is computationally very difficult, as a multi-objective model with a continuum of objectives.

We shall, therefore, concentrate on mean–risk portfolio models, but we shall look for such models whose efficient frontiers consist of stochastically nondominated solutions, at least above a certain modest level of mean return.

The general question of constructing mean–risk models which are in harmony with the stochastic dominance relations has been the subject of the analysis of the recent papers [22, 23, 24]. We shall apply and specialize some of the results obtained there to the portfolio optimization problem. We shall show that the resulting mean–risk models can be formulated as linear programming problems. In this sense, our work has a different motivation than the classical paper by Sharpe [28], where he develops a linear programming approximation to the mean–variance model. We do not want to approximate the mean–variance model, but rather to construct a linear programming model that has better theoretical properties than the mean–variance model and its approximations. Moreover, we develop a highly effective algorithm for recovering the entire efficient frontiers of our models. Our numerical results show that our approach is capable of solving portfolio problems of large sizes in a reasonable time. This, combined with the theoretical property of stochastic efficiency of the solutions obtained constitutes a strong argument for the use of our models in practical portfolio optimization.

## 2 The mean–risk portfolio problem

Let  $R_1, R_2, \dots, R_n$  be random returns of assets  $1, 2, \dots, n$ . We assume that the returns have a discrete joint distribution with realizations  $r_{jt}$ ,  $t = 1, \dots, T$ ,  $j = 1, \dots, n$ , attained

with probabilities  $p_t$ ,  $t = 1, 2, \dots, T$ . Our aim is to invest our capital in these assets in order to obtain some desirable characteristics of the total return on the investment. Denoting by  $x_1, x_2, \dots, x_n$  the fractions of the initial capital invested in assets  $1, 2, \dots, n$  we can easily derive the formula for the total return:

$$R(x) = R_1x_1 + R_2x_2 + \dots + R_nx_n. \quad (1)$$

Clearly, the set of possible asset allocations can be defined as follows:

$$X = \{x \in \mathbb{R}^n : x_1 + x_2 + \dots + x_n = 1, x_j \geq 0, j = 1, 2, \dots, n\}.$$

Our analysis will not depend on the detailed way this set is defined; we shall only use the fact that it is a convex polyhedron. So, in some applications one may introduce the possibility of *short positions*, i.e., allow some  $x_j$ 's to become negative. One can also limit the exposure to particular assets or their groups, by imposing upper bounds on the  $x_j$ 's or on their partial sums. One can also limit the absolute differences between the  $x_j$ 's and some reference investments  $\bar{x}_j$ , which may represent the existing portfolio, etc. All these modifications define some convex polyhedral feasible sets, and are, therefore, covered by our approach.

With each portfolio allocation  $x$  we can associate the mean return

$$\mu(x) = \mathbb{E}[R(x)] = \sum_{j=1}^n \sum_{t=1}^T r_{jt}x_jp_t$$

and some risk measure  $\rho(x)$  representing the variability of the return  $R(x)$ . At this moment we may think of  $\rho(x)$  being the variance of the return, although later we shall work with other risk measures which, as we shall argue, are superior to the variance.

The mean–risk portfolio optimization problem is formulated as follows:

$$\begin{aligned} & \max \mu(x) - \lambda\rho(x) \\ & \text{subject to } x \in X. \end{aligned} \quad (2)$$

Here,  $\lambda$  is a nonnegative parameter representing our desirable exchange rate of mean for risk. If  $\lambda = 0$ , the risk has no value and the problem reduces to the problem of maximizing the mean. If  $\lambda > 0$  we look for a compromise between the mean and the risk, at which further improvements  $\Delta$  in the mean are associated with increases in the risk of at least  $\Delta/\lambda$ , and a reduction  $\delta$  of the risk cannot be achieved without decreasing the mean by at least  $\lambda\delta$ . The results of the mean–risk analysis are usually depicted on the mean–risk graph, as illustrated in Figure 1.

Artzner *et al.* introduced in [3] the concept of *coherent risk measures* by means of several axioms. In our terminology their measures correspond to composite objectives of the form  $-\mu(x) + \lambda\rho(x)$ .

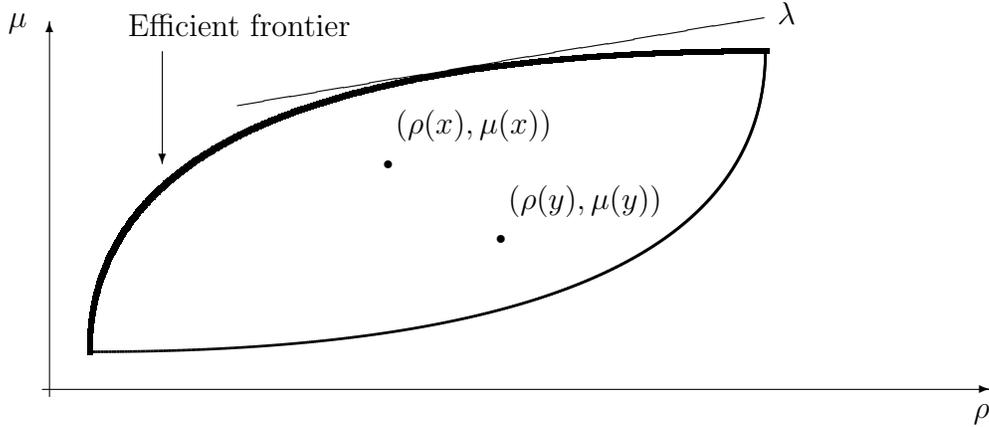


Figure 1: Mean–risk analysis. Portfolio  $x$  is better than portfolio  $y$  in the mean–risk sense, but none of them is efficient.

### 3 Consistency with stochastic dominance

The concept of *stochastic dominance* is related to an axiomatic model of risk-averse preferences [7]. It originated from the theory of majorization [12, 20] for the discrete case and was later extended to general distributions [25, 10, 11, 27]. It is nowadays widely used in economics and finance [4, 15].

In the stochastic dominance approach, random returns are compared by a point-wise comparison of some performance functions constructed from their distribution functions. For a real random variable  $V$ , its first performance function is defined as the right-continuous cumulative distribution function of  $V$ :

$$F_V(\eta) = \mathbb{P}\{V \leq \eta\} \quad \text{for } \eta \in \mathbb{R}.$$

A random return  $V$  is said [14, 25] to *stochastically dominate* another random return  $S$  to the first order, denoted  $V \succeq_{FSD} S$ , if

$$F_V(\eta) \leq F_S(\eta) \quad \text{for all } \eta \in \mathbb{R}.$$

The second performance function  $F^{(2)}$  is given by areas below the distribution function  $F$ ,

$$F_V^{(2)}(\eta) = \int_{-\infty}^{\eta} F_V(\xi) d\xi \quad \text{for } \eta \in \mathbb{R}, \quad (3)$$

and defines the weak relation of the *second-order stochastic dominance* (SSD). That is, random return  $V$  stochastically dominates  $S$  to the second order, denoted  $V \succeq_{SSD} S$ , if

$$F_V^{(2)}(\eta) \leq F_S^{(2)}(\eta) \quad \text{for all } \eta \in \mathbb{R} \quad (4)$$

(see [10, 11]). The corresponding strict dominance relations  $\succ_{FSD}$  and  $\succ_{SSD}$  are defined in the usual way

$$V \succ S \quad \Leftrightarrow \quad V \succeq S \quad \text{and} \quad S \not\succeq V. \quad (5)$$

For portfolios, the random variables in question are the returns defined by (1). To avoid placing the decision vector,  $x$ , in a subscript expression, we shall simply write

$$F(\eta; x) = F_{R(x)}(\eta) \quad \text{and} \quad F^{(2)}(\eta; x) = F_{R(x)}^{(2)}(\eta).$$

It will not lead to any confusion, we believe. Thus, we say that portfolio  $x$  *dominates* portfolio  $y$  *under the FSD rules*, if  $F(\eta; x) \leq F(\eta; y)$  for all  $\eta \in \mathbb{R}$ , where at least one strict inequality holds. Similarly, we say that  $x$  *dominates*  $y$  *under the SSD rules* ( $R(x) \succ_{SSD} R(y)$ ), if  $F^{(2)}(\eta; x) \leq F^{(2)}(\eta; y)$  for all  $\eta \in \mathbb{R}$ , with at least one inequality strict.

Stochastic dominance relations are of crucial importance for decision theory. It is known that  $R(x) \succeq_{FSD} R(y)$  if and only if  $\mathbb{E}[U(R(x))] \geq \mathbb{E}[U(R(y))]$  for any nondecreasing function  $U(\cdot)$  for which these expected values are finite. Also,  $R(x) \succeq_{SSD} R(y)$  if and only if  $\mathbb{E}[U(R(x))] \geq \mathbb{E}[U(R(y))]$  for every nondecreasing and concave  $U(\cdot)$  for which these expected values are finite (see, e.g., [15]).

For a set  $X$  of portfolios, a portfolio  $x \in X$  is called *SSD-efficient* (or *FSD-efficient*) in  $X$  if there is no  $y \in X$  such that  $R(y) \succ_{SSD} R(x)$  (or  $R(y) \succ_{FSD} R(x)$ ).

We shall focus our attention on the SSD relation, because of its consistency with risk-averse preferences: if  $R(x) \succ_{SSD} R(y)$ , then portfolio  $x$  is preferred to  $y$  by all risk-averse decision makers. By changing the order of integration we can express the function  $F^{(2)}(\cdot; x)$  as the expected shortfall [22]: for each target value  $\eta$  we have

$$F^{(2)}(\eta; x) = \mathbb{E}[\max(\eta - R(x), 0)]. \quad (6)$$

The function  $F^{(2)}(\cdot; x)$  is continuous, convex, nonnegative and nondecreasing. Its graph is illustrated in Figure 2.

Our main concern is the following: may the mean-risk efficient frontier, as illustrated in Figure 1, contain portfolios which are dominated in the SSD sense? It is unfortunately true for the mean-risk model using the variance as the risk measure (see, e.g. [31]).

Following [22, 23], we introduce the following definition.

**Definition 1** *The mean-risk model  $(\mu, \rho)$  is consistent with SSD with coefficient  $\alpha > 0$ , if the following relation is true*

$$R(x) \succeq_{SSD} R(y) \quad \Rightarrow \quad \mu(x) - \lambda\rho(x) \geq \mu(y) - \lambda\rho(y)$$

for all  $0 \leq \lambda \leq \alpha$ .

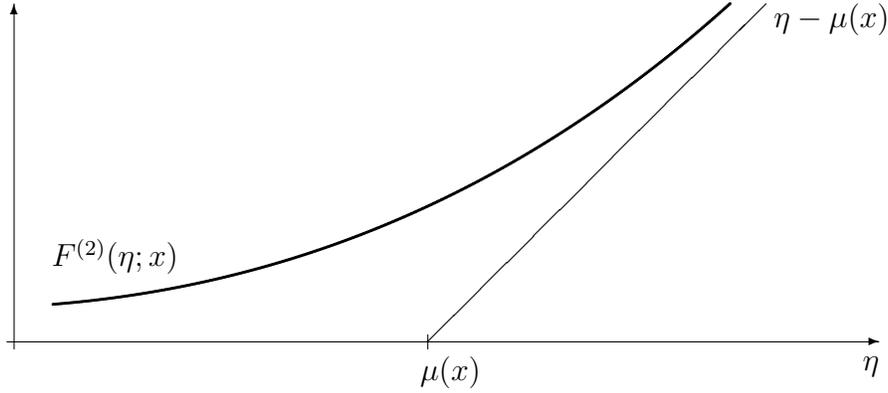


Figure 2: The expected shortfall function.

In fact, as we shall see in the proof below, it is sufficient to have the above inequality satisfied for  $\alpha$ ; its validity for all  $0 \leq \lambda \leq \alpha$  follows from that.

The concept of consistency turns out to be fruitful. In [22] we have proved the following result.

**Theorem 1** *The mean-risk model in which the risk is defined as the absolute semideviation,*

$$\bar{\delta}(x) = \mathbb{E} \{ \max(\mu(x) - R(x), 0) \}, \quad (7)$$

*is consistent with the second-order stochastic dominance relation with coefficient 1.*

We provide an easy alternative proof here.

**Proof.** First, it is clear from (6) that the line  $\eta - \mu(x)$  is the asymptote of  $F^{(2)}(\eta; x)$  for  $\eta \rightarrow \infty$ . Therefore  $R(x) \succeq_{SSD} R(y)$  implies that

$$\mu(x) \geq \mu(y). \quad (8)$$

Secondly, setting  $\eta = \mu(x)$  in (4) we obtain

$$\bar{\delta}(x) \leq \mathbb{E} [\max(0, \mu(x) - R(y))].$$

Since  $\mu(x) - \mu(y) \geq 0$ , we have

$$\begin{aligned} \max(0, \mu(x) - R(y)) &= \max(0, \mu(x) - \mu(y) + \mu(y) - R(y)) \\ &\leq \mu(x) - \mu(y) + \max(0, \mu(y) - R(y)). \end{aligned}$$

Taking the expected value of both sides and combining with the preceding inequality we get

$$\bar{\delta}(x) \leq \mu(x) - \mu(y) + \bar{\delta}(y),$$

which can be rewritten as

$$\mu(x) - \bar{\delta}(x) \geq \mu(y) - \bar{\delta}(y). \quad (9)$$

Combining inequalities (8) and (9) with coefficients  $1 - \lambda$  and  $\lambda$ , where  $\lambda \in [0, 1]$ , we obtain the required result.  $\square$

An identical result (under the condition of finite second moments) has been obtained in [22] for the *standard semideviation*, and further extended in [23] to central semideviations of higher orders and stochastic dominance relations of higher orders (see also [9]).

Elementary calculations show that for any distribution

$$\bar{\delta}(x) = \frac{1}{2}\delta(x),$$

where  $\delta(x)$  is the mean absolute deviation from the mean:

$$\delta(x) = \mathbb{E} |R(x) - \mu(x)|. \quad (10)$$

Thus,  $\delta(x)$  is a consistent risk measure with the coefficient  $\alpha = \frac{1}{2}$ . The mean–absolute deviation model has been introduced as a convenient linear programming mean–risk model by Konno and Yamazaki [13].

Another useful class of risk measures can be obtained by using quantiles of the distribution of the return  $R(x)$ . Let  $q_p(x)$  denote the  $p$ -th quantile<sup>1</sup> of the distribution of the return  $R(x)$ , i.e.,

$$\mathbb{P}[R(x) < q_p(x)] \leq p \leq \mathbb{P}[R(x) \leq q_p(x)].$$

We may define the risk measure

$$\rho_p(x) = \mathbb{E} \left[ \max \left( \frac{1-p}{p}(q_p(x) - R(x)), R(x) - q_p(x) \right) \right]. \quad (11)$$

In the special case of  $p = \frac{1}{2}$  the measure above represents the mean absolute deviation from the median. For small  $p$ , deviations to the left of the  $p$ -th quantile are penalized in a much more severe way than deviations to the right.

Although the  $p$ -th quantile  $q_p(x)$  might not be uniquely defined, the risk measure  $\rho_p(x)$  is a well defined quantity. Indeed, it is the optimal value of a certain optimization problem:

$$\rho_p(x) = \min_z \mathbb{E} \left[ \max \left( \frac{1-p}{p}(z - R(x)), R(x) - z \right) \right]. \quad (12)$$

It is well known that the optimizing  $z$  will be one of the  $p$ -th quantiles of  $R(x)$  (see, e.g., [5]). In [24] we have proved the following result.

---

<sup>1</sup>In the financial literature, the quantity  $-q_p(x)W$ , where  $W$  is the initial investment, is sometimes called the *Value at Risk*.

**Theorem 2** *The mean-risk model with the risk defined as  $\rho_p(x)$  is consistent with the second-order stochastic dominance relation with coefficient 1, for all  $p \in (0, 1)$ .*

Again, we provide here an alternative proof.

**Proof.** Let us consider the composite objective in our mean-risk model (scaled by  $p$ ):

$$G(p; x) = p\mu(x) - p\rho_p(x). \quad (13)$$

It follows from (12) that we can represent it as an optimal value:

$$G(p; x) = \sup_z \left[ p\mu - \mathbb{E} \left[ \max((1-p)(z - R(x)), p(R(x) - z)) \right] \right]. \quad (14)$$

Clearly, we have the identity

$$\max((1-p)(z - R(x)), p(R(x) - z)) = \max(0, z - R(x)) + p(R(x) - z).$$

Using this in (14) we obtain

$$G(p; x) = \sup_z \left[ pz - F^{(2)}(z; x) \right]. \quad (15)$$

Therefore, the function  $G(\cdot; x)$  is the Fenchel conjugate of  $F^{(2)}(\cdot; x)$  (see [6, 26]). Consequently, the second-order dominance  $R(x) \succeq_{SSD} R(y)$  implies that

$$G(p; x) \geq G(p; y)$$

for all  $p \in [0, 1]$ . Recalling (13) we conclude that

$$\mu(x) - \rho_p(x) \geq \mu(y) - \rho_p(y).$$

Since we also have (8), Definition 1 is satisfied with all  $\lambda \in [0, 1]$ . □

Interestingly, the function  $G(\cdot; x)$  can also be expressed as the integral:

$$G(p, x) = \int_0^p q_\alpha(x) d\alpha \quad (16)$$

(non-uniqueness of the quantile does not matter here). Indeed, it follows from (15) that the quantile  $q_p(x)$ , which is the maximizer in (15), is a subgradient of  $G(\cdot, x)$  at  $p$  (see [6, 26]). The integral in (16) is called the *absolute Lorenz curve* and is frequently used (for nonnegative variables and in a normalized form) in income inequality studies (see [2, 8, 16, 21, 24] and the references therein). It is illustrated in Figure 3.

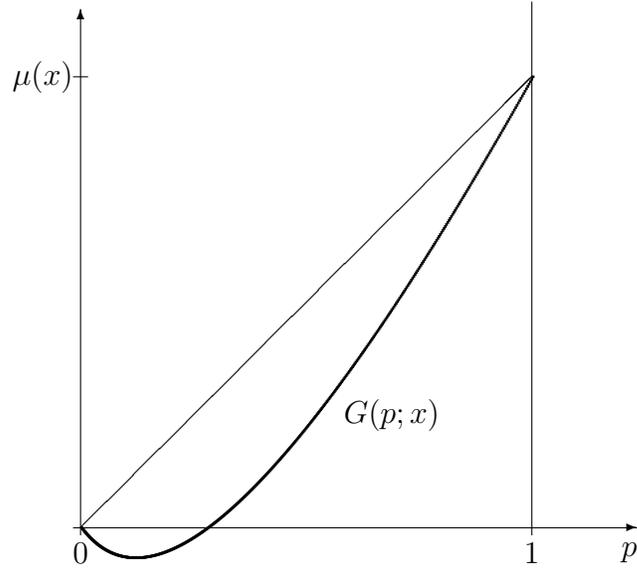


Figure 3: The absolute Lorenz curve.

## 4 Linear programming formulations

The second major advantage of the risk measures discussed in the previous section, in addition to being consistent with second-order stochastic dominance, is the possibility of formulating the models (2) as linear programming problems, if the underlying distributions are discrete.

Let us start from the risk measure defined as the expected absolute deviation from the mean, as defined in (10). The resulting linear programming model takes on the form:

$$\begin{aligned}
 \max \quad & v - \lambda \sum_{t=1}^T p_t u_t \\
 \text{subject to} \quad & v = \sum_{j=1}^n \sum_{t=1}^T r_{jt} p_t x_j, \\
 & u_t \geq \sum_{j=1}^n r_{jt} x_j - v, \\
 & u_t \geq v - \sum_{j=1}^n r_{jt} x_j, \\
 & x \in X,
 \end{aligned} \tag{17}$$

in which the decision variables are  $x_j$ ,  $j = 1, \dots, n$ ,  $v$ , and  $u_t$ ,  $t = 1, \dots, T$ . By construc-

tion,  $v$  is the mean portfolio return, and  $u_t$  represents the absolute deviation of the return from the mean for realization  $t$ .

It follows from Theorem 1 that for every  $\lambda \in (0, \frac{1}{2})$  the set of optimal solutions of this problem contains a portfolio which is nondominated in the SSD sense. So, if the solution is unique, it is nondominated. If it is not unique, there may be another solution  $y$  of (17) that dominates it, but it will have exactly the same values of the mean and the absolute deviation:  $\mu(y) = \mu(x)$  and  $\delta(y) = \delta(x)$ .

To prove it, suppose that the optimal portfolio  $x$  in the above model is dominated (in the SSD sense) by another portfolio  $y$ . It follows from Theorem 1 that

$$\mu(y) - \beta\delta(y) \geq \mu(x) - \beta\delta(x),$$

for all  $\beta \in [0, \frac{1}{2}]$ . For  $\beta = \lambda$  both sides must be equal, because  $x$  is the solution of (17). Since  $\lambda \in (0, \frac{1}{2})$ , the above inequality must in fact be an equation for all  $\beta \in [0, \frac{1}{2}]$ , and the result follows.

The model with weighted absolute deviations from the quantile can be formulated in a similar way

$$\begin{aligned} \max \quad & v - \lambda \sum_{t=1}^T p_t u_t \\ \text{subject to} \quad & v = \sum_{j=1}^n \sum_{t=1}^T r_{jt} p_t x_j, \\ & u_t \geq \sum_{j=1}^n r_{jt} x_j - z, \\ & u_t \geq \frac{1-p}{p} \left[ z - \sum_{j=1}^n r_{jt} x_j \right], \\ & x \in X, \end{aligned} \tag{18}$$

in which the decision variables are  $x_j$ ,  $j = 1, \dots, n$ ,  $v$ ,  $u_t$ ,  $t = 1, \dots, T$ , and  $z$ . Indeed, when  $x$  and  $v$  are fixed, the best value of  $z$  is the  $p$ -th quantile of  $R(x)$ , as follows from (12). Thus, the expression  $\sum_{t=1}^T p_t u_t$  represents the risk measure  $\rho_p(x)$ .

Again, Theorem 2 implies that for every  $\lambda \in (0, 1)$  the set of optimal solutions of this problem contains a portfolio which is nondominated in the SSD sense. If the solution is unique, it is nondominated. If it is not unique, there may be another solution  $y$  of (17) that dominates it, but it will have exactly the same values of the mean, and the average deviation from the  $p$ -th quantile:  $\mu(y) = \mu(x)$  and  $\rho_p(y) = \rho_p(x)$ . The argument that supports this is the same as in the previous case.

## 5 The parametric method for constructing the efficient frontier

There are many variants of the simplex method for solving linear programming problems. Some are so-called *two-phase methods* in which a first phase is used to find a basic feasible solution and a second phase then pivots from one basic feasible solution to another until a basic optimal solution is found. Other, less known variants, are one-phase methods. These methods proceed directly from a basic infeasible solution to a basic optimal solution without placing any particular emphasis on finding a first feasible solution (of course, the optimal solution is required to be feasible).

A particular one-phase method of interest is the *parametric self-dual simplex method*—see e.g. Chapter 7 of [29]. We describe this method in the context of a generic linear programming problem:

$$\begin{aligned} \max \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned} \tag{19}$$

Here,  $A$  is an  $m \times n$  matrix—the dimensions of  $b$ ,  $c$ , and  $x$  are determined accordingly. First, we establish some standard notation one needs to define the simplex method. Consider a partition of the variable indices  $\{1, 2, \dots, n\}$  into two sets:  $\mathcal{B}$  and  $\mathcal{N}$ . Variables  $x_j, j \in \mathcal{B}$ , are called *basic variables* whereas those in  $\mathcal{N}$  are called *nonbasic*. We permute the columns of  $A$  so that the basic variables precede the nonbasic ones:

$$A = [ B \quad N ].$$

The  $x$  vector is permuted accordingly:  $x = [ x_{\mathcal{B}} \quad x_{\mathcal{N}} ]^T$  so that we can write

$$Ax = Bx_{\mathcal{B}} + Nx_{\mathcal{N}}.$$

If the matrix  $B$  is square ( $m \times m$ ) and nonsingular, then the basic variables can be expressed as functions of the nonbasic ones:

$$x_{\mathcal{B}} = \hat{x}_{\mathcal{B}} - B^{-1}Nx_{\mathcal{N}},$$

where

$$\hat{x}_{\mathcal{B}} = B^{-1}b$$

and the *basic solution* associated with this partition is obtained by setting  $x_{\mathcal{N}} = 0$  and reading off the corresponding values for the basic variables:  $x_{\mathcal{B}} = \hat{x}_{\mathcal{B}}$ . If  $\hat{x}_{\mathcal{B}} \geq 0$ , the basic solution is called *primal feasible*.

The function being maximized can also be expressed purely in terms of the nonbasic variables:

$$c^T x = c_{\mathcal{B}}^T x_{\mathcal{B}} + c_{\mathcal{N}}^T x_{\mathcal{N}} = \hat{\xi} - \hat{z}_{\mathcal{N}}^T x_{\mathcal{N}},$$

where

$$\hat{\xi} = c_B^T B^{-1} b \quad \text{and} \quad \hat{z}_N = (B^{-1} N)^T c_B - c_N.$$

If  $\hat{z}_N \geq 0$ , then we say that the basic solution is *dual feasible*. If a basic solution is both primal and dual feasible, then it is *optimal*.

It is an easy task to find a partition for which  $B$  is square and nonsingular but generally the associated solution will be neither primal nor dual feasible. Every variant of the simplex method moves from one basic solution to another until an optimal solution is found. With the parametric self-dual simplex method, one adds to the objective function another auxillary function, which is linear in the original variables and which depends linearly on a parameter  $\lambda$ . Similarly, one adds to the constraint constants another vector of constants also depending linearly on the parameter  $\lambda$ . Given a basis partition, one then has

$$c^T x = \hat{\xi}(\lambda) - (\hat{z}_N + \lambda \bar{z}_N)^T x_N,$$

and

$$x_B = (\hat{x}_B + \lambda \bar{x}_B) - B^{-1} N x_N,$$

where  $\hat{\xi}(\lambda)$  is some specific quadratic expression in  $\lambda$ . The auxillary objective function and constraint constants are chosen with the following two properties:

1. A basic optimal solution to the parametrized problem can be trivially identified for some specific value, say  $\lambda_0$ , (or interval of values) of the parameter.
2. The original problem coincides with the parametrized problem for some specific value, say  $\lambda_1$ , of the parameter.

When used as a method for solving general linear programming problems, the initial range of parameter values for which a basic optimal solution is evident is usually taken to be an interval of the form  $[\lambda_0, \infty)$  and  $\lambda_1$  is taken to be 0. For such a case, it is easy to determine an appropriate auxillary objective function and appropriate constraint constants. Indeed, it suffices to choose them so that  $\bar{z}_N > 0$  and  $\bar{x}_B > 0$ . With such a choice, for  $\lambda$  sufficiently large the initial basis satisfies the two conditions for optimality:

$$\hat{x}_B + \lambda \bar{x}_B \geq 0 \quad \text{and} \quad \hat{z}_N + \lambda \bar{z}_N \geq 0. \quad (20)$$

Suppose for the sake of discussion that  $\lambda_1 < \lambda_0$ . The parametric self-dual method works as follows. It begins by setting the parameter  $\lambda$  to  $\lambda_0$ . The parameter is then decreased as much as possible without violating optimality of the current basis. At the point where  $\lambda$  cannot be further decreased without violating optimality of the current basis, a simplex pivot is performed which allows one to further decrease  $\lambda$ . How one determines the entering and leaving variables for the pivot depends on which inequality in (20) blocks further reduction in  $\lambda$ .

Suppose that the blocking inequality is  $\hat{z}_j + \lambda \bar{z}_j \geq 0$  for some  $j \in \mathcal{N}$ . Then  $x_j$  is the entering variable and the leaving variable  $x_i$  is determined by performing the usual ratio test computed using the current blocked value of  $\lambda = -\hat{z}_j/\bar{z}_j$ :

$$i = \underset{\substack{i \in \mathcal{B} \\ (B^{-1}a_j)_i > 0}}{\operatorname{argmin}} \frac{\hat{x}_i + \lambda \bar{x}_i}{(B^{-1}a_j)_i},$$

where  $a_j$  denotes the  $j$ -th column of  $A$  and  $(B^{-1}a_j)_i$  is the element of the vector  $B^{-1}a_j$  corresponding to the basic variable  $x_i$ .

If, on the other hand, the blocking inequality comes from an inequality on the basic indices then one does a dual pivot in exactly the manner dual to the primal pivot just described.

One can check that for either type of pivot the new basic solution is optimal for an interval of  $\lambda$ 's having the current value as its upper bound (as opposed to its being the lower bound before the pivot). So, after the pivot,  $\lambda$  is again decreased as much as possible until any further reduction would violate the optimality of the current basis. And again, a simplex pivot (either primal or dual as needed) is used to allow further reduction in  $\lambda$ . The method continues in this way until a basis is found whose  $\lambda$ -interval of optimality covers  $\lambda_1$ . At this point,  $\lambda$  is simply set equal to  $\lambda_1$  and the optimal solution to the original problem is obtained.

Although the parametric self-dual simplex method is not widely used in current commercial linear programming solvers, one can argue that it ought to be so employed as it has many advantages. First of all, empirical studies indicate that, for nondegenerate problems, this method takes roughly the same number of pivots to find an optimal solution as the common two-phase methods ([29], Section 12.7). Furthermore, by simply randomizing the auxiliary objective function and constraint constants, one can guarantee that the method will never make any degenerate pivots. Hence, for degenerate problems, the number of pivots of a so-randomized version of this algorithm can be vastly fewer than what one obtains with other variants of the simplex method. Thirdly, this method is more amenable than others to average-case performance analyses. In fact, Adler and Meggido [1] made the first such average-case analysis back in the mid-80s and more recently [29] (Section 12.6) shows that the algorithm can be expected to take, on average,  $n/2$  iterations.

The fourth, and for us the most important, advantage of the parametric self-dual simplex method is that sometimes a parametric representation of the problem is inherent in the original problem. This can be then exploited to great advantage. Such is the case in computing points on the efficient frontier using either (17) or (18).

For both linear programming problems (17) and (18), the solution with  $\lambda = 0$  is trivial:

*put 100% into that security which has the highest expected return.*

To find any specific point on the efficient frontier, say the point corresponding to  $\lambda = \lambda_1$ , one can start at  $\lambda = 0$  and use the parametric self-dual simplex method to walk along

the efficient frontier to the point  $\lambda_1$ . But, of course, in so doing we obtain not only the solution to a single point on the efficient frontier but we also obtain, for free, the solution to every frontier point in between. And, we can continue beyond  $\lambda_1$  and keep going until  $\lambda = \infty$ . In so doing we obtain the entire efficient frontier with about the same amount of computation as would otherwise have been used to determine a single point on the frontier. In fact, *every* step of our parametric simplex method generates an efficient solution.

The only technical issue that needs to be addressed is the identification of the basic and nonbasic variables associated with the initial  $\lambda = 0$  solution. Let us consider first (17). Using the first constraint to eliminate  $v$ , introducing slack variables on the inequalities, and assuming, purely for simplicity, that  $X = \{x \geq 0 : x_1 + x_2 + \cdots + x_n = 1\}$ , we rewrite the problem in the same form as (19):

$$\begin{aligned}
& \max && \sum_{j=1}^n \sum_{t=1}^T r_{jt} p_t x_j - \lambda \sum_{t=1}^T p_t u_t \\
\text{subject to} &&& \sum_{j=1}^n \left( r_{jt} - \sum_{t'=1}^T r_{jt'} p_{t'} \right) x_j - u_t + s_t^+ = 0, && t = 1, 2, \dots, T, \\
&&& - \sum_{j=1}^n \left( r_{jt} - \sum_{t'=1}^T r_{jt'} p_{t'} \right) x_j - u_t + s_t^- = 0, && t = 1, 2, \dots, T, \\
&&& \sum_{j=1}^n x_j = 1, \\
&&& x_j \geq 0, && j = 1, 2, \dots, n, \\
&&& u_t, s_t^+, s_t^- \geq 0, && t = 1, 2, \dots, T.
\end{aligned} \tag{21}$$

There are  $m = 2T + 1$  equality constraints. Hence, we must identify  $2T + 1$  basic variables. The remaining nonbasic variables must, of course, be zero. We use this fact to guide us in determining which variables are basic and which are nonbasic in the initial solution. Let  $j^*$  denote a security with maximal expected return. We can write explicit formulas

for the all of the variables in the  $\lambda = 0$  optimization problem:

$$\begin{aligned} x_j &= \begin{cases} 1, & j = j^*, \\ 0, & \text{else,} \end{cases} \\ u_t &= \left| r_{j^*t} - \sum_t r_{j^*t} p_t \right|, \\ s_t^+ &= \max \left( 0, -r_{j^*t} + \sum_t r_{j^*t} p_t \right), \\ s_t^- &= \max \left( 0, r_{j^*t} - \sum_t r_{j^*t} p_t \right). \end{aligned}$$

One of the  $x_j$ 's is positive, all of the  $u_t$ 's are, and of the  $2T$  slack variables  $T$  of them are positive and  $T$  of them are zero. Hence, we have  $2T + 1$  variables that are positive—the remaining are zero. Clearly, the positive variables are the basic variables and those that are zero are nonbasic.

Now consider the analogous rewrite of (18). If we assume that all of the  $r_{jt}$ 's are nonnegative (which is the case if one understands these numbers in the sense that a 3% return is represented as 1.03), then the  $p$ -th quantile  $z$  will also be nonnegative and we can rewrite (18) as follows:

$$\begin{aligned} \max \quad & \sum_{j=1}^n \sum_{t=1}^T r_{jt} p_t x_j - \lambda \sum_{t=1}^T p_t u_t \\ \text{subject to} \quad & \left( \sum_{j=1}^n r_{jt} x_j - z \right) - u_t + s_t^+ = 0, \quad t = 1, 2, \dots, T, \\ & -(1-p) \left( \sum_{j=1}^n r_{jt} x_j - z \right) - p u_t + s_t^- = 0, \quad t = 1, 2, \dots, T, \\ & \sum_{j=1}^n x_j = 1, \\ & x_j \geq 0, \quad j = 1, 2, \dots, n, \\ & u_t, s_t^+, s_t^- \geq 0, \quad t = 1, 2, \dots, T, \\ & z \geq 0. \end{aligned} \tag{22}$$

As before, the  $\lambda = 0$  solution is easy to write:

$$\begin{aligned} x_j &= \begin{cases} 1, & j = j^*, \\ 0, & \text{else,} \end{cases} \\ z &= r_{j^*t^*}, \\ u_t &= |r_{j^*t} - r_{j^*t^*}|, \\ s_t^+ &= \max(0, -r_{j^*t} + r_{j^*t^*}), \\ s_t^- &= \max(0, r_{j^*t} - r_{j^*t^*}), \end{aligned}$$

where  $t^*$  denotes the sample-space index for which  $R_{j^*}$  attains its  $p$ -th quantile. Again, we must identify  $2T + 1$  basic variables, but this time we have one more variable,  $z$ , and this variable is itself nonzero and hence basic. Regarding the  $u_t$ 's,  $s_t^+$ 's, and  $s_t^-$ 's, the situation is almost the same as before, the only difference being that all of these vanish at  $t^*$ . Hence, we have only  $2T$  variables that must be declared basic and three choices for the final basic variable: either  $u_{t^*}$ ,  $s_{t^*}^+$ , or  $s_{t^*}^-$ . This ambiguity can be resolved in any convenient way. That is, all three choices provide a nonsingular basis matrix  $B$ .

For  $p = 1$ , it is easy to find the optimal solution. Indeed, for  $p = 1$ , the slacked constraints in (22) simplify to the following:

$$\begin{aligned} \left( \sum_{j=1}^n r_{jt} x_j - z \right) &\leq u_t, & t = 1, 2, \dots, T, \\ u_t &\geq 0, & t = 1, 2, \dots, T. \end{aligned}$$

The optimization is over the  $x_j$ 's, the  $u_t$ 's, and  $z$ . We can view this maximization as first a maximization over the  $u_t$ 's and  $z$  for fixed  $x_j$ 's followed by maximization over the  $x_j$ 's:

$$\begin{aligned} \max_{x, u, z} \sum_{j=1}^n \sum_{t=1}^T r_{jt} p_t x_j - \lambda \sum_{t=1}^T p_t u_t &= \max_x \max_{u, z} \left( \sum_{j=1}^n \sum_{t=1}^T r_{jt} p_t x_j - \lambda \sum_{t=1}^T p_t u_t \right) \\ &= \max_x \left( \sum_{j=1}^n \sum_{t=1}^T r_{jt} p_t x_j - \lambda \min_{u, z} \sum_{t=1}^T p_t u_t \right). \end{aligned}$$

The inner minimization over the  $u_t$ 's and  $z$  is solved by setting  $z = \max_t \sum_j r_{jt}$  and  $u_t \equiv 0$ . Hence, the outer maximization becomes simply a maximization of expected return; the risk term vanishes. An alternative way to see that the risk vanishes is to appeal directly to the definition of  $\rho_p(x)$  in (11) with  $p = 1$ . In this case, we get

$$\rho_1(x) = \mathbb{E} \left[ \max(0, R(x) - q_1(x)) \right] = 0.$$

The optimal solution is therefore, as before, obtained by placing the entire portfolio into the security with the greatest expected return.

The optimal solution can also be easily identified for  $p = 0$  and for one particular value of  $\lambda$ , namely  $\lambda = 1$ . Again, appealing to (11), we see that

$$\rho_0(x) = \mathbb{E} \left[ \max ( - \infty, R(x) - q_0(x) ) \right] = \mathbb{E}R(x) - q_0(x).$$

Hence, the objective function is

$$\mu(x) - \lambda\rho_0(x) = (1 - \lambda)\mathbb{E}R(x) + \lambda q_0(x).$$

For  $\lambda = 1$ , the expected return term drops out and the problem simplifies to solving simply for the portfolio that maximizes  $q_0(x)$ . This portfolio consists of just one security—the one which maximizes its lowest return, as the min-max model of [32].

## 6 Numerical illustration

Using a data file consisting of daily return data for 719 securities from January 1, 1990 to March 18, 2002 ( $T = 3080$ ), we computed the entire efficient frontier using both the deviation-from-mean (Figure 4) and the deviation-from-quantile (Figure 5) risk measures. For the deviation from quantile case, we used the  $p = 0.05$  quantile. The clustered marks in each graph represent mean–risk characteristics of individual securities.

Although we don't show it here, we also computed the efficient frontier for the deviation from median risk measure (i.e.,  $p = 0.5$ ). Normally, one would expect the deviation-from-median measure to be superior to the deviation-from-mean measure for two reasons: (i) medians are more robust estimators as evidenced by their fundamental role in non-parametric statistics and (ii) the deviation from median provides stochastically nondominated portfolios for  $0 \leq \lambda \leq 1$  whereas for deviation from the mean the interval is only  $0 \leq \lambda \leq 1/2$ . However, the data set at hand, and perhaps it is true for this type of data in general, does not contain significant outliers that would skew the results obtained using the mean measure. Furthermore, most portfolios fall in the range  $0 \leq \lambda \leq 1/2$  so not much is lost by stopping here. Hence, these two efficient frontiers, while containing different portfolios, look very similar in the usual risk-reward plots such as shown in Figure 4.

### 6.1 Deviation from the mean

For the deviation from the mean, the efficient frontier consists of 23509 distinct nondominated portfolios. At the risky extreme, the first portfolio consists of putting 100% into the highest return security. Other high-risk portfolios consist of mixtures of only a few high return securities. As  $\lambda$  increases, the portfolios get more complicated as they incorporate more and more hedging. The most risk averse portfolio contains a mixture of 80 securities—this represents a typical size for a portfolio at the highly-hedged risk-averse

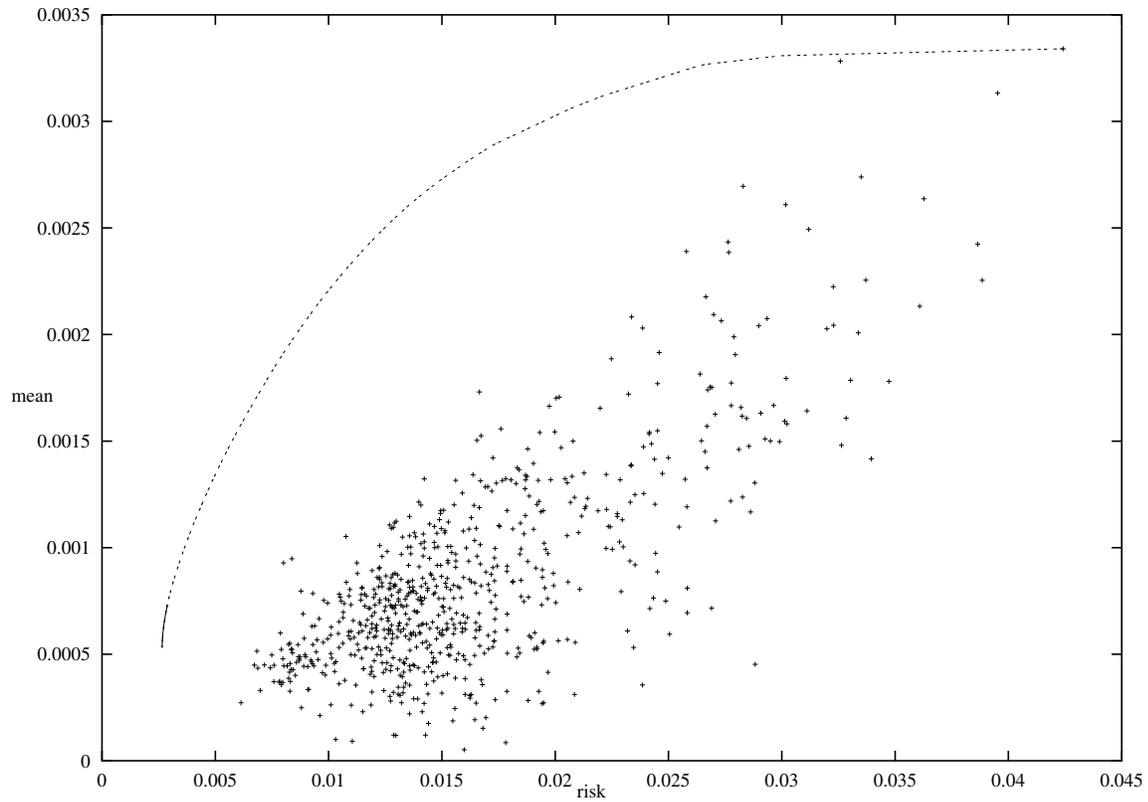


Figure 4: The efficient frontier for the deviation-from-mean risk measure. The dashed portion of the frontier represents stochastically nondominated portfolios ( $\lambda \leq 1/2$ ) whereas the small solid portion in the lower left corner corresponds to  $\lambda > 1/2$ .

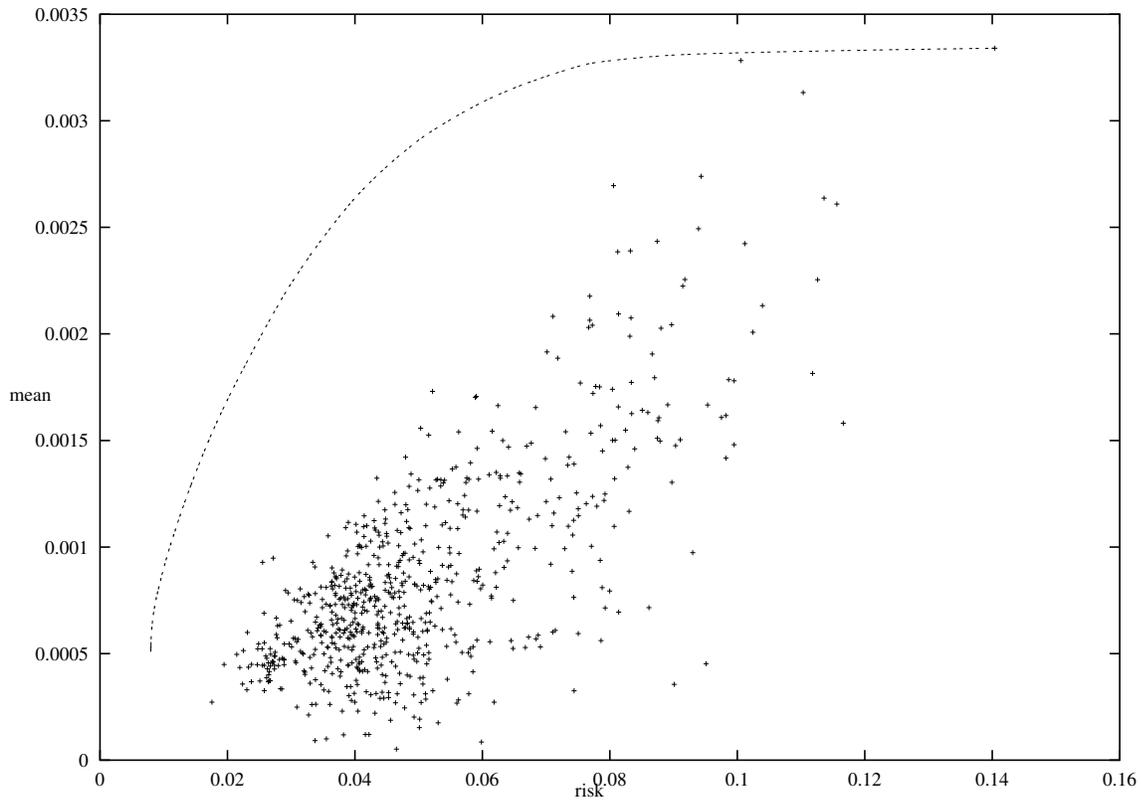


Figure 5: The efficient frontier for the deviation-from-quantile risk measure using the  $p = 0.05$  quantile. The dashed portion of the frontier represents stochastically nondominated portfolios ( $\lambda \leq 1$ ) whereas the small solid portion in the lower left corner corresponds to  $\lambda > 1$ .

end of the efficient frontier. All portfolios corresponding to  $\lambda \in [0, \frac{1}{2})$  are nondominated in the second-order stochastic dominance sense, because each of them is the unique solution for some values of  $\lambda$ . There are 18901 such portfolios and they cover almost the entire efficient frontier displayed in Figure 4.

As we can see from the figure, dramatic improvements in the mean and in the risk are possible, in comparison to the individual securities.

The entire frontier was computed in 1 hour and 46 minutes of cpu time on a Windows 2000 laptop computer having a 1.2 MHz clock. To put this into context, we note that it took an analogous code (i.e., a code built from the same linear algebra subroutines) that implements the usual two-phased simplex method about one and a half hours to compute a single portfolio on the efficient frontier.

## 6.2 Deviation from the quantile

For the deviation from the 0.05-quantile, the efficient frontier consists of 5127 distinct nondominated portfolios and was computed in 18 minutes of cpu time. As with the deviation-from-mean case, risky portfolios contain only a few securities whereas risk-averse portfolios are richer in content. For example, the most risk-averse portfolio consists of 52 securities. All portfolios corresponding to  $\lambda \in [0, 1)$  are nondominated in the second-order stochastic dominance sense. There are 4770 of them and they cover almost the entire efficient frontier displayed in Figure 5.

When the risk is measured by the weighted deviation from the 0.05-quantile, the deviations to the left of it are penalized about 20 times more strongly than the deviations to the right. Improving the shape of the left tail of the distribution has therefore an even more dramatic effect on the risk measure than in the deviation-from-mean case.

## Acknowledgement.

We would like to thank Humbert Suarez at Goldman-Sachs for providing us with a real-world data set on which to base our experiments.

## References

- [1] I. ADLER AND N. MEGIDDO, *A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension*, Journal of the ACM, 32:871–895, 1985.
- [2] B. C. ARNOLD, *Majorization and the Lorenz Order: A Brief Introduction*, Lecture Notes in Statistics 43, Springer-Verlag, Berlin, 1980.
- [3] P. ARTZNER, F. DELBAEN, J.-M. EBER AND D. HEATH, *Coherent measures of risk*, Mathematical Finance, 9 (1999), pp. 203–228.

- [4] V. S. BAWA, *Stochastic dominance: a research bibliography*, Management Science, 28 (1982), pp. 698–712.
- [5] P. BLOOMFIELD AND W. L. STEIGER, *Least Absolute Deviations*, Birkhäuser, Boston 1983.
- [6] W. FENCHEL, *Convex Cones, Sets, and Functions*, lecture notes, Princeton University, 1953.
- [7] P. C. FISHBURN, *Decision and Value Theory*, John Wiley & Sons, New York, 1964.
- [8] J. L. GASTWIRTH, *A general definition of the Lorenz curve*, Econometrica, 39 (1971), pp. 1037–1039.
- [9] J. GOTOH AND H. KONNO, *Third degree stochastic dominance and mean–risk analysis*, Management Science, 46 (2000), pp. 289–301.
- [10] J. HADAR AND W. RUSSELL, *Rules for ordering uncertain prospects*, The American Economic Review, 59 (1969), pp. 25–34.
- [11] G. HANOCH AND H. LEVY, *The efficiency analysis of choices involving risk*, Review of Economic Studies, 36 (1969), pp. 335–346.
- [12] G. H. HARDY, J. E. LITTLEWOOD AND G. PÓLYA, *Inequalities*, Cambridge University Press, Cambridge, MA, 1934.
- [13] H. KONNO AND H. YAMAZAKI, *Mean–absolute deviation portfolio optimization model and its application to Tokyo stock market*, Management Science, 37 (1991), pp. 519–531.
- [14] E. LEHMANN, *Ordered families of distributions*, Annals of Mathematical Statistics, 26 (1955), pp. 399–419.
- [15] H. LEVY, *Stochastic dominance and expected utility: survey and analysis*, Management Science, 38 (1992), pp. 555–593.
- [16] M. O. LORENZ, *Methods of measuring concentration of wealth*, Journal of the American Statistical Association, 9 (1905), pp. 209–219.
- [17] H. M. MARKOWITZ, *Portfolio selection*, Journal of Finance, 7 (1952), pp. 77–91.
- [18] H. M. MARKOWITZ, *Portfolio Selection*, John Wiley & Sons, New York, 1959.
- [19] H. M. MARKOWITZ, *Mean–Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Oxford, 1987.

- [20] A. W. MARSHALL AND I. OLKIN, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, San Diego, 1979.
- [21] P. MULIERE AND M. SCARSINI, *A note on stochastic dominance and inequality measures*, *Journal of Economic Theory*, 49 (1989), pp. 314–323.
- [22] W. OGRYCZAK AND A. RUSZCZYŃSKI, *From stochastic dominance to mean-risk models: semideviations as risk measures*, *European Journal of Operational Research*, 116 (1999), pp. 33–50.
- [23] W. OGRYCZAK AND A. RUSZCZYŃSKI, *On consistency of stochastic dominance and mean-semideviation models*, *Mathematical Programming*, 89 (2001), pp. 217–232.
- [24] W. OGRYCZAK AND A. RUSZCZYŃSKI, *Dual stochastic dominance and related mean-risk models*, *SIAM Journal on Optimization*, to appear in 2002.
- [25] J.P. QUIRK AND R. SAPOSNIK, *Admissibility and measurable utility functions*, *Review of Economic Studies*, 29 (1962), pp. 140–146.
- [26] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970.
- [27] M. ROTHCHILD AND J. E. STIGLITZ, *Increasing risk: I. A definition*, *Journal of Economic Theory*, 2 (1969), pp. 225–243.
- [28] W. F. SHARPE, *A linear programming approximation for the general portfolio analysis problem*, *Journal of Financial and Quantitative Analysis*, 6 (1971), pp. 1263–1275.
- [29] R.J. VANDERBEI, *Linear Programming: Foundations and Extensions*. Kluwer Academic Publishers, 2nd edition, 2001.
- [30] G. A. WHITMORE AND M. C. FINDLAY, eds., *Stochastic Dominance: An Approach to Decision-Making Under Risk*, D.C.Heath, Lexington, MA, 1978.
- [31] S. YITZHAKI, *Stochastic dominance, mean variance, and Gini's mean difference*, *American Economic Review*, 72 (1982), pp. 178–185.
- [32] M. R. YOUNG, *A minimax portfolio selection rule with linear programming solution*, *Management Science*, 44 (1998), pp. 673–683.