Value at Risk in Portfolio Optimization: Properties and Computational Approach

Alexei A. Gaivoronski* and Georg Pflug**

*Department of Industrial Economics and Technology Management
NTNU - Norwegian University of Science and Technology
Alfred Getz vei 1, N-7049 Trondheim, Norway
Alexei.Gaivoronski@iot.ntnu.no
**Department of Statistics
University of Vienna
Universitaetsstrasse 5
A-1010 Vienna, Austria
georg.pflug@univie.ac.at

Abstract

Value at risk (VaR) is an important and widely used measure of the extent to which a given portfolio is subject to risk present in financial markets. Considerable amount of research was dedicated during recent years to development of acceptable methods for evaluation of this risk measure.

In this paper, we present a method of calculating the portfolio which gives the optimal VaR among those, which yield at least some specified expected return. This method allows to calculate the mean-VaR efficient frontier. The method is based on approximation of historic VaR by smoothed VaR (SVaR) which filters out local irregular behavior of historic VaR function.

Moreover, we compare the VaR as a risk measure to other well known measures of risk such as the conditional value at risk (CVaR) and the standard deviation.

It turns out, that the corresponding efficient frontiers are quite different. An investor, who wants to control his VaR should not look at portfolios lying on other than the VaR efficient frontier, although the calculation of this frontier is algorithmically more complex compared to other frontiers.

1 Introduction

Value at Risk (VaR) is an important measure of exposure of a given portfolio of securities to different kinds of risk inherent in financial environment. By now, it became a tool for risk management in financial industry [1] and part of industrial regulatory mechanisms [2]. Considerable amount of research was dedicated recently to development of methods of risk management based on Value at Risk [3, 19, 20, 21, 31]. This literature is dedicated mainly to efficient techniques for computing VaR of a given portfolio.

VaR is a simple and intuitive measure of risk. Some other frequently used, but not equally intuitive measures of risk are the variance, the semi-standard deviation and the conditional value at risk (CVaR, see below). Risk measures can be compared in two ways: Firstly, risk measures

differ in their theoretical properties and the decision maker may choose the measure which fits to his/her preferences. Secondly, efficient frontiers pertaining to different risk measures may be calculated and compared. In this way, the quality of portfolio selected for optimal risk measure 1 can be judged in terms of risk measure 2.

In this paper we compare VaR, CVaR and standard deviation as risk measures both in terms of theoretical properties and efficient frontiers for a concrete data set. Among all possible theoretical properties we concentrate on those which are particularly relevant for optimization problems studied here. Evaluation of different risk measures from the point of view of utility theory is outside the scope of this paper, interesting discussion of this issue is contained in [18]. Efficient frontiers for standard deviation and CVaR are easy to calculate, the VaR case is however much more computationally involved. The problem nonconvex, may exhibit many local minima and is of combinatorial character, i.e. exponential growth in computational complexity. We present a smoothing algorithm, which allows to calculate optimal portfolios in the VaR sense with high accuracy and in reasonable time.

Consider a finite set of assets $i = 1, 2, \ldots, n$ which can be any kind of financial assets, stocks, bonds and options being the most common examples. Within the observation period (e.g. 10 days), the assets generate returns

$$\xi = (\xi_1, \xi_2, \ldots, \xi_n),$$

which are unknown at the time of portfolio allocation and treated as random variables.

The investor has a budget of 1 unit (without loss of generality). He/she may decide on the positions

$$x = (x_1, x_2, \ldots, x_n)$$

in these assets, such that $x_i \geq 0$ (no short sales permitted) and $\sum_{i=1}^{n} x_i = 1$ (budget constraint). Using the vector $\mathbb{1} = (1, \ldots, 1)$ of ones, we may write the budget constraint as $x^T \mathbb{1} = 1$.

The value of the portfolio at the end of the observation period is

$$W = x^T \xi = \sum_{i=1}^{n} x_i \xi_i.$$

The expected return of this portfolio is

$$\mathbb{E}(W) = \mathbb{E}(x^T \xi) = x^T \mathbb{E}(\xi).$$

Suppose $\mathcal{R}$ some risk measure, like those based on VaR, CVaR or the standard deviation (Std). For a given minimal expected return μ, the $\mathcal{R}$-efficient portfolio is the solution of the problem

$$\left\| \begin{array}{ll} \text{Minimize (in } x) & \mathcal{R}(x^T \xi) \\ \text{subject to} & \\ x^T \mathbb{E}(\xi) \geq \mu & \\ x^T \mathbb{1} = 1 & \\ x \geq 0 & \end{array} \right. \tag{1}$$

The curve which represents the dependence of the optimal value of this problem on the parameter μ is called the $\mathcal{R}$-efficient frontier. Essentially, this is a generalization of the classical concept of mean-variance efficient frontier due to Markowitz [23] for the case of an arbitrary risk measure $\mathcal{R}$.

The value W of the portfolio at the end of the observation period is a random variable with distribution function F (say); i.e. $F(u) = \mathbb{P}\{W \leq u\} = \mathbb{P}\{x^T\xi \leq u\}$. Of course, F depends on x.

It is sometimes easier to argue in terms of measures of safety, i.e. characteristic values of a portfolio, which the decision maker wants to make as large as possible. These measures have evident one-to-one relation with risk measures which is clear from examples which follow.

In particular, we will consider the following measures of safety $\mathcal{S}$:

- **The quantile of return of level α.**

$$Q_\alpha(W) = \inf\{u : F(u)) > \alpha\}.$$

- **The conditional expectation of return not exceeding quantile of level α.**

$$C_\alpha(W) = \mathbb{E}(W|W \leq\ Q_\alpha(W)).$$

- **The standard-deviation corrected mean.**

$$\text{SDCM}_\rho(W) = \mathbb{E}(W) - \rho\,\text{Std}(W).$$

  Here ρ is a risk aversion factor.

On the basis of safety measures, risk measures, i.e. values which should be as small as possible, can be defined. In this paper, we consider the following risk measures

- **Value at Risk (VaR) of return**:

$$\text{VaR}_\alpha(W) = 1 -\ Q_\alpha(W).$$

- **Conditional Value at Risk (CVaR) of return**:

$$\text{CVaR}_\alpha(W) = 1 -\ C_\alpha(W).$$

- **Standard deviation**

In this paper we focus on application of Value-at-Risk (VaR) in the context of optimal portfolio selection. This is a relatively novel application of VaR as opposed to utilization of VaR for risk management purposes. One of the reasons is that VaR optimization is inherently more difficult than, for example, variance optimization and efficient solution algorithms for this problem are lacking. The objective of this paper is filling of this gap with the aim of putting VaR on equal footing with variance as portfolio selection criterion. We argue that for investor whose risk preferences are expressed in terms of Value-at-Risk it is important to consider this measure directly because other risk measures like variance, or even seemingly related CVaR, may represent a poor substitute.

The rest of the paper is organized as follows. Section 2 is dedicated to comparison of risk measures considered in this paper based on theoretical properties of widely recognized importance. We argue that no measure posseses uniform advantage over other measures from the point of view of these properties. Formulation of optimization problems used for computation of mean-risk efficient frontiers for different risk measures is presented in section 3. We concentrate here on empirical variants of risk measures derived directly from historic data. Besides, we collect some empirical evidence about behavior of VaR as function of portfolio composition

and perform preliminary comparison between VaR and CVaR based on stock market data. The main point of this section is that behavior of historic VaR is the result of superposition of global and local components. Global component is fairly well behaved, while the local component is nonsmooth and possesses a multitude of local minima. This observaton is in the center of numerical approach for VaR optimization developed in section 4. Here we introduce smoothed VaR (SVaR) which is an approximation of VaR which extracts global component of VaR behavior and filters out irregular local component. Mathematical details of SVaR construction are fairly involved and for this reason we relegate them to Appendix. The reason for introduction of SVaR is that SVaR optimization is much easier and can be performed by efficient nonlinear programming software. The optimal VaR values are recovered from solution of SVaR optimization problem by a few inexpensive postprocessing steps. After developing VaR optimization tools, we are in position to compute and compare mean-risk efficient frontiers for different risk measures in section 5. Result of this comparison is that mean-VaR efficient frontier can not be approximated by efficient frontiers obtained for other risk measures, like variance or CVaR.

Very recently the possibilities to utilize VaR and related measures like CVaR as criterions for optimal portfolio selection started to attract some attention. The relevant literature includes Andersson and Uryasev[4], Basak and Schapiro[6], Gaivoronski and Pflug[16], Grootveld and Hallerbach[18], Medova 98 [25], Uryasev and Rockafellar [32].

VaR optimization can be considered as a stochastic programming problem of special type. In particular, it is related to stochastic programming problems with probability constraints [30]. General references for stochastic programming models and solution techniques are [7], [14],[22],[33]. Different applications of stochastic programming to optimal asset allocation including analysis of different risk measures were considered in [9],[10], [12],[13],[15],[17],[26], [27],[34].

2 Properties of safety measures

A safety measure assigns a numerical value to a random distribution of wealth. Safety measures can be compared on the basis of their properties. In this section, some of these properties are discussed.

First, we collect some definitions of properties of safety measures. Let $\mathcal{S}(W)$ be a safety measure, where W is a random wealth.

(i)$\mathcal{S}$ is *translation-equivariant*, if for all W and all $c > 0$

$$\mathcal{S}(W + c) = \mathcal{S}(W) + c.$$

(ii)$\mathcal{S}$ is *positively homogeneous*, if for all W and all $c > 0$

$$\mathcal{S}(cW) = c\mathcal{S}(W),$$

if $c > 0$.

(iii)$\mathcal{S}$ is *concave* if for all random variables W_1 and W_2 (they may be dependent) and $0 < \lambda < 1$,

$$\mathcal{S}(\lambda W_1 + (1 - \lambda)W_2) \geq \lambda\mathcal{S}(W_1) + (1 - \lambda)\mathcal{S}(W_2).$$

(iv)$\mathcal{S}$ is *comonotone additive*, if

$$\mathcal{S}(W_1 + W_2) = \mathcal{S}(W_1) + \mathcal{S}(W_2)$$

for comonotone random variables W_1 and W_2. Two random variables W_1 and W_2 are *comonotone*, if there is a uniform [0,1] random variable V and two monotone functions g_1 and g_2 such that $W_1 = g_1(V)$, $W_2 = g_2(V)$.

(v) $\mathcal{S}$ is *monotonic w.r.t. the order relation* $\prec$, if $W_1 \prec W_2$ implies that $\mathcal{S}(W_1) \leq \mathcal{S}(W_2)$. In particular, we consider the following order relations.

- *Stochastic dominance of order* 1: We say that the relation

$$W_1 \prec_{SD(1)} W_2$$

holds, iff for all monotonic utility functions U

$$\mathbb{E}[U(W_1)] \leq \mathbb{E}[U(W_2)]$$

- *Stochastic dominance of order* 2: We say that the relation

$$W_1 \prec_{SD(2)} W_2$$

holds iff for all concave, monotonic utility functions U

$$\mathbb{E}[U(W_1)] \leq \mathbb{E}[U(W_2)]$$

- *Monotonic dominance of order* 2: We say that the relation

$$W_1 \prec_{MD(1)} W_2$$

holds iff for all concave utility functions U

$$\mathbb{E}[U(W_1)] \leq \mathbb{E}[U(W_2)]$$

Artzner, Delbaen, Eber and Heath [5] call a risk measure *coherent*, if it is translation-invariant, convex, positively homogeneous and monotonic w.r.t. $\prec_{SD(1)}$.

The following table shows the properties of the three cited measures of safety.

	Q_α	C_α	SDCM$_\rho$
translation equivariant	+	+	+
positively homogeneos	+	+	+
convex	-	+	+
comonotone additive	+	-	-
monotonic w.r.t. SD(1)	+	+	-
monotonic w.r.t. SD(2)	-	+	-
monotonic w.r.t. MD(2)	-	+	+
coherent	-	+	-

Table 1. Properties of VaR, CVaR and SDCM.

The proofs of these properties can be found in [28] resp. [29].

2.1 Relations between the safety measures

If W is distributed according to a Gaussian distribution $N(m, \sigma^2)$, then the following relations hold:

- $$Q_\alpha(W) = m + \sigma\Phi^{-1}(\alpha), \tag{2}$$
  where Φ is the standard normal distribution function.
- $$C_\alpha(W) = m - \sigma\frac{1}{\alpha\sqrt{2\pi}}[1 - \exp(-[\Phi^{-1}(\alpha)]^2/2)]. \tag{3}$$
- $$\text{SDCM}_\rho(W) = m - \rho\sigma. \tag{4}$$
  Here ρ is a risk aversion factor.

For $\alpha < 0.5$, $\Phi^{-1}(\alpha) < 0$ and one sees that all three measures are of the form $m - b\sigma$, for $b > 0$. For distributions other than the normal, the three values measure quite different properties of the distribution.

However, the relation $C_\alpha(W) \leq Q_\alpha(W)$ holds always. Moreover, there is an important property of the $C_\alpha(W)$: It can be represented as the minimal value of some optimization problem [32]:

$$C_\alpha(W) = \sup\{a - \frac{1}{\alpha}\mathbb{E}[W - a]^- : a \in \mathbb{R}\} \tag{5}$$

or, equivalently,

$$-\ C_\alpha(W) = \inf\{-a + \frac{1}{\alpha}\mathbb{E}[W - a]^- : a \in \mathbb{R}\} \tag{6}$$

Here $[u]^- = -\min(u, 0)$.

Properties of risk measures, in particular VaR and CVaR, are obtained from properties of safety measures in an obvious fashion.

3 Optimal portfolios based on historical or simulated data

In this section we start to study the properties of portfolios which exhibit optimal VaR and CVaR properties, develop numerical approaches for finding such portfolios and compare them between themselves and with mean-variance optimal portfolios. There are two possible ways to approach this problem from the data point of view. One possibility is to approximate original historic data by parametrized probabilistic distribution, e.g. normal or lognormal distribution, and use this distribution in order to describe portfolio properties. Another possibility is to use historic data directly by setting up a sample distribution of portfolio returns on the basis of such data. In this paper we choose the second possibility. The strongest argument in favour of this choice is that both VaR and CVaR are sensitive to the tail properties of distributions and many parametrized families are notoriously bad in describing such properties of real data. Indeed, for the case of normal distribution VaR, CVaR and variance will generate the same efficient frontier, which is clear from (2)-(4).

Suppose the optimal portfolio must be found for a sample of historical or simulated return vectors $\xi^1, \xi^2, \ldots, \xi^N$. To put it differently, one may say that we now assume that the distribution of the wealth W takes the values $x^T\xi^i, i = 1, \ldots, N$, each with probability $1/N$. Let $e = \frac{1}{N}\sum_{i=1}^N \xi^i$ be the average return vector.

Using the three different risk measures VaR(W), CVaR(W) and Std(W), leads to quite differently structured optimization problems.

The Std optimization is a convex quadratic program:

$$\left\| \begin{array}{l} \text{Minimize (in } x) \; x^T \left[\frac{1}{N} \sum_{i=1}^{N} (\xi^i - e)(\xi^i - e) \right] x \\ \text{subject to} \\ x^T e \geq \mu \\ x^T \mathbb{1} = 1 \\ x \geq 0 \end{array} \right. \tag{7}$$

Using (6) one may write the CVaR optimization problem as a large scale linear program:

$$\left\| \begin{array}{ll} \text{Minimize (in } x, a \text{ and } z) & -a + \frac{1}{\alpha N} \sum_{i=1}^{N} z^i \\ \text{subject to} & \\ a - x^T \xi^i - z^i \leq 0 & i = 1, \ldots, N \\ x^T e \geq \mu & \\ x^T \mathbb{1} = 1 & \\ z^i \geq 0 & i = 1, \ldots, N \\ x \geq 0 & \end{array} \right. \tag{8}$$

For the definition of the VaR optimization problem introduce the function $M_{[k:N]}(u^1, \ldots, u^N)$ to denote the k-th largest among $u^1, \ldots, u^N$. Thus $M_{[1:N]}$ denotes the minimum and $M_{[N:N]}$ the maximum. The empirical α-quantile of the sample $x^T\xi^1, \ldots x^T\xi^N$ is

$$M_{[\lfloor \alpha N \rfloor :N]}(x^T\xi^1, \ldots, x^T\xi^N).$$

The VaR optimization is a nonconvex program, which may have many local minima.

$$\left\| \begin{array}{ll} \text{Minimize (in } x) & M_{[\lfloor \alpha N \rfloor :N]}(-x^T\xi^1, \ldots, -x^T\xi^N) \\ \text{subject to} & \\ x^T e \geq \mu & \\ x^T \mathbb{1} = 1 & \\ x \geq 0 & \end{array} \right. \tag{9}$$

Efficient commercial software based on several decades of algorithmic and theoretical research exists for solution of problems (7) and (8). Current situation with the problem (9) is the opposite one. This is due to the fact that the objective function in (9), generally speaking, is nonconvex and nondifferentiable. Little is known about properties of such problems and efficient general algorithms for their solution are nonexistent. What is possible, however, is to exploit the structure of (9) in order to develop efficient solution techniques specifically tailored for VaR optimization problem. This is one of the principal aims of the present paper.

In order to understand better the properties of (8) and its relation to CVaR optimization problem (8) we investigated the properties of historic VaR and CVaR as function of portfolio composition using stock market data. Here and in subsequent experiments we utilized the same data set which was used in [11] for evaluation of so-called universal portfolios and which contains more than 20 years of NYSE stock data for representative set of companies from different industry sectors. Typical patterns of VaR dependence on portfolio composition are presented on Figures 1,2.

Two portfolios were selected, say, x^1 and x^2, and a family of portfolios $x(\lambda)$ was considered which is defined by linear combination of these two portfolios:

$$x(\lambda) = \lambda x^1 + (1 - \lambda) x^2 \tag{10}$$

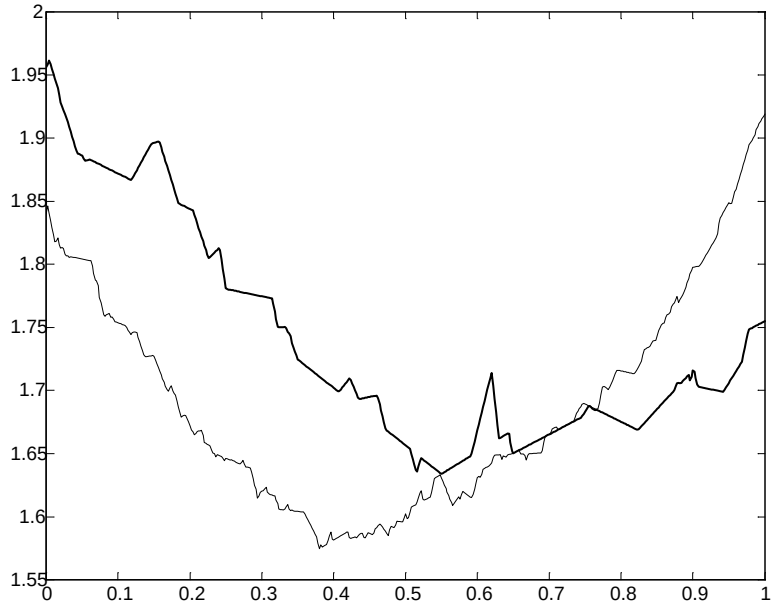

Figure 1: VaR of Ford/IBM portfolio

Both figures show sample VaR of portfolio $x(\lambda)$ as function of λ. Values of λ are shown on horisontal axis. Vertical axis represent values of VaR measured in percentage points of initial portfolio value. Sample VaR is computed after 1 trading day with 0.95 probability. Thick lines show sample VaR computed using data for 500 trading days, while thin lines are computed using 2000 trading days. On Figure 1 portfolio x^1 contains only Ford stock and portfolio x^2 contains only IBM stock. Three stocks were used on Figure 2: Schlumberger, Morris and Commercial Metals. Portfolios x^1 and x^2 were composed from these stocks in the following proportion:

$$x^1 = (0.51283,\ 0,\ 0.48717)\,, \quad x^2 = (0,\ 0.67798,\ 0.32202)\,.$$

The following obsevations about VaR behavior can be made on the basis of these experiments.

1. Sample VaR is very irregular function with multiple local minima and maxima evenly distributed in the function domain. VaR values in overwhelming majority of these minima are very far away from the optimal VaR value. The number of local minima grows with the number of observations and sooner or later a portfolio where VaR attains its locally minimal value can be found in the vicinity of every portfolio. Moreover, VaR function is nondifferentiable in every local minima. This means that straightforward application of standard techniques of nonlinear optimization to solution of problem (9) is practically hopeless. Usually such methods require differentiability of the objective function, which is not the case here. Even if a method can overcome this obstacle, it still will find only a local minima of VaR function. This will not bring us any closer to the solution of (9) because an arbitrary local minimum does not provide any information whatsoever about solution of VaR optimization problem.

2. Closer analysis shows, however, that VaR function possesses pronounced structure which can be exploited for solution of problem (9). In fact, even after a brief look at Figures 1,2 one

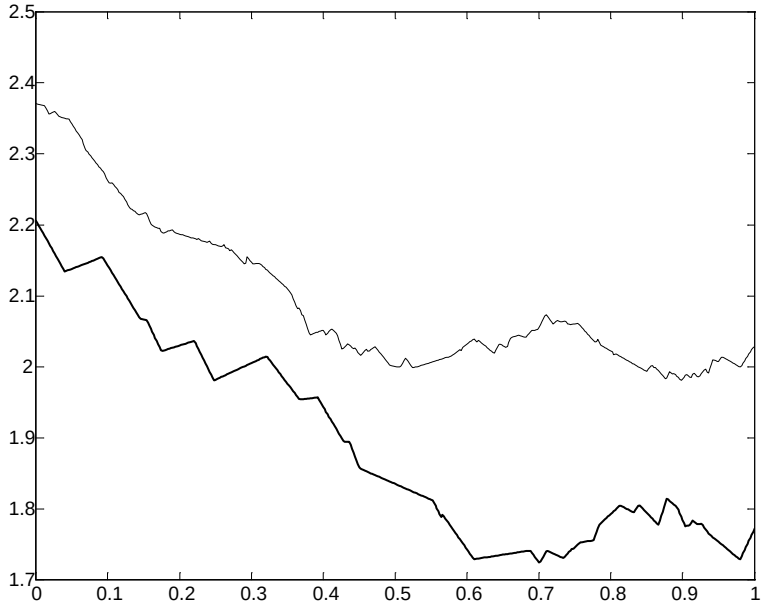

Figure 2: VaR of Schlumberger/Morris/Commercial Metals portfolio

can notice that VaR function is formed by interference of two patterns. The first pattern defines the global behavior of VaR function and it is a fairly regular pattern with pronounced global minimum and smooth behavior which in some cases is close to convex function, like on Figure 1. Even in the cases when its behavior does not seem to be convex, like on Figure 2, it still has only a few local minima with values which are close to the optimal values. The second pattern defines the local behavior. It is a highly irregular pattern which is responsible for almost all local minima and for nondifferentiability. When the number of observations increases, the local pattern becomes more and more irregular, but its magnitude becomes gradually smaller. As a result, the overall behavior of VaR function gradually comes closer to much more regular global pattern.

This observation provides a clue for our approach to solution of problem (9). It is centered around preprocessing of the VaR function. The objective of such preprocessing is to filter out or smooth out the irregular local pattern and extract the global pattern of VaR function. After that nonlinear optimization techniques can be efficiently utilized for solution of problem (9). This approach is explained in the next section.

Before proceeding further let us address another important question: what does empirical evidence say about relation between VaR and CVaR. Answer to this question has profound implication for the ways of approaching the solution of problem (9). If empirical evidence shows that solution of CVaR optimization problem (8) is a good approximation to solution of problem (9) then it is reasonable for all practical purposes to substitute the problem (9) by problem (9). This is because the problem (8) is just a linear programming problem which is much easier to solve than the problem (9). Unfortunately, empirical evidence shows that VaR and CVaR are substantially different risk measures and this substitution may bring about very misleading results.

In order to answer this question we proceeded similarly to (10). Two stocks were selected and two portfolios were considered: portfolio x^1 consisted only of stock 1 and portfolio x^2 consisted only of stock 2. VaR and CVaR were minimized on the set $L(x^1, x^2)$ of all linear combinations of these portfolios (10) parametrized with parameter λ with $0 \leq \lambda \leq 1$. Let us introduce some notations:

VaR(x) - VaR of portfolio x;

x_{VaR} - portfolio from the set $L(x^1, x^2)$ with the minimal value of VaR;

$\mathrm{VaR}_{\min}$=VaR(x_{VaR}) - minimal value of VaR on the set $L(x^1, x^2)$;

$\mathrm{VaR}_{\max}$ - maximal value of VaR on the set $L(x^1, x^2)$;

x_{CVaR} - portfolio from the set $L(x^1, x^2)$ with the minimal value of CVaR;

$R(x)$ - return of portfolio x;

$R_{\min}$ - minimal value of portfolio return on the set $L(x^1, x^2)$;

$R_{\max}$ - maximal value of portfolio return on the set $L(x^1, x^2)$.

Note that in this setting it is very easy to find both VaR-optimal and CVaR-optimal portfolios because in both cases we have one dimensional optimization problem to solve. Suppose now that VaR-optimal portfolio x_{VaR} is substituted by CVaR-optimal portfolio x_{CVaR}. By doing so we make two kinds of substitution errors. The first error E_{VaR} is related to the fact that such substitution leads to a larger value of VaR and is measured as follows:

$$E_{VaR} = \frac{\mathrm{VaR}(x_{\mathrm{CVaR}}) - \mathrm{VaR}_{\min}}{\mathrm{VaR}_{\max} - \mathrm{VaR}_{\min}} 100\%$$

If the cases when CVaR optimal portfolio is a good approximation of VaR optimal portfolio this error will be limited to a few percentage points. On the other hand for randomly selected portfolio such error will have an average value of around 50. The second error E_R measures the difference between returns of VaR-optimal and CVaR-optimal portfolios and is constructed similarly:

$$E_R = \frac{R(x_{\mathrm{CVaR}}) - R(x_{\mathrm{VaR}})}{R_{\max} - R_{\min}} 100\%$$

Our experiments show that in some cases CVaR-optimal portfolio approximates VaR-optimal portfolio reasonably well. On the other hand in many other cases CVaR-optimal portfolio is a poor substitution for VaR-optimal portfolio because the substitution error is not much better compared to the error of randomly selected portfolio. Some such cases are summarized in the following table.

Stocks	VaR error E_{VaR}	Return error E_R
Kodak-Merck	31.5	23.1
General Electric-IBM	27.2	13.9
Sears-Coca Cola	30.5	20.8
Ingersoll-Dow Chemical	17.2	22.3
Dupont-Exxon	20.9	-15.6
AHP-Kimberly-Clark	15.4	22.1
P&G-GTE	25.7	-25.3
Ford-HP	21.2	30.4
GM-P&G	32.7	29

Table 2. Errors committed by substitution of VaR by CVaR.

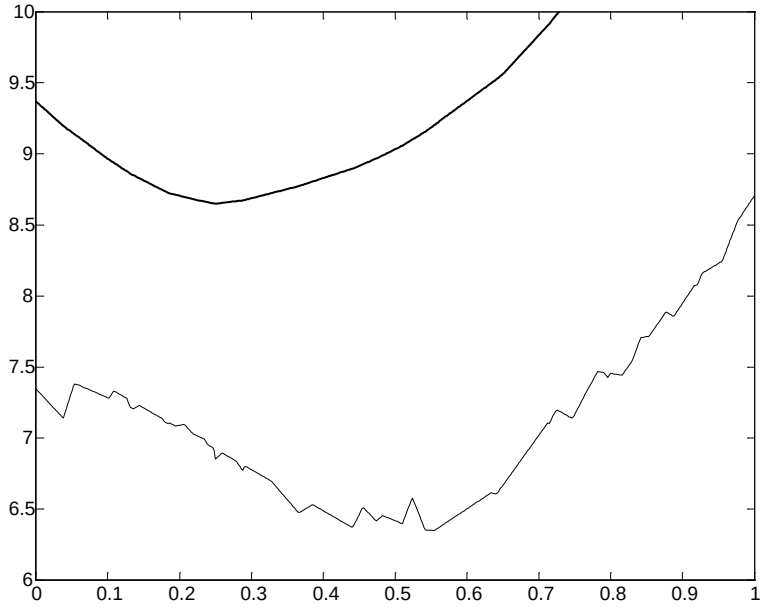

Figure 3: VaR/CVaR comparison for Ford/HP portfolio

In all cases presented in this table sample VaR and CVaR were computed for 10 days period with probability 0.95 using 500 observations. Figure 3 presents in more detail the case of Ford-Hewlett Packard portfolio from this table. This figure is constructed similarly to Figures 1,2 with the difference that on Figure 3 CVaR is presented by the thick line and VaR is presented by the thin line. Values of λ from (10) are shown on the horisontal axis. Note that CVaR is convex function, while VaR is not, while at the same time they obtain respective minima in considerably different regions.

More evidence supporting our claim that VaR and CVaR are two substantially different risk measures is presented in Section 5 where respective efficient frontiers are compared. The evidence presented so far serves as a motivation for the development of numerical approach tailored for computation of VaR-optimal portfolios which is presented in the next section.

4 Computation of VaR efficient portfolios

As we have seen in the last section, VaR function can be described as a sum of two components. The first component defines the global behavior and it is a fairly well behaved component which sometimes approaches convexity. It has pronounced global minimum and few if any local minima. Another component defines the local behavior and it is very noisy, nonsmooth and it has many local minima. It is the local component which makes VaR optimization a difficult task. This observation underlies our approach to numerical solution of VaR optimization problem (9) which consists of three steps.

1. Smooth out or filter out the local noisy component of the VaR function and extract well behaved global component. This is the key feature of our approach. The result is smoothed VaR function (SVAR).

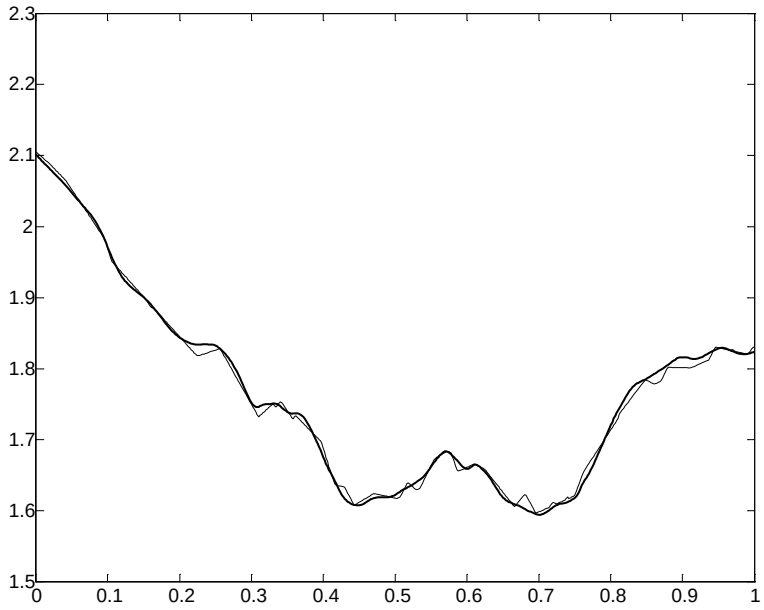

Figure 4: VaR (thin line) and smoothed VaR (thick line) for small value of smoothing parameter ϵ

2. After this the problem (9) becomes treatable by standard off-shelf software developed for solution of linear and nonlinear programming problems. We want to utilize this software as much as possible. This results in approximate solution of VaR optimization problem which in itself will be often sufficient for practical purposes.

3. Postprocessing of approximate solution. This is an optional step where commercial optimization software is used again.

In what follows these three steps are described in more detail.

4.1 Extracting the global behavior of VaR function: SVaR

This is the central part of our approach which aims at achieving two objectives. The first objective is to filter out the numerous local minima produced by noisy local component of VaR function. The second objective is to substitute original nonsmooth VaR function by approximate smooth function which can be treated by standard nonlinear programming techniques. It appears that both these objectives are achieved by special approximation technique developed in this paper.

In principle, general approximation techniques, like spline approximation, can be also used for approximation of VaR function. However, they require computational overhead which grows fast with dimension of portfolio. This makes such approaches difficult to utilize for problems of realistic dimension because the value of approximating function is going to be computed many times during optimization process. For this reason we had to develop approximation approach specifically tailored for exploitation of the special structure of the VaR function. General description of this approach is given in this section and its mathematical background is given in Appendix.

Let us simplify our notations and denote VaR function which is the objective of (9) by $V(x)$. Besides portfolio x it depends on return observations $\xi^1, ..., \xi^N$ and on probability value α, but we omit this dependence where it will not cause confusion.

We approximate original VaR function by the following family of *smoothed VaR functions* $V(\epsilon, x)$ (SVaR) parametrized by *smoothing parameter* ϵ, $\epsilon > 0$:

$$V(\epsilon, x) = \sum_{i=1}^{N} c_i^\epsilon(x) x^T \xi^i. \tag{11}$$

Coefficients $c_i^\epsilon(x)$ in (11) satisfy the following conditions:

1. $c_i^\epsilon(x)$ are twice continuously differentiable for all $\epsilon > 0$.
2. $c_i^\epsilon(x) \to 0$ as $\epsilon \to 0$ for every x and i for which $V(x) \neq x^T \xi^i$.
3. $c_i^\epsilon(x) \to 1/N$ as $\epsilon \to \infty$ for every x and i.
4. $\sum_{i=1}^{N} c_i^\epsilon(x) = 1$.

Under these conditions $V(\epsilon, x)$ is twice continuously differentiable for all $\epsilon > 0$. This property is important because $V(\epsilon, x)$ is going to be minimized by standard nonlinear programming algorithms, and they usually require that the objective function has this property. Another consequence of conditions 1-4 is that $V(\epsilon, x) \to V(x)$ as $\epsilon \to 0$ and

$$V(\epsilon, x) \to \frac{1}{N} \sum_{i=1}^{N} x^T \xi^i$$

as $\epsilon \to \infty$ which is just the average return of portfolio x. Thus, we have the whole range of approximations from very precise approximations for small values of ϵ which, however, leave untouched undesired properties of original VaR function, to very well behaved, almost linear approximations which, however, retain little information about original VaR function. Therefore the choice of smoothing parameter ϵ is governed by tradeoff between two conflicting objectives of having precise approximation and having smooth approximation with as few local minima as possible.

A large family of possible selections of coefficients $c_i^\epsilon(x)$ from (11) is described in Appendix (see (23)-(24) and Theorem 2). It turns out that each specific expression for $c_i^\epsilon(x)$ is defined by a certain function $\varphi_\epsilon(z)$ of one dimensional parameter z which provides a smooth approximation for a unit step function $\varphi(z)$:

$$\varphi(z) = \begin{cases} 1 & \text{if} \quad z \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Function $\varphi_\epsilon(z)$ which we used in our experiments is defined in (25) and shown on Figure 9. This choice is dictated by computational considerations. Resulting smoothed VaR function $V(\epsilon, x)$ requires a moderate computational overhead compared with original VaR function $V(x)$. The important fact is that this overhead does not depend on the number of positions in portfolio x and grows relatively slow with the number of return observations N (see Theorem 5). Figures 4 and 5 show a typical example of smoothed VaR function for different values of smoothing parameter ϵ.

These figures similarly to Figures 1-3 show dependence of VaR function $V(x)$ (thin line) on parameter λ which defines portfolio $x(\lambda)$ as linear combination of portfolios x^1 and x^2 obtained according to (10). Portfolios x^1 and x^2 consisted of three stocks (Schlumberger, Morris,

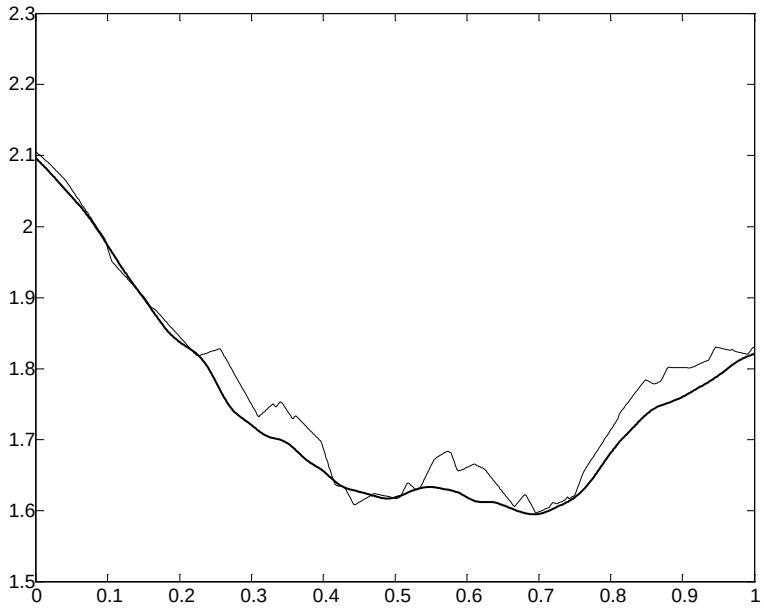

Figure 5: VaR (thin line) and smoothed VaR (thick line) for larger value of smoothing parameter ϵ

Commercial Metals) in the following proportion:

$$x^1 = (0.69668, 0, 0.30332),\ x^2 = (0, 0.64624, 0.35376)$$

VaR was computed on the basis of 500 daily observations for probability 0.95. Thick line on these figures represents smoothed VaR computed for different values of smoothing parameter ϵ. Figure 4 shows the case of small value of smoothing parameter $\epsilon = 0.001$. In this case smoothed VaR $V(\epsilon, x)$ follows regular VaR very well, although even in this case the smoothing managed to cut off many irrelevant local minima. The case of larger value of $\epsilon = 0.004$ is represented on Figure 5. Here local noisy component of VaR is filtered out completely and only global well behaved component remains.

Smoothing approach presented here is implemented in MATLAB environment [24] and presents a robust tool for approximation of sample VaR function by smooth function with pronounced global minimum.

4.2 Minimizing smoothed VaR function

Efficient nonlinear programming software present in the market can be used for finding VaR-efficient portfolios by minimizing smoothed VaR function. In our experiments we used `fmincon` subroutine from MATLAB Optimization Toolbox which is fairly reliable and fast solver for problems of medium size. Another option is to use nonlinear programming solvers available in GAMS environment [8].

4.3 Postprocessing

The result of the previous step is portfolio x_{SVAR} which is either a global minimizer or a good local minimizer of smoothed VaR function. In most cases this will be enough for practical

purposes. However, if one wish to improve on this, it can be done taking as a starting point portfolio $x_{\rm SVAR}$. Let us consider two possibilities for doing this.

1. *Local minimization of VaR function.*

In order to do this is is enough to solve linear programming problem of small to medium size which can be done very fast by current commercial codes. We have found that subroutine `linprog` from MATLAB Optimization Toolbox is sufficient for this purpose. Again, solvers present in GAMS environment present another good choice.

Let us formulate postprocessing LP. According to definition of sample VaR $V(x)$ there exists index j which depends on $x_{\rm SVAR}$ such that $V(x) = x_{\rm SVAR}^T \xi^j$ and inequality

$$x_{\rm SVAR}^T \xi^i \leq x_{\rm SVAR}^T \xi^j \tag{12}$$

is satisfied for at least N^α observations $\xi^i, i \neq j$, where N^α is the smallest integer which is not smaller then $N(1-\alpha)$ and α is a probability for which VaR is computed. Let us denote by Λ an arbitrary set of N^α indices for which inequality 12 is satisfied. Then the solution of the following linear programming problem is a local minimum of VaR function $V(x)$ and provides better value of VaR then $x_{\rm SVAR}$.

$$\min_x \; x^T \xi^j \tag{13}$$

$$x^T \left(\xi^i - \xi^j\right) \leq 0, \; i \in \Lambda \tag{14}$$

$$x^T \left(\xi^i - \xi^j\right) \geq 0, \; 1 \leq i \leq N, \; i \notin \Lambda \tag{15}$$

$$x^T e \geq \mu \tag{16}$$

$$x^T \mathbf{1} = 1, \; x \geq 0 \tag{17}$$

Our experience is that improvement of VaR value obtained through solution of this problem is relatively small.

2. *Global minimization of VaR function.*

Using notations introduced above we can reformulate VaR optimization problem (9) as follows:

$$\min_{j,\Lambda,x} \; x^T \xi^j \tag{18}$$

subject to constraints (14)-(17) where minimization is performed with respect to all x and all pairs (j, Λ) where $1 \leq j \leq N$ and set Λ contains exactly N^α elements among integers $\{1, ..., N\} \setminus \{j\}$. This is a difficult mixed integer programming problem and it is practically impossible to solve it to optimality for problems of realistic dimension without having good information about the optimal solution. However, minimum of smoothed VaR function $x_{\rm SVAR}$ provides just this information. For this reason (18) can be used for improvement of solution $x_{\rm SVAR}$. In our experiments we have not used this approach because minimization of smoothed VaR together with postprocessing according to (13)-(17) gave results which were already good.

5 Comparison of efficient frontiers

Now we have in place all tools necessary for utilization of VaR as a criterion for optimal investment. This is done following classical Markowitz approach [23]: VaR is minimized for different values of return μ by solving problem (3) and mean-VaR feasible set is constructed from which mean-VaR efficient frontier is derived. After this an investor with specific risk preferences can choose the target value of VaR and select portfolio on efficient frontier that provides the best return for a given value of VaR.

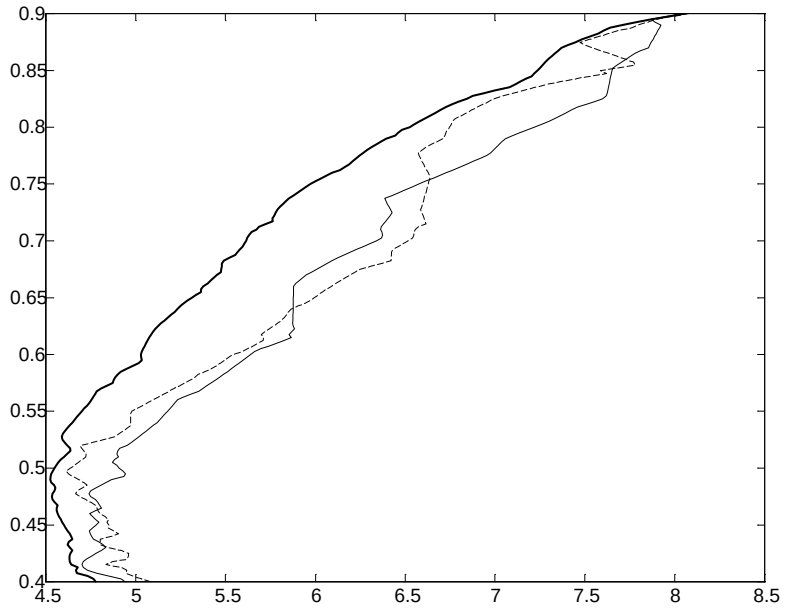

Figure 6: Boundary of mean-VaR feasible set (thick) and images of mean-CVaR boundary (thin) and mean-variance boundary (dashed)

But how such mean-VaR efficient frontier compares with classical mean-variance efficient frontier? If the difference is small there is no point in abandoning classical mean-variance approach in favour of VaR. In order to answer this question we conducted a study with the same set of stock market data which was used in previous experiments. The purpose was to compare mean-variance, mean-CVaR and mean-VaR efficient frontiers obtained on the basis of historical data by solution of problems (7),(8) and (9) for different values of return μ. Each computation of efficient frontier involved 200-500 solutions of respective optimization problem. Problems (7) and (1) are standard quadratic and linear programming problems respectively that can be solved by wide choice of available commercial software. We have found MATLAB optimization toolbox sufficient for our purposes. VaR optimization problem (9) was solved using smoothing described in the previous section. The whole implementation consists of a system of MATLAB M-files which use solvers `fmincon` (nonlinear programming) and `linprog` (LP) from MATLAB Optimization Toolbox.

Results of a typical experiment are presented on Figures 6,7 and 8.

They depict boudaries of sample mean/VaR, mean/CVaR and mean/variance feasible sets computed for 500 ten days return observations for the set of stocks which included Texaco, Ingersoll, Kodak, Fischbach, Gulf and Commercial Metals. Probability 0.95 was used for computation of VaR and CVaR. Efficient frontiers were computed using 201 equidistant return values.

Figure 6 shows boundary of mean-VaR feasible set with thick line. Besides, it shows images of boundaries of mean-CVaR and mean-variance feasible sets in mean-VaR space with thin and dashed lines respectively. In order to obtain these images we computed VaR of portfolios that form mean-CVaR and mean-variance efficient frontiers and placed resulting points on Figure

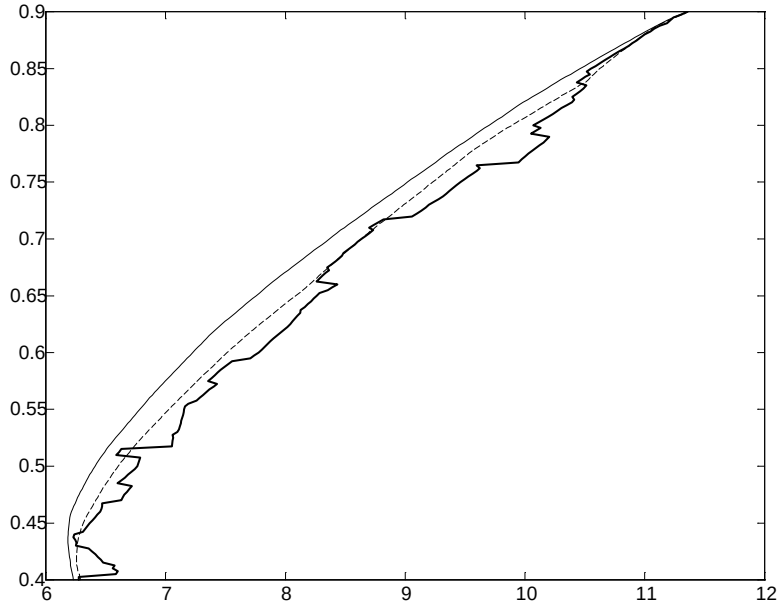

Figure 7: Boundary of mean-CVaR feasible set (thin) and images of mean-VaR boundary (thick) and mean-variance boundary (dashed)

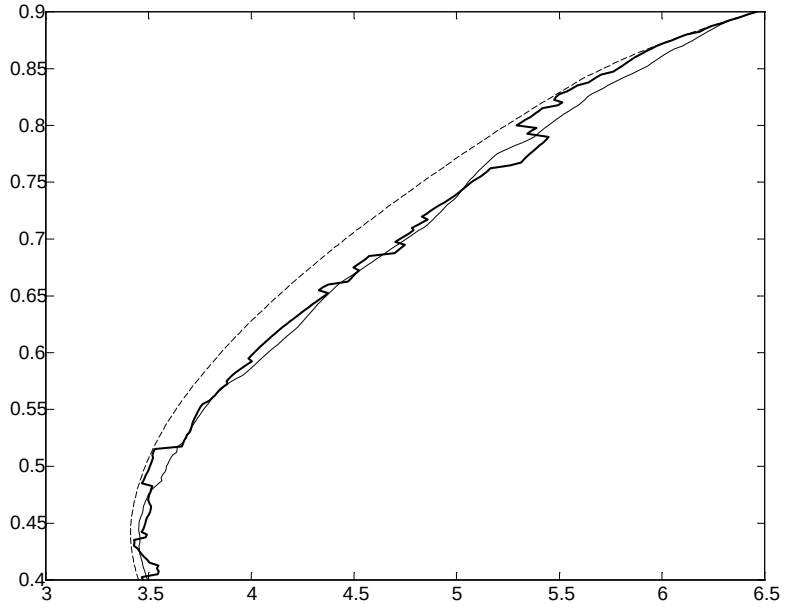

Figure 8: Boundary of mean-variance feasible set (dashed) and images of mean-VaR boundary (thick) and mean-CVaR boundary (thin)

6. Horisontal axis on this figure shows VaR values expressed in percents relative to original portfolio value. Vertical axis shows expected rate of return after 10 trading days expressed in percents. Efficient frontiers are easily obtained from boundaries of feasible sets.

Figure 7 shows how boundaries of feasible sets and efficient frontiers look in mean-CVaR space. Thus, mean-CVaR efficient frontier is depicted with thin line and images of mean-VaR and mean-variance efficient frontiers in mean-CVaR space are shown with thick and dashed lines respectively. The meaning of axes is similar to Figure 6 except that horisontal axis shows CVaR values.

Figure 8 shows how boundaries of feasible sets and efficient frontiers look in mean-standard deviation space. Thus, mean-standard deviation efficient frontier is depicted with dashed line and images of mean-VaR and mean-CVaR efficient frontiers in mean-standard deviation space are shown with thick and thin lines respectively. The meaning of axes is similar to Figure 6 except that horisontal axis shows standard deviation values expressed in percents.

The following conclusions can be drawn from this and other similar experiments.

1. Mean-VaR feasible set and respective efficient frontier differs from traditional efficient frontiers in the sense thatthey are not convex. However, its distance from convex shape is relatively small and as the first approximation it can be considered ”almost” convex, especially if compared to irregular behavior of VaR function itself as shown on Figures 1-5.

2. Taken in optimal portfolio selection context, VaR differs substantially from both CVaR and variance because mean-VaR efficient portfolios may differ considerably from mean-CVaR of mean-variance efficient portfolios. Figures 6-8 show several phenomena which point in this direction. Firstly, the distance between VaR-efficient frontier and images of other frontiers is far larger then distance between frontiers in mean-CVaR space or mean-variance space. In other words, mean-CVaR and mean-variance efficient portfolios provide a poor approximation to mean-VaR efficient portfolios in mean-VaR space, while mean-VaR efficient portfolios provide much better approximation to mean-CVaR and mean-variance portfolios in their respective spaces. Secondly, mean-CVaR efficient portfolios do not approximate mean-VaR efficient portfolios any better than mean-variance efficient portfolios do. Actually, in some cases mean-variance portfolios lie closer to mean-VaR portfolios than mean-CVaR portfolios, as Figure 6 suggests. Vice versa, mean-variance portfolios can approximate mean-CVaR portfolios better than mean-VaR portfolios do (see Figure 7).

3. For high risk portfolios all frontiers approximate each other fairly well, while for medium and small risk portfolios frontiers can differ substantially, especially for medium risk portfolios.

6 Summary

We studied here Value-at-Risk (VaR) in the context of VaR optimization for the purpose of selection of mean-VaR efficient portfolios similar to classical mean-variance approach. Important theme of this paper is comparison of VaR with other important risk measures from optimization point of view, in particular with classical variance and more recent conditional VaR (CVaR). This comparison is performed both on theoretical plane and computational plane using the stock market data.

Our conclusion is that VaR is a substantially different risk measure and investor which expresses her risk preferences in terms of VaR should work with VaR directly in the context of mean-risk tradeoff. In particular, efficient frontiers constructed on the basis of the other risk

measures can be a poor approximation for mean-VaR efficient frontier.

Moreover, we argue that computation of mean-VaR efficient portfolios based on historic data is a feasible task, despite the fact that VaR optimization is more difficult than variance optimization or CVaR optimization. For this purpose we developed a set of VaR optimization tools centered around the notion of smoothed VaR (SVaR) which filters out nonsmooth irregular local behavior of historic VaR. Smoothed VaR can be efficiently optimized using current commercial nonlinear programming software.

7 Appendix: construction of SVaR

Here we provide mathematical details about smoothing technique which was used in section 4 for computing of VaR efficient portfolios.

Let us consider a finite collection of functions $f_i(x)$, $i = 1, ..., N$ defined on some set $X \subseteq \mathbb{R}^n$. Let us fix k, $0 \leq k \leq N-1$ and define function $F(k, x)$ which for fixed x equals function $f_j(x)$:

$$F(k, x) = f_j(x), \, j = j(x) \tag{19}$$

where index $j = j(x)$ is defined by the following two conditions:

1. Inequality $f_j(x) \leq f_i(x)$ is satisfied for at least k functions $f_i(x)$, $i \neq j$.
2. Inequality $f_j(x) \geq f_i(x)$ is satisfied for at least $N-k-1$ functions $f_i(x)$, $i \neq j$.

In these notations

$$F(N-1, x) = \min_{1 \leq i \leq N} f_i(x), \; F(0, x) = \max_{1 \leq i \leq N} f_i(x)$$

Taking $f_i(x) = x^T \xi^i$ we obtain immediately connection with problem (9) since

$$M_{[\lfloor \alpha N \rfloor : N]}(x^T \xi^1, \ldots, x^T \xi^N) = F(N - \lfloor \alpha N \rfloor, x)$$

Function $F(k, x)$ will be nondifferentiable for all k even if functions $f_i(x)$ are arbitrarily smooth. Suppose that $f_i(x)$ are twice continuously differentiable. Our objective is to construct approximation $F_\epsilon(k, x)$ which depends on parameter ϵ such that

1. $F_\epsilon(k, x)$ is twice continuously differentiable for all $\epsilon > 0$;
2. $F_\epsilon(k, x) \to F(k, x)$ as $\epsilon \to 0$.

One possible way to obtain such approximation is a straightforward utilization of general approximation techniques, like spline approximation. However, this approach is impractical because it leads to exponential increase in computational complexity with respect to dimension of x. We develop here another approach which exploits specific structure of the function $F(k, x)$ and results in computational requirements which do not depend on dimension of x once values of $f_i(x)$ are computed. For example, if $f_i(x) = x^T \xi^i$ then the total computational requirements for computing $F_\epsilon(k, x)$ at one point grow linearly with dimension of x. Our approximation is based on the representation of the function $F(k, x)$ as a linear combination of the composite functions $f_i(x)$ where coefficients in the linear combination depend on x:

$$F(k, x) = \sum_{i=1}^{N} c_i(x) f_i(x) \tag{20}$$

In order to derive expressions for coefficients $c_i(x)$ we need the following notations:

M^i - set of all integers from 1 to N except i: $M^i = \{1, 2, ..., N\} \setminus \{i\}$;
Λ_k^i - arbitrary subset of M^i which contains exactly k elements;
$X\left(\Lambda_k^i\right)$ - subset of $\mathbb{R}^n$ associated with set Λ_k^i as follows:

$$X\left(\Lambda_k^i\right) = \left\{x : f_i(x) \leq f_j(x) \text{ for } j \in \Lambda_k^i,\, f_i(x) \geq f_j(x) \text{ for } j \in M^i \setminus \Lambda_k^i\right\}$$

Θ_k^i - family of all different sets Λ_k^i;
$\mathbb{I}_A$ - indicator function of set A, i.e.

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{if} \quad x \in A \\ 0 & \text{otherwise} \end{cases}$$

In order to simplify notations we shall omit dependence of $\mathbb{I}_A$ on x where it will not cause confusion. The following lemma gives required representation of function $F(k, x)$.

Lemma 1 *Suppose that $F(k, x)$ is defined by (19). Then*

$$F(k, x) = \frac{1}{C(x)} \sum_{i=1}^{N} c_i(x) f_i(x), \tag{21}$$

where

$$c_i(x) = \sum_{\Lambda_k^i \in \Theta_k^i} \mathbb{I}_{X\left(\Lambda_k^i\right)},\; C(x) = \sum_{i=1}^{N} c_i(x) \tag{22}$$

Proof.

Suppose that $F(k, x) = f_r(x)$. We are going to prove that also the right hand side of (21) equals $f_r(x)$ provided coefficients $c_i(x)$ are selected according to (22).

Let us consider an arbitrary i for which $f_i(x) > f_r(x)$. Then inequality $f_i(x) \geq f_j(x)$ is satisfied for at least $N - k - 1$ functions $f_j(x)$ due to definition of function $F(k, x)$. Consequently, inequality $f_i(x) \leq f_j(x)$ is satisfied for $k_1 < k$ functions $f_j(x)$ because from $k_1 = k$ would follow $f_i(x) = f_r(x)$. This means that $x \notin X\left(\Lambda_k^i\right)$ for arbitrary Λ_k^i because due to definition of set $X\left(\Lambda_k^i\right)$ we have $f_i(x) \leq f_j(x)$ for at least k functions $f_j(x)$. Therefore $\mathbb{I}_{X\left(\Lambda_k^i\right)} = 0$ for arbitrary Λ_k^i and, consequently, $c_i(x) = 0$. By similar argument we obtain $c_i(x) = 0$ also when $f_i(x) > f_r(x)$. Thus, $c_i(x)$ can differ from zero only if $f_i(x) = f_r(x)$. Therefore

$$\frac{1}{C(x)} \sum_{i=1}^{N} c_i(x) f_i(x) = f_r(x) \frac{1}{C(x)} \sum_{i=1}^{N} c_i(x) = f_r(x)$$

due to (22). The proof is completed.□

We are ready now to formulate our main approximation result which defines a family of smooth approximations of function $F(k, x)$. It is presented in the following theorem.

Theorem 2 *Suppose that $f_i(x)$ are twice continuously differentiable and $\varphi_\epsilon(z)$ is an arbitrary function defined for $z \in \mathbb{R}^1$ and $\epsilon \geq 0$ such that*

1. $\varphi_\epsilon(z)$ is twice continuously differentiable for $\epsilon > 0$.

2. $\varphi_\epsilon(z) \to 1$ as $\epsilon \to 0$ for any fixed $z \leq 0$.

3. $\varphi_\epsilon(z) \to 0$ *as* $\epsilon \to 0$ *for any fixed* $z > 0$.
4. $\varphi_\epsilon(z) \geq 0$ *for all* $\epsilon \geq 0, z$ *and* $\varphi_\epsilon(z) \geq \varkappa_0$ *for some* $\varkappa_0 > 0$ *and all* $\epsilon \geq 0, z \leq 0$.
Then function $F_\epsilon(k, x)$ *defined as follows:*

$$F_\epsilon(k,x) = \frac{1}{C_\epsilon(x)} \sum_{i=1}^{N} c_i^\epsilon(x) f_i(x), \tag{23}$$

$$c_i^\epsilon(x) = \sum_{\Lambda_k^i \in \Theta_k^i} \prod_{j \in \Lambda_k^i} \varphi_\epsilon\left(\Delta_{ij}^x\right) \prod_{j \in M^i \setminus \Lambda_k^i} \varphi_\epsilon\left(-\Delta_{ij}^x\right), \tag{24}$$

$$\Delta_{ij}^x = f_i(x) - f_j(x),\ C_\epsilon(x) = \sum_{i=1}^{N} c_i^\epsilon(x)$$

is twice continuously differentiable for all $\epsilon > 0$ *and such that* $F_\epsilon(k,x) \to F(k,x)$ *as* $\epsilon \to 0$ *for any fixed* x.

Proof.

Let us fix x and suppose that r is such that $F(k,x) = f_r(x)$. Then from (19) follows that there exists set $\Lambda_k^r \in \Theta_k^r$ such that $f_r(x) - f_j(x) \leq 0$ for $j \in \Lambda_k^r$ and $f_j(x) - f_r(x) \geq 0$ for $j \in M^r \setminus \Lambda_k^r$. Therefore due to condition 4 we have:

$$C_\epsilon(x) \geq c_r^\epsilon(x) \geq \prod_{j \in \Lambda_k^i} \varphi_\epsilon\left(\Delta_{ij}^x\right) \prod_{j \in M^i \setminus \Lambda_k^i} \varphi_\epsilon\left(-\Delta_{ij}^x\right) \geq \varkappa_0^{N-1} > 0$$

where the last estimate does not depend on x. This together with differentiability properties of $f_i(x)$ and $\varphi_\epsilon(z)$ yield existence and continuity of gradient and Hessian of $F_\epsilon(k,x)$ for arbitrary x and $\epsilon > 0$.

Observe now that $\varphi_\epsilon(f_i(x) - f_j(x)) \to \mathbb{I}_{\{f_i(x) \leq f_j(x)\}}$ and $\varphi_\epsilon(f_j(x) - f_i(x)) \to \mathbb{I}_{\{f_i(x) \geq f_j(x)\}}$ for arbitrary x and $\epsilon \to 0$ due to conditions 2,3. Therefore

$$\prod_{j \in \Lambda_k^i} \varphi_\epsilon\left(\Delta_{ij}^x\right) \prod_{j \in M^i \setminus \Lambda_k^i} \varphi_\epsilon\left(-\Delta_{ij}^x\right) \to \prod_{j \in \Lambda_k^i} \mathbb{I}_{\{f_i(x) \leq f_j(x)\}} \prod_{j \in M^i \setminus \Lambda_k^i} \mathbb{I}_{\{f_i(x) \geq f_j(x)\}} = \mathbb{I}_{X\left(\Lambda_k^i\right)}$$

which together with Lemma 1 yields $c_i^\epsilon(x) \to c_i(x)$, $C_\epsilon(x) \to C(x)$ and finally $F_\epsilon(k,x) \to F(k,x)$ for $\epsilon \to 0$ and arbitrary x. The proof is completed. □

By selecting a specific function family $\varphi_\epsilon(z)$ specific smooth approximation of nondifferentiable function $F(k,x)$ can be obtained. Under additional technical assumptions this approximation will converge uniformly to the original function. Selection of $\varphi_\epsilon(z)$ should take into account computational considerations. For general functions $\varphi_\epsilon(z)$ computation of $c_i^\epsilon(x)$ from (24) can be a difficult task because the number of terms in the sum from (24) grows exponentially with the number of observations N, more precisely it equals

$$\frac{(N-1)!}{k!\,(N-k-1)!}.$$

The number of nonzero terms can be drastically reduced by special selection of function family $\varphi_\epsilon(z)$, namely by considering only functions with the property $\varphi_\epsilon(z) = 0$ if $z > \epsilon$. The simplest

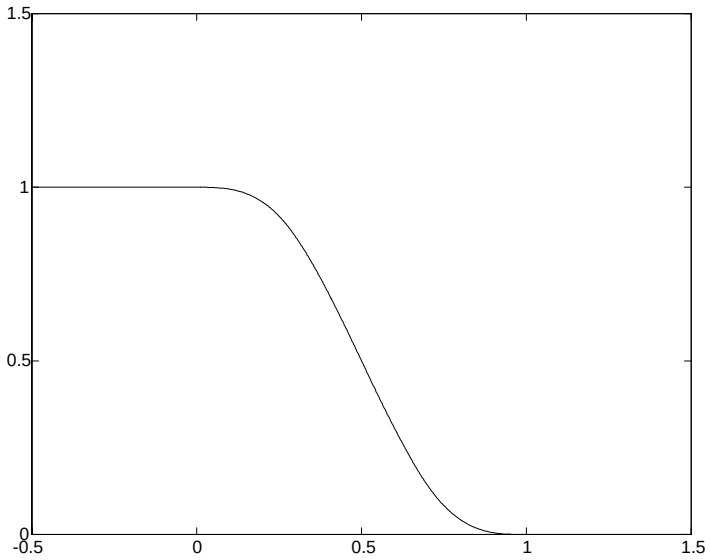

Figure 9: Function $\varphi_\epsilon(z)$ for $\epsilon = 1$.

such function that satisfies conditions of Theorem 2 is the cubic spline:

$$\varphi_\epsilon(z) = \begin{cases} 1 & \text{if} \quad z \le 0 \\ 1 - \frac{16}{3\epsilon^3} z^3 & \text{if} \quad 0 \le z \le \frac{\epsilon}{4} \\ \frac{5}{6} + \frac{2}{\epsilon} z - \frac{8}{\epsilon^2} z^2 + \frac{16}{3\epsilon^3} z^3 & \text{if} \quad \frac{\epsilon}{4} \le z \le \frac{3\epsilon}{4} \\ \frac{16}{3} - \frac{16}{\epsilon} z + \frac{16}{\epsilon^2} z^2 - \frac{16}{3\epsilon^3} z^3 & \text{if} \quad \frac{3\epsilon}{4} \le z \le \epsilon \\ 0 & \text{if} \quad z \ge \epsilon \end{cases} \tag{25}$$

This function is depicted on Figure 9 for $\epsilon = 1$. The benefit of using such functions becomes clear from the following lemmas.

Lemma 3 *Suppose that in addition to conditions of Theorem 2 function $\varphi_\epsilon(z)$ is such that $\varphi_\epsilon(z) = 0$ if $z > \epsilon$. Then the approximation function $F_\epsilon(k, x)$ can be equivalently represented as follows:*

$$F_\epsilon(k, x) = \frac{1}{C_\epsilon(x)} \sum_{i:|F(k,x)-f_i(x)|\le\epsilon} c_i^\epsilon(x) f_i(x) \tag{26}$$

Proof.

It is enough to prove that $c_i^\epsilon(x) = 0$ if $|F(k, x) - f_i(x)| > \epsilon$. Let us fix i and consider the case when $f_i(x) - F(k, x) > \epsilon$. Let us select an arbitrary set of indices $\Lambda_k^i \in \Theta_k^i$. Suppose that $q \in \Lambda_k^i$ and $f_q(x) = \min_{j\in\Lambda_k^i} f_q(x)$. Then $f_q(x) \le F(k, x)$ due to definition of $F(k, x)$. Therefore

$$f_i(x) - f_q(x) \ge f_i(x) - F(k, x) > \epsilon$$

and, consequently, $\varphi_\epsilon(f_i(x) - f_j(x)) = 0$. Since Λ_k^i is arbitrary this means that every term in the sum (24) which defines $c_i^\epsilon(x)$ equals zero and $c_i^\epsilon(x) = 0$. The case when $F(k, x) - f_i(x) > \epsilon$ is treated similarly. □

Thus, in order to compute $F_\epsilon(k, x)$ it is enough to consider only those functions $f_i(x)$ for which $|F(k, x) - f_i(x)| \le \epsilon$ which reduces considerably computational effort. Even though

this approach makes possible to compute the values of $F_\epsilon(k,x)$ for large N, still considerable care is needed in implementation of (23)-(24) taking into account that $F_\epsilon(k,x)$ is going to be computed many times during optimization process. Let us derive equivalent expression for coefficients $c_i^\epsilon(x)$, keeping in mind computational requirements. We shall consider the case when $\varphi_\epsilon(z)=1$ if $z\le 0$ similar to (25).

Lemma 4 *Suppose that in addition to conditions of Lemma 3 function $\varphi_\epsilon(z)$ is such that $\varphi_\epsilon(z)=1$ if $z\le 0$. Then coefficients $c_i^\epsilon(x)$ from (24) can be represented equivalently as follows:*

$$c_i^\epsilon(x)=\sum_{q,r:\ q-r=k-k^i_-} b_q^- b_r^+ \tag{27}$$

$$b_q^-=\sum_{\Lambda_q\subseteq\Lambda^i_{\epsilon-}}\prod_{j\in\Lambda_q}\varphi_\epsilon\left(\Delta_{ij}^x\right),\ b_r^+=\sum_{\Lambda_r\subseteq\ \Lambda^i_{\epsilon+}}\prod_{j\in\Lambda_r}\varphi_\epsilon\left(-\Delta_{ij}^x\right) \tag{28}$$

where Λ_r and Λ_q are arbitrary sets consisting of r and q elements respectively,

$$\Lambda^i_{\epsilon-}=\left\{j:0<f_j(x)-f_i(x)\le\epsilon,\ j\in M^i\right\},\ \Lambda^i_{\epsilon+}=\left\{j:0\le f_i(x)-f_j(x)\le\epsilon,\ j\in M^i\right\}$$

and k^i_- is the number of elements in set $\Lambda^i_{\epsilon-}$.

Proof.

Let us fix x, select i and denote by Λ^i_- the set of all indices $j=1,...,N$ such that $f_i(x)<f_j(x)$. Suppose that k_i is the number of elements in this set. Observe that an arbitrary set $\Lambda^i_k\in\Theta^i_k$ can be represented as follows:

$$\Lambda^i_k=\left(\Lambda^i_-\setminus\Lambda_r\right)\cup\Lambda_q,\ M^i\setminus\Lambda^i_k=\left(\left(M^i\setminus\Lambda^i_-\right)\setminus\Lambda_q\right)\cup\Lambda_r$$

where Λ_r and Λ_q are arbitrary sets containing r and q elements respectively and such that

$$\Lambda_r\subseteq\Lambda^i_-,\ \Lambda_q\subseteq M^i\setminus\Lambda^i_-,\ k_i+q-r=k$$

Since $\varphi_\epsilon\left(\Delta_{ij}^x\right)=1$ if $j\in\Lambda^i_-\setminus\Lambda_r$ and $\varphi_\epsilon\left(-\Delta_{ij}^x\right)=1$ if $j\in M^i\setminus\Lambda^i_-\setminus\Lambda_q$ we can transform expression (24) for coefficient $c_i^\epsilon(x)$ as follows:

$$\begin{aligned}c_i^\epsilon(x) &= \sum_{\substack{\Lambda_r\subseteq\Lambda^i_-\\ \Lambda_q\subseteq M^i\setminus\Lambda^i_-\\ k_i+q-r=k}}\prod_{j\in\Lambda_q}\varphi_\epsilon\left(\Delta_{ij}^x\right)\prod_{j\in\Lambda_r}\varphi_\epsilon\left(-\Delta_{ij}^x\right)\\ &= \sum_{q,r:\ q-r=k-k_i}\left(\sum_{\Lambda_q\subseteq\Lambda^i_-}\prod_{j\in\Lambda_q}\varphi_\epsilon\left(\Delta_{ij}^x\right)\right)\left(\sum_{\Lambda_r\subseteq M^i\setminus\Lambda^i_-}\prod_{j\in\Lambda_r}\varphi_\epsilon\left(-\Delta_{ij}^x\right)\right)\end{aligned}$$

Assertion of lemma is obtained from the last expression repeating the argument of Lemma 3 which leads to substitution of Λ^i_- by $\Lambda^i_{\epsilon-}$, $M^i\setminus\Lambda^i_-$ by $\Lambda^i_{\epsilon+}$ and k_i by k^i_-. □

Expressions (27),(28) allow efficient computation of coefficients $c_i^\epsilon(x)$ because b_q^- and b_r^+ can be computed recursively. Indeed, suppose that $s\in\Lambda^i_{\epsilon-}$. Then

$$b_q^-=\varphi_\epsilon\left(\Delta_{iv}^x\right)\sum_{\Lambda_{q-1}\subseteq\Lambda^i_{\epsilon-}\setminus\{s\}}\prod_{j\in\Lambda_{q-1}}\varphi_\epsilon\left(\Delta_{ij}^x\right)+\sum_{\Lambda_q\subseteq\Lambda^i_{\epsilon-}\setminus\{s\}}\prod_{j\in\Lambda_q}\varphi_\epsilon\left(\Delta_{ij}^x\right)$$

The following algorithm utilizes the last expression in order to compute b_q^- for all $q = 1, ..., q_{\max}$, $q_{\max} \leq k_-^i$.

Algorithm 1.

1. *Initialization.* Select an arbitrary ordering $j_1, ..., j_{k_-^i}$ of elements of set $\Lambda_{\epsilon-}^i$ and denote

$$d_s = \varphi_\epsilon \left(\Delta_{ij_s}^x \right), \ s = 1, ..., k_-^i.$$

Take $e_s = 1$, $s = 1 : k_-^i + 1$.

2. *Computation of* b_q^- . Starting from $q = 1$ perform consecutively for each $q = 1, ..., q_{\max}$:

2a. Starting from $s = q$ compute consecutively for each $s = q, ..., k_-^i$:

- Compute a from expression

$$a = \begin{cases} d_s e_s & \text{if} \quad s = q \\ d_s e_s + \bar{a} & \text{if} \quad s > q \end{cases}$$

- Take $e_s = \bar{a}$ if $s > q$.

- Take $\bar{a} = a$ if $s \geq q$.

2b. Take $b_q^- = a$.

Now we are ready to answer the following important question. How much additional computational work is needed in order to compute $F_\epsilon(k, x)$ compared to computation of $F(k, x)$? Preliminary analysis of expressions (23)-(24) is not encouraging because the number of arithmetic operations necessary for straightforward implementation of (24) grows exponentially with the number of functions N. This will make the computation of $F_\epsilon(k, x)$ problematic even for moderate values of N. However, more refined analysis based on Lemma 4 and Algorithm 1 shows that in reality overhead grows relatively slow which makes computation of $F_\epsilon(k, x)$ an easy task even for large N.

In order to make this statement precise we need to define exactly what we mean by computational overhead. For the purposes of the present discussion we shall measure overhead in a number of arithmetic operations required for computation of $F_\epsilon(k, x)$ after functions $f_i(x)$ are already computed. We do not consider comparisons and memory management operations, but their inclusion will not lead to qualitatively different results. Overhead estimate is contained in the following theorem.

Theorem 5 *Suppose that $\varphi_\epsilon(z)$ is computed according to (25). Then there exists an algorithm for which the number S of additional arithmetic operations necessary for computation of $F_\epsilon(k, x)$ for any fixed k, ϵ and x can be estimated as follows*:

$$S \leq (1 + \rho_N) N^3 \tag{29}$$

where N is the number of functions $f_i(x)$, $\rho_N \to 0$ as $N \to \infty$, $\rho_N \leq 15$ and ρ_N does not depend on dimension of vector x.

Proof.

The proof is based on estimation of the number of arithmetic operations required by Algoritm 1 and implementation of expressions (26)-(28).

Let us denote by k_+^i the number of elements in set $\Lambda_{\epsilon+}^i$ and by k_0 the number of elements in set Λ_ϵ^k,

$$\Lambda_\epsilon^k = \{i : |F(k,x) - f_i(x)| \le \epsilon\}.$$

Computation of all d_s requires k_-^i subtractions in order to get $\Delta_{ij_s}^x$ and another at most $9k_-^i$ operations in order to compute $\varphi_\epsilon\left(\Delta_{ij_s}^x\right)$ if expression (25) is used. Step 2 of Algorithm 1 requires $k_-^i(k_-^i+1)/2$ multiplications and $\left(k_-^i\right)^2/2$ additions to compute all b_q^- for $q = 1, ..., k_-^i$. The same algorithm can be used for computing of coefficients b_r^+. Computation of sum in (27) requires at most $2\min\left\{k_-^i, k_+^i\right\}$ operations. Therefore we have the following estimate for the total amount S_i of arithmetic operations required for computation of $c_i^\epsilon(x)$ for arbitrary fixed i:

$$S_i \le k_-^i(k_-^i+1)+k_+^i(k_+^i+1)+2\min\left\{k_-^i, k_+^i\right\}+10k_-^i+10k_+^i \le \left(k_-^i\right)^2+\left(k_+^i\right)^2+12\left(k_-^i+k_+^i\right).$$

The last inequality yields

$$S_i \le N^2 + 12N$$

because $k_-^i + k_+^i \le N$. After all $c_i^\epsilon(x)$ and $f_i(x)$ are computed it is required another $3k_0 - 1$ operations in order to compute $F_\epsilon(k,x)$ from (26). Therefore the following estimate for S holds:

$$S \le k_0\left(N^2 + 12N\right) + 3k_0 - 1 \le N^3 + 12N^2 + 3N$$

which completes the proof. □

In reality overhead will be much smaller than (29) because for small and moderate ϵ we have

$$k_-^i \ll N,\ k_+^i \ll N,\ k_0 \ll N.$$

Estimates similar to (29) hold also for other functions $\varphi_\epsilon(z)$ which satisfy conditions of Lemma 4.

Algorithm 1 and (26)-(28) were implemented as Matlab M-files and were used as integral part of environment for optimization of nonconvex nonsmooth functions $F(k,x)$ by methods of nonlinear programming designed for twice continuously differentiable functions.

References

[1] RiskMetricsTM. third edition, J.P. Morgan, 1995.

[2] Amendment to the capital accord to incorporate market risks. Bank for International Settlements, 1996.

[3] C.O Alexander and C.T. Leigh. On the covariance matrices used in value at risk models. *The Journal of Derivatives*, pages 50–62, Spring 1997.

[4] F. Andersson and S. Uryasev. Credit risk optimization with Conditional Value-at-Risk criterion. Research Report 99-9, Center for Applied Optimization, University of Florida, Gainesville, 1999.

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

[6] S. Basak and A. Shapiro. Value-at-risk based risk management: Optimal policies and asset prices. Working Paper, Wharton School, University of Pennsylvania, available at http://www.gloriamundi.org, 1998.

[7] John R. Birge and Francois Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

[8] A. Brooke, D. Kendrick, A. Meeraus, and R. Raman. *GAMS, A User's Guide*. GAMS Development Corporation, 1998.

[9] D. R. Carino and W. T. Ziemba. Formulation of the Russell-Yasuda-Kasai financial planning model. *Operations Research*, 46(4):433–449, 1998.

[10] G. Consigli and M.A.H. Dempster. Dynamic stochastic programming for asset-liability management. *Annals of Oper. Res.*, 81:131–161, 1998.

[11] Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

[12] Ron Dembo and Dan Rosen. The practice of portfolio replication. A practical overview of forward and inverse problems. *Annals of Oper. Res.*, 85:267–284, 1999.

[13] J. Dupacova, M. Bertocchi, and V. Moriggia. Postoptimality for scenario based financial planning models with an application to bond portfolio management. In W. Ziemba and J. Mulvey, editors, *Worldwide Asset and Liability Management*, pages 263–285. Cambridge University Press, 1998.

[14] Yu. Ermoliev and R. J.-B. Wets, editors. *Numerical Techniques for Stochastic Optimization*. Springer Verlag, Berlin, 1988.

[15] Alexei A. Gaivoronski and Petter E. de Lange. An asset liability management model for casualty insurers: Complexity reduction vs. parametrized decision rules. *Annals of Oper. Res.*, 2000. to appear.

[16] Alexei A. Gaivoronski and Georg Pflug. Finding optimal portfolios with constraints on Value-at-Risk. In Bo Green, editor, *Proceedings of the Third International Stockholm Seminar on Risk Behaviour and Risk Management*. Stockholm University, 1999.

[17] Alexei A. Gaivoronski and Fabio Stella. Nonstationary optimization approach for finding universal portfolios. Technical Report 1/98, NTNU, Department of Indistrial Economics and Technology Management, 1998.

[18] Henk Grootveld and Winfried G. Hallerbach. Upgrading Value-at-Risk from diagnostic metric to decision variable: A wise thing to do? Working Paper, Erasmus University Rotterdam, 2000.

[19] Thomas S.Y. Ho, Michael Z.H. Chen, and Fred H.T. Eng. VAR analytics: Portfolio struc-

ture, key rate convexities, and VAR betas. *The Journal of Portfolio Management*, pages 89–98, Fall 1996.

[20] Glyn Holton. Simulating value-at-risk. *Risk*, pages 60–63, May 1998.

[21] Philippe Jorion. Risk2: Measuring the risk in value at risk. *Financial Analysts Journal*, pages 47–56, November/December 1996.

[22] P. Kall and S. Wallace. *Stochastic Programming*. John Wiley and Sons, New York, 1994.

[23] H. Markowitz. Portfolio selection. *Journal of Finance*, 8:77–91, 1952.

[24] The MathWorks Inc. *Matlab 5.3*, 1999.

[25] E. Medova. VAR methodology and the limitation of catastrophic or unquantifiable risk. VII International Conference on Stochastic Programming, The University of British Columbia, Vancouver, Canada, 1998.

[26] John M. Mulvey and Hercules Vladimirou. Stochastic network programming for financial planning problems. *Manage. Sci.*, 38(11):1642–1664, 1992.

[27] W. Ogryczak and A. Ruszczynski. From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research*, 116:33–50, 1999.

[28] G. Ch. Pflug. How to measure risk? In *Festschrift to F.Ferschl*. Physica Verlag, 1999.

[29] G. Ch. Pflug. Some remarks on the Value-at-Risk and the Conditional Value-at-Risk. In S. Uryasev, editor, *Probabilistic Constrained Optimization*. Kluwer Academic Publishers, 2000.

[30] A. Prekopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.

[31] Manoj K. Singh. Value at risk using principal components analysis. *The Journal of Portfolio Management*, pages 101–112, Fall 1997.

[32] S. Uryasev and R.T. Rockafellar. Optimization of conditional value-at-risk. Technical Report 99-4, ISE Dept. Univeristy of Florida, 1999.

[33] R.J.-B. Wets. Stochastic programming: Solution techniques and approximation schemes. In A.Bachem, M.Groetshel, and B.Korte, editors, *Mathematical Programming: The State of the Art*. Springer Verlag, Berlin, 1982.

[34] Stavros A. Zenios. A model for portfolio management with mortgage-backed securities. *Ann. Oper. Res.*, 43:337–356, 1993.