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

The Value-at-Risk (V@R) is an important and widely used measure of the extent to which a given portfolio is subject to risk inherent in financial markets. In this paper, we present a method of calculating the portfolio which gives the smallest V@R among those, which yield at least some specified expected return. Using this approach, the complete mean-V@R efficient frontier may be calculated. The method is based on approximating the historic V@R by a smoothed V@R (SV@R) which filters out local irregularities. Moreover, we compare V@R as a risk measure to other well known measures of risk such as the Conditional Value-at-Risk (CV@R) and the standard deviation.

It is shown that the resulting efficient frontiers are quite different. An investor, who wants to control his V@R should not look at portfolios lying on other than the V@R efficient frontier, although the calculation of this frontier is algorithmically more complex. We support these findings by presenting results of a large scale experiment with a representative selection of stock and bond indices from developed and emerging markets which involved the computation of many thousands of V@R-optimal portfolios.

1 Introduction

The Value-at-Risk (V@R) is an important measure of the exposure of a given portfolio of securities to different kinds of risk inherent in financial markets. By now, it became a tool for risk management in financial industry (*RiskMetricsTM* 1995) and part of industrial regulatory mechanisms (*Amendment to the Capital Accord to Incorporate Market Risks* 1996). Considerable

*Journal of Risk, Vol. 7, No. 2, pp. 1-31, Winter 2004-2005

amount of research was dedicated recently to development of methods of risk management based on Value-at-Risk [1].

In this paper we focus on applications of the Value-at-Risk (V@R) concept in the context of optimal portfolio selection. This is a relatively novel application of V@R as opposed to utilization of V@R for risk measurement purposes. One of the reasons is that V@R optimization is inherently more difficult than – for example – variance optimization and efficient solution algorithms for this problem are lacking. The V@R optimization problem is nonconvex, may exhibit many local minima and is of combinatorial character, i.e. exhibits exponential growth in computational complexity. The objective of this paper is to demonstrate how to solve V@R optimization problems and to put V@R on equal footing with the variance as portfolio selection criterion. We present a smoothing algorithm, which allows to calculate optimal portfolios in the V@R sense accurately and in reasonable time.

Recently, the possibilities to utilize V@R and related measures like the Conditional Value-at-Risk (CV@R, see Uryasev & Rockafellar (1999)) as criteria for optimal portfolio selection started to attract some attention. The relevant literature includes Basak & Shapiro (2001), Gaivoronski & Pflug (1999), Krokhmal, Uryasev & Palmquist (2001), Medova (1998), Uryasev & Rockafellar (1999) and sheds an interesting light on the properties of V@R-optimal portfolios while acknowledging considerable computational difficulties which we address in this paper.[2]

Another aim of this paper is to make computational comparisons of the properties of V@R, CV@R and standard deviation as risk measures in context of portfolio selection. In the center of our approach is the comparison of efficient frontiers for different data sets, some of them of considerable interest for institutional investors. This study can be seen as complementary to comparison of theoretical properties as in Artzner, Delbaen, Eber & Heath (1999) or evaluation of risk measures from the utility theory standpoint as in Grootveld & Hallerbach (2000) and Kaplanski & Kroll (2002). We argue that for an 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 standard deviation, or even the seemingly related CV@R may represent a poor substitute. Empirical comparisons of V@R and other risk measures from different angles can be found in Wu & Xiao (2002) and Pfingsten, Wagner & Wolferink (2004).

The rest of the paper is organized as follows. In section 2 we introduce the three risk-return optimization problems for standard deviation (or equivalently variance), V@R and CV@R and discuss their properties illustrating our points by numerical examples. We concentrate here on empirical risk measures derived directly from historic data. In other words, our approach is based on historic simulation. The main finding is that historical V@R can be represented by a superposition of two components: one is relatively well behaved and the other is highly irregular. This observation is exploited in section 3 where we introduce the smoothed Value-at-Risk (SV@R) which is an approximation of V@R, which extracts the global component of

[1] See e.g. Jorion (2001) and Duffie & Pan (1997) for surveys

[2] V@R optimization can be considered as a stochastic programming (SP) problem of special type. It is related to SP problems with probability constraints (Prekopa 1995). General references for stochastic programming models and solution techniques are Birge & Louveaux (1997), Ermoliev & Wets (1988), Kall & Wallace (1994), Wets (1982) among others. Different applications of SP to optimal asset allocation were considered in Cariño & Ziemba (1998), Consigli & Dempster (1998), Dembo & Rosen (1999), Dupacova, Bertocchi & Moriggia (1998), Gaivoronski & de Lange (2000), Gaivoronski & Stella (2000), Mulvey & Vladimirou (1992), Ogryczak & Ruszczynski (1999), Zenios (1993), fast computation of several mean-risk efficient portfolios was considered in Duarte (1999).

the V@R behavior and filters out the irregular local component. The mathematical details of our SV@R construction are fairly involved and therefore we relegate them to Appendix. SV@R optimization is much easier and can be performed by efficient nonlinear programming software. The optimal V@R values are recovered from the solution of SV@R optimization problem by a few inexpensive postprocessing steps. After developing V@R optimization tools[3], we are in position to compute and compare mean-risk efficient frontiers for different risk measures in section 4. An important result of this comparison is that mean-V@R efficient frontier cannot be approximated by efficient frontiers obtained for other risk measures, like the standard deviation or CV@R and therefore V@R-optimal portfolios should be considered directly. We provide additional evidence in support for this recommendation in section 5, where results of a large scale experiment with a representative set of financial data are reported. The data set contains stock and bond indices for several developed and emerging markets and involved computation of more than 80000 V@R-optimal portfolios using the SV@R approximation. Both in sample and out of sample experiments confirm advantage of V@R-optimal portfolios in terms of controlling V@R.

2 Risk-return portfolio optimization

Consider a finite set of arbitrary financial assets $i = 1, 2, \ldots, n$. Within a given observation period these assets generate returns

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

measured as the relative increase (or decrease) of the asset prices during the period under consideration. These returns are unknown at the time of portfolio allocation and are 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 return of the portfolio at the end of the observation period is

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

It is a random variable with distribution function F, i.e. $F(u) = \mathbb{P}\{W \leq u\} = \mathbb{P}\{x^T \xi \leq u\}$. Of course, F depends on x. The expected return of portfolio x is

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

[3]More information about V@R optimization tools developed in this paper, distributable software and results of additional computations is found at http://websterii.iot.ntnu.no/users/alexeig/VaR/

Suppose that $\mathcal{R}$ is some risk measure, like V@R, CV@R or the standard deviation. For a given minimal expected return μ, let us consider the solution of the following optimization problem:

$$\min_x \mathcal{R}(x^T\xi) \tag{1}$$

$$x^T\mathbb{E}(\xi) \geq \mu$$

$$x^T\mathbb{1} = 1$$

$$x \geq 0$$

The curve which represents the dependence of the optimal value of this problem on the parameter μ is the boundary of the feasible set of values of pairs (return, risk). A subset of this boundary forms the $\mathcal{R}$-efficient frontier and individual portfolios on this frontier are $\mathcal{R}$-efficient portfolios. Essentially, this is a generalization of the classical concept of mean-variance efficient frontier due to Markowitz (1952) for the case of an arbitrary risk measure $\mathcal{R}$.

In this paper we are going to compare risk-return optimal portfolios obtained through solution of (1) for the following risk measures.

1. *The Value-at-Risk (V@R)*

$$\mathcal{R}(W) = \text{V@R}(W) = \mathbb{E}(W) - \ Q_\alpha(W) \tag{2}$$

where $Q_\alpha(W)$ is the $\alpha-$quantile of return:

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

The definition which we use describes the largest return underperformance relative to expected portfolio return which is possible in $1-\alpha$ cases of outcomes. We shall refer to $1-\alpha$ as a confidence level[4].

2. *The Conditional Value-at-Risk (CV@R)*

$$\mathcal{R}(W) = \text{CV@R}(W) = \mathbb{E}(W) - C_\alpha(W). \tag{3}$$

where $C_\alpha(W)$ is the expectation of return conditioned on not exceeding the $\alpha-$quantile, or more precisely

$$C_\alpha(W) = \frac{P\{W \leq Q_\alpha(W)\}}{\alpha}\mathbb{E}(W|W \leq \ Q_\alpha(W)) + \left(1 - \frac{P\{W \leq Q_\alpha(W)\}}{\alpha}\right) Q_\alpha(W). \tag{4}$$

Notice that $C_\alpha(W) = \mathbb{E}(W|W \leq \ Q_\alpha(W))$ in case that $P\{W \leq Q_\alpha(W)\} = \alpha$, e.g. if W has no point mass at $Q_\alpha(W)$.

Usually the values of α in these definitions are taken to be small, say, 0.05 or 0.01.

[4]Several slightly different definitions of V@Rcan be found in the literature. Other V@R definitions can be recovered easily from this. In the case of short observation period (few days) one often takes $\text{V@R}_\alpha(W) \simeq -Q_\alpha(W)$ since in this case $\mathbb{E}(W) \simeq 0$ for any reasonable portfolios. Our results are not influenced by difference in these definitions.

3. *The Standard deviation (StDev)*

$$\mathcal{R}(W) = \mathrm{StDev}(W) = \sqrt{\mathbb{E}\left(W - \mathbb{E}(W)\right)^{\top}\left(W - \mathbb{E}(W)\right)}$$

2.1 Optimal portfolios based on historical or simulated data

In order to obtain risk-optimal portfolios through the solution of problem (1) we have essentially two possibilities.

1. *The parametric approach.* It is assumed that the asset returns are governed by a given parametric distribution with known parameters, for example normal or lognormal. Using this knowledge one may obtain analytical expressions or algorithmic procedures for computation of risk measure $\mathcal{R}(x^T\xi)$ and its other quantities necessary for optimization, like derivatives. This is quite possible to do when the risk measure is the standard deviation. In this case problem (1) is equivalent to a quadratic optimization problem for which efficient codes exist. Unfortunately, this approach becomes problematic for the case of V@Rand CV@Rbecause precise computation of these measures for continuously distributed returns may be very difficult unless the return distribution is normal.

2. *The sampling approach.* It works directly with a finite sample $\xi^1, \xi^2, \ldots, \xi^N$ of return observations without assuming any specific distribution for returns. This sample can be actual historic data from which V@R or CV@R is computed through historic simulation or by parametric Monte Carlo. We follow this approach here because it is computationally feasible for considerably larger class of risk-return optimization problems, especially in the case of V@R and CV@R. Besides, both V@R and CV@R are sensitive to the tail properties of distributions and many parameterized families are notoriously bad in describing such properties of real data.

Using the three different risk measures V@R(W), CV@R(W) and StDev(W) in (1), the sampling approach leads to quite differently structured optimization problems.

The StDev optimization problem is equivalent to the variance optimization problem, which is a convex quadratic program:

$$\min_{x \geq 0} x^T \left[\frac{1}{N}\sum_{i=1}^{N}(\xi^i - e)(\xi^i - e)\right] x \tag{5}$$

$$x^T e \geq \mu \tag{6}$$

$$x^T \mathbb{1} = 1 \tag{7}$$

where $e = \frac{1}{N}\sum_{i=1}^{N}\xi^i$ is the average return vector.

The formulation of the sample CV@Roptimization problem requires more work. Observe that for any fixed portfolio x the value of $C_\alpha(W)$ from (4) can be represented as the solution of the following optimization problem (Uryasev & Rockafellar 1999):

$$C_\alpha(W) = -\inf_{a \in \mathbb{R}}\{-a + \frac{1}{\alpha}\mathbb{E}\max\left[W - a, 0\right]\}. \tag{8}$$

Given the relation (3) one may write the CV@R optimization problem as follows:

$$\min_{x\geq 0,a} -a + \frac{1}{\alpha N}\sum_{i=1}^{N} \max[a - x^T\xi^i, 0]$$

with additional constraints (6),(7). Let us introduce here auxiliary variables $z = (z^1, .., z^N)$ such that $z^i \geq \max[a - x^T\xi^i, 0]$. This yields the following large scale linear programming problem:

$$\min_{x\geq 0, z\geq 0, a} -a + \frac{1}{\alpha N}\sum_{i=1}^{N} z^i \tag{9}$$

$$a - x^T\xi^i - z^i \leq 0$$

with constraints (6),(7). Efficient commercial software exists for the solution of this problem. For the definition of the V@R optimization problem let us denote by $\min^k\left\{u^1,\ldots,u^N\right\}$ and $\max^k\left\{u^1,\ldots,u^N\right\}$ the k-th smallest and the k-th largest among $u^1,\ldots,u^N$. Then $\min^1\left\{u^1,\ldots,u^N\right\}$ and $\max^1\left\{u^1,\ldots,u^N\right\}$will be the ordinary minimum and maximum. Besides,

$$\min{}^k\left\{u^1,\ldots,u^N\right\} = \max{}^{N-k+1}\left\{u^1,\ldots,u^N\right\},\ \ \min{}^k\left\{u^1,\ldots,u^N\right\} = -\max{}^k\left\{-u^1,\ldots,-u^N\right\}.$$

In these notations the empirical α-quantile of the sample $x^T\xi^1,\ldots x^T\xi^N$ is $V(x) = \min{}^{\lfloor\alpha N\rfloor+1}\left\{x^T\xi^1,\ldots x^T\xi\right.$ where $\lfloor\alpha N\rfloor$ is the largest integer not exceeding αN. The empirical V@R[5] from definition (2) is

$$\text{V@R}(x) = x^Te - \ \min{}^{\lfloor\alpha N\rfloor+1}\left\{x^T\xi^1,\ldots x^T\xi^N\right\} \tag{10}$$

which yields the following V@Roptimization problem[6].

$$\begin{aligned}&\min_{x\in\mathbb{X}} -V(x) \qquad (11)\\ &\text{where}\\ &V(x) = \min{}^{\lfloor\alpha N\rfloor+1}\left\{x^T\xi^1,\ldots,x^T\xi^N\right\}\\ &\text{and}\\ &\mathbb{X} = \{x\geq 0; x^Te\geq\mu; x^T\mathbb{1}=1\}.\end{aligned}$$

This problem is a nonconvex program which may have many local minima because the function $\max^k$ of N linear functions is convex only for $k = 1$ when it coincides with ordinary maximum. This makes it by far the most difficult problem to solve among the problems (5),(9),(11). While efficient commercial software based on several decades of algorithmic and theoretical research exists for solution of problems (5) and (9), the current situation with the problem (11) is the opposite one. Efficient general algorithms for minimization of functions defined through $\max^k$ are nonexistent. Our aim here is to exploit the structure of (11) and develop efficient

[5]When sample is drawn from historic prices this is called the historic V@R

[6]We assumed here and in (9) that for the set of assets under consideration the less risky portfolios do not yield superior average returns compared to more risky portfolios. Then the first term in (10) will not affect the solution of risk-optimization problem (1) and can be dropped.

solution techniques specifically tailored for V@R optimization problem.

2.2 Properties of V@R and CV@R optimal portfolios

In this section we present an empirical study of the properties of historic V@R and CV@R as functions of portfolio composition using stock market data. Its purpose is to give a motivation for the numerical approach of Section 3 for the solution of problem (11) without making any general claims about the properties of portfolios of equities. Here and in the sections 3, 4 we utilized the same data set which was used in Cover (1991) and a number of other papers for evaluation of so-called universal portfolios. It contains more than 20 years of NYSE stock daily returns for representative set of 36 companies from different industry sectors starting from July 3, 1962 to December 31, 1984. In all experiments reported here we used this data starting from the first day, July 3, 1962.

Typical patterns of V@Rdependence on portfolio composition are presented on Figure 1.

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

$$x(\lambda) = \lambda x^1 + (1-\lambda)x^2,\ 0 \le \lambda \le 1 \tag{12}$$

Figure 1 shows sample V@R(thin line) and CV@R(thick line) of portfolio $x(\lambda)$ as function of λ. The values of λ are shown on horizontal axis, while the vertical axis represents values of V@R and CV@R measured in percentage points of initial portfolio value. These risk measures are computed using 500 observations over a 10 days period, with $\alpha = 0.05$. Portfolio x^1 contains only Ford stock and portfolio x^2 contains only Hewlett Packard stock. These and other experiments provide the following observations about the behavior of V@R.

1. The sample V@R is a very irregular function with multiple local minima and maxima evenly distributed in the function domain. The value of the local minima is in large majority far away from the globally minimal V@R value. We performed many experiments where the number of observations was increasing from 250 to 5000. They show that the number of local minima grows with the number of observations and sooner or later a portfolio where V@R attains its locally minimal value can be found in the vicinity of almost every portfolio. Moreover, the V@R function is nondifferentiable in every local minimum. This means that a straightforward application of standard techniques of nonlinear optimization to the solution of problem (11) is very problematic. 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 could find only a local minimum of the V@R function. This will not bring us any closer to the solution of (11) because an arbitrary local minimum does not provide any information about the solution of the V@R optimization problem.

2. The V@R function possesses a pronounced structure and is composed of two components. The first component exhibits the fairly regular global pattern with a unique global minimum and smooth behavior which in some cases is close to a convex function[7]. The second component is responsible for the local behavior and has a highly irregular pattern which produces almost

[7]In some other examples its behavior did not seem to be convex, but it still had only a few local minima with values which are close to the global minimum, see thick line on Figure 2

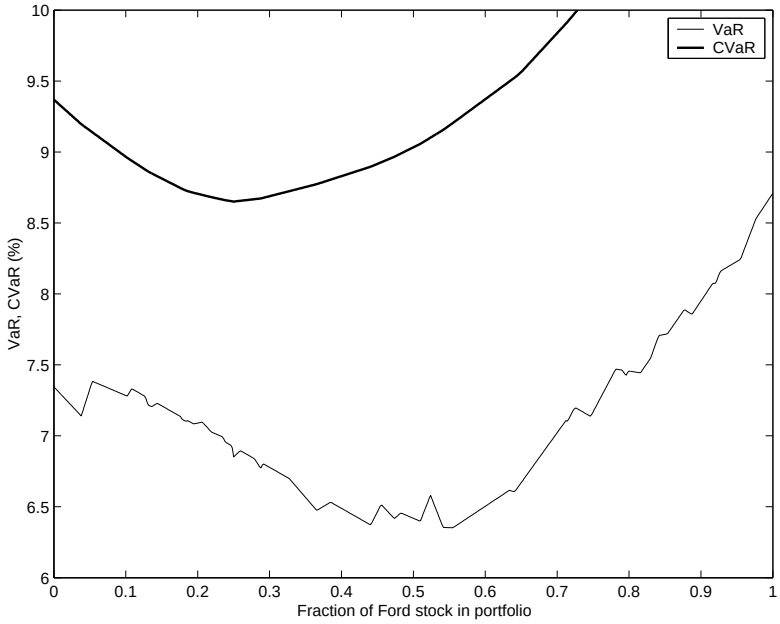

Figure 1: Properties of V@R and CV@R for Ford/HP portfolio

all local minima and nondifferentiability. This dissection into two components is crucial for our approach to the solution of the V@R optimization problem (11).

Before starting with dedicated techniques for optimizing V@R let us explore briefly an alternative approach: the substitution of V@R with some related risk measure which is easier to optimize. If portfolios optimal for such alternative measure have V@R which is close to the optimal V@R then there is no need to develop dedicated techniques for V@R. A natural candidate for such substitution is CV@R which is the smallest convex majorant of V@R and which may be optimized using the solution of linear program (9), a relatively simple issue. In order to get some empirical evidence about the relation between CV@R and V@R optimal portfolios, we conducted a series of experiments with the same data set as before.

For each experiment a pair of stocks was selected and the set of portfolios defined by (12) was considered, where portfolios x^1 and x^2 consisted of pure stocks 1 and 2 respectively. From this set we selected portfolios with minimal V@R and CV@R which are denoted by $x_{\mathrm{V@R}}$ and $x_{\mathrm{CV@R}}$. This is a simple thing to do because we have one-dimensional optimization problem here. Suppose now that the V@R-optimal portfolio $x_{\mathrm{V@R}}$ is substituted by the CV@R-optimal portfolio $x_{\mathrm{CV@R}}$. By doing so we make two kinds of substitution errors: we get a larger $\mathrm{V@R}(x_{\mathrm{CV@R}})$ value and a different value of portfolio return $R(x_{\mathrm{CV@R}})$. These errors are measured as follows:

$$E_{\mathrm{V@R}} = \frac{\mathrm{V@R}(x_{\mathrm{CV@R}}) - \mathrm{V@R}(x_{\mathrm{V@R}})}{\mathrm{V@R}(x_{\mathrm{V@R}})} 100\%, \;\; E_R = \frac{R(x_{\mathrm{V@R}}) - R(x_{\mathrm{CV@R}})}{R(x_{\mathrm{V@R}})} 100\% \qquad (13)$$

Our experiments show that in some cases the CV@R-optimal portfolio approximates the V@R-optimal portfolio reasonably well. In other cases the CV@R-optimal portfolio is a poor substitute for V@R-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 Table 1. V@R errors for randomly selected portfolios formed by the stock pairs from this table are only 60-100% larger compared to the errors of CV@R-optimal portfolios. Differences in returns can be very substantial and considerably different compositions of optimal portfolios may occur. Figure 1 presents in more detail the case of Ford-Hewlett Packard portfolio from this table. Note that CV@Ris a convex function in the portfolio shares, while V@Ris not, and they take their respective minima in considerably different regions.

Stocks	V@R error $E_{\mathrm{V@R}}$	Return error E_R
Kodak-Merck	5.3	5.6
General Electric-IBM	3.4	2.4
Sears-Coca Cola	6.1	11.7
Dupont-Exxon	6.0	-11.4
P&G-GTE	8.1	-0.1
Ford-HP	7.9	30.3
GM-P&G	10.0	5.3

Table 1. Errors committed by substitution of V@R by CV@R, confidence level 0.95, 500 observations over 10 days period, starting day July 3, 1962

This evidence motivates the development of dedicated techniques for the computation of V@R optimal portfolios through the solution of problem (11). More evidence in this direction is presented in Sections 4,5 where respective efficient frontiers are compared.

3 Computation of V@R efficient portfolios using SV@R

The analysis of the previous section is the basis of our approach to the numerical solution of the V@R optimization problem (11) which consists of three steps.

1. Smooth out the local noisy component of the V@R function and extract the well behaved global component. This is the key feature of our approach. The result is the smoothed V@R function (SV@R).

2. After this the problem (11) becomes treatable by standard off-shelf software developed for solution of nonlinear programming problems. We want to utilize this software as much as possible. The result is an approximate solution of the V@R optimization problem which in itself will be often sufficient for practical purposes.

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

1. *Extracting the global behavior of the V@R function: SV@R*

In principle, general approximation techniques, like spline approximation, can be used for smoothing the V@R function. However, they require computational overhead which grows exponentially with the dimension of the portfolio. This makes such approaches impractical because the approximating function is going to be computed repeatedly during the optimization process. For this reason, we had to develop the smoothing techniques which exploits the special structure of the V@R function.

Consider the sample quantile function $V(x)$ from (11). It depends on the portfolio shares x, the observations of return $\xi^1, ..., \xi^N$ and the probability level α. This function is approximated

by the family of *smoothed quantile functions* $V(\epsilon, x)$ parameterized by *smoothing parameter* ϵ, $\epsilon > 0$:

$$V(\epsilon, x) = \sum_{i=1}^{N} c_i^{\epsilon}(x) x^T \xi^i. \tag{14}$$

The coefficients $c_i^{\epsilon}(x)$ in (14) 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[8] for all $\epsilon > 0$. 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 the undesired properties of the original quantile function, to very well behaved, almost linear approximations which, however, retain little information about the original quantile function. Therefore, the choice of smoothing parameter ϵ is governed by trade-off 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 for $c_i^{\epsilon}(x)$ from (14) is described in Appendix (see (26)-(27) and Theorem 2). Each specific expression for $c_i^{\epsilon}(x)$ is defined by a function $\varphi_{\epsilon}(z)$ of one dimensional parameter z which provides a smooth approximation for the unit step function $\varphi(z)$:

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

Smooth approximants $\varphi_{\epsilon}(z)$ of $\varphi(z)$ are defined in (28). This choice is driven by computational considerations. The resulting smoothed quantile function $V(\epsilon, x)$ requires a moderate computational overhead compared with the original quantile function $V(x)$. The 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). Figure 2 shows a typical example of a smoothed quantile function for different values of the smoothing parameter ϵ.

These figures show – similarly to Figure 1 – the dependence of the quantile function $V(x)$ (thick line) on the parameter λ which defines portfolio $x(\lambda)$ as linear combination of portfolios x^1 and x^2 obtained according to (12). Portfolios x^1 and x^2 consisted of three stocks (Schlumberger, Morris, Commercial Metals) in the following proportion:

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

[8]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.

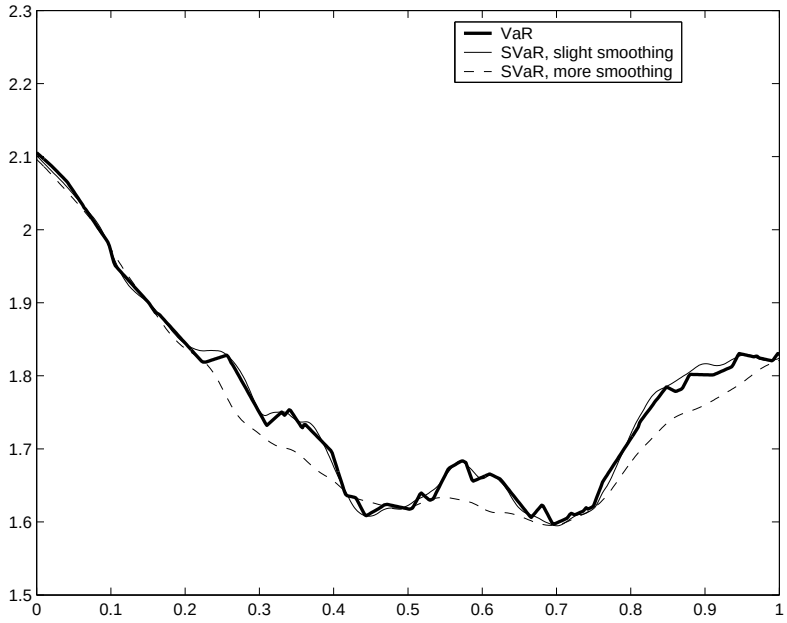

Figure 2: The quantile and the smoothed quantile function for different values of the smoothing parameter ϵ

V@Rwas computed on the basis of 500 daily observations for confidence level 0.95. Thin lines on this figure represents smoothed quantile function computed for different values of smoothing parameter ϵ. The solid thin line shows the case of small value of smoothing parameter $\epsilon = 0.001$. In this case $V(\epsilon, x)$ follows $V(x)$ 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 by a solid dashed line. Here local noisy component of $V(x)$ is filtered out completely and only the global well behaved component remains.

2. Minimizing SV@Rfunction

Efficient nonlinear programming software present in the market can be used for finding V@R-efficient portfolios by minimizing the SV@Rfunction[9].

3. Postprocessing

The result of the previous step is portfolio $x_{\mathrm{SV@R}}$ which is either a global minimizer or a good local minimizer of the smoothed V@R function. In most cases this will be enough for practical purposes. However, if one wish to improve on this, it can be done by local minimization of the V@R function taking as a starting point portfolio $x_{\mathrm{SV@R}}$. For this purpose we solved linear programming problem of small to medium size[10] formulated below.

[9]We implemented SV@Rin MATLAB environment (The Mathworks Inc. 1984-2004) and used `fmincon` subroutine from MATLAB Optimization Toolbox for its minimization. Another option is to use nonlinear programming solvers available in GAMS environment (Brooke, Kendrick, Meeraus & Raman 1998) or in NAG library.

[10]This 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.

According to (11) there exists index j which depends on $x_{\text{SV@R}}$ such that $V(x) = x_{\text{SV@R}}^T \xi^j$ and inequality

$$x_{\text{SV@R}}^T \xi^i \leq x_{\text{SV@R}}^T \xi^j \tag{15}$$

is satisfied for at least $\lfloor \alpha N \rfloor$ observations $\xi^i, i \neq j$. Let us denote by Λ an arbitrary set of $\lfloor \alpha N \rfloor$ indices for which inequality (15) is satisfied. Then the solution of the following linear programming problem is a local minimum of the quantile function $V(x)$ and provides better value of V@R then $x_{\text{SV@R}}$.

$$\min_x \; x^T \xi^j \tag{16}$$

$$x^T \left(\xi^i - \xi^j \right) \leq 0, \; i \in \Lambda \tag{17}$$

$$x^T \left(\xi^i - \xi^j \right) \geq 0, \; 1 \leq i \leq N, \; i \notin \Lambda \tag{18}$$

$$x^T e \geq \mu \tag{19}$$

$$x^T \mathbb{1} = 1, \; x \geq 0 \tag{20}$$

Our experience is that the improvement of the V@R value obtained through solution of this problem is relatively small[11].

4 Comparison of efficient frontiers

Now we have in place all tools necessary for utilization of V@R as a criterion for optimal investment. This is done following the classical Markowitz-type approach (Markowitz 1952): V@R is minimized for different values of minimal return μ by solving the problem (11) and the mean-V@R feasible set is constructed from which the mean-V@R efficient frontier is derived. After this an investor with specific risk preferences can choose the target value of V@R and select portfolio on the efficient frontier that provides the best return for a given value of V@R.

But how such mean-V@R 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 favor of V@R. In order to answer this question we conducted a series of studies with price data of different instruments. Here we report results obtained with the same set of the stock market data which was used in the section 2.1 and present the results graphically. Results of extensive experiments with portfolios of stock and bond indices are presented in the section 5. The purpose was to compare different mean-risk efficient frontiers obtained on the basis of historical data by solution of problems (5),(9) and (11) for different values of return μ. Each computation of efficient frontier involved 200-500 solutions of respective optimization problem.

Results of a typical experiment are presented on Figures 3,4 and 5.

[11] It is possible to go further and minimize (16) subject to constraints (17)-(20) with respect to both portfolio x and all pairs (j, Λ) where $1 \leq j \leq N$ and set Λ contains exactly $\lfloor \alpha N \rfloor$ 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 V@R function $x_{\text{SV@R}}$ provides just this information. In our experiments we have not used this approach because minimization of SV@R together with postprocessing according to (16)-(20) gave very satisfactory results.

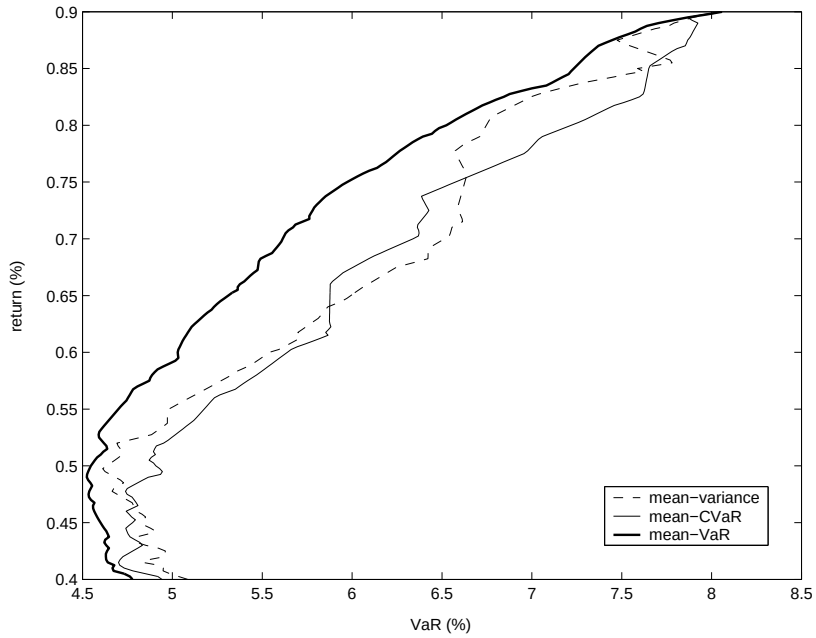

Figure 3: Boundary of mean-V@R feasible set and images of mean-CV@R boundary and mean-variance boundary

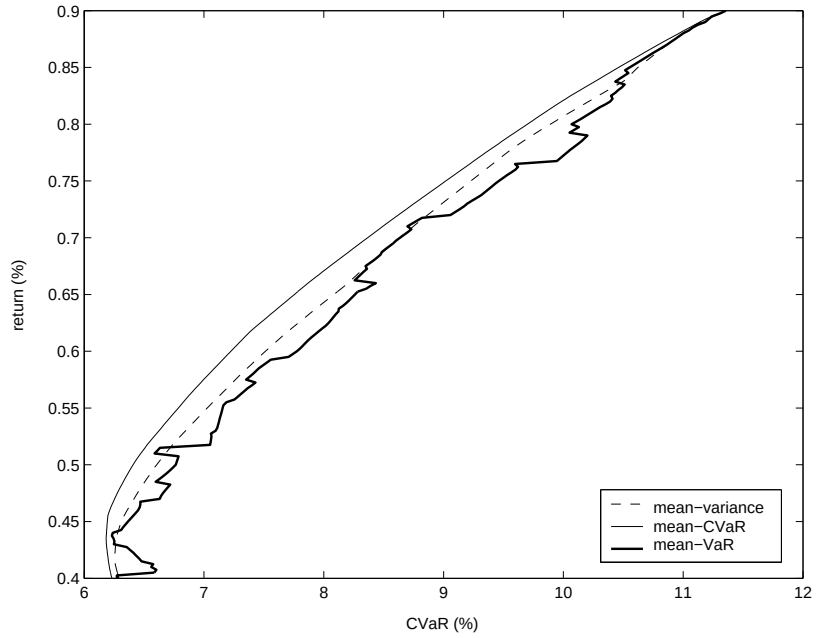

Figure 4: Boundary of mean-CV@R feasible set and images of mean-V@R boundary and mean-variance boundary

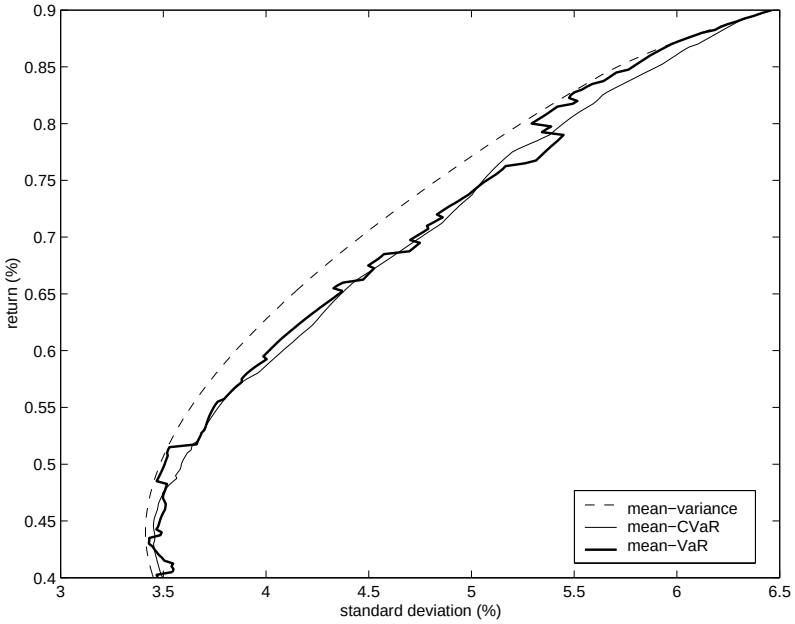

Figure 5: Boundary of mean-variance feasible set and images of mean-V@R boundary and mean-CV@R boundary

They show boundaries of mean/V@R, mean/CV@R 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. A confidence level 0.95 was used for computation of V@R and CV@R. Efficient frontiers were computed using 201 equidistant return values.

Figure 3 shows the boundary of mean-V@R feasible set with a thick line. Besides, it shows the images of boundaries of mean-CV@R and mean-variance feasible sets in mean-V@R space with thin and dashed lines respectively. In order to obtain these images we computed the V@R of portfolios that form mean-CV@R and mean-variance efficient frontiers and placed resulting points on Figure 3. The horizontal axis on this figure shows V@R values expressed in percents relative to original portfolio value. Vertical axis shows expected rate of return after 10 trading days expressed in percents.

Figure 4 shows the boundaries of feasible sets and efficient frontiers in the mean-CV@R space. The thick line represents the mean-CV@R efficient frontier, the images of mean-V@R and mean-variance efficient frontiers are shown with thick and dashed lines respectively. Figure 5 shows the boundaries of feasible sets and efficient frontiers in the mean-standard deviation space. The dashed line represents the mean-standard deviation efficient frontier, images of mean-V@R and mean-CV@R efficient frontiers are shown with thick and thin lines respectively. The meaning of axes on Figures 4, 5 is similar to Figure 3 except that the horizontal axis shows respectively CV@R and standard deviation values expressed in percents.

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

1. The mean-V@R feasible set and the respective efficient frontier are not convex. However,

their 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 the V@R function itself as shown on Figures 1-2.

2. V@Rdiffers substantially from both CV@R and variance because mean-V@R efficient portfolios may differ considerably from mean-CV@R or mean-variance efficient portfolios. The distance between V@R-efficient frontier and images of other frontiers can be considerably larger than the distance between frontiers in mean-CV@Rspace or mean-variance space. Thus mean-CV@R and mean-variance efficient portfolios provide a poor approximation to mean-V@R efficient portfolios in mean-V@R space, while the opposite is not necessarily true. Besides, mean-CV@R efficient portfolios do not approximate mean-V@R efficient portfolios any better than the mean-variance efficient portfolios do. Actually, in some cases mean-variance portfolios lie closer to mean-V@R portfolios than mean-CV@R portfolios, as Figure 3 suggests. Vice versa, mean-variance portfolios can approximate mean-CV@R portfolios better than mean-V@R portfolios do (see Figure 4).

3. As one can expect, all frontiers approximate each other fairly well for high risk portfolios because for large return values all portfolios converge to portfolio consisting of only one asset with largest return. For medium and low risk portfolios the difference between frontiers can be substantial.

The experiments reported in this section represent a useful starting point for further studies. More evidence in the same direction is presented in the next section where we performed a much larger set of experiments.

5 Mean-risk optimal portfolios of stock and bond indices

The techniques developed in the previous sections allow to conduct experimental comparative studies of properties of mean-risk portfolios for different sets of assets with the special emphasis on mean-V@R efficient portfolios. In this section we report results of one such study which involves a representative selection of broad asset classes of interest to an institutional investor. The objective of this study was twofold:

- Gain further insight into the properties of the proposed methodology for computing mean-V@R efficient portfolios with the help of smoothed V@R functions (SV@R).

- Evaluate experimentally the relative properties of mean-risk efficient portfolios. In particular, we wanted to understand what advantage can give mean-V@R efficient portfolios to investor whose risk control targets are formulated in terms of V@R compared to other mean-risk efficient portfolios.

5.1 Summary of results

The data set consisted of 16 market indices: 8 Morgan Stanley Equity Price Indices for USA, UK; Italy, Japan, Russia, Argentina, Brazil and Mexico and 8 J.P. Morgan Bond Indices (Developed Markets and EMBI+ for Emerging Markets) for the same markets from January 19, 1999 to

May 15, 2002 expressed in US dollars[12]. This data set contained representative selection of developed and emerging stock and bond markets and allowed construction of a wide range of portfolios from fairly safe to extremely risky. Moreover, it allowed to evaluate the performance of mean-risk efficient portfolios in times of turbulent stock markets and extreme events in emerging markets, like Argentinian crisis of the second half of 2001. These indices were considered as the composite assets and it was assumed that an institutional investor can choose portfolios which reproduce these indices.

Three types of experiments were performed with this data[13].

1. *In sample experiments.(section 5.2).* These involved computation of mean-risk efficient frontiers from a given windows of historic data, exactly as it was done in section 4, around 1100 of these frontiers were computed for each of the risk measures which required solution of about 250000 portfolio optimization problems (5), (9) and (11). Substantial differences between different frontiers were observed. In particular, the averaged maximal distance between mean-V@R efficient frontier and other frontiers exceeded 30%.

2. *Out of sample experiments (section 5.3).* Portfolios computed on step 1 were hold for a fixed number of trading days following the time windows utilized for their computation (1-60 days). Out of sample V@R was obtained from portfolio returns during the holding periods. V@R-optimal portfolios maintained advantage over the other mean-risk portfolios in terms of V@R which in some cases arrived at the level of 10-17%.

3. *Simulated management of investment fund*[14]. The same portfolios were used for simulated management of a hypothetical investment fund which for each trading day computes risk-efficient portfolios using newly arrived price information and use these portfolios for (possibly partial) rebalancing of its current portfolio. The V@R control advantage of V@R-optimal portfolios over other mean-risk portfolios arrived in some cases at the levels of 14-20%.

5.2 In sample experiments

We report here the results of experiment where the risk measures are computed using return observations over a short time horizon, for example over one day. The whole time horizon $t = 1, ..., T$ is divided into consecutive nonoverlapping periods of length n_1. Asset returns are computed for all these periods which we call observation periods. Observations of return for n_2 consecutive observation periods are used to compute statistical characteristics and risk measures of return. This larger period of length $n_1 n_2$ is called the sampling period. The time horizon $t = 1, ..., T$ contains $T - n_1 n_2$ different sampling periods which start at times $i_t = 1, ..., T - n_1 n_2$. We computed boundaries of mean-V@R, mean-CV@R and mean-variance feasible sets by multiple solving of respective risk minimization problems (11),(9) and (5) for all feasible sampling periods.

[12]We pruned days for which some of the emerging markets information was missing which left us with 829 trading days of full price information.

[13]All experiments described in this section were performed on Dell Latitude 810 laptop with Windows 2000 operation system equipped with Pentium III 1133 Mhz processor. Some critical parts of the Matlab code were reprogrammed in C in order to speed up execution times. More than 90% of the processing time was spent on solving the mean-V@Roptimization problems (11). One solution of such problem took 4-5 seconds of CPU time. The parameters of the algorithm were fixed after few trial runs at the beginning and after that the computation of mean-risk optimal portfolios proceeded without any human intervention.

[14]Details of these experiments are available from the authors upon request

We report here results for $T = 829, n_1 = 1, n_2 = 250$ which yields 579 different mean-risk efficient frontiers for each of the three risk measures[15]. That is, in this experiment we computed one day mean-risk efficient frontiers for 579 days from approximately one year of daily historic return data immediately preceding each given day[16].

The observed behavior of efficient frontiers was similar to one shown on Figures 3-5. Here we focus on quantitative measures of difference between efficient frontiers.

1. Comparison of V@R of mean-V@R efficient portfolios and V@R of other mean-risk efficient portfolios.

In order to access numerically the gain which is obtained by substitution of mean-variance or mean-CV@R efficient portfolios by mean-V@R efficient portfolios we computed the following quantities:

$$E^i_{\mathrm{V@R/CV@R}} = \frac{\mathrm{V@R}(x^i_{\mathrm{CV@R}}) - \mathrm{V@R}(x^i_{\mathrm{V@R}})}{\mathrm{V@R}(x^i_{\mathrm{V@R}})}100\%,\ E^i_{\mathrm{V@R/StDev}} = \frac{\mathrm{V@R}(x^i_{\mathrm{StDev}}) - \mathrm{V@R}(x^i_{\mathrm{V@R}})}{\mathrm{V@R}(x^i_{\mathrm{V@R}})}100\%$$

Here the index i indexes all the risk minimization problems solved, $i = 1 : 35459$. Portfolios $x^i_{\mathrm{V@R}}, x^i_{\mathrm{CV@R}}$ and x^i_{StDev} denote obtained solutions of the i-th instance of the problems (11), (9) and (5) respectively. These measures express relative improvement of V@R estimates expressed in percents. After that we computed the percentage of cases where improvement exceeded threshold $\theta\%$. Results are shown in the table 2.

Threshold θ (%)	-1	-0.1	0.1	1	2	5	10	20	30	40	50
$E^i_{\mathrm{V@R/CV@R}} > \theta$ (%)	99.95	99.79	93.77	90.32	85.90	70.99	43.69	13.56	4.72	0.88	0.15
$E^i_{\mathrm{V@R/StDev}} > \theta$ (%)	99.97	99.90	94.27	91.81	88.45	76.91	57.67	20.21	5.38	1.23	0.35

Table 2. Relative improvement of V@R by mean-V@R efficient portfolios.

We see that in the considerable fraction of cases V@R optimal portfolios constitute a substantial improvement over CV@R and variance optimal portfolios as far as V@R is concerned. Indeed, improvement over 10% compared with variance optimal portfolios was obtained in around 50% of cases in both experiments. Improvement over 30% still occurred in a few percent of cases while in some (rare) cases improvement exceeded 50%. Improvement over CV@R optimal portfolios is somewhat smaller but still substantial.

In a noticeable amount of cases which constituted around 6% the difference between different mean-risk portfolios was within 0.1%. These cases correspond mostly to problems with either

[15]These frontiers were computed as follows. The equidistant grid of *target returns* $\mu_0, ..., \mu_k$ was set up and the problems (5)-(11) were solved for all the values from this grid in the place of the constant μ in the respective return constraints for which these problems were feasible. In this experiment we took $\mu_0 = 0, \mu_k = 0.8\%, k = 200, \mu_{i+1} - \mu_i = \Delta_\mu = 0.004\%$. The total number of the grid points for which solution of problems (5)-(11) existed varied with the position of the data window which supplied the return data and changed between 16 and 158. Totally, 35459 solutions of each of the problems (5)-(11) were computed.

[16]We also performed a different set of experiments with risk measures being computed for return observations over larger time period of approximately one quarter (60 trading days). Observation periods overlapped in these experiments. These experiments showed very similar results and are not reported here. They are available from the authors upon request.

very high or very low values of return constraint where portfolios consisted of only few assets and practically coincided.

Finally, in some extremely rare cases our techniques failed to find the global minimum of the V@R optimization problem and got struck in the local minimum which value exceeded the V@R value of CV@R optimal or variance optimal portfolio. Still, the difference exceeded 1% only in 1-2 cases from the total of 35459 cases reported in this table.

2. Comparison of efficient frontiers formed by mean-risk efficient portfolios.

Here we focus on the whole efficient frontiers and characterize numerically the difference between given mean-V@R efficient frontier and images of mean-CV@Rand mean-variance efficient frontiers into mean-V@R space. Similar characterization is given for mean-CV@Rand mean-variance efficient frontiers. In other words, we give a numeric expression to differences shown on Figures 3-5.

Let us denote by $x^{i\tau}_{\mathrm{V@R}}, x^{i\tau}_{\mathrm{CV@R}}$ and $x^{i\tau}_{\mathrm{StDev}}$ the solutions of problems (11), (9) and (5) respectively computed for value μ_i of target return and using historic price data from the time interval $[\tau-\Delta,\tau]$, we report here results with $\Delta = n_1 n_2 = 250$. Let us consider the following measures of distance between mean-V@R efficient frontier and image of mean-CV@R efficient frontier into mean-V@R space.

$$D^{\tau}_{\mathrm{V@R/CV@R}} = \frac{1}{i^+(\tau)-i^-(\tau)+1}\sum_{i=i^-(\tau)}^{i^+(\tau)} \frac{\mathrm{V@R}(x^{i\tau}_{\mathrm{CV@R}}) - \mathrm{V@R}(x^{i\tau}_{\mathrm{V@R}})}{\mathrm{V@R}(x^{i\tau}_{\mathrm{V@R}})}100\%,$$

$$\hat{D}^{\tau}_{\mathrm{V@R/CV@R}} = \max_{i^-(\tau)\le i\le i^+(\tau)} \frac{\mathrm{V@R}(x^{i\tau}_{\mathrm{CV@R}}) - \mathrm{V@R}(x^{i\tau}_{\mathrm{V@R}})}{\mathrm{V@R}(x^{i\tau}_{\mathrm{V@R}})}100\%$$

where $i^-(\tau)$ and $i^+(\tau)$ are the indexes of the smallest and largest return μ_i from the target return grid for which the solutions of problems (5)-(11) exist. The measure $D^{\tau}_{\mathrm{V@R/CV@R}}$ is the average relative distance between frontiers while $\hat{D}^{\tau}_{\mathrm{V@R/CV@R}}$ is the maximal relative distance between frontiers for a given time τ. We introduce distances $D^{\tau}_{\mathrm{V@R/StDev}}$, $\hat{D}^{\tau}_{\mathrm{V@R/StDev}}$, $D^{\tau}_{\mathrm{CV@R/V@R}}$, $\hat{D}^{\tau}_{\mathrm{CV@R/V@R}}$, $D^{\tau}_{\mathrm{CV@R/StDev}}$, $\hat{D}^{\tau}_{\mathrm{CV@R/StDev}}$, $D^{\tau}_{\mathrm{StDev/V@R}}$, $\hat{D}^{\tau}_{\mathrm{StDev/V@R}}$, $D^{\tau}_{\mathrm{StDev/CV@R}}$, $\hat{D}^{\tau}_{\mathrm{StDev/CV@R}}$ in a similar manner. The values of these distances are reported in Tables 3 and 4.

Distance	$D^{\tau}_{\mathrm{V@R/CV@R}}$	$D^{\tau}_{\mathrm{V@R/StDev}}$	$D^{\tau}_{\mathrm{CV@R/V@R}}$	$D^{\tau}_{\mathrm{CV@R/StDev}}$	$D^{\tau}_{\mathrm{StDev/V@R}}$	$D^{\tau}_{\mathrm{StDev/CV@R}}$
Mean	7.98	9.76	5.25	1.87	3.19	1.20
Max	18.75	22.89	11.51	6.216	8.475	3.740

Table 3. Distances between efficient frontiers measured by average relative difference

Distance	$\hat{D}^{\tau}_{\mathrm{V@R/CV@R}}$	$\hat{D}^{\tau}_{\mathrm{V@R/StDev}}$	$\hat{D}^{\tau}_{\mathrm{CV@R/V@R}}$	$\hat{D}^{\tau}_{\mathrm{CV@R/StDev}}$	$\hat{D}^{\tau}_{\mathrm{StDev/V@R}}$	$\hat{D}^{\tau}_{\mathrm{StDev/CV@R}}$
Mean	30.81	30.44	18.80	7.29	10.85	6.40
Max	59.92	64.85	40.99	14.82	14.82	15.86

Table 4. Distances between efficient frontiers measured by maximal relative difference

These tables show the average and maximal distances between efficient frontiers among 579 triples of frontiers. Results give numeric confirmation to our assertion that V@R, CV@R and variance can produce substantially different risk-optimal portfolios and different mean-risk efficient frontiers. The average V@R difference between frontiers amounted to 8-10% and the

average maximal distance exceeded 30%. This means that at every time τ existed a target return μ for which V@R of mean-CV@R and mean-variance efficient portfolios exceeded V@R of V@R-optimal portfolio in average by more than 30%. The distances in mean-CV@R and mean-variance spaces were smaller, following the pattern present on Figures 3-5 and confirming conclusions of section 4.

5.3 Out of sample experiments

Out of sample experiments permit to evaluate the potential of the mean-risk efficient portfolios for actual risk management purposes. Each of the portfolios computed during experiments described in the previous section is constructed by minimization of a given sample risk measure on the given sample interval $[t-\Delta, t]$ for a given values of t and Δ and for a given value of target return μ. Therefore they possessed the target risk-return properties during this particular interval. Suppose now that we simulate the holding of such portfolio for the time interval $[t, t+\tau]$ of the length τ immediately succeeding the sample period for which portfolios were constructed. Will such portfolios exhibit desired risk-return properties also during this holding period? For example, will V@R-optimal portfolios still possess smaller V@R compared to CV@R-optimal or variance-optimal portfolios? We show in this section that the answer to this question is positive.

The experiments consisted of the following steps.

i. Fix the value of the target return μ, holding period τ and a risk measure among V@R, CV@R and standard deviation.

ii. For each $t = \Delta+1, ..., T-\tau$ take mean-risk optimal portfolios computed for fixed values of t and μ using return observations on the interval $[t-\Delta, t]$. Compute the return of these portfolios on the time interval $[t, t+\tau]$. This gives the set of return observations with $T-\tau-\Delta$ elements[17].

iii. Compute sample V@R for this set of return observations. Let us denote it by $\mathrm{VaR}_V(\mu,\tau)$, $\mathrm{VaR}_C(\mu,\tau)$ and $\mathrm{VaR}_S(\mu,\tau)$ for the cases when the set of returns was generated correspondingly by mean-V@R, mean-CV@R and mean-variance optimal portfolios. It characterizes the risk properties of the whole set of mean-risk portfolios generated for all admissible times t and for a given value of target return μ.

iv. Repeat steps i-iii for all three risk measures, all target returns $\mu = \mu_i$ from the grid of target returns and all holding periods $\tau = 1, ...T_\tau$ with $T_\tau = 60$.

For each value μ_i of target return this experiment gives T_τ values of the relative V@Radvantage $G^{i\tau}_{\mathrm{V@R/CV@R}}$ of V@R-optimal portfolios over CV@R-optimal portfolios

$$G^{i\tau}_{\mathrm{V@R/CV@R}} = \frac{\mathrm{V@R}_C(\mu_i,\tau) - \mathrm{V@R}_V(\mu_i,\tau)}{\mathrm{V@R}_V(\mu_i,\tau)} 100\% \tag{21}$$

Averaging this measure over the set of holding periods τ gives the average V@R advantage $G^{i}_{\mathrm{V@R/CV@R}}$ of V@R-optimal portfolios over CV@R-optimal portfolios for fixed target return μ_i.The measure of advantage $G^{i}_{VaR/StDev}$ over variance-optimal portfolios is defined similarly. The results of computation are presented in Table 5.

[17] provided that portfolio optimization problems (5)-(11) have solutions for this value of μ. In the case of $\tau = 60$ this yielded 519 return observations

max $G^i_{\text{V@R/CV@R}}$	average $G^i_{\text{V@R/CV@R}}$	max $G^i_{\text{V@R/StDev}}$	average $G^i_{\text{V@R/StDev}}$
9.8	4.5	7.1	1.6

Table 5. Out of sample advantage of V@R-optimal portfolios

The maximal and average values of measures $G^i_{\text{V@R/CV@R}}$ and $G^i_{\text{V@R/StDev}}$ presented in this table are taken over the range of target returns which correspond to the range of yearly returns between 5% and 25%. Thus, CV@R-optimal and variance-optimal portfolios do possess larger values of out of sample V@R compared with V@R-optimal portfolios. Although difference averaged among holding times and target returns is between 2-7%, for some values of returns it reaches 10% and in another set of experiments it arrived at 17%.

6 Summary

We studied here Value-at-Risk (V@R) in the context of V@R portfolio optimization for the purpose of selection of mean-V@R efficient portfolios similar to the classical mean-variance approach. Important theme of this paper is comparison of V@R with other important risk measures from the optimization point of view, in particular with classical variance and more recent conditional V@R(CV@R). This comparison is performed on computational plane using the stock market data and the representative selection of stock and bond indices from developed and emerging markets.

Our conclusion is that V@R is a substantially different risk measure and investor which expresses her risk preferences in terms of V@R should work with V@R directly in the context of mean-risk trade-off. In particular, efficient frontiers constructed on the basis of the other risk measures can be a poor approximation for mean-V@R efficient frontier.

Moreover, we argue that computation of mean-V@R efficient portfolios based on historic data is a feasible task, despite the fact that V@R optimization is more difficult than variance optimization or CV@R optimization. For this purpose we developed a set of V@R optimization tools centered around the notion of smoothed V@R(SV@R) which filters out nonsmooth irregular local behavior of historic V@R. Effectiveness of SV@R approach to V@R optimization was confirmed during experiments reported in this paper which involved computation of more than 80000 V@R-optimal portfolios.

7 Acknowledgment

We are grateful to the editor Professor Jorion and the anonymous referee for useful comments which contributed to improvement of the quality of exposition. Besides, we express our gratitude to Dr. Giorgio Consigli who supplied us with financial data utilized for experiments in section 5.

8 Appendix: construction of SV@R

Here we provide mathematical details about the smoothing technique which was used in sections 3-5 for computing of V@Refficient 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

$$F(k,x) = \max_i {}^{k+1} f_i(x) \tag{22}$$

which is equal to the $k+1-$th largest from these functions. In other words, there exists some index $j = j(x)$ such that $F(k,x) = f_j(x)$ and inequality $f_j(x) \leq f_i(x)$ is satisfied for at least k functions $f_i(x)$, $i \neq j$ while inequality $f_j(x) \geq f_i(x)$ is satisfied for at least $N-k-1$ functions $f_i(x)$, $i \neq j$.

Taking $f_i(x) = x^T \xi^i$ we obtain that the minimization of $F(k,x)$ in this case is the same as solution of (11). The function $F(k,x)$ is 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 an 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:

$$F(k,x) = \sum_{i=1}^{N} c_i(x) f_i(x) \tag{23}$$

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 - an arbitrary subset of M^i which contains exactly k elements;

$X(\Lambda_k^i)$ - 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 - the family of all different sets Λ_k^i;

$\mathbb{I}_A$ - the indicator function of set A, i.e. $\mathbb{I}_A = 1$ if $x \in A$ and $\mathbb{I}_A = 0$ otherwise. In order to simplify notations we shall omit the dependence of $\mathbb{I}_A$ on x where it will not cause confusion. The following lemma gives the required representation of function $F(k,x)$.

Lemma 1 *Suppose that $F(k,x)$ is defined by (22). Then*

$$F(k,x) = \frac{1}{C(x)} \sum_{i=1}^{N} c_i(x) f_i(x), \tag{24}$$

where

$$c_i(x) = \sum_{\Lambda_k^i \in \Theta_k^i} \mathbb{I}_{X\left(\Lambda_k^i\right)},\ C(x) = \sum_{i=1}^{N} c_i(x) \tag{25}$$

Proof.

Suppose that $F(k,x) = f_r(x)$. We are going to prove that also the right hand side of (24) equals $f_r(x)$ provided coefficients $c_i(x)$ are selected according to (25).

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 (25). 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 the function $F_\epsilon(k,x)$ defined as follows:

$$F_\epsilon(k,x) = \frac{1}{C_\epsilon(x)} \sum_{i=1}^{N} c_i^\epsilon(x) f_i(x), \tag{26}$$

where

$$c_i^\epsilon(x) = \sum_{\Lambda_k^i \in \Theta_k^i} \prod_{j \in \Lambda_k^i} \varphi_\epsilon\left(\Delta_{ij}^x\right) \prod_{j \in M^i \setminus \Lambda_k^i} \varphi_\epsilon\left(-\Delta_{ij}^x\right), \tag{27}$$

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

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 (22) 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. $\square$

By selecting a family of functions $\varphi_\epsilon(z)$, specific smooth approximation of the nondifferentiable function $F(k,x)$ can be obtained. Under additional technical assumptions this approximation will converge uniformly to the original function. The selection of $\varphi_\epsilon(z)$ should take into account computational considerations. For general functions $\varphi_\epsilon(z)$ the computation of $c_i^\epsilon(x)$ from (27) can be a difficult task because the number of terms in the sum from (27) grows exponentially with the number of observations N, more precisely it equals $(N-1)!/k!\,(N-k-1)!$. 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 such function that satisfies conditions of Theorem 2 is the cubic spline:

$$\varphi_\epsilon(z) = \begin{cases} 1 & \text{if} \quad z \leq 0 \\ 1 - \frac{16}{3\epsilon^3} z^3 & \text{if} \quad 0 \leq z \leq \frac{\epsilon}{4} \\ \frac{5}{6} + \frac{2}{\epsilon} z - \frac{8}{\epsilon^2} z^2 + \frac{16}{3\epsilon^3} z^3 & \text{if} \quad \frac{\epsilon}{4} \leq z \leq \frac{3\epsilon}{4} \\ \frac{16}{3} - \frac{16}{\epsilon} z + \frac{16}{\epsilon^2} z^2 - \frac{16}{3\epsilon^3} z^3 & \text{if} \quad \frac{3\epsilon}{4} \leq z \leq \epsilon \\ 0 & \text{if} \quad z \geq \epsilon \end{cases} \tag{28}$$

The benefit of using such functions becomes clear from the following lemmas.

Lemma 3 *Suppose that in addition to the conditions of Theorem 2 the 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:*

$$F_\epsilon(k,x) = \frac{1}{C_\epsilon(x)} \sum_{i : |F(k,x) - f_i(x)| \leq \epsilon} c_i^\epsilon(x) f_i(x) \tag{29}$$

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 (27) 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 (26)-(27) 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 (28).

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 (27) can be represented equivalently as follows:*

$$c_i^\epsilon(x) = \sum_{q,r:\ q-r=k-k_-^i} b_q^- b_r^+ \tag{30}$$

$$b_q^- = \sum_{\Lambda_q\subseteq\Lambda_{\epsilon-}^i} \prod_{j\in\Lambda_q} \varphi_\epsilon\left(\Delta_{ij}^x\right),\ b_r^+ = \sum_{\Lambda_r\subseteq\ \Lambda_{\epsilon+}^i} \prod_{j\in\Lambda_r} \varphi_\epsilon\left(-\Delta_{ij}^x\right) \tag{31}$$

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

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

and k_-^i is the number of elements in set $\Lambda_{\epsilon-}^i$.

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 $\Lambda_k^i \in \Theta_k^i$ can be represented as follows:

$$\Lambda_k^i = \left(\Lambda_-^i \setminus \Lambda_r\right) \cup \Lambda_q,\ M^i \setminus \Lambda_k^i = \left(\left(M^i \setminus \Lambda_-^i\right) \setminus \Lambda_q\right) \cup \Lambda_r$$

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 (27) for coefficient $c_i^\epsilon(x)$ as follows:

$$c_i^\epsilon(x) = \sum_{\substack{\Lambda_r \subseteq \Lambda_-^i \\ \Lambda_q \subseteq M^i \setminus \Lambda_-^i \\ k_i+q-r=k}} \prod_{j\in\Lambda_q} \varphi_\epsilon\left(\Delta_{ij}^x\right) \prod_{j\in\Lambda_r} \varphi_\epsilon\left(-\Delta_{ij}^x\right)$$

$$= \sum_{q,r:\ q-r=k-k_i} \left(\sum_{\Lambda_q \subseteq \Lambda_-^i} \prod_{j\in\Lambda_q} \varphi_\epsilon\left(\Delta_{ij}^x\right) \right) \left(\sum_{\Lambda_r \subseteq M^i \setminus \Lambda_-^i} \prod_{j\in\Lambda_r} \varphi_\epsilon\left(-\Delta_{ij}^x\right) \right)$$

Assertion of lemma is obtained from the last expression repeating the argument of Lemma 3 which leads to substitution of Λ_-^i by $\Lambda_{\epsilon-}^i$, $M^i \setminus \Lambda_-^i$ by $\Lambda_{\epsilon+}^i$ and k_i by k_-^i. $\square$

Expressions (30),(31) allow efficient computation of coefficients $c_i^\epsilon(x)$ because b_q^- and b_r^+ can be computed recursively. Indeed, suppose that $s \in \Lambda_{\epsilon-}^i$. Then

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

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

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 $s = q, ..., k_-^i$:

- Take $a = d_s e_s$ if $s = q$ and $a = d_s e_s + \bar{a}$ if $s > q$.
- Take $e_s = \bar{a}$ if $s > q$.
- Take $\bar{a} = a$ if $s \ge q$.

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

The important question is 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 (26)-(27) is not encouraging because the number of arithmetic operations necessary for straightforward implementation of (27) 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 (28). 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*:

$$S \leq (1+\rho_N)\, N^3 \tag{32}$$

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 Algorithm 1 and implementation of expressions (29)-(31). 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,

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

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 (28) 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 (30) 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:

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

The last inequality yields

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

because $k_-^i + k_+^i \leq N$. 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 (29). Therefore the following estimate for S holds:

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

which completes the proof. □

In reality overhead will be much smaller than (32) because for small and moderate ϵ we have $k_-^i \ll N$, $k_+^i \ll N$, $k_0 \ll N$. Estimates similar to (32) hold also for other functions $\varphi_\epsilon(z)$ which satisfy conditions of Lemma 4.

Algorithm 1 and (29)-(31) were implemented as Matlab M-files and were used as the 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

Amendment to the Capital Accord to Incorporate Market Risks (1996), Bank for International Settlements.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999), Coherent measures of risk, *Mathematical Finance* **9**, 203 – 228.

Basak, S. & Shapiro, A. (2001), Value-at-risk based risk management: Optimal policies and asset prices, *Rev. Financ. Stud.* **14**, 371–405.

Birge, J. R. & Louveaux, F. (1997), *Introduction to Stochastic Programming*, Springer, New York.

Brooke, A., Kendrick, D., Meeraus, A. & Raman, R. (1998), *GAMS, A User's Guide*, GAMS Development Corporation.

Cariño, D. R. & Ziemba, W. T. (1998), Formulation of the Russell-Yasuda-Kasai financial planning model, *Operations Research* **46**(4), 433–449.

Consigli, G. & Dempster, M. (1998), Dynamic stochastic programming for asset-liability management, *Annals of Oper. Res.* **81**, 131–161.

Cover, T. M. (1991), Universal portfolios, *Mathematical Finance* **1**(1), 1–29.

Dembo, R. & Rosen, D. (1999), The practice of portfolio replication. A practical overview of forward and inverse problems, *Annals of Operations Research* **85**, 267–284.

Duarte, A. M. (1999), Fast computation of efficient portfolios, *Journal of Risk* **1**(4), 1–24.

Duffie, D. & Pan, J. (1997), An overview of value at risk, *Journal of Derivatives* **4**, 7–49.

Dupacova, J., Bertocchi, M. & Moriggia, V. (1998), Postoptimality for scenario based financial planning models with an application to bond portfolio management, *in* W. Ziemba & J. Mulvey, eds, 'Worldwide Asset and Liability Management', Cambridge University Press, pp. 263–285.

Ermoliev, Y. & Wets, R. J.-B., eds (1988), *Numerical Techniques for Stochastic Optimization*, Springer Verlag, Berlin.

Gaivoronski, A. A. & de Lange, P. E. (2000), An asset liability management model for casualty insurers: Complexity reduction vs. parametrized decision rules, *Annals of Operations Research* **99**, 227–250.

Gaivoronski, A. A. & Pflug, G. (1999), Finding optimal portfolios with constraints on Value-at-Risk, *in* B. Green, ed., 'Proceedings of the Third International Stockholm Seminar on Risk Behaviour and Risk Management', Stockholm University.

Gaivoronski, A. A. & Stella, F. (2000), Stochastic nonstationary optimization for finding universal portfolios, *Annals of Operations Research* **100**, 165–188.

Grootveld, H. & Hallerbach, W. G. (2000), Upgrading Value-at-Risk from diagnostic metric to decision variable: A wise thing to do? Working Paper, Erasmus University Rotterdam.

Jorion, P. (2001), *Value at Risk: The New Benchmark for Managing Financial Risk*, second edn, McGraw- Hill, New York.

Kall, P. & Wallace, S. (1994), *Stochastic Programming*, John Wiley and Sons, New York.

Kaplanski, G. & Kroll, Y. (2002), VAR risk measures vs traditional risk measures: An analysis and survey, *Journal of Risk* **4**(3), 1–28.

Krokhmal, P., Uryasev, S. & Palmquist, J. (2001), Portfolio optimization with conditional value-at-risk objective and constraints, *Journal of Risk* **4**(2), 43–68.

Markowitz, H. (1952), Portfolio selection, *Journal of Finance* **8**, 77–91.

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

Mulvey, J. M. & Vladimirou, H. (1992), Stochastic network programming for financial planning problems, *Management Science* **38**(11), 1642–1664.

Ogryczak, W. & Ruszczynski, A. (1999), From stochastic dominance to mean-risk models: Semi-deviations as risk measures, *European Journal of Operational Research* **116**, 33–50.

Pfingsten, A., Wagner, P. & Wolferink, C. (2004), An empirical investigation of the rank correlation between different risk measures, *Journal of Risk* **6**(4), 55–74.

Prekopa, A. (1995), *Stochastic Programming*, Kluwer Academic Publishers.

RiskMetricsTM (1995), third edition, J.P. Morgan.

The Mathworks Inc. (1984-2004), *Matlab 7.0*, The MathWorks Inc.

Uryasev, S. & Rockafellar, R. (1999), Optimization of conditional value-at-risk, Technical Report 99-4, ISE Dept. Univeristy of Florida.

Wets, R.-B. (1982), Stochastic programming: Solution techniques and approximation schemes, *in* A.Bachem, M.Groetshel & B.Korte, eds, 'Mathematical Programming: The State of the Art', Springer Verlag, Berlin.

Wu, G. & Xiao, Z. (2002), An analysis of risk measures, *Journal of Risk* **4**(4), 53–75.

Zenios, S. A. (1993), A model for portfolio management with mortgage-backed securities, *Ann. Oper. Res.* **43**, 337–356.