WORST-CASE VALUE-AT-RISK AND ROBUST PORTFOLIO OPTIMIZATION: A CONIC PROGRAMMING APPROACH

LAURENT EL GHAOUI

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, elghaoui@eecs.berkeley.edu

MAKSIM OKS

Department of Industrial Engineering and Operations Research, University of California, Berkeley, California 94720, max@ieor.berkeley.edu

FRANCOIS OUSTRY

INRIA Rhone-Alpes ZIRST, 655 avenue de l'Europe 38330 Montbonnot Saint-Martin, France, francois.oustry@inria.fr

Classical formulations of the portfolio optimization problem, such as mean-variance or Value-at-Risk (VaR) approaches, can result in a portfolio extremely sensitive to errors in the data, such as mean and covariance matrix of the returns. In this paper we propose a way to alleviate this problem in a tractable manner. We assume that the distribution of returns is partially known, in the sense that only *bounds* on the mean and covariance matrix are available. We define the worst-case Value-at-Risk as the largest VaR attainable, given the partial information on the returns' distribution. We consider the problem of computing and optimizing the worst-case VaR, and we show that these problems can be cast as semidefinite programs. We extend our approach to various other partial information on the distribution, including uncertainty in factor models, support constraints, and relative entropy information.

Received September 2001; revision received July 2002; accepted July 2002.

Subject classifications: Finance, portfolio: Value-at-Risk, portfolio optimization. Programming, nonlinear: semidefinite programming, dualing, robust optimization.

Area of review: Financial Services.

1. INTRODUCTION

We consider a one-period portfolio optimization problem. Over the period, the percentage return of asset i is equal to x_i, with x modeled as a random n-vector. For a given allocation vector w, the total return of the portfolio is the random variable

$$r(w, x) = \sum_{i=1}^{n} w_i x_i = w^{\mathrm{T}} x.$$

The investment policies are constrained. We denote by $\mathcal{W}$ the set of admissible portfolio allocation vectors. We assume that $\mathcal{W}$ is a bounded polytope that does not contain 0.

The basic optimal investment problem is to choose $w \in \mathcal{W}$ to make the return high while keeping the associated risk low. Depending on how we define the risk, we come up with different optimization problems.

1.1. Some Classical Measures of Risk

In the Markowitz approach (see Markowitz 1952, Luenberger 1999), it is assumed that the mean $\hat{x}$ and covariance matrix Γ of the return vector are both known, and risk is defined as the variance of the return. Minimizing the risk subject to a lower bound on the mean return leads to the familiar problem

$$\text{minimize} \quad w^{\mathrm{T}} \Gamma w \quad \text{subject to} \quad \hat{x}^{\mathrm{T}} w \geqslant \mu, \quad w \in \mathcal{W}, \qquad (1)$$

where μ is a pre-defined lower bound on the mean return.

The Value-at-Risk (VaR) framework (see, for example, Linsmeier and Pearson 1996) instead looks at the probability of losses. The VaR is defined as the minimal level γ such that the probability that the portfolio loss $-r(w, x)$ exceeds γ is below ϵ:

$$V(w) = \min \gamma \quad \text{subject to} \quad \mathbf{Prob}\{\gamma \leqslant -r(w, x)\} \leqslant \epsilon,$$

where $\epsilon \in (0, 1]$ is given (say, $\epsilon \simeq 2\%$). In contrast to the Markowitz framework, which requires the knowledge of the first and second moments of the distribution of returns only, the VaR above assumes that the entire distribution is perfectly known. When this distribution is Gaussian, with given mean $\hat{x}$ and covariance matrix Γ, the VaR can be expressed as

$$V(w) = \kappa(\epsilon)\sqrt{w^{\mathrm{T}} \Gamma w} - \hat{x}^{\mathrm{T}} w, \qquad (2)$$

where $\kappa(\epsilon) = -\Phi^{-1}(\epsilon)$.

In practice, the distribution of returns is not Gaussian. One then can use the Chebyshev bound to find an upper bound on the probability that the portfolio loss $-r(w, x)$ exceeds γ. This bound is based on the sole knowledge of the first two moments of the distribution, and results in the formula (2), where now $\kappa(\epsilon) = 1/\sqrt{\epsilon}$. In fact, the classical Chebyshev bound is not exact, meaning that the upper bound is not attained; we can replace it by its exact version, as given by Bertsimas and Popescu (2000) by simply

0030-364X/03/5104-0543 $05.00
1526-5463 electronic ISSN

setting $\kappa(\epsilon) = \sqrt{(1-\epsilon)/\epsilon}$ (we will also obtain this result in §2.1).

In all the above cases, the problem of minimizing the VaR over admissible portfolios adopts the following form:

$$\text{minimize} \quad \kappa\sqrt{w^{\mathrm{T}}\Gamma w} - \hat{x}^{\mathrm{T}} w \quad \text{subject to} \quad w \in \mathcal{W}, \tag{3}$$

where κ is an appropriate "risk factor," which depends on the prior assumptions on the distribution of returns (Gaussian, arbitrary with given moments, etc.). When $\kappa \geqslant 0$ (which in the Gaussian case is true if and only if $\epsilon \in (0, \frac{1}{2}]$), $V(w)$ is a convex function of w, and the above problem can easily be solved globally using, for example, interior-point techniques for convex, second-order cone programming (SOCP; see, e.g., Lobo et al. 1998).

The classical frameworks may not be appropriate for several reasons. Clearly, the variance is not an appropriate measure of risk when the distribution of returns exhibits "fat" tails. On the other hand, the exact computation of VaR requires a complete knowledge of the distribution. Even with that knowledge in hand, the computation of VaR amounts to a numerical integration in a possibly high dimensional space, which is computationally cumbersome. Furthermore, numerical techniques such as Monte-Carlo simulation (Linsmeier and Pearson 1996) are not easily extended to portfolio design.

1.2. The Problem of Data Uncertainty

Despite their shortcomings, the above frameworks do provide elegant solutions to risk analysis and portfolio design. However, they suffer from an important drawback, which is perhaps not so well recognized: These approaches require a perfect knowledge of the data, in our case the mean and covariance matrix. In practice, the data are often prone to errors. Portfolio optimization based on inaccurate point estimates may be highly misleading—meaning, for example, that the true VaR may be widely worse than the optimal computed VaR. This problem is discussed extensively by Black and Litterman (1992), who propose a method to combine the classical Markowitz approach with a priori information or "investor's views" on the market, and by Pearson and Ju (1999).

Errors in the mean and covariance data may have several origins. It may be difficult to obtain statistically meaningful estimates from available historical data; this is often true for the means of stock returns (Black and Litterman 1992). These possibly large estimation errors contribute to a hidden, "numerical" risk not taken into account in the above risk measures. Note that most statistical procedures produce bounds of confidence for the mean vector and covariance matrix; the frameworks above do not use this crucial information.

Another source of data errors comes from modelling itself. To use the variance-covariance approach for complex portfolios, one has to make a number of simplifications, a process referred to as "risk mapping" in Linsmeier and Pearson's (1996) paper. Thus, possibly large modelling errors are almost always present in complex portfolios. We discuss these errors in more detail in §3.

Yet another source of data perturbations could be the user of the Value-at-Risk system. In practice, it is of interest to "stress test" Value-at-Risk estimates, to analyze the impact of different factors and scenarios on these values (Pritsker 1997). It is possible, of course, to come up with a (finite) number of different scenarios and compute the corresponding VaR. (We will return to this problem in §2.3.) However, in many cases, one is interested in analyzing the worst-case impact of possibly continuous changes in the correlation structure, corresponding to an infinite number of scenarios. Such an endeavour becomes quickly intractable using the (finite number of) scenarios approach.

1.3. The Worst-Case VaR

In this paper, our goal is to address some of the issues outlined above in a numerically tractable way. To this end we introduce the notion of worst-case VaR.

Our approach is to assume that the true distribution of returns is only partially known. We denote by $\mathcal{P}$ the set of allowable distributions. For example, $\mathcal{P}$ could consist of the set of Gaussian distributions with mean $\hat{x}$ and covariance matrix Γ, where $\hat{x}$ and Γ are only known up to given componentwise bounds.

For a given loss probability level $\epsilon \in (0, 1]$, and a given portfolio $w \in \mathcal{W}$, we define the *worst-case Value-at-Risk* with respect to the set of probability distributions $\mathcal{P}$ as

$$V_{\mathcal{P}}(w) := \min \gamma \quad \text{subject to} \quad \sup \mathbf{Prob}\{\gamma \leqslant -r(w, x)\} \leqslant \epsilon, \tag{4}$$

where the **sup** in the above expression is taken with respect to all probability distributions in $\mathcal{P}$. The corresponding *robust portfolio optimization* problem is to solve

$$V_{\mathcal{P}}^{\mathrm{opt}} := \min V_{\mathcal{P}}(w) \quad \text{subject to} \quad w \in \mathcal{W}. \tag{5}$$

The VaR based on the (exact) Chebyshev bound is a special case of the above, with $\mathcal{P}$ the set of probability distributions with given mean and covariance.

1.4. Main Results and Paper Outline

Our main result is that, for a large class of allowable probability distribution sets $\mathcal{P}$, the problem of computing and optimizing the worst-case VaR can be solved exactly by solving a *semidefinite programming problem* (SDP). SDPs are convex, finite dimensional problems for which very efficient, polynomial-time interior-point methods, as well as bundle methods for large-scale (sparse) problems, became available recently. For a review of SDP see, for example, Nesterov and Nemirovsky (1994), Vandenberghe and Boyd (1996), or Saigal et al. (2000).

When the mean and covariance matrix are uncertain but bounded, our solution produces not only the worst-case VaR of an optimal portfolio, but at the same time computes a positive semidefinite covariance matrix and a mean vector that satisfy the bounds and that are optimal for our problem. Thus, we select the covariance matrix and the mean vector that is the most prudent for the purpose of computing or optimizing the VaR.

Some of the probability distribution sets $\mathcal{P}$ we consider, specifically those involving support information, lead to seemingly untractable (NP-hard) problems. We show how to compute upper bounds on these problems via SDP.

Lobo and Boyd (1999) were the first to address the issue of worst-case analysis and robustness with respect to second-order moment uncertainty, in the context of the Markowitz framework. They examine the problem of minimizing the worst-case variance with (componentwise or ellipsoidal) bounds on moments. They show that the computation of the worst-case variance is a semidefinite program, and produce an alternative projections algorithm adequate for solving the corresponding portfolio allocation problem. Similarly, the paper by Halldórsson and Tütüncü (2000) discusses a robust formulation of Marikowitz's problem with componentwise bounds on mean and covariance matrix and presents a polynomial time interior-point algorithm for solving it. Our paper extends these results to the context of VaR, with various partial information on the probability distribution. After our first draft was completed, we learned of several more works in the area of robust portfolio optimization. Costa and Paiva (2001) consider the problem of robust portfolio selection for tracking error with the polytopic uncertainty in the mean and covariance matrix of asset returns. They formulate the problem as an SDP, although as we show in §2.3, it is possible to do it using SOCP. In their recent paper, Goldfarb and Iyengar (2001) develop a robust factor model for the asset returns, similarly to our approach in §3. For the uncertainty structures they consider, they're able to formulate several robust portfolio selection problems as SOCPs.

In our approach, we were greatly inspired by the recent work of Bertsimas and Popescu (2000), who also use SDP to find (bounds for) probabilities under partial probability distribution information and apply this approach to option pricing problems (see Berstimas and Popescu 1999). To our knowledge, these papers are the first to make and exploit explicit connections between option pricing and SDP optimization.

The paper is organized as follows. In §2, we consider the problem of worst-case VaR when the mean and covariance matrix are both exactly known, then extend our analysis to cases when the mean and covariance (or second-moment) matrix are known only within a given convex set. We then specialize our results to two kinds of bounds: polytopic and componentwise. In §3, we examine uncertainty structures arising from factor models. We show that uncertainty on the factor's covariance data, as well as on the sensitivity matrix, can be analyzed via SDP, via an upper bound on the worst-case VaR. Section 4 is devoted to several variations on the problems examined in §2: exploiting support information, ruling out discrete probability distributions via relative entropy constraints, handling multiple VaR constraints. We provide a numerical illustration in §5.

2. WORST-CASE VAR WITH MOMENT UNCERTAINTY

In this section, we address the problem of worst-case VaR in the case when the moments of the returns' probability distribution are only known to belong to a given set, and the probability distribution is otherwise arbitrary.

2.1. Known Moments

To lay the ground work for our future developments, we begin with the assumption that the mean vector $\hat{x}$ and covariance matrix Γ of the distribution of returns are known exactly. For two $n \times n$ symmetric matrices $A, B \in \mathcal{S}_n$, $A \succeq B$ (resp. $A \succ B$) means $A - B$ is positive semidefinite (resp. definite). We assume that $\Gamma \succ 0$, although the results can be extended to rank-deficient covariance matrices. We denote by Σ the second-moment matrix:

$$\Sigma := \mathbf{E}\begin{bmatrix} x \\ 1 \end{bmatrix}\begin{bmatrix} x \\ 1 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} S & \hat{x} \\ \hat{x}^{\mathrm{T}} & 1 \end{bmatrix}, \quad \text{where } S := \Gamma + \hat{x}\hat{x}^{\mathrm{T}}. \tag{6}$$

From the assumption $\Gamma \succ 0$, we have $\Sigma \succ 0$.

The following theorem provides several equivalent representations of the worst-case VaR when moments are known exactly. Each one will be useful later for various cases of moment uncertainty. For symmetric matrices of the same size, $\langle A, B\rangle = \mathbf{Tr}(AB)$ denotes the standard scalar product in the space of symmetric matrices.

Theorem 1. *Let $\mathcal{P}$ be the set of probability distributions with mean $\hat{x}$ and covariance matrix $\Gamma \succ 0$. Let $\epsilon \in (0, 1]$ and $\gamma \in \mathbf{R}$ be given. The following propositions are equivalent.*

1. *The worst-case VaR with level ϵ is less than γ, that is,*

$$\sup \mathbf{Prob}\{\gamma \leqslant -r(w, x)\} \leqslant \epsilon,$$

where the **sup** *is taken with respect to all probability distributions in $\mathcal{P}$.*

2. *We have*

$$\kappa(\epsilon)\|\Gamma^{1/2}w\|_2 - \hat{x}^{\mathrm{T}}w \leqslant \gamma, \tag{7}$$

where

$$\kappa(\epsilon) := \sqrt{\frac{1-\epsilon}{\epsilon}}. \tag{8}$$

3. *There exist a symmetric matrix $M \in \mathcal{S}_{n+1}$ and $\tau \in \mathbf{R}$ such that*

$$\langle M, \Sigma\rangle \leqslant \tau\epsilon, \qquad M \succeq 0, \qquad \tau \geqslant 0,$$

$$M + \begin{bmatrix} 0 & w \\ w^{\mathrm{T}} & -\tau + 2\gamma \end{bmatrix} \succeq 0, \tag{9}$$

where Σ is the second-moment matrix defined in Equation (6).

4. *For every* $x \in \mathbf{R}^n$ *such that*

$$\begin{bmatrix} \Gamma & x-\hat{x} \\ (x-\hat{x})^{\mathrm{T}} & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \tag{10}$$

we have $-x^{\mathrm{T}}w \leqslant \gamma$.

5. *There exist* $\Lambda \in \mathscr{S}_n$ *and* $v \in \mathbf{R}$ *such that*

$$\langle \Lambda, \Gamma\rangle + \kappa(\epsilon)^2 v - \hat{x}^{\mathrm{T}} w \leqslant \gamma, \quad \begin{bmatrix} \Lambda & w/2 \\ w^{\mathrm{T}}/2 & v \end{bmatrix} \succeq 0. \tag{11}$$

Let us comment on the above theorem. The equivalence between Propositions 1 and 2 will be proved below, but can also be obtained as an application of the (exact) multivariate Chebyshev bound given in Bertsimas and Popescu (2000). This implies that the problem of optimizing the VaR over $w \in \mathscr{W}$ is equivalent to

$$\text{minimize} \quad \kappa(\epsilon)\sqrt{w^{\mathrm{T}}\Gamma w} - \hat{x}^{\mathrm{T}} w \quad \text{subject to} \quad w \in \mathscr{W}. \tag{12}$$

As noted in the introduction, this problem can be cast as a second-order cone programming problem. SOCPs are special forms of SDPs that can be solved with efficiency close to that of linear programming (see, e.g., Lobo et al. 1998).

The equivalence between Propositions 1 and 3 in the above theorem is a consequence of duality (in an appropriate sense; see proof below).

Note that proposition 4 implies that the worst-case VaR can be computed via the SDP in variable x

$$V_{\mathscr{P}}(w) = \max\ -x^{\mathrm{T}} w, \quad \text{subject to Condition (10).}$$

The above provides a deterministic, or “game-theoretic,” interpretation of the VaR. Indeed, since $\Gamma \succ 0$, Condition (10) is equivalent to $x \in \mathscr{E}$, where $\mathscr{E}$ is the ellipsoid

$$\mathscr{E} = \{x \mid (x-\hat{x})^{\mathrm{T}}\Gamma^{-1}(x-\hat{x}) \leqslant \kappa(\epsilon)^2\}.$$

Therefore, the worst-case VaR can be interpreted as the maximal loss $-x^{\mathrm{T}}w$ when the returns are deterministic, known to belong to $\mathscr{E}$, and are otherwise unknown.

Expression (11) for the worst-case VaR allows us to optimize it by making w a variable. This is a SDP solution to the worst-case VaR optimization problem, which of course is not competitive, in the case of known moments, with the SOCP formulation (12). However, this SDP formulation will prove useful because it can be extended to the more general cases seen in §2.2, while the SOCP approach generally cannot.

Proof of Theorem 1. We first prove the equivalence between Propositions 1 and 3, then show that the latter is equivalent to 2. Proposition 4 is straightforwardly equivalent to the analytical formulation given in Proposition 2. Finally, the equivalence between Propositions 4 and 5 follows from simple SDP duality.

Computing the worst-case probability. We begin with the problem of computing the worst-case probability for a fixed loss level γ. We introduce the Lagrange functional for $(p, M) \in \mathscr{K}(\mathbf{R}^n) \times \mathscr{S}_{n+1}$

$$\mathscr{L}(p, M) = \int_{\mathbf{R}^n} \chi_{\mathscr{S}}(x) p(x)\, dx + \left\langle M, \Sigma - \int_{\mathbf{R}^n} \begin{bmatrix} x \\ 1 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}^{\mathrm{T}} p(x)\, dx \right\rangle,$$

where $\langle A, B\rangle = \mathbf{Tr}(AB)$ denotes the scalar product in the space of symmetric matrices, $M = M^{\mathrm{T}}$ is a Lagrange multiplier matrix, and $\chi_{\mathscr{S}}$ is the indicator function of the set

$$\mathscr{S} = \{x \mid \gamma \leqslant -x^{\mathrm{T}} w\}. \tag{13}$$

Because $\Sigma \succ 0$, strong duality holds (Smith 1995, Bonnans and Shapiro 2000). Thus, the original problem is equivalent to its dual. Hence, the worst-case probability is

$$\theta_{\mathrm{wc}} = \inf_{M=M^{\mathrm{T}}} \theta(M), \tag{14}$$

where $\theta(M)$ is the dual function

$$\theta(M) = \sup_{p\in\mathscr{K}(\mathbf{R}^n)} \mathscr{L}(p, M) = \langle M, \Sigma\rangle + \sup_p \int_{\mathbf{R}^n} (\chi_{\mathscr{S}}(x) - l(x)) p(x)\, dx,$$

and $l(x)$ is the quadratic function

$$l(x) = [x^{\mathrm{T}} \quad 1] M [x^{\mathrm{T}} \quad 1]^{\mathrm{T}}. \tag{15}$$

We have

$$\theta(M) = \sup_p \mathscr{L}(p, M) = \begin{cases} \langle M, \Sigma\rangle & \text{if } \chi_{\mathscr{S}}(x) \leqslant l(x) \text{ for every } x, \\ +\infty & \text{otherwise.} \end{cases}$$

The dual function is finite if and only if

C.1: $l(x) \geqslant 0$ for every $x \in \mathbf{R}^n$;

C.2: $l(x) \geqslant 1$ for every $x \in \mathbf{R}^n$ such that $\gamma + x^{\mathrm{T}} w \leqslant 0$.

Condition C.1 is equivalent to the semidefinite positiveness of the quadratic form: $M \succeq 0$. In addition, Condition C.2 holds if there exist a scalar $\tau \geqslant 0$ such that, for every x, $l(x) \geqslant 1 - 2\tau(\gamma + x^{\mathrm{T}} w)$. Indeed, with condition C.1 in force, an application of the classical strong duality result for convex programs under the Slater assumption (Hiriart-Urruty and Lemaréchal 1993) shows that the above condition is sufficient, provided there exist a x_0 such that $\gamma + x_0^{\mathrm{T}} w < 0$, which is the case here because $w \in \mathscr{W}$ and $\mathscr{W}$ does not contain 0. We obtain that conditions C.1, C.2 are equivalent to:

There exist a $\tau \geqslant 0$ such that: $M \succeq 0$,

$$M + \begin{bmatrix} 0 & \tau w \\ \tau w^{\mathrm{T}} & -1 + 2\tau\gamma \end{bmatrix} \succeq 0.$$

Thus the worst-case probability (14) is the solution to the SDP in variables M, τ:

$$\textbf{inf}\langle M, \Sigma\rangle \quad \text{subject to} \quad \tau \geqslant 0, \qquad M \succeq 0, \qquad M + \begin{bmatrix} 0 & \tau w \\ \tau w^{\mathrm{T}} & -1+2\tau\gamma \end{bmatrix} \succeq 0.$$

Computing the worst-case VaR as an SDP. We obtain that the worst-case VaR can be computed as

$$V_{\mathscr{P}}(w) = \textbf{inf}\ \gamma$$
$$\text{subject to} \quad \langle M, \Sigma\rangle \leqslant \epsilon, \qquad \tau \geqslant 0, \qquad M \succeq 0,$$
$$M + \begin{bmatrix} 0 & \tau w \\ \tau w^{\mathrm{T}} & -1+2\tau\gamma \end{bmatrix} \succeq 0. \tag{16}$$

It can be shown (see El Ghaoui et al. 2000) that the τ-components of optimal solutions (whenever they exist) of Equation (16) are uniformly bounded from below by a positive number. This allows us to divide by τ in the matrix inequality above, replace $1/\tau$ by τ, and M/τ, w/τ by M, w, and obtain the SDP in variables τ, M, γ:

$$V_{\mathscr{P}}(w) = \textbf{inf}\ \gamma,$$
$$\text{subject to} \quad \langle M, \Sigma\rangle \leqslant \tau\epsilon, \qquad \tau \geqslant 0, \qquad M \succeq 0,$$
$$M + \begin{bmatrix} 0 & w \\ w^{\mathrm{T}} & -\tau+2\gamma \end{bmatrix} \succeq 0. \tag{17}$$

Analytical formula for the worst-case VaR. Finally, we show that the SDP (17) yields the analytical formula (7). We first find the dual, in the sense of SDP duality, of the SDP (17). Define the Lagrangian

$$\mathscr{L}(M, \tau, \gamma, \alpha, \mu, X, Y) = \gamma - \alpha(\tau\epsilon - \langle M, \Sigma\rangle) - \mu\tau - \langle X, M\rangle - \left\langle Y, M + \begin{bmatrix} 0 & w \\ w^{\mathrm{T}} & -\tau+2\gamma \end{bmatrix}\right\rangle,$$

so that

$$V_{\mathscr{P}}(w) = \inf_{M=M^{\mathrm{T}},\, \tau,\, \gamma}\ \sup_{\alpha\geqslant 0,\, \mu\geqslant 0,\, X\succeq 0,\, Y\succeq 0} \mathscr{L}(M, \tau, \gamma, \alpha, \mu, X, Y).$$

Partition the dual variable Y as follows:

$$Y = \begin{bmatrix} Z & m \\ m^{\mathrm{T}} & \nu \end{bmatrix}.$$

We obtain the dual problem in variables X, Z, m, ν, α:

$$\textbf{sup} -2m^{\mathrm{T}}w,$$
$$\text{subject to} \quad \nu = 1/2, \qquad \alpha\epsilon + \mu - \nu = 0, \qquad \alpha \geqslant 0,$$
$$\mu \geqslant 0, \qquad \alpha\Sigma = X + Y, \qquad X \succeq 0,$$
$$Y = \begin{bmatrix} Z & m \\ m^{\mathrm{T}} & \nu \end{bmatrix} \succeq 0.$$

The above dual problem is strictly feasible and the feasible set is bounded. Therefore, strong duality holds and both primal and dual values are attained. Eliminating the variables μ, ν, X yields

$$V_{\mathscr{P}}(w) = \max\ -2m^{\mathrm{T}}w,$$
$$\text{subject to} \quad 0 \leqslant \alpha \leqslant \frac{1}{2\epsilon}, \quad \alpha\Sigma \succeq Y = \begin{bmatrix} Z & m \\ m^{\mathrm{T}} & 1/2 \end{bmatrix} \succeq 0.$$

Note that the constraint on Y imply $\alpha \geqslant 1/2 > 0$. Therefore, we make the change of variables $(Z, m, \alpha) \to (V, v, y)$ with $V = Z/\alpha$, $v = m/\alpha$, $y = 1/2\alpha \in [\epsilon, 1]$. We obtain

$$V_{\mathscr{P}}(w) = \max\ -\frac{v^{\mathrm{T}}w}{y},$$
$$\text{subject to} \quad \Sigma \succeq \begin{bmatrix} V & v \\ v^{\mathrm{T}} & y \end{bmatrix} \succeq 0, \qquad \epsilon \leqslant y \leqslant 1. \tag{18}$$

If $y = 1$, we have $v = \hat{x}$ and the objective of the problem is $-\hat{x}^{\mathrm{T}}w$. Assume now that $y < 1$. In view of our partition of Σ given in Equation (6), the matrix inequality constraints in Problem (18) are equivalent to

$$S - \frac{1}{1-y}(\hat{x}-v)(\hat{x}-v)^{\mathrm{T}} \succeq V \succeq \frac{1}{y}vv^{\mathrm{T}}.$$

The above constraint holds for some $V \succeq 0$ if and only if

$$S \succeq \frac{1}{1-y}(\hat{x}-v)(\hat{x}-v)^{\mathrm{T}} + \frac{1}{y}vv^{\mathrm{T}}, \tag{19}$$

or, equivalently,

$$\Gamma = S - \hat{x}\hat{x}^{\mathrm{T}} \succeq \frac{1}{y(1-y)}(v-y\hat{x})(v-y\hat{x})^{\mathrm{T}}. \tag{20}$$

The dual problem now becomes

$$V_{\mathscr{P}}(w) = \max_{v,y} -\frac{v^{\mathrm{T}}w}{y},$$
$$\text{subject to} \quad S - \hat{x}\hat{x}^{\mathrm{T}} \succeq \frac{1}{y(1-y)}(v-y\hat{x})(v-y\hat{x})^{\mathrm{T}},$$
$$y \in [\epsilon, 1].$$

Denote by $\phi(y)$ the objective of the problem with $y < 1$ fixed. We have for $y < 1$

$$\phi(y) = \sqrt{\frac{1-y}{y}}\|\Gamma^{1/2}w\|_2 - \hat{x}^{\mathrm{T}}w.$$

The above expression is valid for $y = 1$. Maximizing over y

$$\max_{\epsilon\leqslant y\leqslant 1} \phi(y) = \sqrt{\frac{1-\epsilon}{\epsilon}}\|\Gamma^{1/2}w\|_2 - \hat{x}^{\mathrm{T}}w.$$

We thus have $y = \epsilon$ at the optimum. This proves expression (7) for the worst-case VaR. □

2.2. Convex Moment Uncertainty

We now turn to the case when $(\Gamma, \hat{x})$ are only known to belong to a given *convex* subset $\mathcal{U}$ of $\mathcal{S}_n \times \mathbf{R}^n$, and the probability distribution is otherwise arbitrary. $\mathcal{U}$ could describe, for example, upper and lower bounds on the components of $\hat{x}$ and Γ. We assume that there is a point $(\Gamma, \hat{x})$ in $\mathcal{U}$ such that $\Gamma \succ 0$. (Checking this assumption can be done easily, as seen later.) We denote by $\mathcal{U}_+$ the set $\{(\Gamma, \hat{x}) \in \mathcal{U} \mid \Gamma \succ 0\}$. Finally, we assume that $\mathcal{U}_+$ is bounded. We denote as before by $\mathcal{P}$ the corresponding set of probability distributions.

In view of the equivalence between Propositions 1 and 3 of Theorem 1, we obtain that the worst-case VaR is less than γ if and only if, for every $x \in \mathbf{R}^n$ and $(\Gamma, \hat{x}) \in \mathcal{U}_+$ such that Condition (10) holds, we have $-x^{\mathrm{T}} w \geqslant \gamma$. It thus suffices to make Γ and $\hat{x}$ variables in the above conditions, to compute the worst-case VaR:

$$V_{\mathcal{P}}(w) = \mathbf{sup} -x^{\mathrm{T}} w \quad \text{subject to} \quad (\Gamma, \hat{x}) \in \mathcal{U}_+, \quad (10).$$

Because $\Gamma \succeq 0$ is implied by Condition (10), and the "sup" over a set (here, $\mathcal{U}_+$) is the same as the "sup" over its closure, we can replace $\mathcal{U}_+$ by $\mathcal{U}$ in the above, and the "sup" then becomes a "max" because $\mathcal{U}$ is bounded. We thus have the following result.

Theorem 2. *When the distribution of returns is only known to have a mean $\hat{x}$ and a covariance matrix Γ belong to a set $\mathcal{U}$ ($(\hat{x}, \Gamma) \in \mathcal{U}$), and is otherwise arbitrary, the worst-case Value-at-Risk is the solution of the optimization problem in variables $\Gamma, \hat{x}, x$:*

$$\max -x^{\mathrm{T}} w \quad \text{subject to} \quad (\Gamma, \hat{x}) \in \mathcal{U}, \quad \begin{bmatrix} \Gamma & x-\hat{x} \\ (x-\hat{x})^{\mathrm{T}} & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \qquad (21)$$

where $\kappa(\epsilon)$ is given in Equation (8).

Solving Problem (21) yields a choice of mean vector $\hat{x}$ and covariance matrix Γ that corresponds to the worst-case choice consistent with the prior information $(\Gamma, \hat{x}) \in \mathcal{U}$. This choice is therefore the most prudent when the mean and covariance matrix are only known to belong to $\mathcal{U}$, and the probability distribution is otherwise arbitrary.

To optimize over the allocation vector w, we consider the problem

$$V_{\mathcal{P}}^{\mathrm{opt}} = \min_{w \in \mathcal{W}} \max_{x, \hat{x}, \Gamma} -x^{\mathrm{T}} w$$
$$\text{subject to} \quad (\hat{x}, \Gamma) \in \mathcal{U}, \qquad \begin{bmatrix} \Gamma & x-\hat{x} \\ (x-\hat{x})^{\mathrm{T}} & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0. \qquad (22)$$

We obtain an alternative expression of the worst-case VaR, using the formulation (11). A given γ is an upper bound on the worst-case VaR if and only if for every $(\Gamma, \hat{x}) \in \mathcal{U}$ with $\Gamma \succ 0$, there exist Λ, v such that Formulation (11) holds. Thus, the worst-case VaR is given by the max-min problem

$$\max_{\Gamma, \hat{x}} \min_{\Lambda, v} \langle \Lambda, \Gamma \rangle + \kappa(\epsilon)^2 v - \hat{x}^{\mathrm{T}} w$$
$$\text{subject to} \quad \begin{bmatrix} \Lambda & w/2 \\ w^{\mathrm{T}}/2 & v \end{bmatrix} \succeq 0,$$
$$(\hat{x}, \Gamma) \in \mathcal{U}, \quad \Gamma \succeq 0. \qquad (23)$$

The feasible set in the above problem is compact and convex, and the objective is linear in $\Gamma, \hat{x}$ for fixed Λ, v (and conversely). It follows that we can exchange the "min" and "max" and optimize (over w) the worst-case VaR by solving the min-max problem

$$\min_{\Lambda, v, w} \max_{\Gamma, \hat{x}} \langle \Lambda, \Gamma \rangle + \kappa(\epsilon)^2 v - \hat{x}^{\mathrm{T}} w$$
$$\text{subject to} \quad \begin{bmatrix} \Lambda & w/2 \\ w^{\mathrm{T}}/2 & v \end{bmatrix} \succeq 0,$$
$$(\hat{x}, \Gamma) \in \mathcal{U}, \quad \Gamma \succeq 0, \quad w \in \mathcal{W}. \qquad (24)$$

The above problem can be interpreted as a game, where the variables Λ, v seek to decrease the VaR while the variables $\Gamma, \hat{x}$ oppose to it. Note that in the case when $\mathcal{U}$ is not convex, the above is an upper bound on the worst-case VaR given by Equation (23).

Theorem 3. *When the distribution of returns is only known have a mean $\hat{x}$ and a covariance matrix Γ such that $(\hat{x}, \Gamma) \in \mathcal{U}$, where $\mathcal{U}$ is convex and bounded and the probability distribution is otherwise arbitrary, the worst-case Value-at-Risk can be optimized by solving the optimization problem in variables $\Gamma, \hat{x}, x$* (22). *Alternatively, we can solve the "min-max" Problem* (24).

The tractability of Problem (22) depends on the structure of sets $\mathcal{U}$ and $\mathcal{W}$. When both sets are described by linear matrix inequalities in $\hat{x}$, Γ and w, the resulting problem can be expressed explicitely as an SDP. The "min-max" formulation is useful when we are able to explicitely solve the inner maximization problem, as will be the case in the next sections.

2.3. Polytopic Uncertainty

As a first example of application of the convex uncertainty model, we discuss the case when the moment pair $(\hat{x}, \Gamma)$ is only known to belong to a given polytope, described by its vertices. Precisely, we assume that $(\hat{x}, \Gamma) \in \mathcal{U}$, where $\mathcal{U}$ is the convex hull of the vertices $(\hat{x}_1, \Gamma_1), \ldots, (\hat{x}_l, \Gamma_l)$

$$\mathcal{U} = \mathbf{Co}\{(\hat{x}_1, \Gamma_1), \ldots, (\hat{x}_l, \Gamma_l)\}, \qquad (25)$$

where the vertices $(\hat{x}_i, \Gamma_i)$ are given. Again, let $\mathcal{P}$ denote the set of probability distributions that have a mean-covariance matrix pair $(\hat{x}, \Gamma) \in \mathcal{U}$, and are otherwise arbitrary.

We can compute the worst-case VaR in this case, and optimize it, as a simple application of the general results of

§2.2. The matrix-vector pair $(\hat{x}, \Gamma)$ is made a variable in the analysis Problem (21) or the portfolio optimization Problem (22). Denoting by ξ the vector containing the independent elements of $\hat{x}$ and Γ, we can express ξ as

$$\xi = \sum_{i=1}^{l} \lambda_i \xi_i, \qquad \sum_{i=1}^{l} \lambda_i = 1, \qquad \lambda \geqslant 0,$$

where ξ_i corresponds to the vertex pair $(\hat{x}_i, \Gamma_i)$. The resulting optimization Problem (21) or (22) is a semidefinite programming problem involving the vector variable λ.

It is interesting to examine the case when the mean and covariance matrix are subject to *independent* polytopic uncertainty. Precisely, we assume that the polytope $\mathcal{U}$ is the direct product of two polytopes: $\mathcal{U} = \mathcal{U}_x \times \mathcal{U}_\Gamma$, where

$$\mathcal{U}_x := \mathbf{Co}\{\hat{x}_1, \ldots, \hat{x}_l\} \subseteq \mathbf{R}^n, \qquad \mathcal{U}_\Gamma := \mathbf{Co}\{\Gamma_1, \ldots, \Gamma_l\} \subseteq \mathcal{S}_n.$$

(We have assumed for simplicity that the number of vertices of each polytope is the same.) Assuming that $\Gamma_i \succeq 0$, $i = 1, \ldots, l$, the worst-case VaR is attained at the vertices. Precisely,

$$\begin{aligned} V_{\mathcal{P}}(w) &= \kappa(\epsilon)\sqrt{\max_{\Gamma \in \mathcal{U}_\Gamma} w^{\mathrm{T}} \Gamma w} - \min_{\hat{x} \in \mathcal{U}_x} \hat{x}^{\mathrm{T}} w \\ &= \max_{1 \leqslant i \leqslant l} \kappa(\epsilon) \|\Gamma_i^{1/2} w\|_2 - \min_{1 \leqslant i \leqslant l} \hat{x}_i^{\mathrm{T}} w. \end{aligned}$$

Thus, the polytopic model yields the same worst-case VaR as in the case when the uncertainty in the mean and covariance matrix consists in a finite number of scenarios.

With the previous polytopic model, the computation of $V_{\mathcal{P}}$ is straightforward. Moreover, its optimization with respect to the portfolio allocation vector w is also very efficient. The optimization problem

$$\min_{w \in \mathcal{W}} V_{\mathcal{P}}(w)$$

can be formulated as the second-order cone program in variables w, α, β:

$$\min_{w \in \mathcal{W}} \alpha - \beta \quad \text{subject to} \quad \kappa(\epsilon)\|\Gamma_i^{1/2} w\|_2 \leqslant \alpha, \quad \beta \leqslant \hat{x}_i^{\mathrm{T}} w, \quad i = 1, \ldots, l.$$

As discussed in Lobo et al. (2000), this problem can be solved in a number of iterations (almost) independent of problem size, and each iteration has a complexity $O(\ln^3)$. Thus, the complexity of the problem grows (almost) linearly with the number of scenarios.

The previous result is useful when the number of different scenarios is moderate; however, the problem becomes quickly intractable if the number of scenarios grows exponentially with the number of assets. This would be the case if we were interested in a covariance matrix whose entries are known only within upper and lower values. In this case, it is more interesting to describe the polytope $\mathcal{U}$ by its facets rather than its vertices, as is done next.

2.4. Componentwise Bounds on Mean and Covariance Matrix

We now specialize the results of §2.2 to the case when $\Gamma, \hat{x}$ are only known within componentwise bounds:

$$x_- \leqslant \hat{x} := \mathbf{E}x \leqslant x_+, \quad \Gamma_- \leqslant \Gamma := \mathbf{E}(x - \hat{x})(x - \hat{x})^{\mathrm{T}} \leqslant \Gamma_+, \quad (26)$$

where x_+, x_- and Γ_+, Γ_- are given vectors and matrices, respectively, and the inequalities are understood componentwise.

The interval matrix $[\Gamma_-, \Gamma_+]$ is not necessarily included in the cone of positive semidefinite matrices: Not all of its members may correspond to an actual covariance matrix. We will, however, assume that there exist at least one nondegenerate probability distribution such that the above moment bounds hold; that is, there exist a matrix $\Gamma \succ 0$ such that $\Gamma_- \leqslant \Gamma \leqslant \Gamma_+$. (Checking if this condition holds, and if so, exhibiting an appropriate Γ, can be solved very efficiently, as seen below.) The problem of computing the worst-case VaR reduces to

$$\begin{aligned} &\text{maximize} \quad -x^{\mathrm{T}} w \\ &\text{subject to} \quad x_- \leqslant \hat{x} \leqslant x_+, \qquad \Gamma_- \leqslant \Gamma \leqslant \Gamma_+, \\ &\qquad \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^{\mathrm{T}} & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \end{aligned} \quad (27)$$

Note that the SDP (27) is strictly feasible if and only if there exist $\Gamma \succ 0$ such that Equation (26) holds. In practice, it may not be necessary to check this strict feasibility condition prior to solving the problem. SDP codes such as SeDuMi (Sturm 1998) produce, in a single phase, either an optimal solution or a certificate of infeasibility (in our case, a proof that no $\Gamma \succ 0$ exists within the given componentwise bounds).

Alternatively, the worst-case VaR is the solution of the min-max problem

$$\begin{aligned} &\min_{\Lambda, v} \max_{\Gamma, \hat{x}} \langle \Lambda, \Gamma \rangle + \kappa(\epsilon)^2 v - \hat{x}^{\mathrm{T}} w \\ &\text{subject to} \quad \begin{bmatrix} \Lambda & w/2 \\ w^{\mathrm{T}}/2 & v \end{bmatrix} \succeq 0, \qquad x_- \leqslant \hat{x} \leqslant x_+, \\ &\qquad \Gamma_- \leqslant \Gamma \leqslant \Gamma_+, \qquad \Gamma \succeq 0. \end{aligned} \quad (28)$$

For fixed Λ, v, the inner maximization Problem (28) is a (particularly simple) SDP in $\hat{x}, \Gamma$. We can write this problem in the dual form of a minimization problem. In fact,

$$\max_{x_- \leqslant \hat{x} \leqslant x_+} -w^{\mathrm{T}} \hat{x} = \min_{\lambda_\pm \geqslant 0,\, w = \lambda_- - \lambda_+} \lambda_+^{\mathrm{T}} x_+ - \lambda_-^{\mathrm{T}} x_-,$$

and a similar result for the term involving Γ:

$$\max_{\Gamma_- \leqslant \Gamma \leqslant \Gamma_+,\, \Gamma \succeq 0} \langle \Lambda, \Gamma \rangle = \min_{\Lambda_\pm \geqslant 0,\, \Lambda \preceq \Lambda_+ - \Lambda_-} \langle \Lambda_+, \Gamma_+ \rangle - \langle \Lambda_-, \Gamma_- \rangle,$$

where we are using that property that both the maximization and minimization problems are strictly feasible, which guarantees that their optimal values are equal.

We obtain that the worst-case VaR is given by the SDP in variables $\lambda_\pm$, $\Lambda_\pm$, v:

$$V_{\mathcal{P}}(w) = \min\langle \Lambda_+, \Gamma_+\rangle - \langle \Lambda_-, \Gamma_-\rangle + \kappa(\epsilon)^2 v + \lambda_+^T x_+ - \lambda_-^T x_-$$

$$\text{subject to} \quad \lambda_+ \geqslant 0, \quad \lambda_- \geqslant 0, \quad \Lambda_+ \geqslant 0, \quad \Lambda_- \geqslant 0,$$

$$\begin{bmatrix} \Lambda_+ - \Lambda_- & w/2 \\ w^T/2 & v \end{bmatrix} \succeq 0, \quad w = \lambda_- - \lambda_+. \quad (29)$$

As noted before, the above formulation allows us to optimize the portfolio over $w \in \mathcal{W}$: It suffices to let w be a variable. Because $\mathcal{W}$ is a polytope, the problem falls in the SDP class.

In the case when the moments are exactly known, that is, $\Gamma_+ = \Gamma_- = \Gamma$ and $\hat{x}_+ = \hat{x}_- = \hat{x}$, the above problem reduces to Problem (3) as expected (with the correct value of κ of course). To see this, note that only the variables v, $\Lambda := \Lambda_+ - \Lambda_-$ and $\lambda_- - \lambda_+ (= w)$ play a role. The optimal value of Λ is easily determined to be $ww^T/4v$, and optimizing over $v > 0$ yields the result.

We should also mention that the worst-case VaR can be similarly computed and optimized when we have componentwise bounds on the mean $\hat{x}$ and second-moment matrix S, specifically,

$$x_- \leqslant \hat{x} := \mathbf{E}x \leqslant x_+, \qquad S_- \leqslant S := \mathbf{E}xx^T \leqslant S_+, \quad (30)$$

where x_+, x_- and S_+, S_- are given vectors and matrices, respectively, and the inequalities are understood componentwise. A derivation similar to above shows that in this case the worst-case VaR can be optimized via the SDP in variables M_+, M_-, λ_+, λ_- and w:

$$V_{\mathcal{P}}^{\text{opt}} = \min\langle M_+, \Gamma_+\rangle - \langle M_-, \Gamma_-\rangle + \kappa(\epsilon)^2 v + \lambda_+^T x_+ - \lambda_-^T x_-$$

$$\text{subject to} \quad \lambda_+ \geqslant 0, \quad \lambda_- \geqslant 0, \quad M_+ \geqslant 0,$$

$$M_- \geqslant 0, \quad M_+ - M_- + \begin{bmatrix} 0 & w/2 \\ w^T/2 & v \end{bmatrix} \succeq 0,$$

$$w = \lambda_- - \lambda_+ \in \mathcal{W}. \quad (31)$$

3. FACTOR MODELS

Factor models arise when modelling the returns in terms of a reduced number of random factors. A factor model expresses the n-vector of returns x as follows:

$$x = Af + u, \quad (32)$$

where f is a m-vector of (random) factors, u contains residuals, and A is a $n \times m$ matrix containing the sensitivities of the returns x with respect to the various factors. If S (resp. $\hat{f}$) is the covariance matrix (resp. mean vector) of the factors, and u is modeled as a zero-mean random variable with diagonal covariance matrix D, uncorrelated with f, then the covariance matrix (resp. mean vector) of the return vector is $\Gamma = D + ASA^T$ (resp. $\hat{x} = A\hat{f}$).

Such models thus impose a structure on the mean and covariance, which in turn imposes structure on the corresponding uncertainty models. In this section, we examine how the impact of uncertainties in factor models can be analyzed (and optimized) via SDP.

3.1. Uncertainty in the Factor's Mean and Covariance Matrix

The simplest case is when we consider errors in the mean-covariances of the factors. Based on a factor model, we may be interested in "stress testing," which amounts to analyzing the impact of simultaneous changes in the correlation structure of the factors, on the VaR. In our model, we will assume (say, componentwise) uncertainty on the factor data S and $\hat{f}$. For a fixed value of the sensitivity matrix A, and of the diagonal marix D, we obtain that the corresponding worst-case VaR can be computed exactly via the SDP

$$\text{maximize} \quad -x^T w$$

$$\text{subject to} \quad \hat{x} = A\hat{f}, \quad \Gamma = D + ASA^T, \quad f_- \leqslant \hat{f} \leqslant f_+,$$

$$S_- \leqslant S \leqslant S_+, \quad \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^T & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0,$$

where $f_\pm$ and $S_\pm$ are componentwise upper and lower bounds on the mean and covariance of factors. A similar analysis can be performed with respect to simultaneous changes in D, S and $\hat{f}$. Likewise, portfolio optimization results are similar to the ones obtained before.

3.2. Uncertainty in the Sensitivity Matrix

One may also be looking at the impact of errors in the sensitivity matrix A, on the VaR. As pointed out by Linsmeier and Pearson (1996), the mean-variance approach to Value-at-Risk can be used to analyze the risk of portfolios containing possibly very complex instruments such as futures contracts, exotic options, etc. This can be done using an approximation called risk mapping, which is a crucial step in any practical implementation of the method.

In general, one can express the return vector of a portfolio containing different instruments as a function of "market factors," such as currency exchange rates, interest rates, underlying asset prices, and so on. Those are quantities for which historical data are available and for which we might have a reasonable confidence in mean and covariance data. In contrast, most complex instruments cannot be directly analyzed in terms of mean and covariance.

The process of risk mapping amounts to approximating the return vector via the decomposition (32). In essence, the factor model is a linearized approximation to the actual return function, which allows use of mean-variance analysis for complex, nonlinear financial instruments.

Because the factor model is a linearized approximation of the reality, it may be useful to keep track of linearization errors via uncertainty in the matrix of sensitivities A. In fact, instead of fitting one linear approximation to the return vector, one may deliberately *choose* to fit linear approximations that serve as upper and lower bounds on the return vector. The risk analysis then proceeds by analyzing both upper and lower bounds, for all the instruments considered. Thus, it is of interest (both for numerical reasons and for more accurate modelling) to take into account uncertainty in the matrix A.

We assume that the statistical data S, D, and $\hat{f}$ are perfectly known, with $S \succeq 0$ and $D \succ 0$, and that the errors in A are modeled by $A \in \mathscr{A}$, where the given set $\mathscr{A}$ describes the possible values for A. We are interested in computing (or finding bounds on) and optimizing with respect to the portfolio weight vector w the worst-case VaR

$$V_{\mathrm{wc}}(w) := \max_{A\in\mathscr{A}} \kappa(\epsilon) \left\| \begin{bmatrix} S^{1/2}A^{\mathrm{T}} \\ D^{1/2} \end{bmatrix} w \right\|_2 - w^{\mathrm{T}} A\hat{f}. \tag{33}$$

Ellipsoidal uncertainty. We first consider the case when A is subject to ellipsoidal uncertainty:

$$\mathscr{A} = \left\{ A_0 + \sum_{i=1}^{l} u_i A_i \,\middle|\, u \in \mathbf{R}^l,\ \|u\|_2 \leqslant 1 \right\}, \tag{34}$$

where the given matrices $A_i \in \mathbf{R}^{n\times m}$, $i = 0, \ldots, l$, determine an ellipsoid in the space of $n \times m$ matrices.

The worst-case VaR then expresses as $-w^{\mathrm{T}} A_0 \hat{f} + \phi(w)$, where

$$\phi(w) := \max_{\|u\|_2\leqslant 1} \kappa(\epsilon) \|C(w)u + d(w)\|_2 + e(w)^{\mathrm{T}} u, \tag{35}$$

for an appropriate matrix $C(w)$ and vectors $d(w)$, $e(w)$, linear functions of w that are defined in Theorem 4 below. Our approach hinges on the following lemma, whose proof can be found in El Ghaoui et al. (2000).

Lemma 1. *Let $C \in \mathbf{R}^{N\times l}$, $d \in \mathbf{R}^N$, $e \in \mathbf{R}^l$ and $\rho \geqslant 0$ be given. An upper bound on the quantity*

$$\phi := \max_{\|u\|_2\leqslant \rho} \|Cu + d\|_2 + e^{\mathrm{T}} u,$$

can be computed via the following SDP:

$$2\phi \leqslant \min \lambda_1 + \rho^2 \lambda_2 + \lambda_3 : \begin{bmatrix} \lambda_1 I_N & C & d \\ C^{\mathrm{T}} & \lambda_2 I_l & e \\ d^{\mathrm{T}} & e^{\mathrm{T}} & \lambda_3 \end{bmatrix} \succeq 0.$$

The following theorem is a direct consequence of the above lemma, in the case $\rho = 1$. It shows that we can not only compute but also optimize an upper bound on the worst-case VaR under our current assumptions on the distribution of returns.

Theorem 4. *When the distribution of returns obeys to the factor model* (32) *and the sensitivity matrix is only known to belong to the set $\mathscr{A}$ given by* (34), *an upper bound on the worst-case VaR given in Equation* (33) *can be computed* (*optimized*) *via the following SDP in variables λ* (*and $w \in \mathscr{W}$*):

$$\begin{aligned} &\min \frac{\kappa(\epsilon)}{2}(\lambda_1 + \lambda_2 + \lambda_3) - w^{\mathrm{T}} A_0 \hat{f}, \\ &\text{subject to } \begin{bmatrix} \lambda_1 I_{n+m} & C(w) & d(w) \\ C(w)^{\mathrm{T}} & \lambda_2 I_l & e(w) \\ d(w)^{\mathrm{T}} & e(w)^{\mathrm{T}} & \lambda_3 \end{bmatrix} \succeq 0, \end{aligned} \tag{36}$$

where

$$C(w) = \begin{bmatrix} M_1 w & \cdots & M_l w \end{bmatrix}, \quad \text{with } M_i = \begin{bmatrix} S^{1/2} A_i^{\mathrm{T}} \\ 0 \end{bmatrix}, \quad 1 \leqslant i \leqslant l,$$

$$d(w) = \begin{bmatrix} S^{1/2} A_0^{\mathrm{T}} \\ D^{1/2} \end{bmatrix} w, \quad e_i(w) = -\frac{w^{\mathrm{T}} A_i \hat{f}}{\kappa(\epsilon)}, \quad 1 \leqslant i \leqslant l.$$

Norm-bound uncertainty. We now consider the case when

$$\mathscr{A} = \{A + L\Delta R \mid \Delta \in \mathbf{R}^{l\times r},\ \|\Delta\| \leqslant 1\}, \tag{37}$$

where $A \in \mathbf{R}^{n\times m}$, $L \in \mathbf{R}^{n\times l}$, and $R \in \mathbf{R}^{r\times m}$ are given and Δ is an uncertain matrix that is bounded by one in maximum singular value norm. The above kind of uncertainty is useful to model "unstructured" uncertainties in some blocks of A, with the matrices L, R specifying which blocks in A are uncertain. For details on unstructured uncertainty in matrices, see Boyd et al. (1994). A specific example obtains by setting $L = R = I$, which corresponds to an additive perturbation of A that is bounded in norm but otherwise unknown. The following result follows quite straightforwardly from Lemma 1.

Theorem 5. *When the distribution of returns obeys to the factor model* (32) *and the sensitivity matrix is only known to belong to the set $\mathscr{A}$ given by Equation* (37), *an upper bound on the worst-case VaR given in Equation* (33) *can be computed* (*optimized*) *via the following SDP in variables λ* (*and $w \in \mathscr{W}$*):

$$\begin{aligned} &\min \frac{\kappa(\epsilon)}{2}(\lambda_1 + t + \lambda_3) - w^{\mathrm{T}} A\hat{f} \\ &\text{subject to } \begin{bmatrix} \lambda_1 I_{n+m} & d(w) \\ d(w)^{\mathrm{T}} & \lambda_3 \end{bmatrix} \succeq \mu \begin{bmatrix} C \\ e^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} C \\ e^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \\ &\qquad \begin{bmatrix} t & w^{\mathrm{T}} L \\ L^{\mathrm{T}} w & \mu I_l \end{bmatrix} \succeq 0, \end{aligned} \tag{38}$$

where

$$d(w) = \begin{bmatrix} S^{1/2} A^{\mathrm{T}} \\ D^{1/2} \end{bmatrix} w, \qquad C = \begin{bmatrix} S^{1/2} R^{\mathrm{T}} \\ 0 \end{bmatrix}, \quad e = \frac{-R\hat{f}}{\kappa(\epsilon)}.$$

Proof of Theorem 5. Defining

$$d(w) = \begin{bmatrix} S^{1/2}A^{\mathrm{T}} \\ D^{1/2} \end{bmatrix} w, \qquad C = \begin{bmatrix} S^{1/2}R^{\mathrm{T}} \\ 0 \end{bmatrix},$$

$$e = \frac{-R\hat{f}}{\kappa(\epsilon)}, \qquad r(w) = L^{\mathrm{T}} w,$$

we may express the worst-case VaR as $-w^{\mathrm{T}}A\hat{f} + \phi(w)$, where

$$\begin{aligned}\phi(w) &= \kappa(\epsilon) \max_{\|\Delta\|\leqslant 1} \|C\Delta^{\mathrm{T}} r(w) + d(w)\|_2 + e^{\mathrm{T}}\Delta^{\mathrm{T}} r(w) \\ &= \kappa(\epsilon) \max_{\|u\|_2 \leqslant \|r(w)\|_2} \|Cu + d(w)\|_2 + e^{\mathrm{T}} u \\ &\leqslant \min \frac{\kappa(\epsilon)}{2}(\lambda_1 + \|r(w)\|_2^2 \lambda_2 + \lambda_3) : \\ &\qquad \begin{bmatrix} \lambda_1 I_{n+m} & C & d(w) \\ C^{\mathrm{T}} & \lambda_2 I_r & e \\ d(w)^{\mathrm{T}} & e^{\mathrm{T}} & \lambda_3 \end{bmatrix} \succeq 0,\end{aligned}$$

where the last inequality is derived from Lemma 1, with $\rho = \|r(w)\|_2$. Using Schur complements we may rewrite the linear matrix inequality in the last line as

$$\begin{bmatrix} \lambda_1 I_{n+m} & d(w) \\ d(w)^{\mathrm{T}} & \lambda_3 \end{bmatrix} \succeq \mu \begin{bmatrix} C \\ e^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} C \\ e^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}},$$

where $\mu = 1/\lambda_2$. Introducing a slack variable $t \geqslant \lambda_2 \cdot \|r(w)\|_2^2 = \|r(w)\|_2^2/\mu$, we can rewrite the objective as $\lambda_1 + t + \lambda_3$, where t is such that $t \geqslant \|r(w)\|_2^2/\mu$. The latter inequality can be written as the linear matrix inequality in the theorem. (We note that this constraint is a second-order cone constraint and its structure should be exploited in a numerical implementation of the theorem.) □

Note: Goldfarb and Iyengar (2001) consider uncertainty structures in which the uncertainty in the mean is independent of the uncertainty in the covariance matrix of the returns. This assumption leads to the terms $e(w)$ in (35) and e in (38) being equal to zero. As a result, the expressions in Lemma 1 and Theorems 4 and 5 become exact, for example, the worst-case VaR can be computed and optimized exactly in this case. Moreover, for the specific uncertainty structures Goldberg and Iyengar consider, they are able to formulate these problems as SOCPs.

4. EXTENSIONS AND VARIATIONS

In this section, we examine extensions and variations on the problem. We assume throughout that $\hat{x}$ and Γ are given, with $\Gamma \succ 0$. The extension to moment uncertainty is straightforward.

4.1. Including Support Information

We now restrict the allowable probability distributions to be of given support $\Omega \subseteq \mathbf{R}^n$ and seek to refine Theorem 1 accordingly.

Hypercube support. First consider the case when the support is the hypercube $\Omega := [x_l \;\; x_u]$, where x_l, x_u are given vectors, with $x_l < \hat{x} < x_u$. Theorem 1 is extended as follows.

Theorem 6. *When the probability distribution of returns has known mean $\hat{x}$ and covariance matrix Γ, its support is included in the hypercube $\Omega := [x_l \;\; x_u]$, and is otherwise arbitrary, we can compute an upper bound on the worst-case Value-at-Risk by solving the semidefinite programming problem in variable x:*

$$\begin{aligned} &\text{maximize} \quad -x^{\mathrm{T}} w \\ &\text{subject to} \quad \begin{bmatrix} \Gamma & x - \hat{x} \\ (x-\hat{x})^{\mathrm{T}} & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \\ &\qquad \kappa(\epsilon)^2 x_l \leqslant \hat{x} - x \leqslant \kappa(\epsilon)^2 x_u, \quad x_l \leqslant x \leqslant x_u, \end{aligned} \tag{39}$$

where $\kappa(\epsilon)$ is given in (8).

We will not present the proof of Theorem 6 here because it is similar to the proof of Theorem 1. Interested readers may refer to El Ghaoui et al. (2000) for details. If we let $x_l = -\infty$ and $x_u = +\infty$, the last inequalities in (39) become void, and the problem reduces to the one obtained in the case of no support constraints. Thus, the above result allows us to refine the condition obtained by simply no taking into account the support constraints. Contrarily to what happens with no support constraints, there is no "closed-form" solution to the VaR, which seems to be hard to compute exactly; but computing an upper bound is easy via SDP.

In fact, Problem (39) can be expressed as an SOCP, which makes it amenable to even faster algorithms. Again, we stress that while this approach is the best in the case of known moments, or with independent polytopic uncertainty (as in §2.3), the SDP formulation obtained above is more useful with general convex uncertainty on the moments. When $\Gamma \succ 0$, the SOCP formulation is

$$\begin{aligned} &\text{maximize} \quad -x^{\mathrm{T}} w, \\ &\text{subject to} \quad \|\Gamma^{-1/2}(x - \hat{x})\|_2 \leqslant \kappa(\epsilon)^2, \\ &\qquad \kappa(\epsilon)^2 x_l \leqslant \hat{x} - x \leqslant \kappa(\epsilon)^2 x_u, \quad x_l \leqslant x \leqslant x_u. \end{aligned} \tag{40}$$

Let us now examine the problem of optimizing the VaR with hypercube support information. We consider the problem of optimizing the upper bound on the worst-case VaR obtained previously:

$$\begin{aligned} &\overline{V}_{\mathscr{P}}^{\mathrm{opt}} := \min_{w \in \mathscr{W}} \max_{x} \; -x^{\mathrm{T}} w \\ &\text{subject to} \quad \begin{bmatrix} \Gamma & x - \hat{x} \\ (x-\hat{x})^{\mathrm{T}} & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \\ &\qquad \kappa(\epsilon)^2 x_l \leqslant \hat{x} - x \leqslant \kappa(\epsilon)^2 x_u, \qquad x_l \leqslant x \leqslant x_u. \end{aligned}$$

We can express the inner maximization problem in a dual form (in SDP sense), as a minimization problem. This leads to the following result.

Theorem 7. *When the distribution of returns has known mean $\hat{x}$ and covariance matrix Γ, its support is included in the hypercube $\Omega := [x_l \ x_u]$, and is otherwise arbitrary, we can optimize an upper bound on the worst-case Value-at-Risk by solving the SDP in variables $w, t, \Lambda, u, v, \lambda_{u,l}, \nu_{u,l}$:*

$$\overline{V}_{\mathscr{P}}^{\text{opt}} = \min\langle\Lambda,\Gamma\rangle+\kappa(\epsilon)^2 v-\hat{x}^{\mathrm{T}}w+\kappa(\epsilon)^2\left(x_u^{\mathrm{T}}\lambda_u-x_l^{\mathrm{T}}\lambda_l\right) + \nu_u^{\mathrm{T}}(x_u-\hat{x})-\nu_l^{\mathrm{T}}(x_l-\hat{x})$$

subject to

$$\begin{bmatrix} \Lambda & (w+\lambda_l-\lambda_u+\nu_u-\nu_l)/2 \\ (w+\lambda_l-\lambda_u+\nu_u-\nu_l)^{\mathrm{T}}/2 & v \end{bmatrix} \succeq 0. \tag{41}$$

In the above, when some of the components of x_u (resp. x_l) are $+\infty$ (resp. $-\infty$), we set the corresponding components of λ_u, ν_u (resp. λ_l, ν_l) to zero.

Again, if we set $\lambda_u=\lambda_l=\nu_u=\nu_l=0$ in (41), we recover the expression of the VaR given in (11), which corresponds to the exact conditions when first and second moments are known, and no support information is used.

Ellipsoidal support. In many statistical approaches, such as maximum-likelihood, the bounds of confidence on the estimates of the mean and covariance matrix take the form of ellipsoids. This motivates us to study the case when the support Ω is an ellipsoid in $\mathbf{R}^n$:

$$\Omega := \{x \mid (x-x_c)^{\mathrm{T}}P^{-1}(x-x_c) \leqslant 1\},$$

where x_c is the center and $P \succ 0$ determines the shape of the ellipsoid. Following the steps taken in the proof of Theorem 1, we obtain the following result.

Theorem 8. *When the distribution of returns has known mean $\hat{x}$ and covariance matrix Γ, and its support is included in the ellipsoid $\Omega := \{x \mid (x-x_c)^{\mathrm{T}}P^{-1}(x-x_c) \leqslant 1\}$, we can compute an upper bound on the worst-case Value-at-Risk by solving the semidefinite programming problem in variables x, X, t:*

$$\text{maximize} \quad -x^{\mathrm{T}}w,$$

$$\text{subject to} \quad \begin{bmatrix} \Gamma-\epsilon X & x-\hat{x} & x \\ (x-\hat{x})^{\mathrm{T}} & \kappa(\epsilon)^2 & 0 \\ x^{\mathrm{T}} & 0 & 1 \end{bmatrix} \succeq 0,$$

$$t/\epsilon = 1-\mathbf{Tr}\,P^{-1}X+2x_c^{\mathrm{T}}P^{-1}x-x_c^{\mathrm{T}}P^{-1}x_c \geqslant 0,$$

$$\begin{bmatrix} 1-t-\mathbf{Tr}\,P^{-1}\Gamma & \hat{x}-x_c \\ (\hat{x}-x_c)^{\mathrm{T}} & P \end{bmatrix} \succeq 0, \tag{42}$$

where $\kappa(\epsilon)$ is given in Equation (8).

4.2. Entropy-Constrained VaR

The worst-case probability distribution arising in Theorem 6, with or without support constraints, is in general discrete (Bertsimas and Popescu 2000). It may be argued that such a worst-case scenario is unrealistic. In this section, we seek to enforce that the worst-case probability distribution has some degree of smoothness. The easiest way to do so is to impose a relative entropy constraint with respect to a given "reference" probability distribution.

We will assume that the probability distribution of returns satisfies the following assumption, and is otherwise arbitrary. We assume that the distribution of returns, while not a Gaussian, is not "too far" from one. Precisely, we assume that the Kullback-Leibler divergence (negative relative entropy) satisfies

$$KL(P,P_0) := \int \log\frac{dP}{dP_0}\,dP \leqslant d, \tag{43}$$

where $d \geqslant 0$ is given, P is the probability distribution of returns, and P_0 is a nondegenerate Gaussian reference distribution, that has given mean $\hat{x}$ and covariance matrix $\Gamma \succ 0$. (Note that a finite d enforces that the distribution of returns P is absolutely continuous with respect to the Gaussian distribution P_0.)

We prove the following theorem.

Theorem 9. *When the probability distribution of returns is only known to satisfy the relative entropy constraint* (43), *and the mean $\hat{x}$ and covariance matrix Γ of the reference Gaussian distribution P_0 are known, the entropy-constrained Value-at-Risk is given by*

$$V_{\mathscr{P}}(w) = \kappa(\epsilon,d)\|\Gamma^{1/2}w\|_2-\hat{x}^{\mathrm{T}}w, \tag{44}$$

where $\kappa(\epsilon,d)$ is given by

$$\kappa(\epsilon,d) := -\Phi^{-1}(f(\epsilon,d)),$$

$$f(\epsilon,d) := \sup_{\lambda>0}\frac{e^{\epsilon/\lambda-d}-1}{e^{1/\lambda}-1} = \sup_{v>0}\frac{e^{-d}(v+1)^{\epsilon}-1}{v}, \tag{45}$$

where Φ is the cumulative distribution function of the standard normal distribution.

The above theorem shows that, by a suitable modification of the "risk factor" $\kappa(\epsilon)$ appearing in Theorem 1, we can handle entropy constraints (however, we do not know how to use support information in this case). For $d=0$, we obtain $\kappa(\epsilon,0)=-\Phi^{-1}(\epsilon)$, as expected, because we are then imposing that the distribution of returns is the Gaussian $P_0=\mathscr{N}(\hat{x},\Gamma)$. The risk factor $\kappa(\epsilon,d)$ increases with d. This is to be expected: As the set of allowable distributions "grows," the worst-case VaR becomes worse, and increases.

Proof of Theorem 9. As before, we begin with the problem of computing the worst-case probability. We address the following problem:

$$\text{maximize} \quad \int_{\mathbf{R}^n}\chi_{\mathscr{S}}(x)p(x)\,dx$$

$$\text{subject to} \quad KL(p,p_0) \leqslant d, \quad \int_{\mathbf{R}^n}p(x)\,dx = 1, \tag{46}$$

where p and p_0 denote the densities of distributions P and P_0. For a given distribution with density p such that $KL(p, p_0)$ is finite, and for given scalars $\lambda_0 \geqslant 0$, λ, we introduce the Lagrangian

$$L(p, \lambda_0, \lambda) = \int_{\mathbf{R}^n} \chi_{\mathcal{S}}(x)p(x)\,dx + \lambda_0\left(1 - \int_{\mathbf{R}^n} p(x)\,dx\right) + \lambda\left(d - \int_{\mathbf{R}^n} \log\frac{p(x)}{p_0(x)}p(x)\,dx\right),$$

where $\chi_{\mathcal{S}}$ is the indicator function of the set $\mathcal{S}$ defined in Equation (13). The dual function is

$$\theta(\lambda_0, \lambda) = \sup_{p\in\mathcal{K}(\mathbf{R}^n)} L(p, \lambda_0, \lambda) = \lambda_0 + \lambda d + \sup_p \int_{\mathbf{R}^n}\left(\chi_{\mathcal{S}}(x) - \lambda_0 - \lambda\log\frac{p(x)}{p_0(x)}\right)p(x)\,dx,$$

and the dual problem is

$$\inf_{(\lambda,\lambda_0)\in\mathbf{R}_+\times\mathbf{R}} \theta(\lambda_0, \lambda).$$

From the assumption that $\Gamma \succ 0$, strong duality holds (Smith 1995). For any pair (λ, λ_0), with $\lambda > 0$, the distribution that achieves the optimum in the "sup" appearing in the expression for Θ above has a density

$$p(x) = p_0(x)\exp\left(\frac{\chi_{\mathcal{S}}(x) - \lambda_0}{\lambda} - 1\right), \tag{47}$$

and the dual function becomes

$$\begin{aligned}\theta(\lambda_0, \lambda) &= \lambda_0 + \lambda d + \lambda\int p_0(x)\exp\left(\frac{\chi_{\mathcal{S}}(x) - \lambda_0}{\lambda} - 1\right)dx\\ &= \lambda_0 + \lambda d + \lambda\big(e^{((1-\lambda_0)/\lambda)-1}\,\mathbf{Prob}\{\gamma \leqslant -x^{\mathrm{T}}w\}\\ &\qquad + e^{-(\lambda_0/\lambda)-1}\,\mathbf{Prob}\{\gamma \geqslant -x^{\mathrm{T}}w\}\big)\\ &= \lambda_0 + \lambda d + \lambda e^{-(\lambda_0/\lambda)-1}\big((e^{1/\lambda} - 1)\phi(\gamma) + 1\big),\end{aligned}$$

where the probabilities above are taken with respect to p_0, and

$$\phi(\gamma) := \mathbf{Prob}\{\gamma \leqslant -x^{\mathrm{T}}w\} = 1 - \Phi\left(\frac{\gamma + w^{\mathrm{T}}\hat{x}}{\sqrt{w^{\mathrm{T}}\Gamma w}}\right).$$

Taking the infimum over λ_0 yields

$$\inf_{\lambda_0\in\mathbf{R}} \theta(\lambda_0, \lambda) = \lambda d + \lambda\log((e^{1/\lambda} - 1)\phi(\gamma) + 1). \tag{48}$$

The worst-case probability is obtained by taking the infimum of the above convex function over $\lambda > 0$.

The constraint $\theta^{\mathrm{opt}} \leqslant \epsilon$ is equivalent to the existence of $\lambda > 0$ such that

$$\lambda d + \lambda\log((e^{1/\lambda} - 1)\phi(\gamma) + 1) \leqslant \epsilon,$$

that is,

$$\gamma \geqslant \kappa(\epsilon, d)\sqrt{w^{\mathrm{T}}\Gamma w} - w^{\mathrm{T}}\hat{x},$$

where $\kappa(\epsilon, d)$ is defined in the Theorem. □

4.3. Multiple VaR Constraints

The framework we used allows us to find a portfolio that satisfies a given level γ of worst-case VaR, for a given probability threshold ϵ:

$$\sup \mathbf{Prob}\{\gamma \leqslant -r(w, x)\} \leqslant \epsilon.$$

We may consider multiple VaR constraints

$$\sup \mathbf{Prob}\{\gamma_i \leqslant -r(w, x)\} \leqslant \epsilon_i, \quad i = 1, \ldots, m,$$

where $\epsilon_1 < \cdots < \epsilon_m$ are given probability thresholds, and $\gamma_1 < \cdots < \gamma_m$ are the corresponding acceptable values of loss. The set of values (γ_i, ϵ_i) therefore determines a "risk profile" chosen by the user.

It is a simple matter to derive SDP conditions, under the assumptions used in this paper, that ensure that the multiple VaR constraints hold robustly with respect to the distribution of returns. For example, in the context of the assumptions of Theorem 2, we have Theorem 10.

Theorem 10. *When the distribution of returns is known only via its first two moments and is otherwise arbitrary, the multiple worst-case Value-at-Risk constraints hold if and only if there exist variables* $\Gamma, \hat{x}, x_1, \ldots, x_m$ *such that*

$$x_i^{\mathrm{T}}w \leqslant -\gamma_i, \quad \begin{bmatrix}\Gamma & x_i - \hat{x}\\ (x_i - \hat{x})^{\mathrm{T}} & \kappa(\epsilon_i)^2\end{bmatrix} \succeq 0, \quad i = 1, \ldots, m.$$

The above theorem allows us to optimize the risk profile by proper choice of the portfolio weights, in several ways: We may minimize an average of the potential losses $\gamma_1 + \cdots + \gamma_m$, for example, or the largest value of the losses, $\max_i \gamma_i$. Such problems fall in the SDP class.

5. NUMERICAL EXAMPLE

In this example we have considered a portfolio involving $n = 13$ assets. Our portfolio weights are restricted to lie in the set

$$\mathcal{W} = \left\{ w \,\middle|\, w \geqslant 0,\ \sum_{i=1}^n w_i = 1\right\}.$$

This kind of set does not usually result in very diversified portfolios. In practice, one can (and should) impose additional linear inequality constraints on w to obtain diversified portfolios; such constraints are discussed in Lobo et al. (2000). In a similar vein, we have not accounted for transaction costs. Our purpose in this paper is not diversification nor transaction costs, but robustness to data uncertainty.

Using historical one-day returns over a period of 261 trading days (from November, 1999 through October, 2000), we have computed the sample mean and covariance matrix of the returns, $\hat{x}^{\mathrm{nom}}$ and Γ^{nom}. With these nominal values and given a risk level ϵ, we can compute a portfolio, using Theorem 3, with $\Gamma_+ = \Gamma_- = \Gamma^{\mathrm{nom}}$ and $\hat{x}_+ = \hat{x}_- = \hat{x}^{\mathrm{nom}}$. We refer to this portfolio—one resulting from the assumption that the data are error-free—as the "nominal" portfolio, against which we can compare a robust portfolio.

We assume that the data (including our mean and covariance estimates) are prone to errors. We denote by ρ a parameter that measures the relative uncertainty on the covariance matrix, understood in the sense of a componentwise, uniform variation. Thus, the uncertainty in the covariance matrix Γ is described by

$$|\Gamma(i,j) - \Gamma^{\mathrm{nom}}(i,j)| \leqslant \rho |\Gamma^{\mathrm{nom}}(i,j)|, \qquad 1 \leqslant i, j \leqslant n.$$

In practice, the mean is harder to estimate than the covariance matrix, so we have put the relative uncertainty on the mean to be ten times that of the covariance matrix, i.e.,

$$|\hat{x} - \hat{x}^{\mathrm{nom}}(i)| \leqslant 10\rho |\hat{x}^{\mathrm{nom}}(i)|, \qquad 1 \leqslant i \leqslant n.$$

We have then examined the worst-case behavior of the nominal portfolio as the uncertainty on the point estimates $\hat{x}^{\mathrm{nom}}$ and Γ^{nom} increase. This worst-case analysis is done via Theorem 2. We have compared the worst-case VaR of the nominal portfolio with that of an optimally robust portfolio, which is computed via Theorem 3. Our results were obtained using the general-purpose semidefinite programming code SP (Vandenberghe and Boyd 1999).

These results are illustrated in Figure 1. The x-axis is the relative uncertainty on the covariance matrix, ρ. The y-axis is the VaR, given as a percentage of the original portfolio value. Figure 2 shows the relative deviation of the worst-case VaR with respect to the nominal VaR, which is obtained by setting $\rho = 0$. For example, for $\rho = 10\%$ the worst-case VaR of the nominal portfolio could be as much as 270% of the nominal VaR, while the VaR of the robust portfolio is about 200% of the nominal VaR.

We see that if we choose the nominal portfolio, data errors can have a dramatic impact on the VaR. Taking into account the uncertainty by solving a robust portfolio allocation problem dampens greatly this potential catastrophic effect. This is even more so as the uncertainty level ρ increases.

In Figure 3, we illustrate the behavior of our portfolios when the probability level ϵ varies. We compare the VaR in three situations: One is the VaR of the optimal nominal portfolio (that is, obtained without taking into account data uncertainty), shown in the lowest curve. The upper curve corresponds to the worst-case analysis of the nominal portfolio. The middle curve shows the worst-case VaR of the robust portfolio. Again, we see a dramatic improvement brought about by the robust portfolio. The latter is less efficient than the nominal portfolio *if* there were no uncertainty; the presence of data uncertainty makes the nominal portfolio a poor choice over the robust one.

Figure 1. Worst-case VaR of the nominal and robust portfolios, as a function of the size of data uncertainty, ρ.

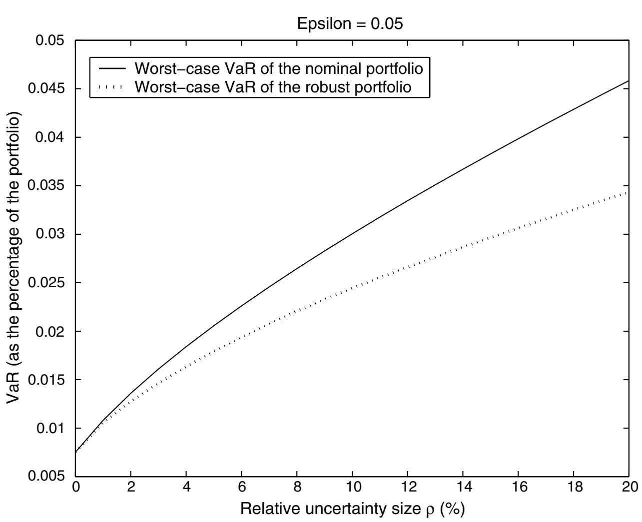


Figure 2. Relative (to the nominal VaR) worst-case VaR of the nominal and robust portfolios, as a function of the size of data uncertainty, ρ.

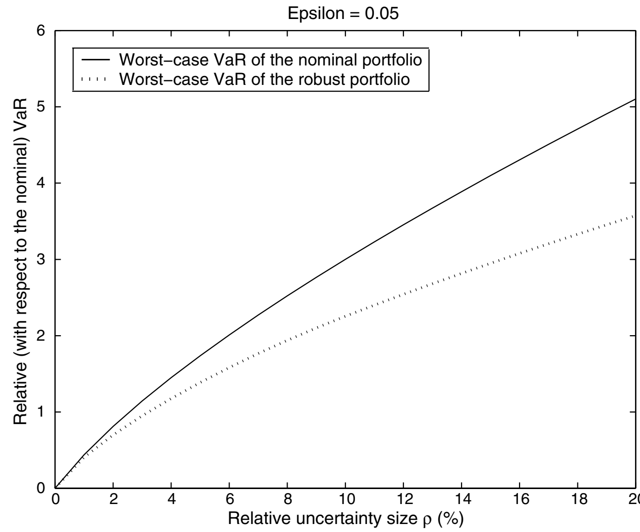
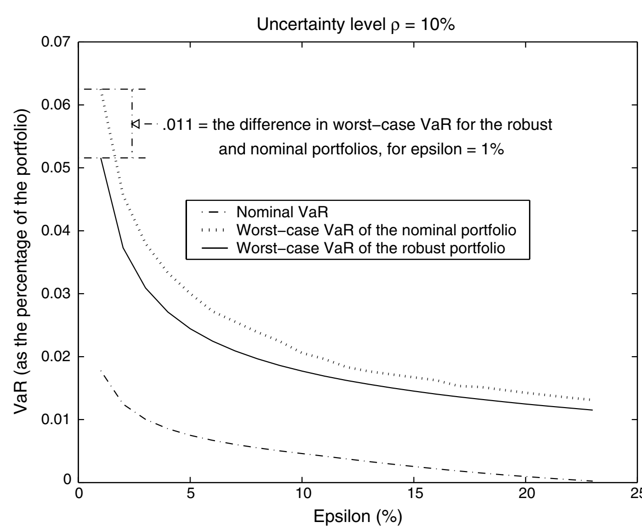


Figure 3. Worst-case VaR of the nominal and robust portfolios, as a function of the probability level ϵ.



6. CONCLUDING REMARKS

Our results can be summarized as follows. The problem of computing the worst-case VaR, or optimizing it, takes the general form

$$\begin{aligned}&\text{minimize} \quad \phi_{\mathscr{W}}(x)\\ &\text{subject to} \quad (x,\hat{x},\Gamma)\in\mathscr{V}, \quad \begin{bmatrix}\Gamma & x-\hat{x}\\ (x-\hat{x})^{\mathrm{T}} & \kappa^2\end{bmatrix}\succeq 0,\end{aligned}$$

where $\phi_{\mathscr{W}}$ is the support function of a convex set, that describes the set of admissible portfolio allocation vectors; the set $\mathscr{V}$ reflects the partial information (moments bounds and support) we have on the distribution of returns, and the risk factor κ depends on the chosen optimization model (entropy-constrained or moment-constrained).

The optimal variables $\Gamma, \hat{x}$ in the above problem are selected to be the most prudent when facing data uncertainty. The duality between the portfolio weights and the worst-case probability distribution information $(\hat{x}, \Gamma)$ is reminiscent of the duality in option pricing problems, between the optimal hedging strategy (for replicating the price of an option) and the risk-neutral probability measure (Musiela and Rutkowski 1997).

As noted in §2.1, the above formulation has a deterministic interpretation, in which the returns are only known to belong to a union of ellipsoids of the form

$$\{x \mid \kappa(\epsilon)^2\Gamma \succeq (x-\hat{x})(x-\hat{x})^{\mathrm{T}}\}$$

where the shape matrix Γ and center $\hat{x}$ are unknown-but-bounded, and the problem is to allocate resources in a "min-max," or game-theoretic, manner. Our SDP solution illustrates a kind of "certainty equivalent principle" by which a problem involving probabilistic uncertainty has an interpretation, and an efficient numerical solution, as a deterministic game.

The numerical tractability of the above problem depends on the structure of the sets $\mathscr{W}, \mathscr{V}$. We have identified some practically interesting cases when these sets result in a tractable, semidefinite programming problem, namely componentwise and ellipsoidal bounds. In the case of support constraints on the distribution of returns the problem does not seem to be tractable, but we have shown how to compute an upper bound on the worst-case VaR via semidefinite programming.

ACKNOWLEDGMENTS

The authors acknowledge the valuable input from Alexandre d'Aspremont, Dimitris Bertsimas, Stephen Boyd, Sanjiv Das, Darrell Duffie, Miguel Lobo, and Ioana Popescu. They thank Xavier Burtschell and Aniruddha Pant for their help with numerical implementations.

REFERENCES

Berstimas, D., I. Popescu. 1999. On the relation between option and stock prices: A convex optimization approach. Working paper #4085, Sloan School of Management, MIT, Cambridge, MA.

——, ——. 2000. Optimal inequalities in probability theory: A convex optimization approach. Working paper, MIT/INSEAD. Submitted to *Oper. Res.*

Black, F., R. Litterman. 1992. Global portfolio optimization. *Financial Analysts J.* 28–43.

Bonnans, J. F., A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer-Verlag, Berlin, Germany.

Boyd, S., L. El Ghaoui, E. Feron, V. Balakrishnan. 1994. *Linear Matrix Inequalities in System and Control Theory*, Vol. 15. *Studies in Applied Mathematics*. SIAM, Philadelphia, PA.

Costa, O., A. Paiva. 2001. Robust portfolio selection using linear-matrix inequalities. *J. Econom. Dynam. Control*. To appear.

El Ghaoui, L., M. Oks, F. Oustry. 2000. Worst-case value-at-risk and robust asset allocation: A semidefinite programming approach. Technical report M00/59, ERL, University of California, Berkeley, CA.

Goldfarb, D., G. Iyengar. 2001. Robust portfolio selection problems. CORC technical report, TR-2001-05. Submitted to *Math. Oper. Res.*

Halldórsson, B. V., R. Tütüncü. 2000. An interior-point method for a class of saddle-point problems. Submitted to *J. Optim. Theory Appl.*

Hiriart-Urruty, J., C. Lemaréchal. 1993. *Convex Analysis and Minimization Algorithms.* Springer-Verlag, Berlin, Germany.

Linsmeier, T. J., N. D. Pearson. 1996. Risk Measurement: An introduction to value-at-risk. Technical report 96-04, OFOR, University of Illinois, Urbana-Champaign, IL.

Lobo, M. S., S. Boyd. 1999. The worst-case risk of a portfolio. In preparation.

——, M. Fazel, S. Boyd. 2000. Portfolio optimization with linear and fixed transaction costs and bounds on risk. Submitted to *Oper. Res.*

——, L. Vandenberghe, S. Boyd, H. Lebret. 1998. Second-order cone programming: Interior-point methods and engineering applications. *Linear Algebra Appl.* **284** 193–228.

Luenberger, D. G. 1999. *Investment Science*. Oxford University Press, Oxford, U.K.

Markowitz, H. 1952. Portfolio selection. *J. Finance* **7**(1) 77–91.

Musiela, M., M. Rutkowski. 1997. *Martingale Methods in Financial Modelling*, Vol. 36. *Applications of Mathematics, Stochastic Modelling and Applied Probability*. Springer, Berlin, Germany.

Nesterov, Y., A. Nemirovsky. 1994. *Interior Point Polynomial Methods in Convex Programming: Theory and Applications.* SIAM, Philadelphia, PA.

Pearson, N. D., X. Ju. 1999. Using value-at-risk to control risk taking: How wrong can you be? *J. Risk* **1**(2).

Pritsker, M. 1997. Evaluating value-at-risk methodologies: Accuracy vs. computational time. *J. Financial Services Res.* **12**(2/3) 201–242.

Saigal, R., L. Vandenberghe, H. Wolkowicz, eds. 2000. *Handbook on Semidefinite Programming and Applications.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Smith, J. 1995. Generalized Chebyshev inequalities: Theory and applications in decision analysis. *Oper. Res.* **43** 807–825.

Sturm, J. 1998. *SeDuMi: A Matlab Toolbox*. Version 1.03.

Vandenberghe, L., S. Boyd. 1996. Semidefinite programming. *SIAM Rev.* **38**(1) 49–95.

——, ——. 1999. **SP** *Version* 1.1: *Software for Semidefinite Programming, User's Guide*.