

Copyright © 2000, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**WORST-CASE VALUE-AT-RISK AND
ROBUST ASSET ALLOCATION: A
SEMIDEFINITE PROGRAMMING APPROACH**

by

Laurent El Ghaoui, Francois Oustry and Maksim Oks

Memorandum No. UCB/ERL M00/59

4 December 2000

**WORST-CASE VALUE-AT-RISK AND
ROBUST ASSET ALLOCATION: A
SEMIDEFINITE PROGRAMMING APPROACH**

by

Laurent El Ghaoui, Francois Oustry and Maksim Oks

Memorandum No. UCB/ERL M00/59

4 December 2000

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Worst-Case Value-at-Risk and Robust Asset Allocation : a Semidefinite Programming Approach*

Laurent El Ghaoui[†] Francois Oustry[‡] Maksim Oks[§]

December 4, 2000

Abstract

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 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.

*Submitted to *Operations Research*, December 2000. Do not distribute.

[†]Author to whom correspondence should be addressed. Department of Electrical Engineering and Computer Sciences, Cory Hall, University of California, Berkeley, CA, 94720, USA. Internet: elghaoui@eecs.berkeley.edu. Research partially supported by a National Science Foundation Career award, a MURI grant, and a gift from BNP-Paribas Group.

[‡]INRIA Rhone-Alpes ZIRST - 655 avenue de l'Europe 38330 Montbonnot Saint-Martin, France. Internet: Francois.Oustry@inria.fr

[§]Department Industrial Engineering and Operations Research, Etcheverry Hall, University of California, Berkeley, CA, 94720, USA. Internet: max@ieor.berkeley.edu.

Contents

- 1 Introduction** **3**
- 1.1 Some classical measures of risk 3
- 1.2 The problem of data uncertainty 4
- 1.3 The worst-case VaR 5
- 1.4 Main results and paper outline 6

- 2 Worst-case VaR with Moment Uncertainty** **7**
- 2.1 Known moments 7
- 2.2 Convex moment uncertainty 9
- 2.3 Polytopic uncertainty 10
- 2.4 Componentwise bounds on mean and covariance matrix 11
- 2.5 Componentwise bounds on mean and second-moment matrix 13

- 3 Factor Models** **14**
- 3.1 Uncertainty in the factor’s mean and covariance matrix 14
- 3.2 Uncertainty in the sensitivity matrix 14

- 4 Extensions and variations** **17**
- 4.1 Including support information 17
- 4.2 Entropy-constrained VaR 19
- 4.3 Imposing constraints on the covariance matrix 19
- 4.4 Multiple VaR constraints 20

- 5 Numerical example** **20**

- 6 Concluding Remarks** **23**

- A Semidefinite and Second-order Cone Programming** **26**

- B Proofs** **27**
- B.1 Proof of theorem 1 27
- B.2 Proof of lemma 1 30
- B.3 Proof of theorem 5 31
- B.4 Proof of theorem 6 32
- B.5 Proof of theorem 9 36

Notation

For two n -vectors x, y , $x \geq y$ (resp. $x < y$) means $x_i \geq y_i$ (resp. $x_i < y_i$), $i = 1, \dots, n$; x_+ denotes the positive part of x , which is the vector with components $\max(x_i, 0)$. e is the n -vector with all components equal to one. The notation \mathcal{S}_n denotes the set of $n \times n$ symmetric matrices. For two $n \times n$ symmetric matrices $A, B \in \mathcal{S}_n$, $A \succeq B$ (resp. $A \succ B$) means $A - B$ positive semidefinite (resp. definite). For $S \succeq 0$, $S^{1/2}$ denotes the symmetric square root of S . For symmetric matrices of the same size, $\langle A, B \rangle = \text{Tr}(AB)$ denotes the standard scalar product in the space of symmetric matrices. For a non empty convex subset Ω of \mathbb{R}^n , we denote by $\mathcal{K}(\Omega)$ the cone of non negative distributions with support included in Ω . Φ denotes the cumulative distribution function of a scalar, zero mean, unit variance Gaussian random variable.

1 Introduction

We consider a one-period asset allocation problem. Over the period, the price of asset i changes by a relative amount 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

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

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 $w \in \mathcal{W}$, in order 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 [14, 13], 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

$$\text{minimize } w^T \Gamma w \text{ subject to } \hat{x}^T w \geq \mu, \quad w \in \mathcal{W} \tag{1}$$

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

The Value-at-Risk framework (see, for example, [18, 9]) instead looks at the probability of low returns. 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 :

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

where $\epsilon \in (0, 1]$ is given (say, $\epsilon \simeq 2\%$). Contrarily 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 expresses as

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

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 in [2], by simply setting $\kappa(\epsilon) = \sqrt{(1-\epsilon)/\epsilon}$ (we will obtain this result in section 2.1).

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

$$\text{minimize } \kappa\sqrt{w^T\Gamma w} - \hat{x}^T w \text{ subject to } w \in \mathcal{W}. \quad (3)$$

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 $\kappa \geq 0$ (which is true iff $\epsilon \in (0, 1/2]$ in the Gaussian case), $V(w)$ is a convex function of w , and the above problem can be easily solved globally using for example interior-point techniques for convex, second-order cone programming (SOCP, see [12, 21]).

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, integration techniques such as Monte-Carlo simulation [9] are not easily extended to portfolio design. Assuming Gaussian returns is a practical, albeit quite rough, approximation. The classical Chebyshev bound does not assume prior knowledge of the distribution, but may be overly pessimistic. The exact version of the Chebyshev bound is not realistic, as the optimal worst-case distribution turns out to be discrete [3]. Moreover, Chebyshev bounds do not take into account support information, stemming for example from the fact that prices are never negative, and hence, returns always greater or equal to -1 .

1.2 The problem of data uncertainty

Despite their shortcomings, the above frameworks do provide elegant solutions to VaR analysis and design. However they suffer from an important drawback, which is perhaps not so well recognized. As pointed out in [4, 17], these approaches require a perfect knowledge of the data, in our case the mean and covariance matrix. In practice, the data is often prone to errors. This hidden risk is compounded by the fact that the point estimates of the covariance matrix sometimes have low (numerical) rank, and this may result in an optimal portfolio

with zero variance, quite an absurd result. In general, portfolio optimization based solely on such 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 in [4], where a method is proposed to combine the classical Markowitz approach with a priori information or “investor’s views” on the market.

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 [4]. 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. In order 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 [9]. Thus, possibly large modelling errors are almost always present in complex portfolios. We discuss these errors in more detail in section 3.

Yet another source of data perturbations could come from 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. It is possible, of course, to come up with a (finite) number of different scenarios (what happens if two usually uncorrelated industry sectors become suddenly highly correlated?), and compute the corresponding VaR. (We will return to this problem in section 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 in 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

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

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

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

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 set \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 recently available [16, 23, 21, 19]. (For more details, see Appendix A.) SDPs are now the subject of intense research and arise in a large number of applications, ranging from statistics to control systems design [19].

When the mean and covariance matrix are uncertain but bounded, our solution produces not only a worst-case VaR or 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 a seemingly untractable (NP-hard) problems. We show how to compute upper bounds on this problem, via SDP.

Lobo and Boyd [10] 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. Our paper extends these results to the context of VaR, with various partial information on the probability distribution.

In our approach, we were greatly inspired by the recent work of Bertsimas and Popescu, who also use SDP to find (bounds for) probabilities under partial probability distribution information [2] and apply this approach to option pricing problems [1]. 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 section 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 only known within a given convex set. We then specialize our results to two kinds of bounds: polytopic and componentwise. In section 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 section 2: exploiting support information, ruling out discrete probability distributions via relative entropy constraints, handling multiple VaR constraints. We provide a numerical illustration in section 5. Proofs are in the Appendix B.

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 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. 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:

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

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. The proof of this result is in Appendix B.1.

Theorem 1 *Let \mathcal{P} be the set of probability distributions that have a given 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,*

$$\sup \text{Prob}\{\gamma \leq -r(w, x)\} \leq \epsilon,$$

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

2. *We have*

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

where

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

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

$$\langle M, \Sigma \rangle \leq \tau\epsilon, \quad M \succeq 0, \quad \tau \geq 0, \quad M + \begin{bmatrix} 0 & w \\ w^T & -\tau + 2\gamma \end{bmatrix} \succeq 0, \quad (9)$$

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

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

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

we have $-x^T w \leq \gamma$.

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

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

Let us comment on the above theorem. The equivalence between propositions 1 and 2 will be proven in Appendix, but can be obtained as a simple application of the (exact) multivariate Chebyshev bound given in [2]. This implies that the problem of optimizing the VaR over $w \in \mathcal{W}$ is equivalent to

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

As noted in the introduction, this problem can be cast as a second-order cone programming problem (SOCP). SOCPs are special forms of SDPs which can be solved with efficiency close to that of linear programming [12].

The equivalence between propositions 1 and 3 in the above theorem is a consequence of duality (in an appropriate sense; see Appendix B.1).

Proposition 4 follows from 3 via duality again, this time in the SDP sense. This connection is not direct, as it involves some elimination of variables. Note that proposition 4 implies that the worst-case VaR can be computed via the SDP in variable x

$$V_{\mathcal{P}}(w) = \max -x^T w \text{ subject to (10).}$$

The above provides a deterministic, or “game-theoretic”, interpretation of the VaR. Indeed, since $\Gamma \succ 0$, condition (4) is equivalent to $x \in \mathcal{E}$, where \mathcal{E} is the ellipsoid

$$\mathcal{E} = \left\{ x \mid (x - \hat{x})^T \Gamma^{-1} (x - \hat{x}) \leq \kappa(\epsilon)^2 \right\},$$

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

Expressions (9) and (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 as it can be extended to the more general cases seen in section 2.2, while the SOCP approach cannot.

2.2 Convex moment uncertainty

We now turn to the case when (Γ, \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 (Γ, \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 (10) holds, we have $-x^T w \geq \gamma$. It thus suffices to make Γ and \hat{x} variables in the above conditions, to compute the worst-case VaR:

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

Since $\Gamma \succeq 0$ is implied by (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” since \mathcal{U} is bounded. We thus have the following result.

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

$$\max -x^T w \quad \text{subject to } (\Gamma, \hat{x}) \in \mathcal{U}, \quad \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^T & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \quad (13)$$

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

Solving problem (13) 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

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

It turns out that we can exchange the “max” and “min” in the above, since \mathcal{U}, \mathcal{W} are convex, bounded and have non empty interiors. This yields the convex optimization problem in variables x, \hat{x} and Γ :

$$-V_{\mathcal{P}}^{\text{opt}} = \min \phi_{\mathcal{W}}(x) \quad \text{subject to } (\hat{x}, \Gamma) \in \mathcal{U}, \quad \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^T & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \quad (14)$$

where $\phi_{\mathcal{W}}$ is a convex function, defined as the support function of the convex set \mathcal{W} :

$$\phi_{\mathcal{W}}(x) := \sup_{w \in \mathcal{W}} x^T w.$$

An optimal portfolio can be recovered from any optimal x by finding a w that achieves the supremum in the above (the existence of which is guaranteed by boundedness of \mathcal{W}); the optimal worst-case VaR is the negative of the optimal objective.

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 (11) holds. Thus, the worst-case VaR is given by the max-min problem

$$\begin{aligned} \max_{\Gamma, \hat{x}} \min_{\Lambda, v} \quad & \langle \Lambda, \Gamma \rangle + \kappa(\epsilon)^2 v - \hat{x}^T w \\ \text{subject to} \quad & \begin{bmatrix} \Lambda & w/2 \\ w^T/2 & v \end{bmatrix} \succeq 0, \quad (\hat{x}, \Gamma) \in \mathcal{U}, \quad \Gamma \succeq 0. \end{aligned} \quad (15)$$

The feasible set in the above problem is compact and convex, and the objective is linear in Γ, \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

$$\begin{aligned} \min_{\Lambda, v, w} \max_{\Gamma, \hat{x}} \quad & \langle \Lambda, \Gamma \rangle + \kappa(\epsilon)^2 v - \hat{x}^T w \\ \text{subject to} \quad & \begin{bmatrix} \Lambda & w/2 \\ w^T/2 & v \end{bmatrix} \succeq 0, \quad (\hat{x}, \Gamma) \in \mathcal{U}, \quad \Gamma \succeq 0, \quad w \in \mathcal{W}. \end{aligned} \quad (16)$$

The above problem can be interpreted as a game, where the variables Λ, v seek to decrease the VaR while the variables Γ, \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 (15).

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 Γ, \hat{x}, x (14). Alternatively, we can solve the “min-max” problem (16).*

The tractability of problem depends on the structure of sets \mathcal{U} and \mathcal{W} . When both sets are described by linear matrix inequalities involving \hat{x}, Γ and w , the resulting problem can be expressed explicitly as an SDP. The “min-max” formulation is useful when we are able to explicitly 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}, Γ) is only known to belong to a given polytope, described by its vertices. Precisely, we assume that $(\hat{x}, \Gamma) \in \mathcal{U}$, where

$$\mathcal{U} = \text{Co} \{(\hat{x}_1, \Gamma_1), \dots, (\hat{x}_l, \Gamma_l)\}, \quad (17)$$

where the vertices (\hat{x}_i, Γ_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 section 2.2. The matrix-vector pair (\hat{x}, Γ) is made a variable in the analysis problem (13) or the portfolio optimization problem (14). Denoting by ξ the vector containing the independent elements of \hat{x} and Γ , we can express ξ as

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

where ξ_i corresponds to the vertex pair (\hat{x}_i, Γ_i) . The resulting optimization problem (13) or (14) 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

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

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

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

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, t :

$$\min_{w \in \mathcal{W}} t \text{ subject to } \kappa(\epsilon) \|\Gamma_i^{1/2} w\|_2 \leq t + \hat{x}_i^T w, \quad i = 1, \dots, l.$$

As seen in appendix A, 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 are interested in a covariance matrix whose entries are only known within two 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 section 2.2 to the case when Γ, \hat{x} are only known within componentwise bounds:

$$x_- \leq \hat{x} := \mathbf{E} x \leq x_+, \quad \Gamma_- \leq \Gamma := \mathbf{E}(x - \hat{x})(x - \hat{x})^T \leq \Gamma_+, \quad (18)$$

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 non-degenerate probability distribution such that the above moment bounds hold, that is, there exist a matrix $\Gamma \succ 0$ such that $\Gamma_- \leq \Gamma \leq \Gamma_+$. (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

$$\begin{aligned} & \text{maximize} && -x^T w \\ & \text{subject to} && x_- \leq \hat{x} \leq x_+, \quad \Gamma_- \leq \Gamma \leq \Gamma_+, \\ & && \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^T & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \end{aligned} \quad (19)$$

Note that the SDP (19) is strictly feasible if and only if there exist $\Gamma \succ 0$ such that (18) holds. In practice, it may not be necessary to check this strict feasibility condition prior to solving the problem. SDP codes such as SeDuMi [21] 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

$$\begin{aligned} & \min_{\Lambda, v} \max_{\Gamma, \hat{x}} \langle \Lambda, \Gamma \rangle + \kappa(\epsilon)^2 v - \hat{x}^T w \\ & \text{subject to} \quad \begin{bmatrix} \Lambda & w/2 \\ w^T/2 & v \end{bmatrix} \succeq 0, \\ & \quad x_- \leq \hat{x} \leq x_+, \quad \Gamma_- \leq \Gamma \leq \Gamma_+, \quad \Gamma \succeq 0. \end{aligned} \quad (20)$$

For fixed Λ, v , the inner maximization problem in (20) is a (particularly simple) SDP in \hat{x}, Γ . We can write this problem in the dual form of a minimization problem. In fact,

$$\max_{x_- \leq \hat{x} \leq x_+} -w^T \hat{x} = \min_{\lambda_{\pm} \geq 0, w = \lambda_- - \lambda_+} \lambda_+^T x_+ - \lambda_-^T x_-,$$

and a similar result for the term involving Γ :

$$\max_{\Gamma_- \leq \Gamma \leq \Gamma_+, \Gamma \succeq 0} \langle \Lambda, \Gamma \rangle = \min_{\Lambda_{\pm} \geq 0, \Lambda \leq \Lambda_+ - \Lambda_-} \langle \Lambda_+, \Gamma_+ \rangle - \langle \Lambda_-, \Gamma_- \rangle,$$

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

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

$$\begin{aligned} V_{\mathcal{P}}^{\text{opt}} = & \min \langle \Lambda_+, \Gamma_+ \rangle - \langle \Lambda_-, \Gamma_- \rangle + \kappa(\epsilon)^2 v + \lambda_+^T x_+ - \lambda_-^T x_- \\ & \text{subject to} \quad \lambda_+ \geq 0, \quad \lambda_- \geq 0, \quad \Lambda_+ \geq 0, \quad \Lambda_- \geq 0, \\ & \quad \begin{bmatrix} \Lambda_+ - \Lambda_- & w/2 \\ w^T/2 & v \end{bmatrix} \succeq 0, \quad w = \lambda_- - \lambda_+. \end{aligned} \quad (21)$$

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. Since \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 $w w^T / 4v$, and optimizing over $v > 0$ yields the result.

2.5 Componentwise bounds on mean and second-moment matrix

We now assume that we have componentwise bounds on the mean \hat{x} and second-moment matrix S , specifically,

$$x_- \leq \hat{x} := \mathbf{E} x \leq x_+, \quad S_- \leq S := \mathbf{E} x x^T \leq S_+, \quad (22)$$

where x_+, x_- and S_+, S_- are given vectors and matrices, respectively, and the inequalities are understood componentwise. We denote by Σ the second-moment matrix defined in (6), and by Σ_+ and Σ_- the corresponding componentwise bounds. Again, we assume that there exist a $\Sigma \succ 0$ within the interval matrix $[\Sigma_- \ \Sigma_+]$. As before this assumption can be checked in a single phase, while computing (or optimizing) the VaR.

We observe that the constraint appearing in formulation (10) for the worst-case VaR can be written as

$$\begin{bmatrix} S & x - \hat{x} & \hat{x} \\ (x - \hat{x})^T & \kappa(\epsilon)^2 & 0 \\ \hat{x}^T & 0 & 1 \end{bmatrix} \succeq 0, \quad (23)$$

which is a linear matrix inequality involving x, \hat{x} , and S . Note that this constraint implies $S \succeq 0$. Thus, we may compute the worst-case VaR by solving the SDP

$$\max -x^T w \text{ subject to } x_- \leq \hat{x} \leq x_+, \quad S_- \leq S \leq S_+, \quad (23).$$

To optimize the VaR, we turn to the formulation (9). The worst-case VaR is less than γ if and only if there exist $\tau \geq 0$ and M such that

$$\sup_{\Sigma} \langle M, \Sigma \rangle \leq \tau \epsilon, \quad M \succeq 0, \quad \tau \geq 0, \quad M + \begin{bmatrix} 0 & w \\ w^T & -\tau + 2\gamma \end{bmatrix} \succeq 0,$$

where Σ is the second-moment matrix defined in (6), and the ‘‘sup’’ is taken with respect to the componentwise bounds on Σ . As in section 2.4, we have

$$\max_{\Sigma_- \leq \Sigma \leq \Sigma_+, \Sigma \succeq 0} \langle M, \Sigma \rangle = \min_{M_{\pm} \succeq 0, M \preceq M_+ - M_-} \langle M_+, \Sigma_+ \rangle - \langle M_-, \Sigma_- \rangle.$$

Therefore, the worst-case VaR can be optimized via the SDP in variables $M_+, M_-, \lambda_+, \lambda_-$ and w :

$$\begin{aligned} V_{\mathcal{P}}^{\text{opt}} = & \min \langle M_+, \Gamma_+ \rangle - \langle M_-, \Gamma_- \rangle + \kappa(\epsilon)^2 v + \lambda_+^T x_+ - \lambda_-^T x_- \\ \text{subject to} & \lambda_+ \geq 0, \lambda_- \geq 0, M_+ \geq 0, M_- \geq 0, \\ & M_+ - M_- + \begin{bmatrix} 0 & w/2 \\ w^T/2 & v \end{bmatrix} \succeq 0, \quad w = \lambda_- - \lambda_+ \in \mathcal{W}. \end{aligned} \quad (24)$$

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

$$x = Af + u, \tag{25}$$

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 matrix D , we obtain that the corresponding worst-case VaR can be computed exactly via the SDP

$$\begin{aligned} & \text{maximize} && -x^T w \\ & \text{subject to} && \hat{x} = A\hat{f}, \quad \Gamma = D + ASA^T, \quad f_- \leq \hat{f} \leq f_+, \quad S_- \leq S \leq S_+, \\ & && \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^T & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \end{aligned}$$

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 in [9], 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 several different instruments as a function of different “market factors”, such as currency exchange rates, interest rates, underlying asset prices, and so on. Those are quantities for which historical data is 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 (25). In essence the factor model is a linearized approximation to the actual return function, which allows to use 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} is perfectly known, with $S \succeq 0$ and $D \succ 0$, and that the errors in A are modeled by $A \in \mathcal{A}$, where the given set \mathcal{A} describes the possible values for A . We are interested in computing, and optimizing with respect to the portfolio weight vector w , the worst-case VaR

$$V_{\text{wc}}(w) := \max_{A \in \mathcal{A}} \left\| \begin{bmatrix} S^{1/2} A^T \\ D^{1/2} \end{bmatrix} w \right\|_2 - w^T A \hat{f}. \quad (26)$$

We will not compute the worst-case VaR exactly, but will come up with upper bounds.

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

$$\mathcal{A} = \left\{ A_0 + \sum_{i=1}^l u_i A_i \mid u \in \mathbb{R}^l, \|u\|_2 \leq 1 \right\} \quad (27)$$

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

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

$$\phi(w) := \max_{\|u\|_2 \leq 1} \|C(w)u + d(w)\|_2 + e(w)^T u,$$

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. This lemma is proved in appendix B.2.

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

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

can be computed via the SDP

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

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 (25), and the sensitivity matrix is only known to belong to the set \mathcal{A} given by (27), an upper bound on the worst-case VaR given in (26) can be computed (optimized) via the following SDP in variables λ (and $w \in \mathcal{W}$)*

$$\min \frac{1}{2}(\lambda_1 + \lambda_2 + \lambda_3) - w^T A_0 \hat{f} \text{ subject to } \begin{bmatrix} \lambda_1 I_{n+m} & C(w) & d(w) \\ C(w)^T & \lambda_2 I_l & e(w) \\ d(w)^T & e(w)^T & \lambda_3 \end{bmatrix} \succeq 0, \quad (28)$$

where

$$C(w) = [M_1 w \quad \dots \quad M_l w], \text{ with } M_i = \begin{bmatrix} S^{1/2} A_i^T \\ 0 \end{bmatrix}, \quad 1 \leq i \leq l,$$

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

Norm-bound uncertainty. We now consider the case when

$$\mathcal{A} = \{ A + L\Delta R \mid \Delta \in \mathbf{R}^{l \times r}, \|\Delta\| \leq 1 \}, \quad (29)$$

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. (See [6] for details on unstructured uncertainty in matrices.) 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, and is proved in appendix B.3.

Theorem 5 *When the distribution of returns obeys to the factor model (25), and the sensitivity matrix is only known to belong to the set \mathcal{A} given by (29), an upper bound on the worst-case VaR given in (26) can be computed (optimized) via the following SDP in variables λ (and $w \in \mathcal{W}$)*

$$\min \frac{1}{2}(\lambda_1 + t + \lambda_3) - w^T A \hat{f} \text{ subject to } \begin{bmatrix} \lambda_1 I_{n+m} & d(w) \\ d(w)^T & \lambda_3 \end{bmatrix} \succeq \mu \begin{bmatrix} C \\ e^T \end{bmatrix} \begin{bmatrix} C \\ e^T \end{bmatrix}^T, \quad (30)$$

$$\begin{bmatrix} t & w^T L \\ L^T w & \mu I_l \end{bmatrix} \succeq 0,$$

where

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

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. Its proof is in Appendix B.4.

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 variables x, κ^2 :*

$$\begin{aligned} & \text{maximize} && -x^T w \\ & \text{subject to} && \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^T & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \\ & && \kappa(\epsilon)^2 x_l \leq \hat{x} - x \leq \kappa(\epsilon)^2 x_u, \quad x_l \leq x \leq x_u. \end{aligned} \tag{31}$$

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

If we let $x_l = -\infty$ and $x_u = +\infty$, the last inequalities in (31) 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 (31) 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 dealt with in section 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

$$\begin{aligned} & \text{maximize} && -x^T w \\ & \text{subject to} && \|\Gamma^{-1/2}(x - \hat{x})\|_2 \leq \kappa(\epsilon)^2, \\ & && \kappa(\epsilon)^2 x_l \leq \hat{x} - x \leq \kappa(\epsilon)^2 x_u, \quad x_l \leq x \leq x_u. \end{aligned} \tag{32}$$

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:

$$\bar{V}_{\mathcal{P}}^{\text{opt}} := \min_{w \in \mathcal{W}} \max_x -x^T w \text{ subject to } \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^T & \kappa(\epsilon)^2 \end{bmatrix} \succeq 0, \\ \kappa(\epsilon)^2 x_l \leq \hat{x} - x \leq \kappa(\epsilon)^2 x_u, \quad x_l \leq x \leq x_u,$$

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 $w, t, \Lambda, u, v, \lambda_{u,l}, \nu_{u,l}$:*

$$\begin{aligned} \bar{V}_p^{\text{opt}} = \min \quad & \langle \Lambda, \Gamma \rangle + \kappa(\epsilon)^2 v - \hat{x}^T w \\ & + \kappa(\epsilon)^2 (x_u^T \lambda_u - x_l^T \lambda_l) + \nu_u^T (x_u - \hat{x}) - \nu_l^T (x_l - \hat{x}) \\ \text{subject to} \quad & \begin{bmatrix} \Lambda & (w + \lambda_l - \lambda_u + \nu_u - \nu_l)/2 \\ (w + \lambda_l - \lambda_u + \nu_u - \nu_l)^T/2 & v \end{bmatrix} \succeq 0. \end{aligned} \quad (33)$$

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 (33), 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.

As before, we can express the above problem as an SOCP. To see this, it suffices to express the dual, in the SOCP sense, of problem (32).

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 of \mathbb{R}^n , given by

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

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 6, 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 $\Omega := \{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$, we can compute an upper bound on the worst-case Value-at-Risk by solving the semidefinite programming problem in variables x, X, t :*

$$\begin{aligned} \text{maximize} \quad & -x^T w \text{ subject to} \quad \begin{bmatrix} \Gamma - \epsilon X & x - \hat{x} & x \\ (x - \hat{x})^T & \kappa(\epsilon)^2 & 0 \\ x^T & 0 & 1 \end{bmatrix} \succeq 0, \\ & t/\epsilon = 1 - \text{Tr } P^{-1} X + 2x_c^T P^{-1} x - x_c^T P^{-1} x_c \geq 0, \\ & \begin{bmatrix} 1 - t - \text{Tr } P^{-1} \Gamma & \hat{x} - x_c \\ (\hat{x} - x_c)^T & P \end{bmatrix} \succeq 0, \end{aligned} \quad (34)$$

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

4.2 Entropy-constrained VaR

The worst-case probability distribution arising in theorem 6, with or without support constraints, is in general discrete [3]. 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

$$KL(P, P_0) := \int \log \frac{dP}{dP_0} dP \leq d \quad (35)$$

where $d \geq 0$ is given, P is the probability distribution of returns, and P_0 is a non-degenerate 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 .)

The proof of the following theorem is in Appendix B.5.

Theorem 9 *When the probability distribution of returns is only known to satisfy the relative entropy constraint (35), 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*

$$V_{\mathcal{P}}(w) = \kappa(\epsilon, d) \|\Gamma^{1/2} w\|_2 - \hat{x}^T w, \quad (36)$$

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

$$\kappa(\epsilon, d) := -\Phi^{-1} f(\epsilon, d), \quad f(\epsilon, d) := \sup_{\lambda > 0} \frac{e^{\epsilon/\lambda - d} - 1}{e^{1/\lambda} - 1} = \sup_{v > 0} \frac{e^{-d}(v+1)^\epsilon - 1}{v}, \quad (37)$$

where Φ is the conditional distribution function of the zero-mean, unit-variance Gaussian 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, since we are then imposing that the distribution of returns is the Gaussian $P_0 = \mathcal{N}(\hat{x}, \Gamma)$. 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.

4.3 Imposing constraints on the covariance matrix

We have seen that the covariance matrix Γ , which is assumed to be exactly known in the classical approaches, becomes an optimization variable here: our approach amounts to selecting the most prudent one in the face of data uncertainty. We may impose additional constraints

on the matrix Γ , if prior information is available. For example, we may impose some bounds on the condition number of Γ , allowing us to rule out near-degenerate distributions if those are deemed unrealistic. A condition number constraint has the form

$$\frac{1}{\mu} \cdot I \preceq \Gamma \preceq \mu \cdot I,$$

where $\mu > 0$ is given. The above is a simple linear matrix inequality constraint on Γ , which can be added to problems such as (2) or (31).

4.4 Multiple VaR constraints

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

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

Such a constraint does not account for what happens to “extreme” cases when the loss exceeds γ . To overcome this difficulty, we may consider multiple VaR constraints

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

where $\epsilon_1 < \dots < \epsilon_m$ are given probability thresholds, and $\gamma_1 < \dots < \gamma_m$ 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 *When the distribution of returns is only known 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 $\Gamma, \hat{x}, x_1, \dots, x_m$ such that*

$$\begin{aligned} x_- \leq \hat{x} \leq x_+, \quad \Gamma_- \leq \Gamma \leq \Gamma_+, \\ x_i^T w \leq -\gamma_i, \quad \begin{bmatrix} \Gamma & x_i - \hat{x} \\ (x_i - \hat{x})^T & \kappa(\epsilon_i)^2 \end{bmatrix} \succeq 0, \quad i = 1, \dots, m. \end{aligned}$$

The above theorem allows to optimize the risk profile by proper choice of the portfolio weights, in several ways: we may minimize an average of the potential losses $\gamma_1 + \dots + \gamma_m$, 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

$$\mathcal{W} = \left\{ w \mid w \geq 0, \sum_{i=1}^n w_i = 1 \right\}.$$

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 in order to obtain diversified portfolios; such constraints are discussed in [11]. 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}^{nom} and Γ^{nom} . With these nominal values, and given a risk level ϵ , we can compute a portfolio, using theorem 3, with $\Gamma_+ = \Gamma_- = \Gamma^{\text{nom}}$ and $\hat{x}_+ = \hat{x}_- = \hat{x}^{\text{nom}}$. We refer to this portfolio—one resulting from the assumption that the data is 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) is 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

$$|\Gamma(i, j) - \Gamma^{\text{nom}}(i, j)| \leq \rho |\Gamma^{\text{nom}}(i, j)|, \quad 1 \leq i, j \leq n.$$

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.

$$|\hat{x} - \hat{x}^{\text{nom}}(i)| \leq 10\rho |\hat{x}^{\text{nom}}(i)|, \quad 1 \leq i \leq n.$$

We have then examined the worst-case behavior of the nominal portfolio as the uncertainty on the point estimates \hat{x}^{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 [22].

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. The curves shown are for $\epsilon = 5\%$, which is the level of probability we want to impose on potential losses.

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

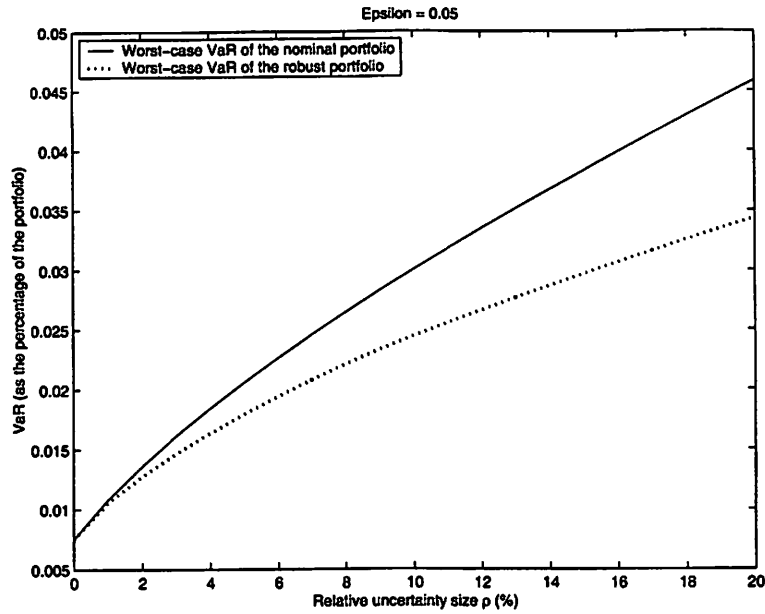


Figure 1: Comparison between the nominal and robust portfolios, as a function of the size of data uncertainty, ρ . Dotted line: worst-case VaR of the nominal portfolio. Solid line: worst-case VaR of the robust portfolio.

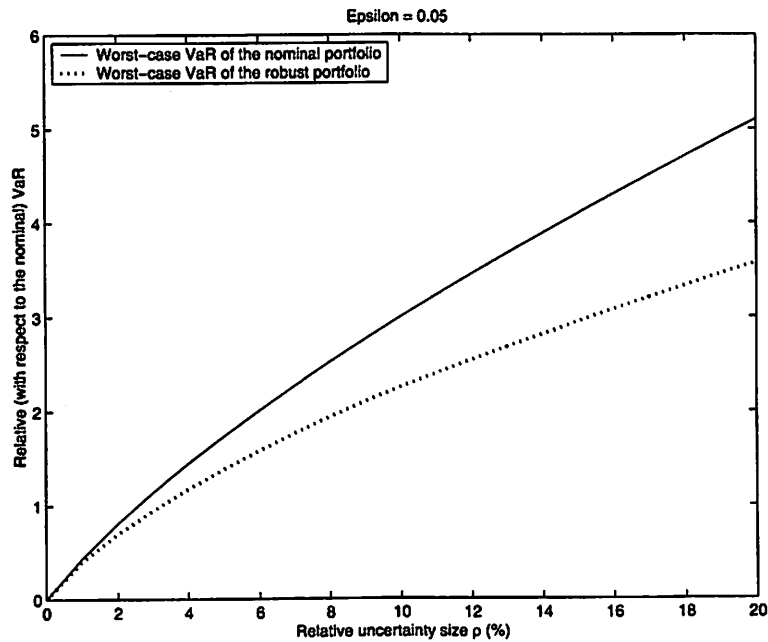


Figure 2: Comparison between the nominal and robust portfolios, as a function of the size of data uncertainty, ρ . Dotted line: deviation of worst-case VaR of the nominal portfolio, relative to the nominal VaR ($\rho = 0$). Solid line: relative deviation of worst-case VaR of the robust portfolio.

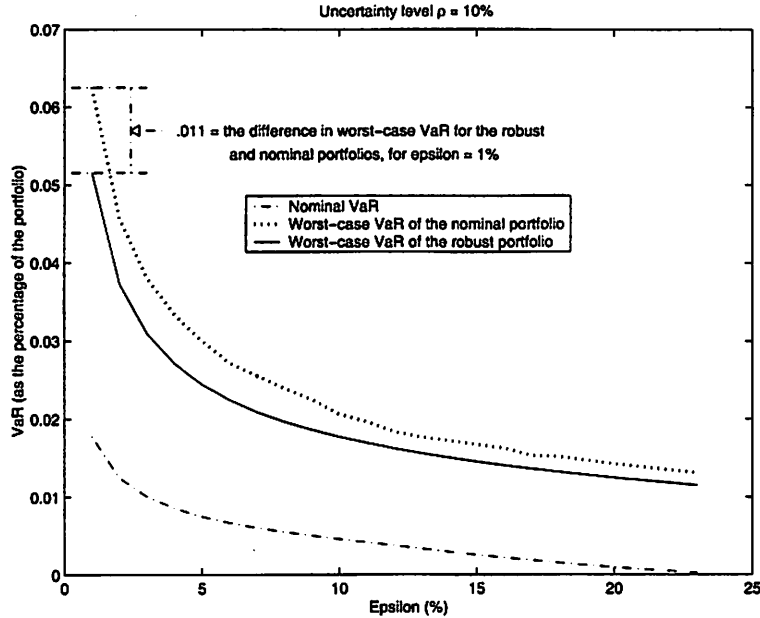


Figure 3: Comparison between the nominal and robust portfolios, as a function of the probability level ϵ . Lower curve: nominal VaR of the nominal portfolio. Middle curve: worst-case VaR of the robust portfolio. Upper curve: worst-case VaR of the nominal portfolio.

improvement brought about by the robust portfolio. The latter is less performant 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.

Note that the benefit of using a robust portfolio increases as the confidence level ϵ decreases: it turns out that the middle curve becomes closer to the lowest curve when ϵ tends to zero. This can seem paradoxical: in a “nominal” context, that is, with error-free data, decreasing ϵ is a way to bring more certainty about “good” events. This is only seemingly true: in our example, when data uncertainty is present, decreasing ϵ has a *negative* effect on the nominal portfolio. This effect is tamed only by taking data uncertainty into account *explicitely*, via robust optimization.

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

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

where $\phi_{\mathcal{W}}$ is the support function of a convex set, that describes the set of admissible portfolio allocation vectors; the set \mathcal{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 Γ, \hat{x} in the above problem are selected to be the most prudent when facing data uncertainty. The portfolio weights (w) are recovered as dual variables; in a sense, they come “for free” if we use a primal-dual interior point method for solving SDP, such as the one described in [23]. The duality between the portfolio weights and the worst-case probability distribution information (\hat{x}, Γ) 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 [15].

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

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

where the shape matrix Γ and center \hat{x} are unknown-but-bounded, and the problem is to best 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 \mathcal{W}, \mathcal{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.

Acknowledgements

We wish to acknowledge the valuable input from Alexandre d’Aspremont, Dimitris Bertsimas, Stephen Boyd, Sanjiv Das, Darrell Duffie, Miguel Lobo, and Ioana Popescu. Finally, we would like to thank Xavier Burtschell and Aniruddha Pant for their help with numerical implementations.

References

- [1] D. Bertsimas and I. Popescu. On the relation between option and stock prices: a convex optimization approach. Sloan working paper #4085, June 1999.
- [2] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: a convex optimization approach. MIT/INSEAD working paper, available at <http://faculty.insead.edu/popescu/ioana>. Submitted to *Operations Research*.
- [3] D. Bertsimas and J. Sethuraman. Moment problems and semidefinite optimization. In R. Saigal, L. Vandenberghe, and H. Wolkowicz, editors, *Semidefinite Programming and Applications*. Kluwer Academic, 2000.

- [4] F. Black and R. Litterman. Global portfolio optimization. *Financial Analysts Journal*, pages 28–43, Sep-Oct 1992.
- [5] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer-Verlag, 2000.
- [6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.
- [7] J.B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Springer-Verlag, 1993. Two volumes.
- [8] K. Isii. On the sharpness of Chebyshev-type inequalities. *Ann. Inst. Stat. Math.*, 14:185–197, 1963.
- [9] T. J. Linsmeier and N. D. Pearson. Risk measurement: An introduction to value at risk. Technical Report 96-04, Office for Futures and Options Research (OFOR) at the University of Illinois, Urbana-Champaign, 1996. Available at <http://ideas.uqam.ca/ideas/data/Papers/wopiluca9604.html>.
- [10] M. S. Lobo and S. Boyd. The worst-case risk of a portfolio. In preparation.
- [11] M. S. Lobo, M. Fazel, and S. Boyd. Portfolio optimization with linear and fixed transaction costs and bounds on risk. Available at <http://www.Stanford.EDU/~boyd/portfolio.html>.
- [12] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Second-order cone programming: interior-point methods and engineering applications. *Linear Algebra & its Applications*, 284:193–228, November 1998.
- [13] D. G. Luenberger. *Investment Science*. Oxford University Press, 1999.
- [14] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.
- [15] M. Musiela and M. Rutkowski. *Martingale methods in financial modelling*, volume 36 of *Applications of Mathematics, stochastic modelling and applied probability*. Springer, 1997.
- [16] Yu. Nesterov and A. Nemirovsky. *Interior point polynomial methods in convex programming: Theory and applications*. SIAM, Philadelphia, PA, 1994.
- [17] N. D. Pearson and X. Ju. Using value-at-risk to control risk taking: How wrong can you be ? *Journal of Risk*, 1(2), 1999.
- [18] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

- [19] R. Saigal, L. Vandenberghe, and H. Wolkowicz, editors. *Handbook on Semidefinite Programming and Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
- [20] J. Smith. Generalized Chebyshev inequalities: theory and applications in decision analysis. *Oper. Research*, 43:807–825, 1995.
- [21] J. Sturm. *SeDuMi: a Matlab toolbox*. available at <http://members.tripod.lycos.nl/SeDuMi>, April 1998. Version 1.03.
- [22] L. Vandenberghe and S. Boyd. *SP version 1.1: Software for semidefinite programming, user's guide*, May 1999. Available via anonymous ftp to [isl.stanford.edu](ftp://isl.stanford.edu) under `/pub/boyd/semidef_prog`.
- [23] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

A Semidefinite and Second-order Cone Programming

A semidefinite programming problem takes the form

$$\text{minimize } c^T z \text{ subject to } F(z) := F_0 + \sum_{i=1}^m z_i F_i \succeq 0, \quad (38)$$

where z is the m -vector of variables, $c \in \mathbf{R}^m$ represents the linear objective, the F_i 's are given $p \times p$ symmetric matrices, and the constraint (called a linear matrix inequality on z) means that $F(z)$ must be positive semidefinite. The main conceptual tool used in this paper is convex duality, in particular, SDP duality. The dual of the above problem is also an SDP in a $p \times p$ symmetric matrix variable Z :

$$\text{maximize } -\text{Tr } F_0 Z \text{ subject to } Z \succeq 0, \quad \text{Tr } F_i Z = c_i, \quad i = 1, \dots, m. \quad (39)$$

When either primal or dual problem are strictly feasible, strong duality holds, and both problems share the same optimal value [23]. If both problems are strictly feasible, then any pair x, Z such that x (resp. Z) is feasible for the primal (resp. dual) problem is optimal if and only if

$$F(x)Z = 0.$$

A second-order cone programming problem is one of the form

$$\begin{aligned} &\text{minimize } c^T x \\ &\text{subject to } \|C_i x + d_i\| \leq e_i^T x + f_i, \quad i = 1, \dots, L, \end{aligned} \quad (40)$$

where $C_i \in \mathbf{R}^{n_i \times m}$, $d_i \in \mathbf{R}^{n_i}$, $e_i \in \mathbf{R}^m$, $f_i \in \mathbf{R}$, $i = 1, \dots, L$. The dual problem of problem (40) is an SOCP as well:

$$\begin{aligned} &\text{maximize } -\sum_{i=1}^L (d_i^T z_i + f_i s_i), \\ &\text{subject to } \sum_{i=1}^L (C_i^T z_i + e_i s_i) = c, \quad \|z_i\| \leq s_i, \quad i = 1, \dots, L, \end{aligned} \quad (41)$$

where $z_i \in \mathbf{R}^{n_i}$, $s_i \in \mathbf{R}$, $i = 1, \dots, L$ are the dual variables. Optimality conditions similar to those for SDPs can be obtained for SOCPs. SOCPs can be expressed as SDPs, therefore they can be solved in polynomial-time using interior-point methods for SDPs. However the SDP formulation is not the most efficient numerically, as special interior-point methods can be devised for SOCPs [16, 12].

Precise complexity results on interior-point methods for SOCPs and SDPs are given by Nesterov and Nemirovsky [16, p.224,236]. In practice, it is observed that the number of iterations is almost constant, independent of problem size [23]. For the SOCP, each iteration has complexity $O((n_1 + \dots + n_L)m^2 + m^3)$; for the SDP, we refer the reader to [16].

B Proofs

B.1 Proof of theorem 1

We first prove the equivalence between propositions 1 and 3, then show that the latter is equivalent to 2. As noted in the comments following the theorem, equivalence between propositions 4, and 5, follows from simple SDP duality, and both are straightforwardly equivalent to the analytical formulation given in proposition 2.

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 $(p, M) \in \mathcal{K}(\mathbf{R}^n) \times \mathcal{S}_{n+1}$

$$\mathcal{L}(p, M) = \int_{\mathbf{R}^n} \chi_S(x) p(x) dx + \langle M, \Sigma - \int_{\mathbf{R}^n} \begin{bmatrix} x \\ 1 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}^T p(x) dx \rangle.$$

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

$$S = \{x \mid \gamma \leq -x^T w\}. \quad (42)$$

Since $\Sigma \succ 0$, strong duality holds (see [8, 3] or [5]). Thus, the original problem is equivalent to its dual. Hence, the worst-case probability is

$$\theta_{\text{wc}} = \inf_{M=M^T} \theta(M), \quad (43)$$

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

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

and $l(x)$ is the quadratic function

$$l(x) = \begin{bmatrix} x \\ 1 \end{bmatrix}^T M \begin{bmatrix} x \\ 1 \end{bmatrix}. \quad (44)$$

We have

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

The dual function is finite if and only if:

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

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

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 $\tau \geq 0$ such that, for every x , $l(x) \geq 1 - 2\tau(\gamma + x^T w)$. Indeed, with condition **C.1** in force, an application of the classical strong duality result for convex programs under the Slater assumption [7] shows that the above condition is sufficient, provided there exist a x_0 such that $\gamma + x_0^T w < 0$, which is the case here since $w \in \mathcal{W}$ and \mathcal{W} does not contain 0.

We obtain that conditions **C.1**, **C.2** are equivalent to

$$\text{There exist a } \tau \geq 0 \text{ such that: } M \succeq 0, \quad M + \begin{bmatrix} 0 & \tau w \\ \tau w^T & -1 + 2\tau\gamma \end{bmatrix} \succeq 0.$$

We obtain that the worst-case probability (43) is the solution to the SDP in variables M, τ :

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

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

$$V_{\mathcal{P}}(w) = \inf \gamma \quad \text{subject to } \langle M, \Sigma \rangle \leq \epsilon, \quad \tau \geq 0, \quad M \succeq 0, \\ M + \begin{bmatrix} 0 & \tau w \\ \tau w^T & -1 + 2\tau\gamma \end{bmatrix} \succeq 0. \quad (45)$$

We now show that the τ -components of optimal solutions (whenever they exist) of (45) are uniformly bounded from below by a positive number. This will allow us to divide by τ in the matrix inequality above, replace $1/\tau$ by τ , and $M/\tau, w/\tau$ by M, w , and obtain the SDP in variables τ, M, γ :

$$V_{\mathcal{P}}(w) = \inf \gamma \quad \text{subject to } \langle M, \Sigma \rangle \leq \tau\epsilon, \quad \tau \geq 0, \quad M \succeq 0, \\ M + \begin{bmatrix} 0 & w \\ w^T & -\tau + 2\gamma \end{bmatrix} \succeq 0. \quad (46)$$

To prove our claim we first note that setting $M_0 = (\epsilon/\text{Tr } \Sigma) \cdot I$, with I the $n \times n$ identity matrix, $\tau_0 = 1$ and γ_0 such that $\gamma_0 + w^T \hat{x} > 0$ and $2\gamma_0 \geq \|w\|^2(\text{Tr } \Sigma)/\epsilon + 1$ yields a triple (M_0, τ_0, γ_0) that is feasible for problem (46). Therefore we can add the constraint $\gamma \leq \gamma_0$ in

(45) without changing its optimal value. When $\gamma \leq \gamma_0$, and M, τ, γ are feasible for (45), the last matrix inequality in (45) implies

$$\begin{aligned} 0 &\leq \langle M, \Sigma \rangle + \langle \Sigma, \begin{bmatrix} 0 & \tau w \\ \tau w^T & -1 + 2\tau\gamma \end{bmatrix} \rangle \\ &\leq 2\tau(w^T \hat{x} + \gamma) + \epsilon - 1 \\ &\leq 2\tau(w^T \hat{x} + \gamma_0) + \epsilon - 1 \end{aligned}$$

Since $w^T \hat{x} + \gamma_0 > 0$, we conclude that optimal τ are uniformly bounded from below, as claimed. Conditions in (46) are exactly those appearing in proposition 3. We have thus shown the equivalence between propositions 1 and 3.

Analytical formula for the worst-case VaR. Finally, we show that the SDP (46) yields the analytical formula (7). We first find the dual, in the sense of SDP duality, of the SDP (46). Define the Lagrangian

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

so that

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

Partition the dual variable Y as follows:

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

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

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

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:

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

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

$$V_{\mathcal{P}}(w) = \max -\frac{v^T w}{y} \quad \text{subject to} \quad \Sigma \succeq \begin{bmatrix} V & v \\ v^T & y \end{bmatrix} \succeq 0, \quad \epsilon \leq y \leq 1. \quad (47)$$

If $y = 1$, we have $v = \hat{x}$ and the objective of the problem is $-\hat{x}^T w$.

Assume now that $y < 1$. In view of our partition of Σ given in (6), the matrix inequality constraints in problem (47) are equivalent to

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

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

$$S \succeq \frac{1}{1-y}(\hat{x} - v)(\hat{x} - v)^T + \frac{1}{y}vv^T, \quad (48)$$

or, equivalently,

$$\Gamma = S - \hat{x}\hat{x}^T \succeq \frac{1}{y(1-y)}(v - y\hat{x})(v - y\hat{x})^T. \quad (49)$$

The dual problem now becomes

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

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

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

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

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

We thus have $y = \epsilon$ at the optimum. This proves the expression of the worst-case VaR given by (7). \diamond

B.2 Proof of lemma 1

We have

$$\begin{aligned} \phi &:= \max_{\|u\|_2 \leq \rho} \|Cu + d\|_2 + e^T u, \\ &= \max_{\|u\|_2 \leq \rho, \|v\|_2 \leq 1} v^T (Cu + d) + e^T u. \end{aligned}$$

Defining

$$M := \begin{bmatrix} 0 & C & d \\ C^T & 0 & e \\ d^T & e^T & 0 \end{bmatrix},$$

we can express ϕ as

$$2\phi = \max \begin{bmatrix} v \\ u \\ 1 \end{bmatrix}^T M \begin{bmatrix} v \\ u \\ 1 \end{bmatrix} : \|u\|_2 \leq \rho, \quad \|v\|_2 \leq 1,$$

Thus, we can express ϕ via an SDP with a rank constraint:

$$2\phi = \max\langle M, X \rangle : X = \begin{bmatrix} X_{11} & X_{12} & x_1 \\ X_{12}^T & X_{22} & x_2 \\ x_1^T & x_2^T & 1 \end{bmatrix} \succeq 0, \quad (50)$$

$$\text{Tr } X_{11} \leq 1, \quad \text{Tr } X_{22} \leq \rho^2, \quad \text{rank } X = 1,$$

where X_{11} is $N \times N$, X_{22} is $l \times l$.

Now consider the following upper bound on 2ϕ , obtained by relaxing the rank constraint:

$$\psi = \max\langle M, X \rangle : X = \begin{bmatrix} X_{11} & X_{12} & x_1 \\ X_{12}^T & X_{22} & x_2 \\ x_1^T & x_2^T & 1 \end{bmatrix} \succeq 0, \quad \text{Tr } X_{11} \leq 1, \quad \text{Tr } X_{22} \leq \rho^2. \quad (51)$$

Problem (51) is an SDP, the dual of which is

$$\psi' = \min \lambda_1 + \rho^2 \lambda_2 + \lambda_3 : M(\lambda) := \begin{bmatrix} \lambda_1 I_N & C & d \\ C^T & \lambda_2 I_l & e \\ d^T & e^T & \lambda_3 \end{bmatrix} \succeq 0. \quad (52)$$

We obviously have $2\phi \leq \psi \leq \psi'$. Since both SDPs (51) and (52) are strictly feasible, therefore strong duality holds. Thus, $\psi' = \psi$ is an upper bound on 2ϕ . \diamond

B.3 Proof of theorem 5

Defining

$$d(w) = \begin{bmatrix} S^{1/2} A^T \\ D^{1/2} \end{bmatrix} w, \quad C = \begin{bmatrix} S^{1/2} R^T \\ 0 \end{bmatrix}, \quad e = -R\hat{f}, \quad r(w) = L^T w,$$

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

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

where the last inequality is derived from lemma 1, with $\rho = \|r(w)\|_2$. Using Schur complements [6], we may rewrite the linear matrix inequality in the last line as

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

where $\mu = 1/\lambda_2$. Introducing a slack variable $t \geq \lambda_2 \|r(w)\|_2^2 = \|r(w)\|_2^2/\mu$, we can rewrite the objective as $\lambda_1 + t + \lambda_3$, where t is such that $t \geq \|r(w)\|_2^2/\mu$. 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.) \diamond

B.4 Proof of theorem 6

The proof proceeds in the same three steps as in the proof of theorem B.1. Again, we denote by Σ the second-moment matrix defined in (6). From the assumption $\Gamma \succ 0$, we have $\Sigma \succ 0$.

Computing the worst-case probability. We begin with the problem of computing the worst-case probability. We introduce the Lagrange functional: for $(p, \Lambda) \in \mathcal{K}(\Omega) \times \mathcal{S}_{n+1}$

$$\mathcal{L}(p, \Lambda) = \int_{\Omega} \chi_S(x) p(x) dx + \langle \Lambda, \Sigma - \int_{\Omega} \begin{bmatrix} x \\ 1 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}^T p(x) dx \rangle.$$

Again, since $\Sigma \succ 0$, strong duality holds, and the original problem is equivalent to its dual. Hence, the worst-case probability is

$$\theta_{\mathcal{P}} = \inf_{\Lambda \in \mathcal{S}_{n+1}} \theta(\Lambda),$$

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

$$\theta(\Lambda) = \sup_{p \in \mathcal{K}(\Omega)} \mathcal{L}(p, \Lambda) = \langle \Lambda, \Sigma \rangle + \sup_{p \in \mathcal{K}(\Omega)} \int_{\Omega} (\chi_S(x) - l(x)) p(x) dx,$$

where $l(x)$ is the quadratic function defined in (44). We have

$$\theta(\Lambda) = \begin{cases} \langle \Lambda, \Sigma \rangle & \text{if } \chi_S(x) \leq l(x) \text{ for every } x \in \Omega, \\ +\infty & \text{otherwise.} \end{cases}$$

First, we examine the condition **C.1**:

$$l(x) \geq 0 \text{ for every } x, x_l \leq x \leq x_u.$$

The condition is true if, there exist two vectors $\lambda_u \geq 0$, $\lambda_l \geq 0$ such that

$$\text{for every } x, l(x) \geq 2\lambda_l^T(x - x_l) + 2\lambda_u^T(x_u - x).$$

In turn, the above is true if and only if

$$\Lambda \succeq \begin{bmatrix} 0 & \lambda_l - \lambda_u \\ (\lambda_l - \lambda_u)^T & 2(\lambda_u^T x_u - \lambda_l^T x_l) \end{bmatrix}. \quad (53)$$

Note that, in the case when no support information is available, the above condition is exact, provided we set $\lambda_u = \lambda_l = 0$. Indeed, condition **C.1** then simply says that $l(x) \geq 0$ for every $x \in \mathbf{R}^n$, while (53) reduces to $\Lambda \succeq 0$.

Similarly, the condition **C.2**:

$$l(x) \geq 1 \text{ for every } x \in \Omega, x^T w + \gamma \leq 0$$

holds if there exist a vector $\nu \geq 0$ and a scalar $\tau \geq 0$ such that

$$\text{for every } x, l(x) \geq 1 - 2\tau(\gamma + x^T w) + 2\nu_l^T(x - x_l) + 2\nu_u^T(x_u - x).$$

In turn, the above is true if and only if

$$\Lambda \succeq \begin{bmatrix} 0 & -\tau w \\ -\tau w^T & 1 - 2\tau\gamma \end{bmatrix} + \begin{bmatrix} 0 & \nu_l - \nu_u \\ (\nu_l - \nu_u)^T & 2(\nu_u^T x_u - \nu_l^T x_l) \end{bmatrix}. \quad (54)$$

Again, in the case when no support information is available, the above condition is exact, provided we set $\nu_u = \nu_l = 0$. Indeed, in this case C.2 reduces to the condition: $l(x) \geq 1$ for every $x \in \mathbf{R}^n$ such that $\gamma + x^T w \leq 0$. This holds if *and only if* there exist a scalar $\tau \geq 0$ such that

$$\text{For every } x \ l(x) \geq 1 - 2\tau(\gamma + x^T w).$$

This fact, with condition C.1 in force, is an application of the classical strong duality result for convex programs under the Slater assumption [7], which shows that the above condition is sufficient, provided there exist a x_0 such that $\gamma + x_0^T w < 0$, which is the case here since $w \in \mathcal{W}$ and \mathcal{W} does not contain 0. The above condition is the same as (54) with $\nu_u = \nu_l = 0$.

To summarize, an upper bound on the dual problem

$$\inf_{\Lambda \in \mathcal{S}_{n+1}} \theta(\Lambda)$$

is obtained by solving the SDP in variables Λ, τ, μ, ν :

$$\begin{aligned} \inf \langle \Lambda, \Sigma \rangle \quad & \text{subject to} \\ & \tau \geq 0, \quad \lambda_u \geq 0, \quad \lambda_l \geq 0, \quad \nu_u \geq 0, \quad \nu_l \geq 0, \\ & \Lambda \succeq \begin{bmatrix} 0 & \lambda_l - \lambda_u \\ (\lambda_l - \lambda_u)^T & 2(\lambda_u^T x_u - \lambda_l^T x_l) \end{bmatrix}, \\ & \Lambda \succeq \begin{bmatrix} 0 & -\tau w \\ -\tau w^T & 1 - 2\tau\gamma \end{bmatrix} + \begin{bmatrix} 0 & \nu_l - \nu_u \\ (\nu_l - \nu_u)^T & 2(\nu_u^T x_u - \nu_l^T x_l) \end{bmatrix}. \end{aligned}$$

The above SDP results in a sufficient condition (expressed in an SDP form) ensuring that the worst-case probability is below a given level ϵ . This condition is exact when no support information is available, if we set $\lambda_u = \lambda_l = 0$ and $\nu_u = \nu_l = 0$.

Computing the worst-case VaR as an SDP. We obtain that (an upper bound on) the worst-case VaR can be computed as

$$\begin{aligned} V_{\mathcal{P}}(w) = \inf \gamma \quad & \text{subject to} \\ & \langle \Lambda, \Sigma \rangle \leq \epsilon, \quad \tau \geq 0, \quad \lambda_u \geq 0, \quad \lambda_l \geq 0, \quad \nu_u \geq 0, \quad \nu_l \geq 0, \\ & \Lambda \succeq \begin{bmatrix} 0 & \lambda_l - \lambda_u \\ (\lambda_l - \lambda_u)^T & 2(\lambda_u^T x_u - \lambda_l^T x_l) \end{bmatrix}, \\ & \Lambda \succeq \begin{bmatrix} 0 & -\tau w \\ -\tau w^T & 1 - 2\tau\gamma \end{bmatrix} + \begin{bmatrix} 0 & \nu_l - \nu_u \\ (\nu_l - \nu_u)^T & 2(\nu_u^T x_u - \nu_l^T x_l) \end{bmatrix}. \end{aligned}$$

We can assume τ to be uniformly bounded away from 0. To see this, we first choose a set of variables $(\Lambda_0, \tau_0, \gamma_0)$ and $\lambda_{u,l} = \nu_{u,l} = 0$ that is feasible for the above problem, and such

that $\gamma_0 + w^T \hat{x} > 0$. When $\gamma \leq \gamma_0$, and $\Lambda, \tau, \gamma, \lambda_{u,l}, \mu_{u,l}$ are feasible for the above problem, we have

$$\begin{aligned} 0 &\leq \langle \Lambda, \Sigma \rangle + \langle \Sigma, \begin{bmatrix} 0 & \tau w - \nu_l + \nu_u \\ (\tau w - \nu_l + \nu_u)^T & -1 + 2\tau\gamma - 2(\nu_u^T x_u - \nu_l^T x_l) \end{bmatrix} \rangle \\ &\leq 2\tau(w^T \hat{x} + \gamma) + \epsilon - 1 + 2(\nu_u^T(\hat{x} - x_u) + \nu_l^T(x_l - \hat{x})) \\ &\leq 2\tau(w^T \hat{x} + \gamma_0) + \epsilon - 1, \end{aligned}$$

in view of our assumption that $x_l \leq x_- = \hat{x} = x_+ \leq x_u$. Since $w^T \hat{x} + \gamma_0 > 0$, we conclude that τ is uniformly bounded, as claimed.

Dividing by τ , we obtain the SDP in variables $\tau, \lambda_u, \lambda_l, \nu_u, \nu_l, \Lambda, \gamma$

$$\begin{aligned} V_{\mathcal{P}}(w) \leq V_{\text{wc-up}}(w) := \inf \gamma \quad \text{subject to} \\ \langle \Lambda, \Sigma \rangle \leq \tau\epsilon, \quad \tau \geq 0, \quad \lambda_u \geq 0, \quad \lambda_l \geq 0, \quad \nu_u \geq 0, \quad \nu_l \geq 0, \\ \Lambda \succeq \begin{bmatrix} 0 & \lambda_l - \lambda_u \\ (\lambda_l - \lambda_u)^T & 2(\lambda_u^T x_u - \lambda_l^T x_l) \end{bmatrix}, \\ \Lambda \succeq \begin{bmatrix} 0 & -w \\ -w^T & \tau - 2\gamma \end{bmatrix} + \begin{bmatrix} 0 & \nu_l - \nu_u \\ (\nu_l - \nu_u)^T & 2(\nu_u^T x_u - \nu_l^T x_l) \end{bmatrix}. \end{aligned} \quad (55)$$

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. When no support information is used, all components of $\lambda_{u,l}, \nu_{u,l}$ are zero, and the upper bound is equal to the worst-case VaR.

Dual form and analytical expression. Finally, we show that the SDP (55) can be equivalently expressed in the form (31), and reduces to the analytical formula (7) when no support information is used.

We first find the dual, in the sense of SDP duality, of the SDP (55). Define the Lagrangian

$$\begin{aligned} \mathcal{L}(\Lambda, \lambda_{u,l}, \nu_{u,l}, \tau, \gamma, \alpha, \mu, p_{u,l}, q_{u,l}, X, Y) = \\ \gamma - \alpha(\tau\epsilon - \langle \Lambda, \Sigma \rangle) - \mu\tau - p_l^T \lambda_l - p_u^T \lambda_u - q_l^T \nu_l - q_u^T \nu_u \\ - \langle X, \Lambda - \begin{bmatrix} 0 & \lambda_l - \lambda_u \\ (\lambda_l - \lambda_u)^T & 2(\lambda_u^T x_u - \lambda_l^T x_l) \end{bmatrix} \rangle \\ - \langle Y, \Lambda - \begin{bmatrix} 0 & -w \\ -w^T & \tau - 2\gamma \end{bmatrix} + \begin{bmatrix} 0 & \nu_l - \nu_u \\ (\nu_l - \nu_u)^T & 2(\nu_u^T x_u - \nu_l^T x_l) \end{bmatrix} \rangle, \end{aligned}$$

so that

$$V_{\text{wc-up}}(w) = \inf_{\Lambda = \Lambda^T, \lambda_{u,l}, \nu_{u,l}, \tau, \gamma} \sup_{\alpha \geq 0, \mu \geq 0, p_{u,l} \geq 0, q_{u,l} \geq 0, X \succeq 0, Y \succeq 0} \mathcal{L}$$

Partition the dual variables X, Y as follows:

$$X = \begin{bmatrix} X_{11} & x_{12} \\ x_{12}^T & x_{22} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_{11} & y_{12} \\ y_{12}^T & y_{22} \end{bmatrix}.$$

We obtain the dual problem in variables X, Y, α :

$$\begin{aligned} \sup -2y_{12}^T w \quad \text{subject to} \\ y_{22} = 1/2, \quad \alpha \epsilon \leq y_{22}, \quad \alpha \geq 0, \\ \alpha \Sigma = X + Y, \\ X = \begin{bmatrix} X_{11} & x_{12} \\ x_{12}^T & x_{22} \end{bmatrix} \succeq 0, \quad Y = \begin{bmatrix} Y_{11} & y_{12} \\ y_{12}^T & y_{22} \end{bmatrix} \succeq 0, \\ x_l x_{22} \leq x_{12} \leq x_u x_{22}, \quad x_l y_{22} \leq y_{12} \leq x_u y_{22}. \end{aligned}$$

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 variable X yields:

$$\begin{aligned} V_{\mathcal{P}}(w) = \max -2y_{12}^T w \quad \text{subject to} \\ 0 \leq \alpha \leq \frac{1}{2\epsilon}, \quad \alpha \Sigma \succeq Y = \begin{bmatrix} Y_{11} & y_{12} \\ y_{12}^T & y_{22} \end{bmatrix} \succeq 0, \\ x_l \leq 2y_{12} \leq 2x_u, \quad (2\alpha - 1)x_l \leq 2(\alpha \hat{x} - y_{12}) \leq (2\alpha - 1)x_u. \end{aligned}$$

Note that the constraint on Y implies $\alpha \geq 1/2 > 0$. Therefore, we make the change of variables $(Y_{11}, y_{12}, \alpha) \rightarrow (V, v, y)$ with $V = Y_{11}/\alpha$, $v = y_{12}/\alpha$, $y = 1/2\alpha \in [\epsilon, 1]$. We obtain

$$\begin{aligned} V_{\mathcal{P}}(w) = \max -\frac{v^T w}{y} \quad \text{subject to} \quad \Sigma \succeq \begin{bmatrix} V & v \\ v^T & y \end{bmatrix} \succeq 0, \quad \epsilon \leq y \leq 1, \\ (1-y)x_l \leq \hat{x} - v \leq (1-y)x_u, \quad yx_l \leq v \leq yx_u. \end{aligned} \quad (56)$$

If $y = 1$, we have $v = \hat{x}$ and the objective of the problem is $-\hat{x}^T w$.

Assume now that $y < 1$. In view of our partition of Σ given in (6), the matrix inequality constraints in problem (56) are equivalent to

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

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

$$S \succeq \frac{1}{1-y}(\hat{x} - v)(\hat{x} - v)^T + \frac{1}{y}vv^T, \quad (57)$$

or, equivalently,

$$\Gamma = S - \hat{x}\hat{x}^T \succeq \frac{1}{y(1-y)}(v - y\hat{x})(v - y\hat{x})^T. \quad (58)$$

The dual problem now becomes

$$\begin{aligned} V_{\mathcal{P}}(w) = \max_{v,y} -\frac{v^T w}{y} \quad \text{subject to} \quad S - \hat{x}\hat{x}^T \succeq \frac{1}{y(1-y)}(v - y\hat{x})(v - y\hat{x})^T, \quad y \in [\epsilon, 1], \\ (1-y)x_l \leq \hat{x} - v \leq (1-y)x_u, \quad yx_l \leq v \leq yx_u. \end{aligned}$$

Introduce the new variables $x = v/y$, $\kappa^2 = 1/y - 1 \in [0, \kappa(\epsilon)^2]$, where $\kappa(\epsilon)$ is defined in (8). Since $\Gamma = S - \hat{x}\hat{x}^T$, using Schur complements we can rewrite the above as

$$V_{\mathcal{P}}(w) = \max_{x, \kappa} -x^T w \text{ subject to } \begin{bmatrix} \Gamma & x - \hat{x} \\ (x - \hat{x})^T & \kappa^2 \end{bmatrix} \succeq 0, \quad \kappa^2 \in [0, \kappa(\epsilon)^2], \quad (59)$$

$$\kappa^2 x_l \leq \hat{x} - x \leq \kappa^2 x_u, \quad x_l \leq x \leq x_u.$$

The above problem is strictly feasible and its feasible set is bounded. We can form the dual, in the SDP sense, to this problem, and there will be no duality gap. The dual is

$$\begin{aligned} \min \quad & \langle \Lambda, \Gamma \rangle + \kappa(\epsilon)^2 v - \hat{x}^T w \\ & + \kappa(\epsilon)^2 (x_u^T \lambda_u - x_l^T \lambda_l) + \nu_u^T (x_u - \hat{x}) - \nu_l^T (x_l - \hat{x}) + t \kappa(\epsilon)^2 \\ \text{subject to } \quad & \begin{bmatrix} \Lambda & (w + \lambda_l - \lambda_u + \nu_u - \nu_l)/2 \\ (w + \lambda_l - \lambda_u + \nu_u - \nu_l)^T/2 & v \end{bmatrix} \succeq 0, \quad t \geq 0. \end{aligned}$$

At the optimum, we have $t = 0$. The problem thus obtained is exactly the one we would have obtained by setting $\kappa = \kappa(\epsilon)$ in (59). This proves the formulation (31).

If we let $x_l = -\infty$ and $x_u = +\infty$, the last inequalities in the above problem become void, and the VaR is exact. Its expression reduces to the one with no support constraints given in (7). \diamond

B.5 Proof of theorem 9

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

$$\text{maximize } \int_{\mathbf{R}^n} \chi_{\mathcal{S}}(x) p(x) dx \text{ subject to } KL(p, p_0) \leq d, \quad \int_{\mathbf{R}^n} p(x) dx = 1, \quad (60)$$

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 $\lambda_0 \geq 0$, 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 as before, $\chi_{\mathcal{S}}$ is the indicator function of the set \mathcal{S} defined in (42). 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 [20]. 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

$$p(x) = p_0(x) \exp \left(\frac{\chi_{\mathcal{S}}(x) - \lambda_0}{\lambda} - 1 \right), \quad (61)$$

and the dual function expresses as

$$\begin{aligned}
\theta(\lambda_0, \lambda) &= \lambda_0 + \lambda d + \lambda \int p_0(x) \exp\left(\frac{\chi_S(x) - \lambda_0}{\lambda} - 1\right) dx \\
&= \lambda_0 + \lambda d + \lambda \left(e^{-\frac{\lambda_0}{\lambda} - 1} \mathbf{Prob}\{\gamma \leq -x^T w\} + e^{-\frac{\lambda_0}{\lambda} - 1} \mathbf{Prob}\{\gamma \geq -x^T w\} \right) \\
&= \lambda_0 + \lambda d + \lambda e^{-\frac{\lambda_0}{\lambda} - 1} \left((e^{1/\lambda} - 1) \phi(\gamma) + 1 \right),
\end{aligned}$$

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

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

Taking the infimum over λ_0 yields

$$\inf_{\lambda_0 \in \mathbf{R}} \theta(\lambda_0, \lambda) = \lambda d + \lambda \log\left((e^{1/\lambda} - 1) \phi(\gamma) + 1\right). \quad (62)$$

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

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

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

that is,

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

where $\kappa(\epsilon, d)$ is defined in the theorem. \diamond