Robust portfolio selection problems*

D. Goldfarb [†] G. Iyengar [‡]

December 26, 2001

Abstract

In this paper we show how to formulate and solve robust portfolio selection problems. The objective of these robust formulations is to systematically combat the sensitivity of the optimal portfolio to statistical and modeling errors in the estimates of the relevant market parameters. We introduce "uncertainty structures" for the market parameters and show that the robust portfolio selection problems corresponding to these uncertainty structures can be reformulated as second-order cone programs and, therefore, the computational effort required to solve them is comparable to that required for solving convex quadratic programs. Moreover, we show that these uncertainty structures correspond to confidence regions associated with the statistical procedures used to estimate the market parameters. We demonstrate a simple recipe for efficiently computing robust portfolios given raw market data and a desired level of confidence.

Keywords: Robust optimization, Mean-variance portfolio selection, Value-at-risk portfolio selection, Second-order cone programming, Linear regression.

*Submitted to *Mathematics of Operations Research*. Do not distribute

[†]IEOR Department, Columbia University, Email: `gold@ieor.columbia.edu`. Research partially supported by DOE grant GE-FG01-92ER-25126, and NSF grants DMS-94-14438, CDA-97-26385 and DMS-01-04282

[‡]IEOR Department, Columbia University, Email: `garud@ieor.columbia.edu`. Research partially supported by NSF grants CCR-00-09972 and DMS-01-04282.

1 Introduction

Portfolio selection is the problem of allocating capital over a number of available assets in order to maximize the "return" on the investment while minimizing the "risk". Although the benefits of diversification in reducing risk have been appreciated since the inception of financial markets, the first mathematical model for portfolio selection was formulated by Markowitz [19, 20]. In the Markowitz portfolio selection model, the "return" of a portfolio is measured by the expected value of the random portfolio return, and the associated "risk" is quantified by the variance of the portfolio return. Markowitz showed that, given either an upper bound on the risk that the investor is willing to take or a lower bound on the return the investor is willing to accept, the optimal portfolio can be obtained by solving a convex quadratic programming problem. This mean-variance model has had a profound impact on the economic modeling of financial markets and the pricing of assets – the Capital Asset Pricing Model (CAPM) developed primarily by Sharpe [26], Lintner [17] and Mossin [23] was an immediate logical consequence of the Markowitz mean-variance portfolio theory. In 1990, Sharpe and Markowitz shared the Nobel Memorial Prize in Economic Sciences for their work on portfolio allocation and asset pricing.

In spite of the theoretical success of the mean-variance model, practitioners have shied away from this model. The following quote from [22], a well regarded text among practitioners, summarizes the problem: "Although Markowitz efficiency is a convenient and useful theoretical framework for portfolio optimality, in practice it is an error prone procedure that often results in *error-maximized* and *investment-irrelevant* portfolios." This behavior is a reflection of the fact that solutions of optimization problems are often very sensitive to perturbations in the parameters of the problem; since the estimates of the market parameters are subject to statistical errors, the results of the subsequent optimization are not very reliable. Currently, practitioners attempt to mitigate the effect of parameter uncertainty by resampling the mean returns $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of the assets from a confidence region around a nominal set of parameters, and then aggregating the portfolios obtained by solving a Markowitz problem for each sample [22]. Like any sampling based method, this technique does not provide any hard guarantees and becomes very inefficient as the number of assets grows.

In this paper we propose alternative deterministic models that are robust to parameter uncertainty and estimation errors. In this framework, the perturbations in the market parameters are modeled as unknown, but bounded, and optimization problems are solved assuming worst case behavior of these perturbations. This robust optimization framework was introduced by Nemirovski and Ben-Tal for linear programming [5] and general convex programming [4] (see also [6]). There is also a parallel literature on robust formulations of optimization problems originating from robust control [9, 11, 10].

Our contributions in this paper are as follows:

(a) We develop a robust factor model for the asset returns. In this model the vector of random asset returns $\mathbf{r} \in \mathbf{R}^n$ is given by

$$\mathbf{r} = \boldsymbol{\mu} + \mathbf{V}^T\mathbf{f} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\mu} \in \mathbf{R}^n$ is the vector of mean returns, $\mathbf{f} \in \mathbf{R}^m$ is the vector of random returns of the $m(< n)$ factors that drive the market, $\mathbf{V} \in \mathbf{R}^{m \times n}$ is the factor loading matrix and $\boldsymbol{\epsilon}$ is the vector of residual returns. The mean return vector $\boldsymbol{\mu}$, the factor loading matrix $\mathbf{V}$, and the covariance matrices of the factor return vector $\mathbf{f}$ and the residual error vector $\boldsymbol{\epsilon}$ are known to lie within suitably defined uncertainty sets. For this market model, we formulate robust analogs of classical mean-variance and value-at-risk portfolio selection problems.

(b) We show that the natural uncertainty sets for the market parameters are defined by the statistical procedures employed to estimate these parameters from market return data. This class of uncertainty sets is completely parametrized by the market data and a parameter ω that controls the confidence level; thereby allowing one to provide probabilistic guarantees on the performance of the robust portfolios. In all previous work on robust optimization [4, 5, 9, 11, 13], the uncertainty sets for the parameters are assumed to have a certain structure without any explicit justification. Also, there is no discussion of how these sets are parametrized from raw data.

(c) We show that the robust optimization problems corresponding to the natural class of uncertainty sets (defined by the estimation procedures) can be reformulated as second-order cone programs (SOCPs). SOCPs can be solved very efficiently using interior point algorithms [24, 18, 27]. In fact, both the worst case and practical computational effort required to solve an SOCP is comparable to that for solving a convex quadratic program of similar size and structure, i.e. in practice the computational effort required to solve these robust portfolio selection problems is comparable to that required to solve the classical Markowitz mean-variance portfolio selection problems.

In a recent related paper [13], it was shown that if the uncertain mean return vector $\boldsymbol{\mu}$ and the uncertain covariance matrix $\boldsymbol{\Sigma}$ of the asset returns $\mathbf{r}$ belong to the component-wise uncertainty sets $S_m = \{\boldsymbol{\mu} : \boldsymbol{\mu}^L \leq \boldsymbol{\mu} \leq \boldsymbol{\mu}^U\}$ and $S_v = \{\boldsymbol{\Sigma} : \boldsymbol{\Sigma} \succeq 0, \boldsymbol{\Sigma}^L \leq \boldsymbol{\Sigma} \leq \boldsymbol{\Sigma}^U\}$ respectively, the robust problem reduces to a nonlinear saddle-point problem that involves semidefinite constraints. Here $\mathbf{A} \succeq 0$ (resp. $\succ 0$) denotes that the matrix $\mathbf{A}$ is symmetric and positive semidefinite (resp. definite). This approach has several shortcomings when applied to practical problems - the model is not a factor model and does not allow for a rank deficient covariance matrix (which is most often the case in practice), no procedure is provided for specifying the extreme values $(\boldsymbol{\mu}^L, \boldsymbol{\mu}^U)$ and $(\boldsymbol{\Sigma}^L, \boldsymbol{\Sigma}^U)$ defining the uncertainty structure and, moreover, the solution algorithm, although polynomial, is not practicable when the number of assets is large. A multi-period robust model, where the uncertainty sets are finite sets, was proposed in [3].

The organization of the paper is as follows. In Section 2 we introduce the robust factor model and uncertainty sets for the mean return vector, the factor loading matrix, and the covariance matrix of the residual return. We also formulate robust counterparts of the mean-variance optimal portfolio selection problem [19], the maximum Sharpe ratio portfolio selection problem [26] and the value-at-risk (VaR) portfolio selection problem. The uncertainty sets introduced in this section are ellipsoidal (intervals in the one dimensional case), and may appear quite arbitrary. Before demonstrating that these sets are, indeed, natural, we first establish in Section 3 that the robust mean-variance portfolio selection problem in markets where the factor loading and mean returns are uncertain, but the factor covariance is known and fixed, can be reformulated as an SOCP. An SOCP formulation of the robust maximum Sharpe ratio problem in such markets follows as a corollary. In Section 4 we develop an SOCP reformulation for the robust VaR portfolio selection problem. In Section 5 we justify the uncertainty sets introduced in Section 2 by relating them to linear regression. Specifically, we show that the uncertainty sets correspond to confidence regions around the least-squares estimate of the market parameters and can be constructed to reflect any desired confidence level. In this section, we collect the results from the preceding sections and present a recipe for robust portfolio allocation that closely parallels the classical one. In Section 6 we improve the factor model by allowing uncertainty in the factor covariance matrix and show that, for natural classes of uncertainty sets, all robust portfolio allocation problems continue to be SOCPs. We also show that these natural classes of uncertainty sets correspond to the confidence regions associated with maximum likelihood estimation of the covariance matrix. In Section 7 we present results of some of preliminary computational experiments with our robust portfolio allocation framework.

2 Market model and robust investment problems

We assume that the market opens for trading at discrete instants in time and has n traded assets. The vector of asset returns over a single market period is denoted by $\mathbf{r} \in \mathbf{R}^n$, with the interpretation that asset i returns $(1 + r_i)$ dollars for every dollar invested in it. The returns on the assets in different market periods are assumed to be independent. The single period return $\mathbf{r}$ is assumed to be a random variable given by

$$\mathbf{r} = \boldsymbol{\mu} + \mathbf{V}^T\mathbf{f} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\mu} \in \mathbf{R}^n$ is the vector of mean returns, $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{F}) \in \mathbf{R}^m$ is the vector of returns of the factors that drive the market, $\mathbf{V} \in \mathbf{R}^{m \times n}$ is the matrix of factor loadings of the n assets and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$ is the vector of residual returns. Here $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes that $\mathbf{x}$ is a multivariate Normal random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

In addition, we assume that the vector of residual returns $\boldsymbol{\epsilon}$ is independent of the vector of factor returns $\mathbf{f}$, the covariance matrix $\mathbf{F} \succ \mathbf{0}$ and the covariance martix $\mathbf{D} = \mathbf{diag}(\mathbf{d}) \succ \mathbf{0}$, i.e. $d_i > 0$, $i = 1, \ldots, n$. Thus, the vector of asset returns $\mathbf{r} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V}^T\mathbf{F}\mathbf{V} + \mathbf{D})$.

Except in Section 6, the covariance matrix $\mathbf{F}$ of the factor returns $\mathbf{f}$ is assumed to be stable and known exactly. The individual diagonal elements d_i of the covariance matrix $\mathbf{D}$ are assumed to lie in an interval $[\underline{d}_i, \overline{d}_i]$, i.e. the uncertainty set S_d for the matrix $\mathbf{D}$ is given by

$$S_d = \{\mathbf{D} : \mathbf{D} = \mathbf{diag}(\mathbf{d}), d_i \in [\underline{d}_i, \overline{d}_i], i = 1, \ldots, n\}. \tag{2}$$

The columns of the matrix $\mathbf{V}$, i.e. the factor loadings of the individual assets, are also assumed to be known approximately. In particular, $\mathbf{V}$ belongs to the elliptical uncertainty set S_v given by

$$S_v = \left\{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}_i\|_g \leq \rho_i, i = 1, \ldots, n\right\}, \tag{3}$$

where $\mathbf{W}_i$ is the i-th column of $\mathbf{W}$ and $\|\mathbf{w}\|_g = \sqrt{\mathbf{w}^T \mathbf{G} \mathbf{w}}$ denotes the elliptic norm of $\mathbf{w}$ with respect to a symmetric, positive definite matrix $\mathbf{G}$.

The mean returns vector $\boldsymbol{\mu}$ is assumed to lie in the uncertainty set set S_m given by

$$S_m = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\xi}, |\xi_i| \leq \gamma_i, i = 1, \ldots, n\}; \tag{4}$$

i.e. each component of $\boldsymbol{\mu}$ is assumed to lie within a certain interval. The choice of the uncertainty sets is motivated by the fact that the factor loadings and the mean returns of assets are estimated by linear regression. The justification of the uncertainty structures and suitable choices for the matrix $\mathbf{G}$, and the bounds ρ_i, γ_i, $\overline{d}_i$, $\underline{d}_i$, $i = 1, \ldots, n$, are discussed in Section 5.

An investor's position in this market is described by a portfolio $\boldsymbol{\phi} \in \mathbf{R}^n$, where the i-th component ϕ_i represents the fraction of total wealth invested in asset i. The return r_ϕ on the portfolio $\boldsymbol{\phi}$ is given by

$$r_\phi = \mathbf{r}^T \boldsymbol{\phi} = \boldsymbol{\mu}^T \boldsymbol{\phi} + \mathbf{f}^T \mathbf{V} \boldsymbol{\phi} + \boldsymbol{\epsilon}^T \boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\phi}^T \boldsymbol{\mu}, \boldsymbol{\phi}^T (\mathbf{V}^T \mathbf{F} \mathbf{V} + \mathbf{D}) \boldsymbol{\phi}).$$

The objective of the investor is to choose a portfolio that maximizes the "return" on the investment subject to some constraints on the "risk" of the investment. The mathematical model for portfolio selection proposed by Markowitz [19, 20] assumes that the expected value $\mathbf{E}[\mathbf{r}]$ of the asset returns and the covariance $\mathbf{Var}[\mathbf{r}]$ are known with certainty. In this model investment "return" is the expected value $\mathbf{E}[r_\phi]$ of the portfolio return and the associated "risk" is the variance $\mathbf{Var}[\mathbf{r}_\phi]$. The objective of the investor is to choose a portfolio $\boldsymbol{\phi}^*$ that has the minimum variance amongst those that have expected return at least α, i.e. $\boldsymbol{\phi}^*$ is the optimal solution of the convex quadratic optimization problem

$$\begin{array}{rl} \text{minimize} & \mathbf{Var}[\mathbf{r}_\phi] \\ \text{subject to} & \mathbf{E}[r_\phi] \geq \alpha, \\ & \mathbf{1}^T \boldsymbol{\phi} = 1. \end{array} \tag{5}$$

As was pointed out in the introduction, the primary criticism leveled against the Markowitz model is that the optimal portfolio $\boldsymbol{\phi}^*$ is extremely sensitive to the market parameters $(\mathbf{E}[\mathbf{r}], \mathbf{Var}[\mathbf{r}])$ – since these parameters are estimated from noisy data, $\boldsymbol{\phi}^*$ often amplifies noise. By introducing "measures of uncertainty" in the market models, we are attempting to correct this sensitivity to perturbations. The uncertainty sets S_m, S_v and S_d represent the uncertainty of our limited (inexact) information of the market parameters and we wish to select portfolios that performs well for all parameter values that are consistent with this limited information. Such portfolios are solutions of appropriately defined min-max optimization problems called *robust portfolio selection problems.*

The robust analog of the Markowitz mean-variance optimization problem (5) is given by

$$\begin{array}{rl} \text{minimize} & \max_{\{\mathbf{V}\in S_v,\mathbf{D}\in S_d\}} \mathbf{Var}\left[\mathbf{r}_\phi\right] \\ \text{subject to} & \min_{\{\boldsymbol{\mu}\in S_m\}} \mathbf{E}[r_\phi] \geq \alpha, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1. \end{array} \tag{6}$$

The objective of the *robust minimum variance portfolio selection problem* (6) is to minimize the worst case variance of the portfolio subject to the constraint that the worst case expected return on the portfolio is at least α. We expect that the sensitivity of the optimal solution of this mathematical program to parameter fluctuations will be significantly smaller than it would be for its classical counterpart (5).

A closely related problem, the *robust maximum return problem*, is the dual of (6). In this problem, the objective is to maximize the worst case expected return subject to a constraint on the worst case variance, i.e. to solve the mathematical program

$$\begin{array}{rl} \text{maximize} & \min_{\{\boldsymbol{\mu}\in S_m\}} \mathbf{E}[r_\phi] \\ \text{subject to} & \max_{\{\mathbf{V}\in S_v,\mathbf{D}\in S_d\}} \mathbf{Var}\left[\mathbf{r}_\phi\right] \leq \lambda, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1. \end{array} \tag{7}$$

Another variant of the robust optimization problem (6) is the *robust maximum Sharpe ratio problem.* Here the objective is to choose a portfolio that maximizes the worst case ratio of the expected excess return on the portfolio, i.e. the return in excess of the risk-free rate r_f, to the standard deviation of the return. The corresponding max-min problem is given by,

$$\text{maximize}_{\{\boldsymbol{\phi}:\mathbf{1}^T\boldsymbol{\phi}=1\}} \quad \min_{\{\boldsymbol{\mu}\in S_m,\mathbf{V}\in S_v,\mathbf{D}\in S_d\}} \left\{ \frac{\mathbf{E}[r_\phi]-r_f}{\sqrt{\mathbf{Var}[\mathbf{r}_\phi]}} \right\} \tag{8}$$

All these variants are studied in Section 3. We show that for the uncertainty sets S_d, S_m and S_v, defined in (2)-(4) above, all of these problems reduce to SOCPs.

In Section 4 we study robust portfolio selection with value-at-risk (VaR) constraints where the objective is to maximize the worst case expected return of the portfolio subject to the constraint that the probability of the return r_ϕ falling below a threshold α is less than a prescribed limit, i.e. the objective is to solve the following mathematical program.

$$\begin{array}{rl} \text{maximize} & \min_{\{\boldsymbol{\mu}\in S_m\}} \mathbf{E}[r_\phi] \\ \text{subject to} & \max_{\{\mathbf{V}\in S_v,\boldsymbol{\mu}\in S_m,\mathbf{D}\in S_d\}} \mathbf{P}(r_\phi \leq \alpha) \leq \beta. \end{array} \tag{9}$$

VaR was introduced as a performance analysis tool in the context of risk management. Recently there has been a growing interest in imposing VaR-type constraints while optimizing credit risk [15, 21, 2]. We show that, for the uncertainty sets defined in (2)-(4), this problem can also be reduced to an SOCP and, hence, can be solved very efficiently. This result extends to another closely related performance measure called the conditional VaR [21].

The optimization problems of interest here fall in the general class of robust convex optimization problems. A generic problem of this class has the following structure,

$$\begin{array}{rll} \text{minimize} & \mathbf{c}^T\mathbf{x} & \\ \text{subject to} & F(\mathbf{x}, \boldsymbol{\xi}) \in \mathcal{K} \subset \mathbf{R}^m, & \forall \boldsymbol{\xi} \in \mathcal{U}, \end{array} \tag{10}$$

where $\boldsymbol{\xi}$ are the uncertain parameters in the problem, $\mathbf{x} \in \mathbf{R}^n$ is the decision vector, $\mathcal{K}$ is a convex cone and, for fixed $\boldsymbol{\xi} \in \mathcal{U}$, the function $F(\mathbf{x}, \boldsymbol{\xi})$ is $\mathcal{K}$-concave [4, 6].

The essential ideas leading to this formalism were developed in robust control (see [28] and references therein). Robustness was introduced to mathematical programming by Ben-Tal and Nemirovski [4, 5, 6]. They established that for suitably defined uncertainty sets $\mathcal{U}$, the robust counterparts of linear programs, quadratic programs, and general convex programs are themselves tractable optimization problems. Robust least squares problems and robust semidefinite programs were independently studied by El Ghaoui and his collaborators [9, 11]. Ben-Tal, Margalit, and Nemirovski have studied robust modeling of multi-stage portfolio problems [3]. Halldórsson and Tütüncü [13] study robust investment problems that reduce to saddle point problems.

3 Robust mean-variance portfolio selection

This section begins with a detailed analysis of the robust minimum variance problem (6). It is shown that for uncertainty sets defined in (2)-(4) this problem reduces to an efficiently solvable SOCP. This result is subsequently extended to the maximum return problem (7) and the robust maximum Sharpe ratio problem (8).

3.1 Robust minimum variance problem

Since the return $r_\phi \sim \mathcal{N}(\boldsymbol{\mu}^T\boldsymbol{\phi}, \boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V} + \mathbf{D})\boldsymbol{\phi})$, the robust minimum variance portfolio selection problem (6) is given by

$$\begin{array}{rl} \text{minimize} & \max_{\{\mathbf{V}\in S_v\}} \left\{\boldsymbol{\phi}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi}\right\} + \max_{\{\mathbf{D}\in S_d\}} \left\{\boldsymbol{\phi}^T\mathbf{D}\boldsymbol{\phi}\right\} \\ \text{subject to} & \min_{\{\boldsymbol{\mu}\in S_m\}} \boldsymbol{\mu}^T\boldsymbol{\phi} \geq \alpha, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1. \end{array} \tag{11}$$

The bounds $\underline{d}_i \leq d_i \leq \overline{d}_i$ imply that $\boldsymbol{\phi}^T\mathbf{D}\boldsymbol{\phi} \leq \boldsymbol{\phi}^T\overline{\mathbf{D}}\boldsymbol{\phi}$, where $\overline{\mathbf{D}} = \mathbf{diag}(\overline{\mathbf{d}})$. Also, since the covariance matrix $\mathbf{F}$ of the factor $\mathbf{f}$ is assumed to be strictly positive definite, the function $\|\mathbf{x}\|_f : x \mapsto \sqrt{\mathbf{x}^T\mathbf{F}\mathbf{x}}$ defines a norm on $\mathbf{R}^m$. Thus, (11) is equivalent to the robust augmented least squares problem (see [9])

$$
\begin{array}{ll}
\text{minimize} & \max_{\{\mathbf{V}\in S_v\}} \|\mathbf{V}\boldsymbol{\phi}\|_f^2 + \boldsymbol{\phi}^T\overline{\mathbf{D}}\boldsymbol{\phi} \\
\text{subject to} & \min_{\{\boldsymbol{\mu}\in S_m\}} \boldsymbol{\mu}^T\boldsymbol{\phi} \geq \alpha, \\
& \mathbf{1}^T\boldsymbol{\phi} = 1.
\end{array}
\tag{12}
$$

By introducing auxiliary variables ν and δ, the robust problem (12) can be reformulated as

$$
\begin{array}{llcl}
\text{minimize} & \nu + \delta & & \\
\text{subject to} & \max_{\{\mathbf{V}\in S_v\}} \|\mathbf{V}\boldsymbol{\phi}\|_f^2 & \leq & \nu, \\
& \boldsymbol{\phi}^T\overline{\mathbf{D}}\boldsymbol{\phi} & \leq & \delta, \\
& \min_{\{\boldsymbol{\mu}\in S_m\}} \boldsymbol{\mu}^T\boldsymbol{\phi} & \geq & \alpha, \\
& \mathbf{1}^T\boldsymbol{\phi} & = & 1.
\end{array}
\tag{13}
$$

If the uncertainty sets S_v and S_m are finite, i.e. $S_v = \{\mathbf{V}_1, \ldots, \mathbf{V}_s\}$ and $S_m = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_r\}$, then (13) reduces to the convex quadratically constrained problem

$$
\begin{array}{llcll}
\text{minimize} & \lambda + \delta & & & \\
\text{subject to} & \|\mathbf{V}_k\boldsymbol{\phi}\|_f^2 & \leq & \lambda, & \text{for all } k = 1, \ldots, s, \\
& \boldsymbol{\phi}^T\overline{\mathbf{D}}\boldsymbol{\phi} & \leq & \delta, & \\
& \boldsymbol{\phi}^T\boldsymbol{\mu}_k & \geq & \alpha, & \text{for all } k = 1, \ldots, r, \\
& \mathbf{1}^T\boldsymbol{\phi} & = & 1. &
\end{array}
\tag{14}
$$

This problem can be easily converted to an SOCP (see [18] or [24] for details). El Ghaoui and Lebret [9] show that for

$$
S_v = \left\{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}\| = \sqrt{\mathbf{Tr}(\mathbf{W}^T\mathbf{W})} \leq \rho\right\},
$$

the problem can still be reformulated as an SOCP. Methodologically speaking, our results can be viewed as an extension of the work in [9] to other classes of uncertainty sets more suited to the application at hand.

When the uncertainty in $\mathbf{V}$ and $\boldsymbol{\mu}$ is specified by (3) and (4) respectively, the worst case mean return of a fixed portfolio $\boldsymbol{\phi}$ is given by,

$$
\min_{\{\boldsymbol{\mu}\in S_m\}} \boldsymbol{\mu}^T\boldsymbol{\phi} = \boldsymbol{\mu}_0^T\boldsymbol{\phi} - \boldsymbol{\gamma}^T|\boldsymbol{\phi}|;
\tag{15}
$$

and the worst case variance is given by,

$$
\begin{array}{ll}
\text{maximize} & \|(\mathbf{V}_0 + \mathbf{W})\boldsymbol{\phi}\|_f^2 \\
\text{subject to} & \|\mathbf{W}_i\|_g \leq \rho_i, \quad i = 1, \ldots, n.
\end{array}
\tag{16}
$$

Since the constraints $\|\mathbf{W}_i\|_g \leq \rho_i$, $i = 1, \ldots, n$ imply the bound,

$$
\|\mathbf{W}\boldsymbol{\phi}\|_g = \left\|\sum_{i=1}^n \phi_i\mathbf{W}_i\right\|_g \leq \sum_{i=1}^n |\phi_i|\,\|\mathbf{W}_i\|_g \leq \sum_{i=1}^n \rho_i\,|\phi_i|,
\tag{17}
$$

the optimization problem,

$$\begin{array}{ll} \text{maximize} & \|\mathbf{V}_0\boldsymbol{\phi} + \mathbf{w}\|_f^2 \\ \text{subject to} & \|\mathbf{w}\|_g \leq r, \end{array} \tag{18}$$

where $r = \boldsymbol{\rho}^T|\boldsymbol{\phi}| = \sum_{i=1}^n \rho_i|\phi_i|$, is a relaxation of (16), i.e. the optimal value of (18) is at least as large as that of (16).

The objective function in (18) is convex; therefore, the optimal solution $\mathbf{w}^*$ lies on the boundary of the feasible set, i.e. $\|\mathbf{w}^*\|_g = r$. For $i = 1, \ldots, n$, define

$$\mathbf{W}_i^* = \begin{cases} \frac{\phi_i}{|\phi_i|}\frac{\rho_i}{r}\mathbf{w}^*, & \phi_i \neq 0, \\ \frac{\rho_i}{r}\mathbf{w}^*, & \text{otherwise.} \end{cases} \tag{19}$$

Then $\|\mathbf{W}_i^*\|_g = \rho_i$, for all $i = 1, \ldots, n$, i.e. $\mathbf{W}^*$ is feasible for (16), and $\mathbf{W}^*\boldsymbol{\phi} = \sum_{i=1}^n \phi_i\mathbf{W}_i^* = \mathbf{w}^*$. Therefore, the optimal value of (16) and (18) are, in fact, the same.

Thus, for a fixed portfolio $\boldsymbol{\phi}$, the worst-case variance is less than ν if, and only if,

$$\max_{\{\mathbf{y}:\|\mathbf{y}\|_g \leq r\}} \|\mathbf{y}_0 + \mathbf{y}\|_f^2 \leq \nu, \tag{20}$$

where $\mathbf{y}_0 = \mathbf{V}_0\boldsymbol{\phi}$ and $r = \boldsymbol{\rho}^T|\boldsymbol{\phi}|$.

The following lemma reformulates this constraint as a collection of linear equalities, linear inequalities and restricted hyperbolic constraints (i.e. constraints of the form: $\mathbf{z}^T\mathbf{z} \leq xy$, $\mathbf{z} \in \mathbf{R}^n$, $x, y \in \mathbf{R}$ and $x, y \geq 0$).

Lemma 1 *Let $r, \nu > 0$, $\mathbf{y}_0, \mathbf{y} \in \mathbf{R}^m$ and $\mathbf{F}, \mathbf{G} \in \mathbf{R}^{m\times m}$ be positive definite matrices. Then the constraint*

$$\max_{\{\mathbf{y}:\|\mathbf{y}\|_g \leq r\}} \|\mathbf{y}_0 + \mathbf{y}\|_f^2 \leq \nu \tag{21}$$

is equivalent to either of the following:

(i) there exist $\tau, \sigma \geq 0$, and $\mathbf{t} \in \mathbf{R}_+^m$ that satisfy

$$\begin{array}{rcl} \nu & \geq & \tau + \mathbf{1}^T\mathbf{t} \\ \sigma & \leq & \frac{1}{\lambda_{\max}(\mathbf{H})}, \\ r^2 & \leq & \sigma\tau, \\ w_i^2 & \leq & (1 - \sigma\lambda_i)t_i, \quad i = 1, \ldots, m, \end{array} \tag{22}$$

where $\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$ is the spectral decomposition of $\mathbf{H} = \mathbf{G}^{-\frac{1}{2}}\mathbf{F}\mathbf{G}^{-\frac{1}{2}}$, $\boldsymbol{\Lambda} = \mathbf{diag}(\lambda_i)$, and $\mathbf{w} = \mathbf{Q}^T\mathbf{H}^{\frac{1}{2}}\mathbf{G}^{\frac{1}{2}}\mathbf{y}_0$;

(ii) there exist $\tau \geq 0$ and $\mathbf{s} \in \mathbf{R}_+^m$ that satisfy

$$\begin{array}{rcl} r^2 & \leq & \tau(\nu - \mathbf{1}^T\mathbf{s}), \\ u_i^2 & \leq & (1 - \tau\theta_i)s_i, \quad i = 1, \ldots, m, \\ \tau & \leq & \frac{1}{\lambda_{\max}(\mathbf{K})}, \end{array} \tag{23}$$

where $\mathbf{P}\boldsymbol{\Theta}\mathbf{P}^T$ *is the spectral decomposition of* $\mathbf{K} = \mathbf{F}^{\frac{1}{2}}\mathbf{G}^{-1}\mathbf{F}^{\frac{1}{2}}$, $\boldsymbol{\Theta} = \mathbf{diag}(\theta_i)$, *and* $\mathbf{u} = \mathbf{P}^T\mathbf{F}^{\frac{1}{2}}\mathbf{y}_0$.

Proof: By setting $\mathbf{y} = r\bar{\mathbf{y}}$, we have that (21) is equivalent to

$$(\nu - \mathbf{y}_0^T\mathbf{F}\mathbf{y}_0) - 2r\mathbf{y}_0^T\mathbf{F}\bar{\mathbf{y}} - r^2\bar{\mathbf{y}}^T\mathbf{F}\bar{\mathbf{y}} \geq 0, \tag{24}$$

for all $\bar{\mathbf{y}}$ such that $1 - \bar{\mathbf{y}}^T\mathbf{G}\bar{\mathbf{y}} \geq 0$. Before proceeding further, we need the following:

Lemma 2 ($\mathcal{S}$-procedure) *Let* $F_i(\mathbf{x}) = \mathbf{x}^T\mathbf{A}_i\mathbf{x} + 2\mathbf{b}_i^T\mathbf{x} + c_i$, $i = 0, \ldots, p$ *be quadratic functions of* $\mathbf{x} \in \mathbf{R}^n$. *Then* $F_0(\mathbf{x}) \geq 0$ *for all* $\mathbf{x}$ *such that* $F_i(\mathbf{x}) \geq 0$, $i = 1, \ldots, p$, *if there exist* $\tau_i \geq 0$ *such that*

$$\begin{bmatrix} c_0 & \mathbf{b}_0^T \\ \mathbf{b}_0 & \mathbf{A}_0 \end{bmatrix} - \sum_{i=1}^{p} \tau_i \begin{bmatrix} c_i & \mathbf{b}_i^T \\ \mathbf{b}_i & \mathbf{A}_i \end{bmatrix} \succeq 0.$$

Moreover, if $p = 1$ *then the converse holds if there exists* $\mathbf{x}_0$ *such that* $F_1(\mathbf{x}_0) > 0$.

For a discussion of the $\mathcal{S}$-procedure and its applications, see [7]. Since $\bar{\mathbf{y}} = \mathbf{0}$ is strictly feasible for $1 - \bar{\mathbf{y}}^T\mathbf{G}\bar{\mathbf{y}} \geq 0$, the $\mathcal{S}$-procedure implies that (24) holds for all $1 - \bar{\mathbf{y}}^T\mathbf{G}\bar{\mathbf{y}} \geq 0$ if and only if there exists a $\tau \geq 0$ such that

$$\mathbf{M} = \begin{bmatrix} \nu - \tau - \mathbf{y}_0^T\mathbf{F}\mathbf{y}_0 & -r\mathbf{y}_0^T\mathbf{F} \\ -r\mathbf{F}\mathbf{y}_0 & \tau\mathbf{G} - r^2\mathbf{F} \end{bmatrix} \succeq 0. \tag{25}$$

Let the spectral decomposition of $\mathbf{H} = \mathbf{G}^{-\frac{1}{2}}\mathbf{F}\mathbf{G}^{-\frac{1}{2}}$ be $\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$, where $\boldsymbol{\Lambda} = \mathbf{diag}(\boldsymbol{\lambda})$, and define $\mathbf{w} = \mathbf{Q}^T\mathbf{H}\mathbf{G}^{\frac{1}{2}}\mathbf{y}_0 = \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Q}^T\mathbf{G}^{\frac{1}{2}}\mathbf{y}_0$. Observing that $\mathbf{y}_0^T\mathbf{F}\mathbf{y}_0 = \mathbf{w}^T\mathbf{w}$, we have that the matrix $\mathbf{M} \succeq \mathbf{0}$ if and only if

$$\bar{\mathbf{M}} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{Q}^T\mathbf{G}^{\frac{1}{2}} \end{bmatrix} \mathbf{M} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{G}^{\frac{1}{2}}\mathbf{Q} \end{bmatrix} = \begin{bmatrix} \nu - \tau - \mathbf{w}^T\mathbf{w} & -r\mathbf{w}^T\boldsymbol{\Lambda}^{\frac{1}{2}} \\ -r\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{w} & \tau\mathbf{I} - r^2\boldsymbol{\Lambda} \end{bmatrix} \succeq \mathbf{0}.$$

The matrix $\bar{\mathbf{M}} \succeq \mathbf{0}$ if and only if $\tau \geq r^2\lambda_i$, for all $i = 1, \ldots, m$ (i.e. $\tau \geq r^2\lambda_{\max}(\mathbf{H})$), $w_i = 0$ for all i such that $\tau = r^2\lambda_i$, and Schur complement of the nonzero rows and columns of $\tau\mathbf{I} - r^2\boldsymbol{\Lambda}$

$$\nu - \tau - \mathbf{w}^T\mathbf{w} - r^2\left(\sum_{i:\tau\neq r^2\lambda_i} \frac{\lambda_i w_i^2}{\tau - r^2\lambda_i}\right) = \nu - \tau - \sum_{i:\sigma\lambda_i\neq 1} \frac{w_i^2}{1-\lambda_i} \geq 0,$$

where $\sigma = \frac{r^2}{\tau}$. It follows that

$$\max_{\{\mathbf{y}:\|\mathbf{y}\|_g\leq r\}} \|\mathbf{y}_0 + r\mathbf{y}\|_f^2 \leq \nu,$$

if and only if there exists $\tau, \sigma \geq 0$ and $\mathbf{t} \in \mathbf{R}_+^m$ satisfying,

$$\begin{array}{rcl} \nu & \geq & \tau + \mathbf{1}^T\mathbf{t}, \\ r^2 & = & \sigma\tau, \\ w_i^2 & = & (1 - \sigma\lambda_i)t_i, \quad i = 1, \ldots, m, \\ \sigma & \leq & \frac{1}{\lambda_{\max}(\mathbf{H})}. \end{array} \tag{26}$$

It is easy to establish there exist $\tau, \sigma \geq 0$, and $\mathbf{t} \in \mathbf{R}_+^m$ that satisfy (26) if and only if there exist $\tau, \sigma \geq 0$, and $\mathbf{t} \in \mathbf{R}_+^m$ that satisfy (26) with the equalities replaced by inequalities, i.e. that satisfy (22).

To establish the second representation, note that (21) holds if and only if

$$\mathbf{y}^T \mathbf{G} \mathbf{y} \geq r^2, \tag{27}$$

for all $\mathbf{y}$ such that $\|\mathbf{y}_0 + \mathbf{y}\|_f^2 \geq \nu$. Since $\|\mathbf{y}_0 + \mathbf{y}\|_f^2 > \nu$ for all sufficiently large $\mathbf{y}$, the $\mathcal{S}$-procedure implies that (27) holds for all $\|\mathbf{y}_0 + \mathbf{y}\| \geq \nu$ if and only if there exists a $\tau \geq 0$ such that

$$\mathbf{M} = \begin{bmatrix} -r^2 - \tau(\mathbf{y}_0^T \mathbf{F} \mathbf{y}_0 - \nu) & -\tau \mathbf{y}_0^T \mathbf{F} \\ -\tau \mathbf{F} \mathbf{y}_0 & \mathbf{G} - \tau \mathbf{F} \end{bmatrix} \succeq 0. \tag{28}$$

Let $\mathbf{P\Theta P}^T$ be the spectral decomposition of $\mathbf{K} = \mathbf{F}^{\frac{1}{2}} \mathbf{G}^{-1} \mathbf{F}^{\frac{1}{2}}$, where $\mathbf{\Theta} = \mathbf{diag}(\boldsymbol{\theta})$, and $\mathbf{u} = \mathbf{P}^T \mathbf{F}^{\frac{1}{2}} \mathbf{y}_0$. Then $\mathbf{M} \succeq \mathbf{0}$ if and only if

$$\bar{\mathbf{M}} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{P}^T \mathbf{G}^{-\frac{1}{2}} \end{bmatrix} \mathbf{M} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{G}^{-\frac{1}{2}} \mathbf{P} \end{bmatrix} = \begin{bmatrix} r^2 - \tau(\mathbf{u}^T \mathbf{u} - \nu) & -\tau \mathbf{u}^T \\ -\tau \mathbf{u} & \mathbf{\Theta}^{-1} - \tau \mathbf{I} \end{bmatrix} \succeq \mathbf{0}.$$

But $\bar{\mathbf{M}} \succeq \mathbf{0}$ if and only if $\tau \leq \frac{1}{\theta_i}$ (i.e. $\tau \leq 1/\lambda_{\max}(\mathbf{K})$), $u_i = 0$ for all i such that $\tau\theta_i = 1$, and the Schur complement of the nonzero row and columns of $\mathbf{\Theta}^{-1} - \tau \mathbf{I}$

$$r^2 - \tau(\mathbf{u}^T \mathbf{u} - \nu) - \tau^2 \left(\sum_{i:\tau\theta_i \neq 1} \frac{u_i^2}{\theta_i^{-1} - \tau} \right) = r^2 + \tau \left(\nu - \sum_{i:\tau\theta_i \neq 1} \frac{u_i^2}{1 - \tau\theta_i} \right) \geq 0.$$

Hence,

$$\max_{\{\mathbf{y}:\|\mathbf{y}\|_g \leq r\}} \|\mathbf{y}_0 + r\mathbf{y}\|_f^2 \leq \nu,$$

if and only if there exists $\tau \geq 0$ and $\mathbf{s} \in \mathbf{R}_+^m$ satisfying,

$$\begin{array}{rcl} r^2 & \leq & \tau(\nu - \mathbf{1}^T \mathbf{s}), \\ u_i^2 & = & (1 - \tau\theta_i) s_i, \quad i = 1, \ldots, m, \\ \tau & \leq & \frac{1}{\lambda_{\max}(\mathbf{K})}. \end{array} \tag{29}$$

Completely analogous to the proof of Part (i), we have that there exists $\tau \geq 0$ and $\mathbf{s} \in \mathbf{R}_+^m$ satisfying (29) if and only if there exist $\tau \geq 0$ and $\mathbf{s} \in \mathbf{R}_+^m$ satisfying (23). This proves the second result. ■

To illustrate the equivalence of parts (i) and (ii) of Lemma 1, we note that if $\mathbf{F} = \kappa \mathbf{G}$ then $\mathbf{K} = \mathbf{H} = \kappa \mathbf{I}$ and $\mathbf{w} = \mathbf{u}$. Hence, if $(\tau, \mathbf{s})$ satisfies (23) then $(\tilde{\sigma}, \tilde{\tau}, \tilde{\mathbf{t}}) = (\tau, \nu - \mathbf{1}^T \mathbf{s}, \mathbf{s})$ satisfies (22).

The restricted hyperbolic constraints, $\mathbf{z}^T \mathbf{z} \leq xy$, $x, y \geq 0$, can be reformulated as second-order cone constraints as follows (e.g., see [24], Section 6.2.3)

$$\mathbf{z}^T \mathbf{z} \leq xy \quad \Leftrightarrow \quad 4\mathbf{z}^T \mathbf{z} \leq (x+y)^2 - (x-y)^2 \quad \Leftrightarrow \quad \left\| \begin{bmatrix} 2\mathbf{z} \\ x - y \end{bmatrix} \right\| \leq x + y. \tag{30}$$

The above result and Lemma 1 motivate the following definition.

Definition 1 *Given* $\mathbf{V}_0 \in \mathbf{R}^{m\times n}$*, and* $\mathbf{F}, \mathbf{G} \in \mathbf{R}^{m\times m}$*, define* $\mathcal{H}(\mathbf{V}_0, \mathbf{F}, \mathbf{G})$ *to be the set of all vectors* $(r;\nu;\delta;\boldsymbol{\phi}) \in \mathbf{R}\times\mathbf{R}\times\mathbf{R}\times\mathbf{R}^n$ *such that* $(r, \nu, \mathbf{y}_0 = \mathbf{V}_0\boldsymbol{\phi})$ *satisfy (22), i.e.*

(i) there exist $\sigma, \tau \geq 0$ *and* $\mathbf{t} \in \mathbf{R}^m_+$ *that satisfy*

$$
\begin{array}{rcl}
\tau + \mathbf{1}^T\mathbf{t} & \leq & \nu, \\
\sigma & \leq & \frac{1}{\lambda_{\max}(\mathbf{H})}, \\
\left\| \begin{bmatrix} 2r \\ \sigma - \tau \end{bmatrix} \right\| & \leq & \sigma + \tau, \\
\left\| \begin{bmatrix} 2w_i \\ 1 - \sigma\lambda_i - t_i \end{bmatrix} \right\| & \leq & 1 - \sigma\lambda_i + t_i, \quad i = 1, \ldots, m,
\end{array}
$$

where $\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$ *is the spectral decomposition of* $\mathbf{H} = \mathbf{G}^{-\frac{1}{2}}\mathbf{F}\mathbf{G}^{-\frac{1}{2}}$*,* $\boldsymbol{\Lambda} = \mathbf{diag}(\lambda_i)$ *and* $\mathbf{w} = \mathbf{Q}^T\mathbf{H}^{\frac{1}{2}}\mathbf{G}^{\frac{1}{2}}\mathbf{V}_0\boldsymbol{\phi}$*;*

(ii) and $\boldsymbol{\phi}^T\overline{\mathbf{D}}\boldsymbol{\phi} \leq \delta$*, or equivalently*

$$
\left\| \begin{bmatrix} 2\overline{\mathbf{D}}^{\frac{1}{2}}\boldsymbol{\phi} \\ 1 - \delta \end{bmatrix} \right\| \leq 1 + \delta.
$$

From (15), (20), and Definition 1, it follows that (13) can be reformulated as

$$
\begin{array}{lrcl}
\text{minimize} & \nu + \delta, & & \\
\text{subject to} & \boldsymbol{\mu}_0^T\boldsymbol{\phi} - \boldsymbol{\gamma}^T|\boldsymbol{\phi}| & \geq & \alpha, \\
& \boldsymbol{\rho}^T|\boldsymbol{\phi}| & \leq & r, \\
& \mathbf{1}^T\boldsymbol{\phi} & = & 1, \\
& (r;\nu;\delta;\boldsymbol{\phi}) & \in & \mathcal{H}(\mathbf{V}_0, \mathbf{F}, \mathbf{G}).
\end{array}
\tag{31}
$$

where the equality $r = \boldsymbol{\rho}^T|\boldsymbol{\phi}|$ has been relaxed by recognizing that the relaxed constraint will always be tight at an optimal solution. Although (31) is a convex optimization problem, it is not an SOCP. However, replacing $|\boldsymbol{\phi}|$ by a new variable $\boldsymbol{\psi} \in \mathbf{R}^n$ and adding the $2n$ linear constraints, $\psi_i \geq |\phi_i|$, $i = 1, \ldots, n$, leads to the following SOCP formulation for the robust minimum variance portfolio selection problem (11)

$$
\begin{array}{lrcll}
\text{minimize} & \nu + \delta, & & & \\
\text{subject to} & \boldsymbol{\mu}_0^T\boldsymbol{\phi} - \boldsymbol{\gamma}^T\boldsymbol{\psi} & \geq & \alpha, & \\
& r & \geq & \boldsymbol{\rho}^T\boldsymbol{\psi}, & \\
& \psi_i & \geq & \phi_i, & i = 1, \ldots, n, \\
& \psi_i & \geq & -\phi_i, & i = 1, \ldots, n, \\
& \mathbf{1}^T\boldsymbol{\phi} & = & 1, & \\
& (r;\nu;\delta;\boldsymbol{\phi}) & \in & \mathcal{H}(\mathbf{V}_0, \mathbf{F}, \mathbf{G}). &
\end{array}
\tag{32}
$$

Another possible transformation leading to an SOCP is to replace $\boldsymbol{\phi}$ by $\boldsymbol{\phi} = \boldsymbol{\phi}_+ - \boldsymbol{\phi}_-$, $\boldsymbol{\phi}_+, \boldsymbol{\phi}_- \in \mathbf{R}^n_+$. An alternative SOCP formulation is obtained if one employs Part (ii) of Lemma to characterize the worst case variance. The derivation of these results is left to the reader.

To keep the exposition simple, in the rest of this paper it will assumed that short sales are not allowed, i.e. $\boldsymbol{\phi} \geq \mathbf{0}$. In each case the result can be extended to general $\boldsymbol{\phi}$ by employing the above transformations.

3.2 Robust maximum return problem

The robust maximum return problem is given by

$$\begin{array}{rl} \text{maximize} & \min_{\{\boldsymbol{\mu} \in S_m\}} \boldsymbol{\mu}^T \boldsymbol{\phi}, \\ \text{subject to} & \max_{\{\mathbf{V} \in S_v, \mathbf{D} \in S_d\}} \boldsymbol{\phi}^T (\mathbf{V}^T \mathbf{F} \mathbf{V} + \mathbf{D}) \boldsymbol{\phi} \leq \lambda, \\ & \mathbf{1}^T \boldsymbol{\phi} = 1, \\ & \boldsymbol{\phi} \geq \mathbf{0}, \end{array}$$

or equivalently, from (15), by

$$\begin{array}{rrcl} \text{maximize} & (\boldsymbol{\mu} - \boldsymbol{\gamma})^T \boldsymbol{\phi}, & & \\ \text{subject to} & \max_{\{\mathbf{V} \in S_v\}} \boldsymbol{\phi}^T \mathbf{V}^T \mathbf{F} \mathbf{V} \boldsymbol{\phi} & \leq & \lambda - \delta, \\ & \boldsymbol{\phi}^T \overline{\mathbf{D}} \boldsymbol{\phi} & \leq & \delta, \\ & \mathbf{1}^T \boldsymbol{\phi} & = & 1, \\ & \boldsymbol{\phi} & \geq & \mathbf{0}. \end{array} \tag{33}$$

From Definition 1, it follows that the robust maximum return problem is equivalent to the SOCP

$$\begin{array}{rrcl} \text{maximize} & (\boldsymbol{\mu} - \boldsymbol{\gamma})^T \boldsymbol{\phi}, & & \\ \text{subject to} & \mathbf{1}^T \boldsymbol{\phi} & = & 1, \\ & \boldsymbol{\phi} & \geq & \mathbf{0}, \\ & (\boldsymbol{\rho}^T \boldsymbol{\phi}; \lambda - \delta; \delta; \boldsymbol{\phi}) & \in & \mathcal{H}(\mathbf{V}_0, \mathbf{F}, \mathbf{G}). \end{array} \tag{34}$$

3.3 Robust maximum Sharpe ratio problem

The robust maximum Sharpe ratio problem is given by,

$$\max_{\{\boldsymbol{\phi} : \boldsymbol{\phi} \geq \mathbf{0}, \mathbf{1}^T \boldsymbol{\phi} = 1\}} \min_{\{\mathbf{V} \in S_v, \boldsymbol{\mu} \in S_m, \mathbf{D} \in S_d\}} \left\{ \frac{\boldsymbol{\mu}^T \boldsymbol{\phi} - r_f}{\sqrt{\mathbf{Var}[\boldsymbol{\phi}]}} \right\}, \tag{35}$$

where r_f is the risk-free rate of return. We will assume that the optimal value of this max-min problem is strictly positive, i.e. there exists a portfolio with finite worst case variance whose worst case return is strictly greater than the risk-free rate r_f. This is not an unreasonable assumption in practice.

Since the components of the portfolio vector $\boldsymbol{\phi}$ add up to 1, the objective of the max-min problem (35),

$$\frac{\boldsymbol{\mu}^T \boldsymbol{\phi} - r_f}{\sqrt{\mathbf{Var}[\boldsymbol{\phi}]}} = \frac{(\boldsymbol{\mu} - r_f \mathbf{1})^T \boldsymbol{\phi}}{\sqrt{\mathbf{Var}[\boldsymbol{\phi}]}},$$

is a homogeneous function of the portfolio $\boldsymbol{\phi}$. This implies that the normalization condition $\mathbf{1}^T\boldsymbol{\phi} = 1$ can be dropped and the constraint $\min_{\{\boldsymbol{\mu}\in S_m\}}(\boldsymbol{\mu} - r_f\mathbf{1})^T\boldsymbol{\phi} = 1$ added to (35) without any loss of generality. With this transformation, (35) reduces to minimizing the worst case variance; i.e. (35) is equivalent to

$$\begin{array}{lrcl} \text{minimize} & \multicolumn{3}{l}{\max_{\{\mathbf{V}\in S_v\}} \boldsymbol{\phi}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi} + \boldsymbol{\phi}^T\overline{\mathbf{D}}\boldsymbol{\phi}} \\ \text{subject to} & (\boldsymbol{\mu}_0 - \boldsymbol{\gamma} - r_f\mathbf{1})^T\boldsymbol{\phi} & \geq & 1, \\ & \boldsymbol{\phi} & \geq & \mathbf{0}, \end{array} \tag{36}$$

where the constraint $\min_{\{\boldsymbol{\mu}\in S_m\}}(\boldsymbol{\mu} - r_f\mathbf{1})^T\boldsymbol{\phi} = 1$ has been relaxed by recognizing that the relaxed constraint will always be tight at an optimal solution. Consequently, the robust maximum Sharpe ratio problem is equivalent to a robust minimum variance problem with $\alpha = 1$, $\boldsymbol{\mu}$ replaced by $\boldsymbol{\mu} - r_f\mathbf{1}$, and $\boldsymbol{\phi}$ no longer normalized. Exploiting this relationship, we have that (35) is equivalent to the SOCP

$$\begin{array}{lrcl} \text{minimize} & & \nu + \delta, & \\ \text{subject to} & (\boldsymbol{\mu}_0 - \boldsymbol{\gamma} - r_f\mathbf{1})^T\boldsymbol{\phi} & \geq & 1, \\ & (\boldsymbol{\rho}^T\boldsymbol{\phi}; \nu; \delta; \boldsymbol{\phi}) & \in & \mathcal{H}(\mathbf{V}_0, \mathbf{F}, \mathbf{G}). \end{array} \tag{37}$$

The crucial step in the reduction of (35) to (37) is the realization that the objective function in (35) is homogeneous in $\boldsymbol{\phi}$, and therefore, its numerator can be restricted to 1 without any loss of generality. This homogenization goes through even when there are additional inequality constraints on the portfolio choices.

Suppose the portfolio $\boldsymbol{\phi}$ is constrained to satisfy $\mathbf{A}\boldsymbol{\phi} \geq \mathbf{b}$ ($\boldsymbol{\phi} \geq \mathbf{0}$ is assumed to be subsumed in $\mathbf{A}\boldsymbol{\phi} \geq \mathbf{b}$). Then, the robust maximum Sharpe ratio problem,

$$\max_{\{\boldsymbol{\phi}:\mathbf{A}\boldsymbol{\phi}\geq\mathbf{b},\mathbf{1}^T\boldsymbol{\phi}=1\}} \min_{\{\mathbf{V}\in S_v,\boldsymbol{\mu}\in S_m,\mathbf{D}\in S_d\}} \left\{ \frac{\boldsymbol{\mu}^T\boldsymbol{\phi} - r_f}{\sqrt{\mathbf{Var}[\boldsymbol{\phi}]}} \right\}, \tag{38}$$

is equivalent to the robust minimum variance problem

$$\begin{array}{lrcl} \text{minimize} & \multicolumn{3}{l}{\max_{\{\mathbf{V}\in S_v\}} \boldsymbol{\phi}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi} + \boldsymbol{\phi}^T\overline{\mathbf{D}}\boldsymbol{\phi}} \\ \text{subject to} & (\boldsymbol{\mu}_0 - \boldsymbol{\gamma} - r_f\mathbf{1})^T\boldsymbol{\phi} & \geq & 1, \\ & \mathbf{A}\boldsymbol{\phi} & \geq & \zeta\mathbf{b}, \\ & \mathbf{1}^T\boldsymbol{\phi} & = & \zeta, \\ & \zeta & \geq & 0, \end{array} \tag{39}$$

where ζ is an auxiliary variable that has been introduced to homogenize the constraints $\mathbf{A}\boldsymbol{\phi} \geq \mathbf{b}$ and $\mathbf{1}^T\boldsymbol{\phi} = 1$. The problem (39) can easily be converted into a second-order cone problem by using the techniques developed above.

4 Robust Value-At-Risk (VaR) portfolio selection

The robust VaR portfolio selection problem is given by

$$\begin{array}{ll} \text{maximize} & \min_{\boldsymbol{\mu}\in S_m} \mathbf{E}[r_\phi] \\ \text{subject to} & \max_{\{\boldsymbol{\mu}\in S_m, \mathbf{V}\in S_v, \mathbf{D}\in S_d\}} \{\mathbf{P}(r_\phi \le \alpha)\} \le \beta, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1, \\ & \boldsymbol{\phi} \ge \mathbf{0}. \end{array} \tag{40}$$

This optimization problem maximizes the expected return subject to the constraint that the shortfall probability is less than β. Note that for ease of exposition we again assume that no short sales are allowed, i.e. $\boldsymbol{\phi} \ge \mathbf{0}$. The solution in the general case can be obtained by employing a transformation identical to the one detailed at the end of Section 3.1.

For fixed $(\boldsymbol{\mu}, \mathbf{V}, \mathbf{D})$, the return vector $r_\phi = \mathcal{N}(\boldsymbol{\mu}^T\boldsymbol{\phi}, \boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V}+\mathbf{D})\boldsymbol{\phi})$. Therefore

$$\begin{aligned} \mathbf{P}(r_\phi \le \alpha) \le \beta \quad & \Leftrightarrow \quad \mathbf{P}\left(\boldsymbol{\mu}^T\boldsymbol{\phi} + \mathcal{Z}\sqrt{\boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V}+\mathbf{D})\boldsymbol{\phi}} \le \alpha\right) \le \beta, \\ & \Leftrightarrow \quad \mathbf{P}\left(\mathcal{Z} \le \frac{\alpha - \boldsymbol{\mu}^T\boldsymbol{\phi}}{\sqrt{\boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V}+\mathbf{D})\boldsymbol{\phi}}}\right) \le \beta, \\ & \Leftrightarrow \quad \frac{\alpha - \boldsymbol{\mu}^T\boldsymbol{\phi}}{\sqrt{\boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V}+\mathbf{D})\boldsymbol{\phi}}} \le \mathcal{F}_{\mathcal{Z}}^{-1}(\beta), \end{aligned} \tag{41}$$

where $\mathcal{Z} \sim \mathcal{N}(0,1)$ is the standard Normal random variable and $\mathcal{F}_{\mathcal{Z}}(\cdot)$ is its cumulative density function. In typical VaR applications $\beta << 1$, therefore $\mathcal{F}_{\mathcal{Z}}^{-1}(\beta) < 0$. Thus, the probability constraint for fixed $(\boldsymbol{\mu}, \mathbf{V}, \mathbf{D})$ is equivalent to the second-order cone constraint

$$-\mathcal{F}_{\mathcal{Z}}^{-1}(\beta)\sqrt{\|\mathbf{F}^{\frac{1}{2}}\mathbf{V}\boldsymbol{\phi}\|^2 + \|\mathbf{D}^{\frac{1}{2}}\boldsymbol{\phi}\|^2} \le \boldsymbol{\mu}^T\boldsymbol{\phi} - \alpha. \tag{42}$$

On incorporating (42), the problem (40) can be rewritten as

$$\begin{array}{ll} \text{maximize} & \min_{\boldsymbol{\mu}\in S_m} \boldsymbol{\mu}^T\boldsymbol{\phi} \\ \text{subject to} & \max_{\{\boldsymbol{\mu}\in S_m, \mathbf{V}\in S_v, \mathbf{D}\in S_d\}} \left\{-\mathcal{F}_{\mathcal{Z}}^{-1}(\beta)\sqrt{\|\mathbf{F}^{\frac{1}{2}}\mathbf{V}\boldsymbol{\phi}\|^2 + \|\mathbf{D}^{\frac{1}{2}}\boldsymbol{\phi}\|^2} - \boldsymbol{\mu}^T\boldsymbol{\phi}\right\} \le \alpha, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1, \\ & \boldsymbol{\phi} \ge \mathbf{0}. \end{array} \tag{43}$$

Assuming that the uncertainty sets S_d, S_v and S_m are given by (2)-(4), (43) reduces to

$$\begin{array}{llcl} \text{maximize} & (\boldsymbol{\mu}_0 - \boldsymbol{\gamma})^T\boldsymbol{\phi} & & \\ \text{subject to} & -\mathcal{F}_{\mathcal{Z}}^{-1}(\beta)\left\|\begin{bmatrix}\nu \\ \delta\end{bmatrix}\right\| & \le & (\boldsymbol{\mu}_0 - \boldsymbol{\gamma})^T\boldsymbol{\phi} - \alpha, \\ & \left\|\overline{\mathbf{D}}^{\frac{1}{2}}\boldsymbol{\phi}\right\| & \le & \delta, \\ & \max_{\{\mathbf{V}\in S_v\}} \left\|\mathbf{F}^{\frac{1}{2}}\mathbf{V}\boldsymbol{\phi}\right\| & \le & \nu, \\ & \mathbf{1}^T\boldsymbol{\phi} & = & 1, \\ & \boldsymbol{\phi} & \ge & \mathbf{0}. \end{array} \tag{44}$$

From Part (ii) of Lemma 1, it follows that

$$\max_{\{\mathbf{V}\in S_v\}} \|\mathbf{F}^{\frac{1}{2}}\mathbf{V}\boldsymbol{\phi}\| \leq \nu,$$

if and only if there exist $\bar{\tau} \geq 0$ and $\bar{\mathbf{s}} \in \mathbf{R}^m_+$ such that,

$$\begin{array}{rcl} r^2 & \leq & \bar{\tau}(\nu^2 - \mathbf{1}^T\bar{\mathbf{s}}), \\ u_i^2 & \leq & (1-\bar{\tau}\theta_i)\bar{s}_i, \quad i = 1,\ldots,m, \\ \lambda_{\max}(\mathbf{K})\bar{\tau} & \leq & 1, \end{array} \tag{45}$$

where $\mathbf{K} = \mathbf{P}\boldsymbol{\Theta}\mathbf{P}^T$ is the spectral decomposition of $\mathbf{K} = \mathbf{F}^{\frac{1}{2}}\mathbf{G}^{-1}\mathbf{F}^{\frac{1}{2}}$, $\boldsymbol{\Theta} = \mathbf{diag}(\theta_i)$, and $\mathbf{u} = \mathbf{P}^T\mathbf{F}^{\frac{1}{2}}\mathbf{V}_0\boldsymbol{\phi}$. On introducing the change of variables, $\tau = \nu\bar{\tau}$ and $\mathbf{s} = \frac{1}{\nu}\bar{\mathbf{s}}$, (45) becomes

$$\begin{array}{rcl} r^2 & \leq & \tau(\nu - \mathbf{1}^T\mathbf{s}), \\ u_i^2 & \leq & (\nu - \tau\theta_i)s_i, \quad i = 1,\ldots,m, \\ \lambda_{\max}(\mathbf{K})\tau & \leq & \nu. \end{array} \tag{46}$$

Therefore, the robust VaR portfolio selection problem (40) is equivalent to the SOCP

$$\begin{array}{lrcl} \text{maximize} & (\boldsymbol{\mu}_0 - \boldsymbol{\gamma})^T\boldsymbol{\phi}, & & \\ \text{subject to} & \mathbf{u} & = & \mathbf{P}^T\mathbf{F}^{\frac{1}{2}}\mathbf{V}_0^T\boldsymbol{\phi}, \\ & -\mathcal{F}_{\mathcal{Z}}^{-1}(\beta)\left\|\begin{bmatrix}\nu \\ \delta\end{bmatrix}\right\| & \leq & (\boldsymbol{\mu}_0 - \boldsymbol{\gamma})^T\boldsymbol{\phi} - \alpha, \\ & \left\|\mathbf{D}^{\frac{1}{2}}\boldsymbol{\phi}\right\| & \leq & \delta, \\ & \left\|\begin{bmatrix}2\boldsymbol{\rho}^T\boldsymbol{\phi} \\ (\tau - \nu + \mathbf{1}^T\mathbf{s})\end{bmatrix}\right\| & \leq & (\tau + \nu - \mathbf{1}^T\mathbf{s}), \\ & \left\|\begin{bmatrix}2u_i \\ (\nu - \tau\theta_i - s_i)\end{bmatrix}\right\| & \leq & (\nu - \tau\theta_i + s_i), \qquad i = 1,\ldots,m, \\ & \nu - \tau\lambda_{\max}(\mathbf{K}) & \geq & 0, \\ & \mathbf{1}^T\boldsymbol{\phi} & = & 1, \\ & \boldsymbol{\phi} & \geq & \mathbf{0}, \\ & \tau & \geq & 0, \end{array} \tag{47}$$

As in the case of the minimum variance problem, an alternative SOCP formulation for the robust VaR problem follows from Part (i) of Lemma 1.

5 Multivariate regression and norm selection

In this section results from the statistical theory of multivariate linear regression are used justify the uncertainty structures S_d, S_v and S_m proposed in Section 2 and motivate natural choices for the matrix $\mathbf{G}$ defining the elliptic norm $\|\cdot\|_g$ and the bounds ρ_i, γ_i, $\overline{d}_i$, $i = 1,\ldots,n$.

In Section 2, the return vector $\mathbf{r}$ is assumed to be given by the linear model,

$$\mathbf{r} = \boldsymbol{\mu} + \mathbf{V}^T\mathbf{f} + \boldsymbol{\epsilon}, \tag{48}$$

where $\boldsymbol{\epsilon}$ is the residual return. In practice, the parameters $(\boldsymbol{\mu}, \mathbf{V})$ of this linear model are estimated by linear regression. Given market data consisting of samples of asset return vectors and the corresponding factor returns, the linear regression procedure computes the least squares estimates $(\boldsymbol{\mu}_0, \mathbf{V}_0)$ of $(\boldsymbol{\mu}, \mathbf{V})$. In addition, the procedure also gives multi-dimensional confidence regions around these estimates with the property that the true value of the parameters lie in these regions with a prescribed confidence level. The structure of the uncertainty sets introduced in Section 2 is motivated by these confidence regions. The rest of this section describes the steps involved in parameterizing the uncertainty structures, i.e. computing $\boldsymbol{\mu}_0$, $\mathbf{V}_0$, $\mathbf{G}$, ρ_i, γ_i, $\overline{d}_i$, $i = 1, \ldots, n$.

Suppose the market data consists of asset returns, $\{\mathbf{r}^t : t = 1, \ldots, p\}$, for p periods and the corresponding factor returns $\{\mathbf{f}^t : t = 1, \ldots, p\}$. Then the linear model (48) implies that

$$r_i^t = \mu_i + \sum_{j=1}^{n} V_{ji} f_j^t + \epsilon_i^t, \quad i = 1, \ldots, n, \quad t = 1, \ldots, p.$$

In linear regression analysis, typically, in addition to assuming that the vector of residual returns $\boldsymbol{\epsilon}^t$ in period t is composed of independent normals, it is assumed that the residual returns of different market periods are independent. Thus, $\{\epsilon_i^t\}$ are all independent normal random variables and $\epsilon_i^t \sim \mathcal{N}(0, \sigma_i^2)$, for all $t = 1, \ldots, p$, i.e. the variance of the residual return of the i-th asset is σ_i^2.

Let $\mathbf{S} = [\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^p] \in \mathbf{R}^{n \times p}$ be the matrix of asset returns and $\mathbf{B} = [\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^p] \in \mathbf{R}^{m \times p}$ be the matrix of factor returns. Collecting together terms corresponding to a particular asset i over all the periods $t = 1, \ldots, p$, we get the following linear model for the returns $\{r_i^t : t = 1, \ldots, p\}$

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \boldsymbol{\epsilon}_i,$$

where

$$\mathbf{y}_i = [r_i^1 \ r_i^2 \ \ldots \ r_i^p]^T, \quad \mathbf{A} = [\mathbf{1} \ \ \mathbf{B}^T], \quad \mathbf{x}_i = [\mu_i \ V_{1i} \ V_{2i} \ \ldots \ V_{mi}]^T,$$

and $\boldsymbol{\epsilon}_i = [e_i^1, \ldots, e_i^p]^T$ is the vector of residual returns corresponding to asset i.

The least squares estimate $\bar{\mathbf{x}}_i$ of the true parameter $\mathbf{x}_i$ is given by the solution of the normal equations

$$\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}}_i = \mathbf{A}^T\mathbf{y}_i;$$

i.e.

$$\bar{\mathbf{x}}_i = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}^i, \tag{49}$$

if $\mathbf{rank}(A) = m + 1$. Substituting $\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \boldsymbol{\epsilon}_i$, we get

$$\bar{\mathbf{x}}_i - \mathbf{x}_i = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\epsilon}_i = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\mathbf{\Sigma} = \sigma_i^2(\mathbf{A}^T\mathbf{A})^{-1}$. Hence

$$\mathcal{X} = \frac{1}{\sigma_i^2}(\bar{\mathbf{x}}_i - \mathbf{x}_i)^T(\mathbf{A}^T\mathbf{A})(\bar{\mathbf{x}}_i - \mathbf{x}_i) \sim \chi^2_{m+1}, \tag{50}$$

is a χ^2 random variable with $(m+1)$ degrees of freedom. Since the true variance σ_i^2 is unknown, (50) is not of much practical value. However, a standard result in regression theory states if σ_i^2 in the quadratic form (50) is replaced by $(m+1)s_i^2$, where s_i^2 is the unbiased estimate of σ_i^2 given by

$$s_i^2 = \frac{\|\mathbf{y}_i - \mathbf{A}\bar{\mathbf{x}}_i\|^2}{p-m-1}, \tag{51}$$

then the resulting random variable

$$\mathcal{Y} = \frac{1}{(m+1)s_i^2} \cdot (\bar{\mathbf{x}}_i - \mathbf{x}_i)^T(\mathbf{A}^T\mathbf{A})(\bar{\mathbf{x}}_i - \mathbf{x}_i) \tag{52}$$

is distributed according to the F-distribution with $(m+1)$ degrees of freedom in the numerator and $(p-m-1)$ degrees of freedom in the denominator [1, 12].

Let $0 < \omega < 1$, $\mathcal{F}_J$ denote the cumulative distribution function of the F-distribution with J degrees of freedom in the numerator and $p-m-1$ degrees of freedom in the denominator and $c_J(\omega)$ be the ω-critical value, i.e. the solution of the equation

$$\mathcal{F}_J(c_J(\omega)) = \omega.$$

Then the probability $\mathcal{Y} \leq c_{m+1}(\omega)$ is ω, or equivalently,

$$\mathbf{P}\left((\bar{\mathbf{x}}_i - \mathbf{x}_i)^T(\mathbf{A}^T\mathbf{A})(\bar{\mathbf{x}}_i - \mathbf{x}_i) \leq (m+1)c_{m+1}(\omega)s_i^2\right) = \omega. \tag{53}$$

Define

$$S_i(\omega) = \left\{\mathbf{x} : (\bar{\mathbf{x}}_i - \mathbf{x}_i)^T(\mathbf{A}^T\mathbf{A})(\bar{\mathbf{x}}_i - \mathbf{x}_i) \leq (m+1)c_{m+1}(\omega)s_i^2\right\}. \tag{54}$$

Then, (53) implies that $S_i(\omega)$ is a ω-confidence set for the parameter vector $\mathbf{x}_i$ corresponding to asset i. And since the residual errors $\{\boldsymbol{\epsilon}_i : i = 1, \ldots, n\}$ are assumed to be independent, it follows that

$$S(\omega) = S_1(\omega) \times S_2(\omega) \times \ldots \times S_n(\omega), \tag{55}$$

is a ω^n-confidence set for $(\boldsymbol{\mu}, \mathbf{V})$.

Let $S_m(\omega)$ denote the projection of S along the vector $\boldsymbol{\mu}$, i.e.

$$S_m(\omega) = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\nu}, |\nu_i| \leq \gamma_i, i = 1, \ldots, n\},$$

where

$$\mu_{0,i} = \bar{\mu}_i, \quad \gamma_i = \sqrt{(m+1)(\mathbf{A}^T\mathbf{A})^{-1}_{11}c_{m+1}(\omega)s_i^2}, \quad i = 1, \ldots, n. \tag{56}$$

Then (55) implies that $S_m(\omega)$ is a ω^n-confidence set for the mean vector $\boldsymbol{\mu}$. Note that the uncertainty structure (4) for the mean $\boldsymbol{\mu}$ assumed in Section 2 is identical to $S_m(\omega)$.

Let $\mathbf{Q} = [\mathbf{e}_2, \mathbf{e}_3, \ldots, \mathbf{e}_{m+1}]^T \in \mathbf{R}^{m\times(m+1)}$ be projection matrix that projects $\mathbf{x}_i$ along $\mathbf{V}_i$. Define the projection $S_v(\omega)$ of $S(\omega)$ along $\mathbf{V}$ as follows

$$S_v(\omega) = \left\{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}_i\|_g \leq \rho_i, i = 1, \ldots, n\right\},$$

where

$$\begin{aligned}
\mathbf{V}_0 &= [\bar{\mathbf{V}}_1 \ \ldots \ \bar{\mathbf{V}}_n], \\
\mathbf{G} &= (\mathbf{Q}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{Q}^T)^{-1} = \mathbf{B}\mathbf{B}^T - \tfrac{1}{p}(\mathbf{B}\mathbf{1})(\mathbf{B}\mathbf{1})^T, \\
\rho_i &= \sqrt{(m+1)c_{m+1}(\omega)s_i^2}, \quad i = 1, \ldots, n.
\end{aligned} \tag{57}$$

Then $S_v(\omega)$ is a ω^n-confidence set for the factor loading matrix $\mathbf{V}$. As in the case of $S_m(\omega)$, the uncertainty structure (3) for the factor covariance $\mathbf{V}$ assumed in Section 2 has precisely the same structure as $S_v(\omega)$ defined above.

The construction of the confidence regions can be done in the reverse direction as well, i.e. individual confidence regions can be suitably combined to yield joint confidence regions. Let $\tilde{S}_m \subset \mathbf{R}^n$ and $\tilde{S}_v \subset \mathbf{R}^{m\times m}$ be any $\tilde{\omega}$-confidence regions for $\boldsymbol{\mu}$ and $\mathbf{V}$ respectively. Then

$$\begin{aligned}
\mathbf{P}((\boldsymbol{\mu}, \mathbf{V}) \in \tilde{S}_m \times \tilde{S}_v) &= 1 - \mathbf{P}((\boldsymbol{\mu}, \mathbf{V}) \notin \tilde{S}_m \times \tilde{S}_v), \\
&\geq 1 - \mathbf{P}(\boldsymbol{\mu} \notin \tilde{S}_m) - \mathbf{P}(\mathbf{V} \notin \tilde{S}_v), \\
&= 2\tilde{\omega} - 1,
\end{aligned} \tag{58}$$

i.e. Cartesian products of individual confidence regions are joint confidence regions. This leads to an alternative means of constructing joint confidence regions for market parameters $(\boldsymbol{\mu}, \mathbf{V})$.

Let $\mathbf{Q} \in \mathbf{R}^{J\times m}$. Then

$$\mathcal{Y} = \frac{1}{Js_i^2}(\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\bar{\mathbf{x}}_i)^T(\mathbf{Q}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{Q}^T)^{-1}(\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\bar{\mathbf{x}}_i), \tag{59}$$

is distributed according to the F-distribution with J degrees of freedom in the numerator and $p-m-1$ degrees of freedom in the denominator, i.e. the probability $\mathcal{Y} \leq c_J(\tilde{\omega})$ is $\tilde{\omega}$, or equivalently,

$$\mathbf{P}\left((\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\bar{\mathbf{x}}_i)^T(\mathbf{Q}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{Q}^T)^{-1}(\mathbf{Q}\mathbf{x}_i - \mathbf{Q}\bar{\mathbf{x}}_i) \leq Jc_J(\tilde{\omega})s_i^2\right) = \tilde{\omega}. \tag{60}$$

Set $\mathbf{Q} = \mathbf{e}_1^T$. Then, $\mathbf{Q}\bar{\mathbf{x}}_i = \bar{\mu}_i$, the least squares estimate of the mean return of asset i and $\mathbf{Q}\mathbf{x}_i = \mu_i$, the true mean return of asset i. Therefore, (60) implies that

$$\mathbf{P}\left(|\bar{\mu}_i - \mu_i| \leq \sqrt{(\mathbf{A}^T\mathbf{A})_{11}^{-1}c_1(\tilde{\omega})s_i^2}\right) = \tilde{\omega}. \tag{61}$$

Since the residual errors $\boldsymbol{\epsilon}_i$ are assumed to be independent, it follows that

$$\tilde{S}_m(\tilde{\omega}) = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\nu}, |\nu_i| \leq \gamma_i, i = 1, \ldots, n\},$$

where

$$\mu_{0,i} = \bar{\mu}_i, \quad \gamma_i = \sqrt{(\mathbf{A}^T\mathbf{A})_{11}^{-1}c_1(\tilde{\omega})s_i^2}, \quad i = 1, \ldots, n, \tag{62}$$

is a $\tilde{\omega}^n$-confidence set for the mean vector $\boldsymbol{\mu}$.

Next set $Q = [\mathbf{e}_2, \mathbf{e}_3, \ldots, \mathbf{e}_{m+1}]^T \in \mathbf{R}^{m\times(m+1)}$. Then, the projection $\mathbf{Q}\bar{\mathbf{x}}_i = [\bar{V}_{1i}\ \bar{V}_{2i}\ \ldots\ \bar{V}_{im}]^T = \bar{\mathbf{V}}_i$, the least squares estimate of the true factor loading $\mathbf{Q}\mathbf{x}_i = \mathbf{V}_i$, and (60) implies that

$$\mathbf{P}\left((\bar{\mathbf{V}}_i - \mathbf{V}_i)^T(\mathbf{Q}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{Q}^T)^{-1}(\bar{\mathbf{V}}_i - \mathbf{V}_i) \le mc_m(\tilde{\omega})s_i^2\right) = \tilde{\omega}.$$

As in the case of the mean vector $\boldsymbol{\mu}$, it follows that

$$\begin{aligned}\tilde{S}_v(\tilde{\omega}) &= \left\{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}_i\|_g \le \rho_i, i = 1, \ldots, n\right\}, \\ &\mathbf{V}_0 = \bar{\mathbf{V}}, \quad \mathbf{G} = (\mathbf{Q}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{Q}^T)^{-1} = \mathbf{B}\mathbf{B}^T - \tfrac{1}{p}(\mathbf{B}\mathbf{1})(\mathbf{B}\mathbf{1})^T, \\ &\rho_i = \sqrt{mc_m(\tilde{\omega})s_i^2}, \quad i = 1, \ldots, n,\end{aligned} \tag{63}$$

is $\tilde{\omega}^n$-confidence set for the factor loading matrix $\mathbf{V}$. From (58), it follows that $\tilde{S}(\omega) = \tilde{S}_m(\tilde{\omega})\times\tilde{S}_v(\tilde{\omega})$ is a joint confidence region with $(2\tilde{\omega}^n - 1)$ confidence. The confidence region defined by (56)-(57) typically contains the one defined by (62)-(63).

The two families of uncertainty sets defined by (56)–(57) and (62)–(63) are not entirely comparable. Since the critical value $c_{J,m}(\omega)$ is an increasing function of J, it follows that, for a fixed $\omega \in (0, 1)$, $\tilde{S}(\omega) \subset S(\omega)$. On the other hand, if $\tilde{\omega}$ is chosen such that $2\tilde{\omega}^n - 1 = \omega^n$, the direction of inclusion depends on the dimension m.

After determining the uncertainty sets for $\boldsymbol{\mu}$ and $\mathbf{V}$, all that remains to be established is the bound $\overline{d}_i$ on the variance of regression error σ_i^2. From the analysis above, it follows that the natural choices for the bounds are given by the confidence interval around the least squares estimate s_i^2 of the error variance σ_i^2. Unfortunately, the regression procedure yields only a single unbiased sample s_i^2 of the error variance. A possible solution is to construct bootstrap confidence intervals [8]. Since the bootstrapping step can often be computationally expensive and since the robust optimization problems only require an estimate of the worst case error variance, any reasonable estimate of the worst case error variance could alternatively be used as the bound $\overline{d}_i$.

Incorporating the results developed in this section, we have the following recipe for solving the robust portfolio selection problem:

1. Collect data on the returns $\mathbf{r}$ of the assets and the returns $\mathbf{f}$ of the factors.
2. Using (49) one asset at a time, evaluate the least squares estimates $\boldsymbol{\mu}_0$ and $\mathbf{V}_0$ of the mean $\boldsymbol{\mu}$ and the factor loading matrix $\mathbf{V}$, respectively.
3. Choose a confidence threshold ω.
   (a) Construct a bootstrap ω-confidence interval around σ_i^2. Alternatively use any estimate of the worst case error variance.
   (b) Define S_m and S_v using (56) and (57) respectively (or (62) and (63) respectively).

4. Solve the robust problem of interest.

A word of caution about the choice of ω. If ω is chosen very high, the uncertainty sets will be very large, i.e. implicitly one is demanding robustness with respect to a very large set of parameter values. The resulting portfolio will be very conservative, and therefore its performance for a particular set of parameters $(\boldsymbol{\mu}_0, \mathbf{V}_0)$ will be significantly worse than the portfolio designed for that set of parameters. On the other hand, if ω is chosen too low, the portfolio choice will not be robust enough. The typical choices of ω lie in the range 0.95-0.99. See Section 7 for more details on the implications of the choice of ω and Appendix A for a discussion of the probabilistic guarantees on the performance of the optimal robust portfolio.

Recall that the maximum likelihood estimate $\mathbf{F}_{ml}$ of covariance matrix $\mathbf{F}$ of the factors is given by,

$$\mathbf{F}_{ml} = \frac{1}{p-1}\left[\mathbf{B}\mathbf{B}^T - \frac{1}{p}(\mathbf{B}\mathbf{1})(\mathbf{B}\mathbf{1})^T\right], \tag{64}$$

i.e. $\mathbf{G} = (p-1)\mathbf{F}_{ml}$. Suppose the true covariance matrix $\mathbf{F}$ is approximated by the maximum likelihood estimate $\mathbf{F}_{ml}$. Then $\mathbf{F} = \kappa\mathbf{G}$, where $\kappa = 1/(p-1)$. Recall that the worst case variance

$$\max_{\mathbf{V}\in S_v}\left\{\boldsymbol{\phi}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi}\right\} \leq \nu^2, \tag{65}$$

if and only if

$$\max_{\{\mathbf{y}:\|\mathbf{y}\|_g\leq r\}} \|\mathbf{V}_0\boldsymbol{\phi} + \mathbf{y}\|_f \leq \nu,$$

where $r = \boldsymbol{\rho}^T\boldsymbol{\phi}$. Since $\mathbf{F} = \kappa\mathbf{G}$, the above norm constraint is equivalent to

$$\max_{\{\mathbf{u}:\|\mathbf{u}\|\leq r\sqrt{\kappa}\}} \left\|\mathbf{F}^{\frac{1}{2}}\mathbf{V}_0\boldsymbol{\phi} + \mathbf{u}\right\| \leq \nu, \tag{66}$$

where $\|\cdot\|$ is now the usual Euclidean norm. It is easy to see that the maximum in (66) is attained at

$$\mathbf{u} = \frac{r\sqrt{\kappa}}{\|\mathbf{F}^{\frac{1}{2}}\mathbf{V}_0\boldsymbol{\phi}\|}\cdot\mathbf{F}^{\frac{1}{2}}\mathbf{V}_0\boldsymbol{\phi}$$

and the corresponding maximum value is $\left\|\mathbf{F}^{\frac{1}{2}}\mathbf{V}_0\boldsymbol{\phi}\right\| + r\sqrt{\kappa}$. Thus, the worst case variance constraint (65) is equivalent to the second-order cone constraint

$$\left\|\mathbf{F}^{\frac{1}{2}}\mathbf{V}_0\boldsymbol{\phi}\right\| \leq \nu - \sqrt{\kappa}\boldsymbol{\rho}^T\boldsymbol{\phi}. \tag{67}$$

From the fact that $\nu^2 \leq \tau$ is equivalent to a second-order cone constraint, it follows that in the special case $\mathbf{F} = \kappa\mathbf{G}$ the robust minimum variance reduces to the following simple SOCP

$$\begin{array}{lrcl}
\text{minimize} & \tau + \delta, & & \\
\text{subject to} & (\boldsymbol{\mu} - \boldsymbol{\gamma})^T \boldsymbol{\phi} & \geq & \alpha, \\
& \mathbf{1}^T \boldsymbol{\phi} & = & 1, \\
& \left\| \mathbf{F}^{\frac{1}{2}} \mathbf{V}_0 \boldsymbol{\phi} \right\| & \leq & \nu - \sqrt{\kappa} \boldsymbol{\rho}^T \boldsymbol{\phi}, \\
& \left\| \begin{bmatrix} 2\overline{\mathbf{D}}^{\frac{1}{2}} \boldsymbol{\phi} \\ 1 - \delta \end{bmatrix} \right\| & \leq & 1 + \delta, \\
& \left\| \begin{bmatrix} 2\nu \\ 1 - \tau \end{bmatrix} \right\| & \leq & 1 + \tau, \\
& \boldsymbol{\phi} & \geq & \mathbf{0}.
\end{array} \tag{68}$$

The robust maximum return problem, the robust maximum Sharpe ratio, as well as the robust VaR problem can all be simplified in a similar manner.

In practice, however, the covariance matrix $\mathbf{F}$ is assumed to be stable, and is, typically, estimated from a much larger data set and by taking several extraneous macroeconomic indicators into account [16]. If this is the case, then the above simplification cannot be used and one would have to revert back to the formulation in (31).

6 Robust portfolio allocation with uncertain covariance matrices

The market model introduced in Section 2 assumes that the covariance matrix $\mathbf{F}$ of the factors is completely known and stable. While this is a good first approximation, a more complete market model is one that allows some uncertainty in the covariance matrix and optimizes over it. In order for any uncertainty structure for covariance matrices to be useful it must be flexible enough to model a variety of perturbations, while at the same time restricted enough to admit fast parameterization and efficient optimization. Our goal in this section is to develop such an uncertainty structure for covariance matrices. We assume the market structure is the one introduced in Section 2, except that the covariance matrix is no longer fixed.

6.1 Uncertainty structure for covariance inverse

Consider the following uncertainty structure for the factor covariance matrix $\mathbf{F}$:

$$S_{f^{-1}} = \left\{ \mathbf{F} : \mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \boldsymbol{\Delta} \succeq 0, \boldsymbol{\Delta} = \boldsymbol{\Delta}^T, \|\mathbf{F}_0^{\frac{1}{2}} \boldsymbol{\Delta} \mathbf{F}_0^{\frac{1}{2}}\| \leq \eta \right\}, \tag{69}$$

where $\mathbf{F}_0 \succ \mathbf{0}$. The norm $\|\mathbf{A}\|$ in (69) is given by $\|\mathbf{A}\| = \max_i |\lambda_i(A)|$, where $\{\lambda_i(A)\}$ are the eigenvalues of $\mathbf{A}$.

The uncertainty set $S_{f^{-1}}$ restricts the perturbations $\boldsymbol{\Delta}$ of the covariance matrix to be symmetric, bounded in norm relative to a nominal covariance matrix $\mathbf{F}_0$ and subject to $\mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \boldsymbol{\Delta} \succeq 0$.

Clearly this uncertainty structure is not the most general – one could, for example, allow for each element of the covariance matrix to lie in some uncertainty interval subject to the constraint that the matrix is positive semidefinite. However, the robust optimization problem corresponding to this uncertainty structure is not very tractable [13]. On the other hand, if the covariance uncertainty is given by (69), all the robust portfolio allocation problems introduced in Section 2 can be reformulated as SOCPs. Moreover, as in the case of the uncertainty structures for $\boldsymbol{\mu}$ and $\mathbf{V}$, the uncertainty structure $S_{f^{-1}}$ for $\mathbf{F}$ corresponds to the confidence region associated with statistical procedure used to estimate $\mathbf{F}$. In particular, we show that the maximum likelihood estimation of the covariance matrix $\mathbf{F}$ provides a confidence region of the form $S_{f^{-1}}$ and yields a value of η that reflects any desired confidence level.

All the robust portfolio selection problems introduced in Section 2 can be formulated for this market as well. For example, the robust Sharpe ratio problem in this market model is given by,

$$\text{maximize}_{\{\boldsymbol{\phi}:\mathbf{1}^T\boldsymbol{\phi}=1\}} \quad \min_{\{\boldsymbol{\mu}\in S_m,\mathbf{V}\in S_v,\mathbf{D}\in S_d,\mathbf{F}\in S_f\}} \left\{ \frac{\boldsymbol{\mu}^T\boldsymbol{\phi}-r_f}{\sqrt{\boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V}+\mathbf{D})\boldsymbol{\phi}}} \right\}.$$

Lemma 3, below, establishes that the worst case variance for a fixed portfolio is given by a collection of linear and restricted hyperbolic constraints – the critical step in reformulating robust portfolio selection problems as SOCPs.

Lemma 3 *Fix a portfolio $\boldsymbol{\phi}$ and let $S_{f^{-1}} = \left\{\mathbf{F}\middle|\mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \boldsymbol{\Delta} \succeq \mathbf{0}, \boldsymbol{\Delta} = \boldsymbol{\Delta}^T, \|\mathbf{F}_0^{\frac{1}{2}}\boldsymbol{\Delta}\mathbf{F}_0^{\frac{1}{2}}\| \leq \eta\right\}$, where $\mathbf{F}_0 \succ \mathbf{0}$.*

(i) If the bound $\eta \geq 1$, the worst-case variance is unbounded, i.e.

$$\max_{\{\mathbf{V}\in S_v,\mathbf{F}\in S_{f-1}\}} \boldsymbol{\phi}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi} = \infty.$$

(ii) If $\eta < 1$, the worst-case variance $\max_{\{\mathbf{V}\in S_v,\mathbf{F}\in S_{f-1}\}} \boldsymbol{\phi}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi} \leq \nu$ if and only if $(\boldsymbol{\rho}^T\boldsymbol{\phi}; (1-\eta)\nu; \boldsymbol{\phi}) \in \mathcal{H}(\mathbf{V}_0, \mathbf{F}_0, \mathbf{G})$, where $\mathcal{H}(\cdot,\cdot,\cdot)$ is defined in Definition 1.

Proof: Define $\tilde{\boldsymbol{\Delta}} = \mathbf{F}_0^{\frac{1}{2}}\boldsymbol{\Delta}\mathbf{F}_0^{\frac{1}{2}}$. Then

$$S_{f^{-1}} = \left\{\mathbf{F} : \mathbf{F}^{-1} = \mathbf{F}_0^{-\frac{1}{2}}(\mathbf{I}+\tilde{\boldsymbol{\Delta}})\mathbf{F}_0^{-\frac{1}{2}} \succeq \mathbf{0}, \tilde{\boldsymbol{\Delta}} = \tilde{\boldsymbol{\Delta}}^T, \|\tilde{\boldsymbol{\Delta}}\| \leq \eta\right\}.$$

Fix $\mathbf{V} \in S_v$ and define $\mathbf{x} = \mathbf{V}^T\boldsymbol{\phi}$. Then

$$\sup_{\mathbf{F}\in S_{f-1}} \{\mathbf{x}^T\mathbf{F}\mathbf{x}\} = \sup\left\{(\mathbf{F}_0^{\frac{1}{2}}\mathbf{x})^T(\mathbf{I}+\tilde{\boldsymbol{\Delta}})^{-1}(\mathbf{F}_0^{\frac{1}{2}}\mathbf{x}) : \ \tilde{\boldsymbol{\Delta}} = \tilde{\boldsymbol{\Delta}}^T, \ \|\tilde{\boldsymbol{\Delta}}\| \leq \eta, \ \mathbf{I}+\tilde{\boldsymbol{\Delta}} \succeq \mathbf{0}\right\}.$$

First consider the case $\eta \geq 1$. Choose $\tilde{\boldsymbol{\Delta}} = -\beta\mathbf{I}$ where $0 \leq \beta \leq 1$. Then $\|\tilde{\boldsymbol{\Delta}}\| = \beta \leq \eta$ and $\mathbf{I}+\tilde{\boldsymbol{\Delta}} = (1-\beta)\mathbf{I} \succeq \mathbf{0}$. Thus,

$$\sup_{\mathbf{F}\in S_{f-1}} \{\mathbf{x}^T\mathbf{F}\mathbf{x}\} \geq \sup_{\{\beta:0\leq\beta\leq1\}} \frac{1}{1-\beta}\left(\mathbf{x}^T\mathbf{F}_0\mathbf{x}\right) = \infty.$$

Next, suppose $\eta < 1$. Since $\eta < 1$ implies that $\mathbf{I} + \tilde{\mathbf{\Delta}} \succeq \mathbf{0}$ for all $\|\tilde{\mathbf{\Delta}}\| \leq \eta$,

$$S_{f^{-1}} = \left\{ \mathbf{F} : \mathbf{F}^{-1} = \mathbf{F}_0^{-\frac{1}{2}}(\mathbf{I} + \tilde{\mathbf{\Delta}})\mathbf{F}_0^{-\frac{1}{2}}, \tilde{\mathbf{\Delta}} = \tilde{\mathbf{\Delta}}^T, \|\tilde{\mathbf{\Delta}}\| \leq \eta \right\}.$$

Therefore

$$\begin{aligned}
\sup_{\mathbf{F} \in S_{f-1}} \{\mathbf{x}^T \mathbf{F} \mathbf{x}\} &= \sup \left\{ (\mathbf{F}_0^{\frac{1}{2}}\mathbf{x})(\mathbf{I} + \tilde{\mathbf{\Delta}})^{-1}(\mathbf{F}_0^{\frac{1}{2}}\mathbf{x}) : \tilde{\mathbf{\Delta}} = \tilde{\mathbf{\Delta}}^T, \|\tilde{\mathbf{\Delta}}\| \leq \eta \right\} \\
&= \sum_i \frac{(\mathbf{q}_i^T(\mathbf{F}_0^{\frac{1}{2}}\mathbf{x}))^2}{1 + \lambda_i(\tilde{\mathbf{\Delta}})},
\end{aligned}$$

where $\{\mathbf{q}_i, i = 1, \ldots, m\}$ is the set of eigenvectors of $\tilde{\mathbf{\Delta}}$. Since $|\lambda_i(\tilde{\mathbf{\Delta}})| \leq \eta$, it follows that

$$\sup_{\mathbf{F} \in S_{f-1}} \{\mathbf{x}^T \mathbf{F} \mathbf{x}\} \leq \frac{1}{1-\eta} \cdot \sum_i \left(\mathbf{q}_i^T(\mathbf{F}_0^{\frac{1}{2}}\mathbf{x}) \right)^2 = \frac{1}{1-\eta} \cdot \mathbf{x}^T \mathbf{F}_0 \mathbf{x}.$$

Moreover, the supremum is achieved by $\tilde{\mathbf{\Delta}} = -\eta \mathbf{I}$. Thus,

$$\max_{\{\mathbf{V} \in S_v, \mathbf{F} \in S_{f-1}\}} \boldsymbol{\phi}^T(\mathbf{V}^T \mathbf{F} \mathbf{V})\boldsymbol{\phi} = \frac{1}{1-\eta} \cdot \max_{\mathbf{V} \in S_v} \boldsymbol{\phi}^T \mathbf{V}^T \mathbf{F}_0 \mathbf{V} \boldsymbol{\phi}.$$

From Part (i) of Lemma 1, we have that

$$\frac{1}{1-\eta} \cdot \max_{\mathbf{V} \in S_v} \boldsymbol{\phi}^T \mathbf{V}^T \mathbf{F}_0 \mathbf{V} \boldsymbol{\phi} \leq \nu$$

if and only if $(\boldsymbol{\rho}^T \boldsymbol{\phi}; (1-\eta)\nu; \boldsymbol{\phi}) \in \mathcal{H}(\mathbf{V}_0, \mathbf{F}_0, \mathbf{G})$. ■

From Lemma 3 it follows that if $\eta < 1$ then all robust portfolio problems introduced in Section 2 can be reformulated as SOCPs.

The next step is to justify the uncertainty structure $S_{f^{-1}}$ and show how it can be parametrized from market data. In Section 5 we show that the natural uncertainty structures for $(\boldsymbol{\mu}, \mathbf{V})$ and their parameterizations are implied by the confidence regions associate with the statistical procedures used to compute the point estimates. Here, we show that the uncertainty structure $S_{f^{-1}}$ and its parametrization, i.e. the choice of the nominal matrix $\mathbf{F}_0$ and the bound η, arise naturally from maximum-likelihood estimation typically used to compute the point estimate of $\mathbf{F}$.

Suppose the covariance matrix of the factor returns is estimated from the data $\mathbf{B} = [\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^p]$ over p market periods. Then the maximum likelihood estimate $\mathbf{F}_{ml}$ of the factor covariance matrix $\mathbf{F}$ is given by (64).

Suppose one is in a Bayesian setup and assumes a non-informative conjugate prior distribution for the true covariance $\mathbf{F}$. Then the posterior distribution for $\mathbf{F}$ conditioned on the data $\mathbf{B}$ is given by

$$\mathbf{F}|\mathbf{B} \sim W_{p-1}^{-1}((p-1)\mathbf{F}_{ml}), \tag{70}$$

Figure 1: $\omega_{\max}$ vs p ($m = 40$)

where $W_q^{-1}(\mathbf{A})$ denotes an inverse Wishart distribution with q degrees of freedom and parameter $\mathbf{A}$ [1, 25], i.e. the density $f(\mathbf{F}|\mathbf{B})$ is given by

$$f(\mathbf{F}|\mathbf{B}) = \begin{cases} c|\mathbf{F}_{ml}|^{(p-m-2)/2}|\mathbf{F}^{-1}|^{(p-1)/2}e^{-\frac{(p-1)}{2}\mathbf{Tr}(\mathbf{F}^{-1}\mathbf{F}_{ml})}, & \mathbf{F} \succ \mathbf{0}, \\ 0, & \text{otherwise}, \end{cases} \tag{71}$$

where $|\mathbf{A}| = \mathbf{det}(\mathbf{A})$ and c is a normalization constant. This density can be rewritten as follows.

$$f(\mathbf{F}|\mathbf{B}) = \begin{cases} c|\mathbf{F}_{ml}^{-1}|^{(m-1)/2}|\mathbf{F}_{ml}^{\frac{1}{2}}\mathbf{F}^{-1}\mathbf{F}_{ml}^{\frac{1}{2}}|^{(p-1)/2}e^{-\frac{(p-1)}{2}\mathbf{Tr}(\mathbf{F}_{ml}^{\frac{1}{2}}\mathbf{F}^{-1}\mathbf{F}_{ml}^{\frac{1}{2}})}, & \mathbf{F} \succ \mathbf{0}, \\ 0, & \text{otherwise}. \end{cases} \tag{72}$$

From (72) it follows that the natural choice for the nominal covariance matrix $\mathbf{F}_0$ in the definition of $S_{f^{-1}}$ is $\mathbf{F}_0 = \mathbf{F}_{ml}$.

Let $\tilde{\boldsymbol{\lambda}} \in \mathbf{R}^m$ be the vector of eigenvalues of $\mathbf{F}_0^{\frac{1}{2}}\mathbf{F}^{-1}\mathbf{F}_0^{\frac{1}{2}}$. Then (72) implies that the density of $\tilde{\boldsymbol{\lambda}}$ is given by

$$f(\tilde{\boldsymbol{\lambda}}|\mathbf{B}) = \begin{cases} c_\lambda \prod_{i=1}^m \tilde{\lambda}_i^{(p-1)/2}e^{-\frac{(p-1)}{2}\tilde{\lambda}_i}, & \tilde{\boldsymbol{\lambda}} \geq \mathbf{0}, \\ 0, & \text{otherwise}, \end{cases} \tag{73}$$

where c_λ is a normalization constant, i.e. the eigenvalues $\tilde{\lambda}_i$, $i = 1, \ldots, m$, are IID $\Gamma\left(\frac{p+1}{2}, \frac{p-1}{2}\right)$ distributed random variables.

Let $\boldsymbol{\Delta} = \mathbf{F}^{-1} - \mathbf{F}_0^{-1}$ be the deviation of the $\mathbf{F}^{-1}$ from the nominal inverse $\mathbf{F}_0^{-1}$ and let $\tilde{\boldsymbol{\Delta}} = \mathbf{F}_0^{\frac{1}{2}}\boldsymbol{\Delta}\mathbf{F}_0^{\frac{1}{2}} = \mathbf{F}_0^{\frac{1}{2}}\mathbf{F}^{-1}\mathbf{F}_0^{\frac{1}{2}} - \mathbf{I}$. Then $\|\tilde{\boldsymbol{\Delta}}\| \leq \eta$ if and only if $1 - \eta \leq \tilde{\lambda}_i \leq 1 + \eta$, i.e.

$$\mathbf{P}(\|\tilde{\boldsymbol{\Delta}}\| \leq \eta) = (\mathbf{P}(1 - \eta \leq \mathcal{G}_p \leq 1 + \eta))^m = \Big(\mathcal{F}_{\Gamma_p}(1 + \eta) - \mathcal{F}_{\Gamma_p}(1 - \eta)\Big)^m, \tag{74}$$

where $\mathcal{G}_p \sim \Gamma\left(\frac{p+1}{2}, \frac{p-1}{2}\right)$ and $\mathcal{F}_{\Gamma_p}$ is the corresponding CDF.

If $\eta < 1$, $\mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \boldsymbol{\Delta} \succ \mathbf{0}$ for all $\|\tilde{\boldsymbol{\Delta}}\| \leq \eta$, and therefore (74) implies that

$$\mathbf{P}(\mathbf{F} \in S_{f^{-1}}) = (\mathcal{F}_{\Gamma_p}(1 + \eta) - \mathcal{F}_{\Gamma_p}(1 - \eta))^m. \tag{75}$$

Thus, in order to satisfy a desired confidence level ω^m, the parameter η must be set equal to the unique solution of

$$\mathcal{F}_{\Gamma_p}(1 + \eta) - \mathcal{F}_{\Gamma_p}(1 - \eta) = \omega. \tag{76}$$

Since the problem is unbounded for $\eta \geq 1$, (76) implicitly states that the confidence level ω^m that can be supported by p return samples is at most $\omega_{\max}(p) = \Big(\mathcal{F}_{\Gamma_p}(2)\Big)^m$. Figure 1 plots $\omega_{\max}$ as a function of the data length p for $m = 40$ (the plot begins at $p = m + 1$ since at least $m + 1$ observations are needed to estimate the covariance matrix). From the plot, one can observe that for $p \geq 50$, $\omega_{\max}(p) \geq 0.995$. Thus, the restriction $\omega^m \leq \omega_{\max}(p)$ is not likely to be restrictive.

This methodology can be modified to accommodate prior information about the structure of $\mathbf{F}$. Suppose the prior distribution on $\mathbf{F}$ is the *informative* conjugate prior $W_k^{-1}((k-1)\bar{\mathbf{F}})$, $k \geq 1$. Then, the posterior distribution $\mathbf{F}|\mathbf{B} \sim W_{k+p}^{-1}((p-1)\mathbf{F}_{ml} + (k-1)\bar{\mathbf{F}})$. In this case, the nominal covariance matrix $\mathbf{F}_0 = \frac{1}{k+p}((p-1)\mathbf{F}_{ml} + (k-1)\bar{\mathbf{F}})$, and η is given by

$$\mathcal{F}_{\Gamma_{k+p}}(1+\eta) - \mathcal{F}_{\Gamma_{k+p}}(1-\eta) = \omega,$$

where $\mathcal{F}_{\Gamma_{k+p}}$ denotes the CDF of a $\Gamma\Big(\frac{k+p+2}{2}, \frac{k+p}{2}\Big)$ random variable.

From the results in this section, it follows that, as far as the portfolio selection problems are concerned, all that the uncertainty in the covariance matrix does is "shrink" the MLE $\mathbf{F}_0^{-1}$ to $(1-\eta)\mathbf{F}_0^{-1}$, where the shrinkage factor $(1-\eta)$ is a function of the desired confidence level ω. Thus, this procedure has the flavor of robust statistics [14].

6.2 Uncertainty structure for the covariance

Another uncertainty structure for the factor covariance matrix is

$$S_f = \Big\{\mathbf{F} : \mathbf{F} = \mathbf{F}_0 + \mathbf{\Delta} \succeq 0, \mathbf{\Delta} = \mathbf{\Delta}^T, \|\mathbf{N}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{-\frac{1}{2}}\| \leq \zeta\Big\}, \tag{77}$$

where $\mathbf{F}_0 \succeq \mathbf{0}$. The norm $\|\mathbf{A}\|$ in (77) is given by $\|\mathbf{A}\| = \max_i |\lambda_i(A)|$, or $\|\mathbf{A}\| = \sqrt{\sum_i \lambda_i^2(\mathbf{A})}$, where $\{\lambda_i(A)\}$ are the eigenvalues of $\mathbf{A}$.

In order for (77) to be a viable uncertainty structure, the corresponding robust problem should be efficiently solvable and the structure ought to be easily parameterizable from raw market data. The first requirement is established by the following lemma.

Lemma 4 *Fix a portfolio $\boldsymbol{\phi} \in \mathbf{R}_+^n$. Then $\max_{\{\mathbf{V}\in S_v, \mathbf{F}\in S_f\}} \boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V})\boldsymbol{\phi} \leq \nu$ if and only if $(\boldsymbol{\rho}^T\boldsymbol{\phi}; \nu; \boldsymbol{\phi}) \in \mathcal{H}(\mathbf{V}_0, (\mathbf{F}_0 + \zeta\mathbf{N}), \mathbf{G})$, where $\mathcal{H}(\cdot,\cdot,\cdot)$ is defined in Definition 1.*

Proof: Define $\tilde{\mathbf{\Delta}} = \mathbf{N}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{-\frac{1}{2}}$. Then

$$S_f = \Big\{\mathbf{F} : \mathbf{F} = \mathbf{F}_0 + \mathbf{N}^{\frac{1}{2}}\tilde{\mathbf{\Delta}}\mathbf{N}^{\frac{1}{2}} \succeq \mathbf{0}, \tilde{\mathbf{\Delta}} = \tilde{\mathbf{\Delta}}^T, \|\tilde{\mathbf{\Delta}}\| \leq \zeta\Big\}.$$

Fix $\mathbf{V} \in S_v$ and define $\mathbf{x} = \mathbf{V}^T\boldsymbol{\phi}$. Then

$$\begin{aligned}
\max_{\mathbf{F}\in S_f}\{\mathbf{x}^T\mathbf{F}\mathbf{x}\} &= \max\Big\{\mathbf{x}^T\mathbf{F}_0\mathbf{x} + (\mathbf{N}^{\frac{1}{2}}\mathbf{x})^T\tilde{\mathbf{\Delta}}(\mathbf{N}^{\frac{1}{2}}\mathbf{x}) : \ \tilde{\mathbf{\Delta}} = \tilde{\mathbf{\Delta}}^T, \|\tilde{\mathbf{\Delta}}\| \leq \zeta, \mathbf{F}_0 + \mathbf{N}^{\frac{1}{2}}\tilde{\mathbf{\Delta}}\mathbf{N}^{\frac{1}{2}} \succeq \mathbf{0}\Big\}, \\
&\leq \Big\{\mathbf{x}^T\mathbf{F}_0\mathbf{x} + (\mathbf{N}^{\frac{1}{2}}\mathbf{x})^T\tilde{\mathbf{\Delta}}(\mathbf{N}^{\frac{1}{2}}\mathbf{x}) : \ \tilde{\mathbf{\Delta}} = \tilde{\mathbf{\Delta}}^T, \|\tilde{\mathbf{\Delta}}\| \leq \zeta\Big\}, & (78) \\
&\leq \mathbf{x}^T\mathbf{F}_0\mathbf{x} + \zeta(\mathbf{N}^{\frac{1}{2}}\mathbf{x})^T(\mathbf{N}^{\frac{1}{2}}\mathbf{x}), & (79)
\end{aligned}$$

where (79) follows from the properties of the matrix norm.

Since $\|\tilde{\mathbf{\Delta}}\| = \max\{|\lambda_i(\tilde{\mathbf{\Delta}})|\}$ or $\sqrt{\sum_i \lambda_i^2(\tilde{\mathbf{\Delta}})}$ and $\mathbf{N} \succ \mathbf{0}$, the bound (79) is achieved by

$$\tilde{\mathbf{\Delta}}^* = \zeta \cdot \frac{(\mathbf{N}^{\frac{1}{2}}\mathbf{x})(\mathbf{N}^{\frac{1}{2}}\mathbf{x})^T}{\|\mathbf{N}^{\frac{1}{2}}\mathbf{x}\|^2},$$

unless $\mathbf{x} = \mathbf{0}$. Thus, the right hand side of (78) is given by $\mathbf{x}^T(\mathbf{F}_0 + \zeta\mathbf{N})\mathbf{x}$ and is achieved by $\tilde{\mathbf{\Delta}}^* = \zeta \cdot \frac{\mathbf{N}\mathbf{x}\mathbf{x}^T\mathbf{N}}{\mathbf{x}^T\mathbf{N}\mathbf{x}}$, unless $\mathbf{x} = \mathbf{0}$. Since $\mathbf{F}_0 + \mathbf{N}^{\frac{1}{2}}\tilde{\mathbf{\Delta}}^*\mathbf{N}^{\frac{1}{2}} \succeq \mathbf{0}$, it follows that the inequality (78) is, in fact, an equality, i.e.

$$\max_{\mathbf{F}\in S_f} \{\mathbf{x}^T\mathbf{F}\mathbf{x}\} = \mathbf{x}^T(\mathbf{F}_0 + \zeta\mathbf{N})\mathbf{x}.$$

Thus,

$$\max_{\{\mathbf{V}\in S_v, \mathbf{F}\in S_f\}} \boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V})\boldsymbol{\phi} = \max_{\mathbf{V}\in S_v} \boldsymbol{\phi}^T\mathbf{V}^T(\mathbf{F}_0 + \zeta\mathbf{N})\mathbf{V}\boldsymbol{\phi}.$$

From Lemma 1, we have that

$$\max_{\mathbf{V}\in S_v} \boldsymbol{\phi}^T\mathbf{V}^T(\mathbf{F}_0 + \zeta\mathbf{N})\mathbf{V}\boldsymbol{\phi} \le \nu$$

if and only if $(\boldsymbol{\rho}^T\boldsymbol{\phi}; \nu; \boldsymbol{\phi}) \in \mathcal{H}(\mathbf{V}_0, (\mathbf{F}_0 + \zeta\mathbf{N}), \mathbf{G})$ where $\mathcal{H}(\cdot,\cdot,\cdot)$ is defined in Definition 1. ■

Notice that, unlike in the case $S_{f^{-1}}$, here we allow $\mathbf{N} \neq \mathbf{F}_0$ and do not require $\zeta < 1$.

The next step is to devise a method for parameterizing this uncertainty structure. The parametrization will once again be implied by the statistical procedure used to estimate $\mathbf{F}$.

Let $\mathbf{F}$ be the true (unknown) covariance matrix of the factor returns and let $\mathbf{F}_{ml}$ be the MLE of $\mathbf{F}$ computed from the return data $\mathbf{B}$. Then it is well known (see [1]) that if $p > m$ (i.e. the number of observations is larger than the dimension of the matrix) then $\mathbf{F}_{ml} \succ \mathbf{0}$ with probability 1 and

$$\mathbf{F}_{ml} \sim W_{p-1}((p-1)\mathbf{F}), \tag{80}$$

where $W_q(\mathbf{A})$ denotes a Wishart distribution with q degrees of freedom centered at the matrix $\mathbf{A}$, i.e. the density $f(\mathbf{F}_{ml} \mid \mathbf{F})$ is given by

$$f(\mathbf{F}_{ml}|\mathbf{F}) = \begin{cases} c|(\mathbf{F})^{-1}|^{(p-1)/2}|\mathbf{F}_{ml}|^{(p-m-2)/2}e^{-\frac{(p-1)}{2}\mathbf{Tr}(\mathbf{F}^{-1}\mathbf{F}_{ml})}, & \mathbf{F}_{ml} \succ \mathbf{0}, \\ 0, & \text{otherwise.} \end{cases} \tag{81}$$

An argument very similar to the one in the previous section establishes that $\mathbf{F}_{ml}$ belongs to the set

$$\tilde{\mathcal{S}} = \{\tilde{\mathbf{F}} : \tilde{\mathbf{F}} = \mathbf{F} + \mathbf{\Delta} \succ \mathbf{0}, \mathbf{\Delta} = \mathbf{\Delta}^T, \|\mathbf{F}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{F}^{-\frac{1}{2}}\| \le \beta\}, \tag{82}$$

with probability ω^m if the bound β is set equal to the solution of

$$\mathcal{F}_{\Gamma_p}(1+\beta) - \mathcal{F}_{\Gamma_p}(1-\beta) = \omega, \tag{83}$$

where $\mathcal{F}_{\Gamma_p}$ denotes the CDF of a $\Gamma((p+1)/2, (p-1)/2)$ random variable. Note that this equation is identical to (76), i.e. the solution β of (83) is equal to the solution η of (76).

Suppose the bound $\beta < 1$, or equivalently $\omega^m < \omega_{\max}(p) = (\mathcal{F}_{\Gamma_p}(2))^m$. Then we have that

$$\begin{aligned} \|\mathbf{F}_{ml}^{-\frac{1}{2}}(\mathbf{F}-\mathbf{F}_{ml})\mathbf{F}_{ml}^{-\frac{1}{2}}\| &= \|\mathbf{F}_{ml}^{-\frac{1}{2}}\mathbf{F}^{\frac{1}{2}}\mathbf{F}^{\frac{1}{2}}(\mathbf{F}-\mathbf{F}_{ml})\mathbf{F}^{-\frac{1}{2}}\mathbf{F}^{\frac{1}{2}}\mathbf{F}_{ml}^{-\frac{1}{2}}\|, \\ &\leq \|\mathbf{F}_{ml}^{-\frac{1}{2}}\mathbf{F}^{\frac{1}{2}}\|\|\mathbf{F}^{-\frac{1}{2}}(\mathbf{F}-\mathbf{F}_{ml})\mathbf{F}^{-\frac{1}{2}}\|\|\mathbf{F}^{\frac{1}{2}}\mathbf{F}_{ml}^{-\frac{1}{2}}\|, \\ &\leq \frac{\beta}{1-\beta}, \end{aligned} \tag{84}$$

where (84) follows from the fact that $\mathbf{F}_{ml} \in \tilde{\mathcal{S}}$ defined in (82).

From (84) we have that

$$\mathcal{S} = \left\{ \mathbf{F} : \mathbf{F}_{ml} + \boldsymbol{\Delta} \succeq \mathbf{0}, \boldsymbol{\Delta} = \boldsymbol{\Delta}^T, \|\mathbf{F}_{ml}^{-\frac{1}{2}}\boldsymbol{\Delta}\mathbf{F}_{ml}^{-\frac{1}{2}}\| \leq \frac{\beta}{1-\beta} \right\}, \tag{85}$$

contains a ω-confidence set for the (true) covariance matrix $\mathbf{F}$, i.e. a natural parametrization of the uncertainty structure in (77) is $\mathbf{F}_0 = \mathbf{N} = \mathbf{F}_{ml}$ and $\zeta = \frac{\beta}{1-\beta}$. Thus, the uncertainty structure (77) is completely parametrized by considering the confidence regions around the MLE.

Note that for $\mathbf{N} = \mathbf{F}_0$ and $\zeta = \frac{\beta}{1-\beta}$, Lemma 4 implies that

$$\max_{\{\mathbf{V} \in S_v, \mathbf{F} \in S_f\}} \boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V})\boldsymbol{\phi} \leq \nu$$

if and only if

$$(\boldsymbol{\rho}^T\boldsymbol{\phi}; \nu; \boldsymbol{\phi}) \in \mathcal{H}\Big(\mathbf{V}_0, (1+\beta/(1-\beta))\mathbf{F}_0, \mathbf{G}\Big) = \mathcal{H}\Big(\mathbf{V}_0, 1/(1-\beta)\mathbf{F}_0, \mathbf{G}\Big).$$

As noted above, the solution β of (83) is equal to the solution η of (76), and from Definition 1 it follows that

$$(\boldsymbol{\rho}^T\boldsymbol{\phi}; \nu; \boldsymbol{\phi}) \in \mathcal{H}\Big(\mathbf{V}_0, 1/(1-\beta)\mathbf{F}_0, \mathbf{G}\Big) \Leftrightarrow (\boldsymbol{\rho}^T\boldsymbol{\phi}; (1-\eta)\nu; \boldsymbol{\phi}) \in \mathcal{H}\Big(\mathbf{V}_0, \mathbf{F}_0, \mathbf{G}\Big).$$

Thus, $\max_{\{\mathbf{V} \in S_v, \mathbf{F} \in S_f\}} \boldsymbol{\phi}^T(\mathbf{V}^T\mathbf{F}\mathbf{V})\boldsymbol{\phi} \leq \nu$ if and only if $(\boldsymbol{\rho}^T\boldsymbol{\phi}; (1-\eta)\nu; \boldsymbol{\phi}) \in \mathcal{H}\Big(\mathbf{V}_0, \mathbf{F}_0, \mathbf{G}\Big)$. Since this is the precisely the condition in Part (ii) of Lemma 3, it follows that, although the two uncertainty sets (69) and (77) are not the same, they imply the same worst-case variance constraint. However, unlike the parametrization of $S_{f^{-1}}$, the parametrization of S_f cannot be biased to reflect prior knowledge about $\mathbf{F}$.

7 Computational results

In this section we report the results of our preliminary computational experiments with the robust portfolio selection framework proposed in paper. The objective of these computational experiments was to contrast the performance of the classical portfolio selection strategies with that of the robust portfolio selection strategies. We conducted two types of computational tests. The first set of tests compared performance on simulated data and the second set compared sample path performance on real market data. In our tests both the classical and robust portfolios were selected by solving the corresponding maximum Sharpe ratio problem. The details of the experimental procedure and the results are given below.

Figure 2: Performance as a function of ω (Run 1)

Figure 3: Performance as a function of ω (Run 2)

7.1 Computational results for simulated data

For our computational tests on simulated data, we fixed the number of assets $n = 500$ and the number of factors $m = 40$. A symmetric positive definite factor covariance matrix $\mathbf{F}$ was randomly generated except that we ensured that the condition number of $\mathbf{F}$, i.e. $\lambda_{\max}(\mathbf{F})/\lambda_{\min}(\mathbf{F})$, was at most 20 by adding a suitable multiple of the identity. This factor covariance matrix was assumed to be known and fixed. The nominal factor loading matrix $\mathbf{V}$ was also randomly generated. The covariance matrix $\mathbf{D}$ of the residual returns $\boldsymbol{\epsilon}$ was assumed to be certain (i.e. $\mathbf{D} = \overline{\mathbf{D}} = \underline{\mathbf{D}}$) and set to $\mathbf{D} = 0.1\,\mathbf{diag}(\mathbf{V}^T\mathbf{F}\mathbf{V})$, i.e. it was assumed that the linear model explains 90% of the asset variance.

The risk-free rate r_f was set to $r_f = 3$ and the nominal asset returns μ_i were chosen independently according to a uniform distribution on $[r_f - 2, r_f + 2]$. Next, we generated a sequence of asset and factor return vectors according to the market model (1) in Section 2 for an investment period of length $p = 90$ and used equations (62) and (63) in Section 5 to set the parameters $\mathbf{V}_0$, $\boldsymbol{\mu}_0$, $\mathbf{G}$, $\boldsymbol{\rho}$ and $\boldsymbol{\gamma}$. (Note that we did not estimate $\mathbf{D}$ from data.) Next, the robust and the classical portfolios $\boldsymbol{\phi}_r$ and $\boldsymbol{\phi}_m$ were computed by solving the robust maximum Sharpe ratio problem (8) and its classical counterpart, respectively. Although the precise numbers used in these simulation experiments were arbitrary, we expect that the qualitative aspects of the results are not dependent on the precise values.

In the first set of simulation experiments we compared the performance of the robust and classical portfolios as the confidence threshold ω (see Section 5 for details) was increased from 0.01 to 0.95. The experimental results for three independent runs are shown in Figure 2 - Figure 4. In each of the three figures, the top plot is the ratio of the mean Sharpe ratio of the robust portfolio to that of the classical portfolio. The mean Sharpe ratio of any portfolio $\boldsymbol{\phi}$ is given by

$$\frac{(\boldsymbol{\mu}_0 - r_f\mathbf{1})^T\boldsymbol{\phi}}{\sqrt{\boldsymbol{\phi}^T(\mathbf{V}_0^T\mathbf{F}\mathbf{V}_0)\boldsymbol{\phi}}}.$$

The classical portfolio $\boldsymbol{\phi}_m$ maximizes the mean Sharpe ratio. The bottom plot in Figure 2 - Figure 4 is the ratio of the worst case Sharpe ratio of the robust portfolio to that of the classical portfolio. The worst case Sharpe ratio of a portfolio $\boldsymbol{\phi}$ is given by

$$\min_{\{\mathbf{V}\in S_v, \boldsymbol{\mu}\in S_m, \mathbf{D}\in S_d\}} \frac{(\boldsymbol{\mu} - r_f\mathbf{1})^T\boldsymbol{\phi}}{\sqrt{\boldsymbol{\phi}^T(\mathbf{V}\mathbf{F}\mathbf{V} + \mathbf{D})\boldsymbol{\phi}}},$$

Figure 4: Performance as a function of ω (Run 3)

Figure 5: Performance as a function of σ (Run 1)

Figure 6: Performance as a function of σ (Run 2)

where S_d, S_v and S_m are given by (2)-(4). The robust portfolio $\boldsymbol{\phi}_r$ maximizes the worst case Sharpe ratio. The performance on the three runs is almost identical – the ratio of the mean Sharpe ratios drops from approximately 1 to approximately 0.8 as w increases from 0.01 to 0.95 while the ratio of the worst case Sharpe ratios increases from approximately 1 to approximately 2. Thus, at a modest 20% reduction in the mean performance the robust framework delivers an impressive 200% increase in the worst case performance. Notice that the drop in mean performance in run 3 (see Figure 4) is close to 25% but the corresponding improvement in the worst case performance is close to 260%.

The second set of simulation experiments compared the performance of robust and classical portfolios as a function of the upper bound on $\mathbf{D}$. For this set of experiments, we set $\omega = 0.95$ and $\mathbf{D} = \sigma^2 \,\mathbf{diag}(\mathbf{V}_0^T \mathbf{F} \mathbf{V}_0)$, where σ^2 increased from 0.01 to 1. Since the performance of both the robust and classical portfolio was very sensitive to the sample path – particularly for large values of σ^2 – the results shown in Figure 5 - Figure 7 were averaged over 10 runs for every value of σ^2. As before, the top plot is the ratio of the mean performances of the robust and classical portfolios and the bottom plot is the ratio of the worst case performances. (The three plots truncate at different values of σ^2 because we only compare performance for values of σ^2 for which the worst case Sharpe ratio of the classical portfolio is non-negative). Again the essential features of the performance were independent of the particular run – the mean performance of the robust portfolio did not degrade significantly with the increase in noise variance, if it did so at all; and the worst case performance of the robust portfolio was significantly superior to that of the classical portfolio as the data became noisy. This is not unexpected since the robust portfolios were designed to combat noisy data.

7.2 Computational results for real market data

In this section we compare the performance of our robust approach with the classical approach on real market data. The universe of assets that were chosen for investment were those currently ranked at the top of each of 10 industry categories by Dow Jones in August, 2000. In total there were $n = 43$ assets in this set (see Table 1). The base set of factors were five major market indices (see Table 2), to which we added the eigenvectors corresponding to the 5 largest eigenvalues of the covariance matrix of the asset returns, i.e. the total number of factors used was $m = 10$. This choice of factors was made to ensure that the linear model (1) would have good predictability, i.e. small values for the residual return variances. (Selecting appropriate factors to explain the

Figure 7: Performance as a function of σ (Run 3)

Aerospace industry		Telecommunication	
AIR	AAR corporation	T	AT&T
BA	Boeing Corp.	LU	Lucent Technologies
LMT	Lockheed Martin	NOK	Nokia
UTX	United Technologies	MOT	Motorola
Semiconductor		Computer Software	
AMD	Applied Materials	ARBA	Ariba
INTC	Intel Corp.	CMRC	Commerce One Inc.
HIT	Hitachi	MSFT	Microsoft
TXN	Texas Instruments	ORCL	Oracle
Computer Hardware		Internet and Online	
DELL	Dell Computer Corp.	AKAM	Akamai
PALM	Palm Inc.	AOL	AOL Corp.
HWP	Hewlett Packard	CSCO	Cisco Systems
IBM	IBM Corp.	NT	Northern Telecom
SUNW	Sun Microsystems	PSIX	PsiNet Inc.
Biotech and Pharmaceutical		Utilities	
BMV	Bristol-Myers-Squibb	ENE	Enron Corporation
CRA	Applera Corp.- Celera	DUK	Duke Energy Company
CHIR	Chiron Corp.	EXC	Exelon Corp.
LLV	Eli Lilly and Co.	PNW	Pinnacle West
MRK	Merck and Co.		
Chemicals		Industrial goods	
AVY	Avery Denison Corp.	FMC	FMC Corp.
DD	Du Pont	GE	General Electric
DOW	Dow Chemical	HON	Honeywell
EMN	Eastman Chemical Co.	IR	Ingersoll Rand

Table 1: Assets

covariance structure of the returns is a sophisticated industry, and our choice of factors is by no means claimed to be the most appropriate.) The data sequence was daily asset returns from January 2, 1997 through December 29, 2000.

A complete description of the experimental procedure is as follows. The entire data sequence was divided into investment periods of length $p = 90$ days. For each investment period t, we first estimated the covariance matrix $\mathbf{\Sigma}_R$ of the asset returns based on the market data of the previous p trading days and extracted the eigenvectors corresponding to the 5 largest eigenvalues (if an asset did not exist during the entire period it was removed from consideration). These eigenvectors together with the base market indices defined the factors for a particular period, and their returns were used to estimate the factor covariance matrix $\mathbf{F}$. Next, the equations (62) and (63) in Section 5 were used to set $\mathbf{V}_0$, $\boldsymbol{\mu}_0$, $\mathbf{G}$, $\boldsymbol{\rho}$ and $\boldsymbol{\gamma}$. The bound $\overline{d}_i$ on the variance of the residual return was

$\hat{\text{D}}$JA	Dow Jones Composite Average
$\hat{\text{N}}$DX	Nasdaq 100
$\hat{\text{S}}$PC	Standard and Poor's 500 Index (S&P500)
$\hat{\text{R}}$UT	Russell 2000
$\hat{\text{T}}$YX	30-year bond

Table 2: Base factors

Figure 8: Evolution of relative wealth for $\omega = 0.95$

set to $\overline{d}_i = s_i^2$, where s_i^2 is given by (51), and the risk-free rate r_f was set to zero. Once all the parameters were set, the robust portfolio $\boldsymbol{\phi}_r^t$ (resp. classical portfolio $\boldsymbol{\phi}_m^t$) was computed by solving the robust (resp. classical) maximum Sharpe ratio problem. The portfolio $\boldsymbol{\phi}_r^t$ and $\boldsymbol{\phi}_m^t$ were held constant for the period t and then rebalanced to the portfolio $\boldsymbol{\phi}_r^{(t+1)}$ and $\boldsymbol{\phi}_m^{(t+1)}$ in period $t+1$.

Let w_r^t (resp. w_m^t) denote the wealth at the end of period t of an investor with initial wealth w_0 and employing the robust (resp. classical) portfolio strategy. Then

$$\begin{aligned} w_r^{(t+1)} &= \left[\left(\prod_{tp<k\leq(t+1)p}(\mathbf{1}+\mathbf{r}_k)^T\right)\boldsymbol{\phi}_r^t\right]\cdot w_r^t, \\ w_m^{(t+1)} &= \left[\left(\prod_{tp<k\leq(t+1)p}(\mathbf{1}+\mathbf{r}_k)^T\right)\boldsymbol{\phi}_m^t\right]\cdot w_m^t. \end{aligned} \tag{86}$$

Since both the robust and classical strategies require a block of data of length $p = 90$ to estimate the parameters, the first investment period labeled $t = 1$ starts from the time instant $p+1$. The time period Jan 2, 1997-Dec 29, 2000 contains 11 periods of length $p = 90$, i.e. in all there are 10 investment periods.

Figure 8 is a plot of the relative performance of robust strategy with respect to the classical strategy, i.e. it plots $100(\frac{w_r^t}{w_m^t} - 1)$ as a function of t, for a confidence threshold of $\omega = 0.95$. At the end of 11 periods the wealth generated by the robust investment strategy is 40% greater than that generated by the classical investment strategy. Notice that at $t = 7$ there is a precipitous drop in the relative performance of the robust strategy. This drop is probably because the model estimate in period $t = 6$ is remarkably close to the realization in period $t = 7$, i.e. the robust strategy is being unnecessarily conservative. Therefore, one would expect that this problem would be somewhat mitigated if the threshold ω were reduced. Figure 9 plots the relative performance of the robust and classical investment strategies for a confidence threshold of $\omega = 0.7$. It is obvious that the drop in period 7 is reduced; however, the final wealth generated by the robust strategy is, in fact, 9% less than that generated by the classical portfolio ! Thus, it is not guaranteed that the robust strategy will always outperform the classical portfolio. Figure 10 plots the time evolution of the relative

Figure 9: Evolution of relative wealth for $\omega = 0.7$

Figure 10: Evolution of relative wealth as a function of ω

Figure 11: Final relative wealth as a function of ω

performance for various values of ω. As one would expect for small values of ω the performance of the robust and classical strategies is quite close. For intermediate values, the dip at $t = 7$ is reduced but unimpressive performance over the other periods drags the overall relative performance of the robust strategy down. The performance improves as $\omega \uparrow 1$ (for $\omega = 0.99$ the robust strategy generates a final wealth that is 50% larger than that generated by the classical strategy). Figure 11 plots the final wealth ratio as a function of ω (results were obtained for $\omega = 0.1, 0.2, \ldots, 0.9, 0.95, 0.99$). It is clear that the performance is not monotonic in ω.

An important aspect of any investment strategy is the cost of implementing it. Since we are interested in comparing the costs of implementing the robust strategy with that of implementing the classical strategy we will quantify the transaction costs by $\|\boldsymbol{\phi}^t - \boldsymbol{\phi}^{(t-1)}\|_1$. In Figure 12 we plot the ratio of the costs, i.e. $\|\boldsymbol{\phi}_r^t - \boldsymbol{\phi}_r^{(t-1)}\|_1 / \|\boldsymbol{\phi}_m^t - \boldsymbol{\phi}_m^{(t-1)}\|_1$, for a confidence threshold of $\omega = 0.95$. The average cost is 0.9623, i.e. the transaction costs incurred by the robust strategy were approximately 4% less that that incurred by the classical strategy. Figure 13 plots the same quantity for $\omega = 0.7$ and now the average cost is 1.0057, i.e. as the robust strategy become less conservative it pays more in transaction costs. Figure 14 shows the average cost as a function of ω. The average cost remains almost constant at a value slightly greater than 1 until $\omega > 0.9$ and then it decreases monotonically.

7.3 Summary of computation results

The summary of our simulation experiments and the experiments with real market data is as follows:

(a) The mean performance of the robust portfolios does not significantly degrade as the confidence level ω is increased. Even at $\omega = 0.95$ the relative loss of the robust portfolios is only about 20%. On the other hand the worst case performance of the robust portfolios is about 200% better.

(b) Robust portfolios are able to withstand noisy data considerably better than classical portfolios.

(c) Summarizing the performance of robust strategies on the real market data sequence is not easy. Based on the simulation data one would expect a monotonic improvement in performance as the threshold ω is increased. However, the experimental results do not uphold this hypothesis.

Figure 12: Relative cost per period for $\omega = 0.95$

Figure 13: Relative cost per period for $\omega = 0.7$

Figure 14: Average cost vs ω

> For the particular data sequence used in our experiments, the robust strategy was clearly superior; i.e. it generated a larger wealth at a smaller cost, when ω was sufficiently large; whereas for small ω there was no discernible trend.

These computational experiments, particularly those with the real data sequence, are by no means comprehensive. For one the problem size – ($m = 40, n = 500$) for the simulation experiments and ($m = 10, n = 43$) for market data experiments – was small. For the robust strategy to be acceptable one would have to ensure that the complexity of the robust optimization problems is not significantly higher than that of the classical problems. A simple comparison of run times suggests that this is indeed the case.

The experimental results on the real data suggests (see Figure 10) that if one wants the robust strategy to consistently outperform the classical strategy one would want to choose $\omega \approx 1$. But such a choice of ω makes the robust strategy extremely conservative which would hamper its performance if the noise in the model was low. Since the noise is not known *a priori*, the correct choice of ω remains a vexing problem. Our preliminary experiments suggest that the factor model we employed was probably noisy. More extensive experiments have to be conducted before one can assert that this is almost always the case (e.g., using "better" factors may significantly change the results), and therefore, setting $\omega \approx 1$ is a good choice. We would expect that in practice one would have to adjust ω dynamically by comparing the robust strategy with the classical benchmark. Also, we have not tested robust portfolios based on our model in Section 6 which incorporates uncertainty in the factor covariance matrix. These and other experimental studies are planned.

References

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, 2nd edition, 1984.

[2] F. Andersson, H. Mausser, D. Rosen, and S. Uryasev. Credit Risk Optimization with Conditional Value-At-Risk Criterion. *Math. Prog., Series B*, 89:273–291, 2001.

[3] A. Ben-Tal, T. Margalit, and A. Nemirovski. Robust modeling of multi-stage portfolio problems. In *High performance optimization*, pages 303–328. Kluwer Acad. Publ., Dordrecht, 2000.

[4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805, 1998.

[5] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, 25(1):1–13, 1999.

[6] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Analysis, algorithms, and engineering applications.

[7] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear matrix inequalities in system and control theory.* SIAM, Philadelphia, PA, 1994.

[8] B. Efron and R. Tibshirani. *An Introduction to Bootstrap.* Number 57 in Monographs on statistics and applied probability. Chapman & Hall, New York, 1993.

[9] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.*, 18(4):1035–1064, 1997.

[10] L. El Ghaoui and N. Niculescu, editors. *Recent advances on LMI methods in control.* SIAM, 1999.

[11] L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1):33–52, 1998.

[12] W. H. Greene. *Econometric analysis.* Macmillan Pub. Co., 1990.

[13] B. V. Halldórsson and R. H. Tütüncü. An interior-point method for a class of saddle point problems. Technical report, Carnegie Mellon University, April 2000.

[14] P. J. Huber. *Robust Statistics.* John Wiley & Sons, 1981.

[15] R. Kast, E. Luciano, and L. Peccati. VaR and Optimization. In *Workshop on Preferences and Decisions*, July 1998.

[16] O. Ledoit. A well-conditioned estmiator for large dimensional covariance matrices. Technical report, Anderson School, UCLA, 1996.

[17] J. Lintner. Valuation of risky assets adnd the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47:13–37, 1965.

[18] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra Appl.*, 284(1-3):193–228, 1998.

[19] H. M. Markowitz. Portfolio selection. *J. Finance*, 7:77–91, 1952.

[20] H. M. Markowitz. *Portfolio Selection.* Wiley, New York, 1959.

[21] H. Mausser and D. Rosen. Beyond VaR: From measuring risk to managing risk. *ALGO Research Quarterly*, 1(2):5–20, 1999.

[22] R. O. Michaud. *Efficient Asset Management: a practical guide to stock portfolio management and asset allocation.* Financial Management Association survey and synthesis series. HBS Press, 1998.

[23] J. Mossin. Equilibrium in capital asset markets. *Econometrica*, 34(4):768–783, 1966.

[24] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming.* SIAM, Philadelphia, 1993.

[25] J. L. Schafer. *Analysis of Incomplete Multivariate Data.* Chapman and Hall, New York, 1997.

[26] W. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

[27] J. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11–12:625–653, 1999.

[28] K. Zhou, J. Doyle, and K. Glover. *Robust and Optimal Control,.* Prentice-Hall, 1996.

A Probabilistic interpretation

In this section, we interpret the choice of the uncertainty sets S_m, S_v and S_f in terms of the implied probabilistic guarantees on the risk-return performance.

First consider the case where the factor covariance $\mathbf{F}$ is fixed. Define $S_m(\omega)$ and $S_v(\omega)$ for a confidence level ω using (56) and (57). Let $\boldsymbol{\phi}^*$ be the optimal solution of the robust maximum Sharpe ratio problem (37) and let $s^*(\omega)$ be the corresponding Sharpe ratio.

Since $\boldsymbol{\mu}^T\boldsymbol{\phi}^* \geq 1$ for all $\boldsymbol{\mu} \in S_m(\omega)$, it follows that $\boldsymbol{\mu}^T\boldsymbol{\phi}^* \geq 1$ for $(\boldsymbol{\mu}, \mathbf{V})$ in the joint uncertainty set $S(\omega)$ defined in (55). Similarly, we have that

$$\sqrt{(\boldsymbol{\phi}^*)^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi}^*} \leq \frac{1}{s^*(\omega)}, \forall\ \mathbf{V} \in S_v(\omega) \quad \Rightarrow \quad \sqrt{(\boldsymbol{\phi}^*)^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi}^*} \leq \frac{1}{s^*(\omega)},\ \forall(\boldsymbol{\mu}, \mathbf{V}) \in S(\omega).$$

Therefore, we have that Sharpe ratio of the portfolio $\boldsymbol{\phi}^*$

$$\frac{\boldsymbol{\mu}^T\boldsymbol{\phi}^*}{\sqrt{(\boldsymbol{\phi}^*)^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi}^*}} \geq s^*(\omega), \quad \forall(\boldsymbol{\mu}, \mathbf{V}) \in S(\omega),$$

i.e. the Sharpe ratio is at least as large as $s^*(\omega)$ with confidence ω^n. Similar probabilistic guarantees can be provided for the performance of the optimal solutions of robust minimum variance, maximum return and VaR problems. Computing probabilistic guarantees on the performance of the robust portfolio when the uncertainty sets are defined by (62)–(63) is left to the reader.

Thus, in contrast to the classical Markowitz portfolio selection, the robust formulation allows one to impose confidence thresholds on the value of the optimization problem. Setting a low value for

ω results in a high value for the corresponding robust Sharpe ratio $s^*(\omega)$ but the confidence that the corresponding optimal portfolio $\boldsymbol{\phi}^*$ would indeed achieve the Sharpe ratio is low. On the other hand, a high value for ω would imply a lower Sharpe ratio $s^*(\omega)$ but it would ensure that the corresponding optimal portfolio will achieve the Sharpe ration with a higher confidence. Thus, the parameter ω can be viewed as a surrogate for risk aversion.

Next, consider the case where the factor covariance matrix $\mathbf{F}$ is uncertain. Suppose one assumes the following Bayesian setup. The market parameters $(\boldsymbol{\mu}, \mathbf{V})$ and $\mathbf{F}$ are assumed to be *a priori* independent and distributed according to a non-informative conjugate prior (see [12, 1] for appropriate choices for the conjugate prior). Suppose also that the residual return $\boldsymbol{\epsilon}$ is independent of $\boldsymbol{\mu}$, $\mathbf{V}$ and $\mathbf{F}$. Then the market model (1) implies that the conditional distribution $f(\mathbf{S}, \mathbf{R} \mid \boldsymbol{\mu}, \mathbf{V}, \mathbf{F})$ of the asset returns $\mathbf{S} = [\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^p]$ and factor returns $\mathbf{B} = [\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^p]$ can be factored as follows

$$f(\mathbf{S}, \mathbf{R} \mid \mu, \mathbf{V}, \mathbf{F}) = f(\mathbf{B} \mid \mathbf{F}) f(\mathbf{S} \mid \boldsymbol{\mu}, \mathbf{V}, \mathbf{B}). \tag{87}$$

The *posterior* distribution $f(\boldsymbol{\mu}, \mathbf{V}, \mathbf{F} \mid \mathbf{S}, \mathbf{B})$ is, therefore, given by

$$\begin{aligned} f(\boldsymbol{\mu}, \mathbf{V}, \mathbf{F} \mid \mathbf{S}, \mathbf{B}) &= \frac{f(\mathbf{B} \mid \mathbf{F}) f(\mathbf{F}) f(\mathbf{S} \mid \boldsymbol{\mu}, \mathbf{V}, \mathbf{B}) f(\boldsymbol{\mu}, \mathbf{V})}{f(\mathbf{S}, \mathbf{R})}, \\ &= \Big(c_1 f(\mathbf{B} \mid \mathbf{F}) f(\mathbf{F})\Big) \cdot \Big(c_2 f(\mathbf{S} \mid \boldsymbol{\mu}, \mathbf{V}, \mathbf{B}) f(\boldsymbol{\mu}, \mathbf{V})\Big), \end{aligned} \tag{88}$$

where c_1, c_2 are suitable normalizing constants. From (88) it follows that $\mathbf{F}$ and $(\boldsymbol{\mu}, \mathbf{V})$ are *a posteriori* independent. Therefore, the uncertainty sets for $(\boldsymbol{\mu}, \mathbf{V}, \mathbf{F})$ can be represented as a Cartesian product $S \times S_{f^{-1}}$.

Suppose the confidence threshold is set to $\omega \leq (\mathcal{F}_{\Gamma_p}(2))^m$, where $\mathcal{F}_{\Gamma_p}$ is the CDF of a $\Gamma\Big(\frac{p+1}{2}, \frac{p-1}{2}\Big)$ random variable (for a discussion of the upper bound on the achievable confidence see Section 6). Set $\mathbf{F}_0 = \mathbf{F}_{ml}$ and set η by solving (76). Then $\mathbf{F} \in S_{f^{-1}}(\eta)$ with confidence ω^m. As before, let $S(\omega)$ be the ω^n confidence set for $(\boldsymbol{\mu}, \mathbf{V})$. Let $\boldsymbol{\phi}^*$ be the optimal solution of the robust Sharpe ratio problem with uncertain covariance and let $s^*(\omega)$ be the corresponding value. Then the confidence that $(\boldsymbol{\mu}, \mathbf{V}, \mathbf{F}) \in S(\omega) \times S_{f^{-1}}(\eta)$ is $\omega^{(m+n)}$, i.e. with a confidence level $\omega^{(m+n)}$, the realized Sharpe ratio of $\boldsymbol{\phi}^*$ is at least as large as $s^*(\omega)$.

The latter development relied on the fact that an independent Bayesian prior ensures that the joint confidence set for $(\boldsymbol{\mu}$, $\mathbf{V}$, $\mathbf{F})$ is the Cartesian product of the individual confidence sets for $(\boldsymbol{\mu}, \mathbf{V})$ and $\mathbf{F}$. Such a partition is not immediately obvious in the non-Bayesian setup, and, consequently, it is not clear how the uncertainty set S_f given by (77) could lead to probabilistic guarantees.