Submitted to *INFORMS Journal on Optimization*
manuscript (Please, provide the manuscript number!)

Portfolio optimization in the presence of estimation errors on the expected asset returns

Gérard Cornuéjols
Carnegie Mellon University, Tepper School of Business 5000 Forbes Ave, Pittsburgh, PA 15213, gc0v@andrew.cmu.edu

Özgün Elçi
Carnegie Mellon University, Tepper School of Business 5000 Forbes Ave, Pittsburgh, PA 15213, elci@cmu.edu

Matthias Köppe
University of California, Davis, Department of Mathematics One Shields Avenue, Davis, CA 95616
mkoeppe@math.ucdavis.edu

It is well known that the classical Markowitz model for portfolio optimization is extremely sensitive to estimation errors on the expected asset returns. Robust optimization mitigates this issue. We focus on ellipsoidal uncertainty sets around a point estimate of the expected asset returns. An important issue is the choice of the matrix that specifies this ellipsoid. In this paper we investigate the performance of diagonal estimation-error matrices. We show that the class of diagonal estimation-error matrices can achieve an arbitrarily small loss in the expected portfolio return as compared to the optimum. We then formulate the problem of finding the best estimation error matrix as a bilevel program. Finally we analyze the use of an identity matrix as the estimation-error matrix. The results of our simulation show that robust portfolio models featuring an identity matrix as an estimation-error matrix outperform the classical Markowitz model when the size of the uncertainty set is chosen properly.

Key words: portfolio optimization; robust optimization; estimation error; bilevel programming

History:

1. Introduction

Consider a portfolio optimization problem where we want to invest in n assets. If the return vector $\mathbf{r} \in \mathbb{R}^n$ is given, we formulate the problem as maximize$\{\mathbf{r}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \Delta\}$ where $\mathbf{x} \in \mathbb{R}^n$ denotes the fraction of investment in each asset and Δ denotes the feasible region. In this paper we consider $\Delta := \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, \mathbf{x} \geq 0\}$. The constraints $\mathbf{x} \geq 0$ restrict the model to portfolios with long positions only. This problem has a trivial optimal solution: Only invest in the asset that has the highest return.

In practice, however, the assets are risky and the return vector $\mathbf{r}$ is random. The classical mean-variance portfolio optimization problem introduced by Markowitz (1952) addresses this issue by maximizing the expected return of the portfolio subject to a constraint on the risk modeled as the variance of the portfolio return.

$$\underset{\mathbf{x}}{\text{maximize}} \quad \boldsymbol{\mu}^\mathsf{T}\mathbf{x} \tag{1a}$$

$$\text{subject to} \quad \mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{x} \leq v \tag{1b}$$

$$\mathbf{x} \in \Delta. \tag{1c}$$

Here, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the expectation vector and covariance matrix of the asset returns, respectively. In practice, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are estimated, and it has been observed that the Markowitz model tends to amplify estimation errors. In particular, small errors in $\boldsymbol{\mu}$ may produce large changes in portfolio holdings (see, e.g., Best and Grauer 1991, Chopra and Ziemba 1993, Michaud 2008). Approaches to mitigate the effect of estimation errors in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ on portfolio construction have lead to a vast literature. Among these we name Horst et al. (2001), Chopra (1993), Jagannathan and Ma (2003), Goldfarb and Iyengar (2003), Tutuncu and Koenig (2004), Ceria and Stubbs (2006), Scherer (2007), Garlappi et al. (2007), Kan and Zhou (2007), DeMiguel et al. (2009a), DeMiguel et al. (2009b), Lim et al. (2012), and Ban et al. (2018). In this paper, we pursue this line of work, focusing on the uncertainty in the $\boldsymbol{\mu}$ estimates.

We assume that the expected return vector $\boldsymbol{\mu}$ is unknown and belongs to an ellipsoidal uncertainty set given by

$$\mathcal{U} := \left\{\boldsymbol{\mu} \in \mathbb{R}^n \ : \ (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\mathsf{T}\boldsymbol{\Xi}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \kappa^2\right\} \tag{2}$$

where $\hat{\boldsymbol{\mu}}$ is the estimated expected return and the positive definite matrix $\boldsymbol{\Xi}$ is referred to as the estimation-error matrix. Using the uncertainty set $\mathcal{U}$, we formulate the robust portfolio optimization problem as follows

$$\max_{\mathbf{x}\in\mathcal{X}} \min_{\boldsymbol{\mu}\in\mathcal{U}} \left\{ \boldsymbol{\mu}^{\mathsf{T}}\mathbf{x} \right\} \tag{3}$$

where $\mathcal{X}$ denotes the feasible set (1b) – (1c). We assume that the covariance matrix $\boldsymbol{\Sigma}$ is known. Consequently, $\mathcal{X}$ is the intersection of an ellipsoid with a simplex.

This problem can be reformulated following Ben-Tal and Nemirovski (1999) as

$$\max_{\mathbf{x}\in\mathcal{X}} \left\{ \hat{\boldsymbol{\mu}}^{\mathsf{T}}\mathbf{x} - \kappa\sqrt{\mathbf{x}^{\mathsf{T}}\boldsymbol{\Xi}\mathbf{x}} \right\}. \tag{4}$$

We refer the reader to Goldfarb and Iyengar (2003), Tutuncu and Koenig (2004) and Fabozzi et al. (2007) for more detailed discussions on robust portfolio optimization. The *risk-like* term $\kappa\sqrt{\mathbf{x}^{\mathsf{T}}\boldsymbol{\Xi}\mathbf{x}}$ in formulation (4) can be interpreted as an estimation risk that must be considered by risk-averse investors on top of the risk caused by the variance $\mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{x}$ of the portfolio return (Fabozzi et al. 2007, p 371).

Our contribution. In this paper, we provide a theoretical analysis on the choice for the estimation-error matrix $\boldsymbol{\Xi}$ in robust portfolio optimization. The literature on selecting/constructing an estimation-error matrix is scarce (Gotoh and Takeda 2011). Stubbs and Vance (2005) provide a comprehensive overview on practical approaches for computing estimation-error matrices. They also recommend the use of diagonal estimation-error matrices to practitioners. On the other hand, there are several studies in which a scalar multiple of $\boldsymbol{\Sigma}$ is used as the estimation-error matrix (see, e.g., Scherer 2007, Garlappi et al. 2007, Olivares-Nadal and DeMiguel 2018). Among these, Scherer (2007) has a skeptical take on robust optimization and shows that such a choice for $\boldsymbol{\Xi}$ is equivalent to some other well known shrinkage approaches.

In our work, we begin by discussing the difference between true and actual frontier - a well known approach that is used to quantify the impact of the estimation errors. We show that when $\boldsymbol{\Xi}$ is a multiple of $\boldsymbol{\Sigma}$, robust optimization cannot improve the actual performance of the Markowitz model. This analysis supports the recommendation of Stubbs and Vance (2005) where they claim that such a choice is not used in practice and is not recommended. We then focus on the use of diagonal estimation-error matrices and show that the class of diagonal estimation-error matrices can achieve an arbitrarily small loss in the expected portfolio return as compared to the optimum. We accomplish this by reformulating the optimality conditions of the robust portfolio problem when we have a single expected return vector.

We then investigate diagonal estimation-error matrices in the presence of multiple expected returns. To this end, we propose a bilevel model that computes the loss and we show that the diagonal estimation-error matrices can achieve an arbitrarily small loss even when there are multiple estimates for the expected return. The proposed bilevel model also provides a principled way to construct estimation-error matrices using data, though it is a non-convex optimization problem and it is difficult to solve in practice.

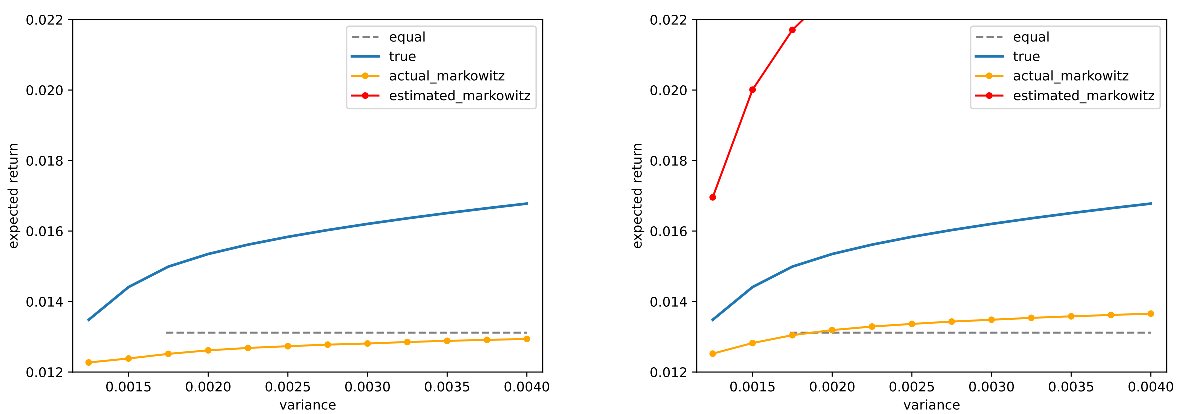
We finally focus on the use of an identity matrix for $\boldsymbol{\Xi}$. This choice requires calibrating a single parameter (κ) for the robust problem. We perform computational experiments to test whether robust optimization can perform better than the Markowitz model. Our results demonstrate that a good choice for κ can significantly improve the performance of the Markowitz model, especially when the expected return estimates are not reliable.

The rest of the paper is organized as follows. We introduce true, estimated and actual frontiers in Section 2. In Section 3, we show that robust optimization can improve on the actual performance. We present the analysis of the robust optimization problem in Section 4. Section 5 generalizes the analysis of the robust model to incorporate multiple expected returns. We analyze the use of an identity matrix as estimation-error matrix in Section 6. Section 7 concludes the paper.

2. True, estimated, and actual frontiers

The sensitivity of mean-variance portfolio optimization models to estimation errors on the expected asset returns is well documented in the literature (see, e.g., Best and Grauer 1991, Chopra 1993, Michaud 2008). It is sometimes referred to as the error maximization tendency of mean-variance optimization. In order to quantify the effect of estimation errors, Broadie (1993) made a distinction between true, estimated, and actual frontiers. A *frontier* plots the maximum expected return of a portfolio of assets as a function of the risk threshold (Markowitz 1952). The *true frontier* is computed by using the true expected returns of the assets, a quantity in fact unknown to the decision maker. The *estimated frontier* is computed by using the estimated expected returns. It describes what appears will be the expected return of a portfolio optimized based on the estimated parameters. The *actual frontier* plots the expected return one actually achieves (using the true expected returns) when one invests in the above portfolio (constructed using estimated expected returns). We next describe how we compute these three frontiers.

We solve the problem maximize$\{\boldsymbol{\mu}^\mathsf{T}\mathbf{x}:\ (1b)-(1c)\}$ to obtain the optimal solution $\mathbf{x}^*$ where $\boldsymbol{\mu}$ is the true (but unknown to the investor) vector of expected asset returns. Using $\mathbf{x}^*$, we construct the true frontier $\boldsymbol{\mu}^\mathsf{T}\mathbf{x}^*$. Let $\hat{\boldsymbol{\mu}}$ be the vector of estimated expected asset returns. We solve maximize$\{\hat{\boldsymbol{\mu}}^\mathsf{T}\mathbf{x}:\ (1b)-(1c)\}$ to obtain the *Markowitz solution estimate* $\hat{\mathbf{x}}^M$. Using $\hat{\mathbf{x}}^M$, we construct the *estimated Markowitz frontier* $\hat{\boldsymbol{\mu}}^\mathsf{T}\hat{\mathbf{x}}^M$ and the *actual Markowitz frontier* $\boldsymbol{\mu}^\mathsf{T}\hat{\mathbf{x}}^M$.



(a) sample size = 1 (b) sample size = 24

Figure 1 True, actual Markowitz, and estimated Markowitz frontiers.

In Figure 1, we illustrate *the gap* between the true and the actual frontiers on real-world data. Details on the data set can be found in Appendix EC.1. We assume that the asset returns are normally distributed with distribution Normal$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The vector $\hat{\boldsymbol{\mu}}$ is the sample average of N

random samples. Equivalently, we generate $\hat{\boldsymbol{\mu}}$ from Normal$(\boldsymbol{\mu}, \boldsymbol{\Sigma}/N)$. We illustrate the cases $N = 1$ and $N = 24$. Each point in the figures represents the average value over 10000 trials. The dashed-line represents the performance of the equal-weight portfolio where we invest equally in all the assets. The expected return of the equal-weight portfolio is 0.0131. Another interesting portfolio for comparison is the minimum variance portfolio, which has an expected return of 0.0122.

We see in Figure 1a that the actual performance of the Markowitz solution estimate is poor when the sample size is equal to 1. In fact, the average expected return values are worse than the equal-weight portfolio. On the other hand, we see that the performance of the Markowitz solution estimate improves significantly in Figure 1b with more accurate return estimates with a sample size of 24. As N goes to $+\infty$, the actual Markowitz frontier converges to the true frontier. Furthermore, we plot the estimated frontier in Figures 1a and 1b. We see that the estimated frontier is far away from the true frontier so much that it is off the chart in Figure 1a.

Similarly, solving the robust optimization problem (4), we obtain the *robust solution estimate* $\hat{\mathbf{x}}^R$. Using $\hat{\mathbf{x}}^R$, we construct the *estimated robust frontier* $\hat{\boldsymbol{\mu}}^\mathsf{T}\hat{\mathbf{x}}^R$ and the *actual robust frontier* $\boldsymbol{\mu}^\mathsf{T}\hat{\mathbf{x}}^R$. A key aspect here is the modeler's choice of the estimation-error matrix $\boldsymbol{\Xi}$.

3. Robust optimization can improve the actual performance

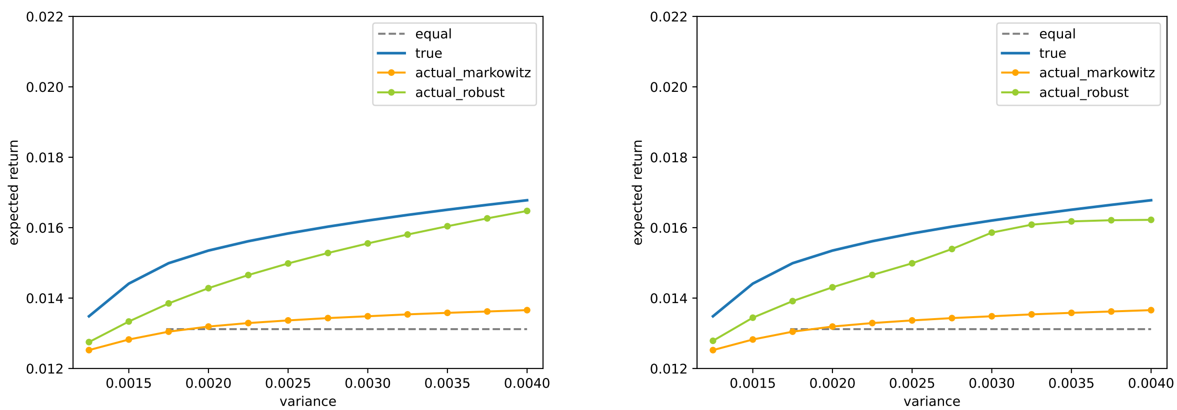
A possible choice for $\boldsymbol{\Xi}$ is to make this matrix proportional to $\boldsymbol{\Sigma}$, namely $\boldsymbol{\Xi} = \rho\boldsymbol{\Sigma}$. Such an idea has been proposed by Horst et al. (2001) and Garlappi et al. (2007). Notice however that, for solutions $\mathbf{x}$ that satisfy the risk constraint $\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{x} \le v$ at equality (which is the most interesting case), the objective $\hat{\boldsymbol{\mu}}^\mathsf{T}\mathbf{x} - \kappa\sqrt{\mathbf{x}^\mathsf{T}\boldsymbol{\Xi}\mathbf{x}}$ of the robust problem (4) becomes

$$\hat{\boldsymbol{\mu}}^\mathsf{T}\mathbf{x} - \kappa\sqrt{\mathbf{x}^\mathsf{T}\rho\boldsymbol{\Sigma}\mathbf{x}} = \hat{\boldsymbol{\mu}}^\mathsf{T}\mathbf{x} - \kappa\sqrt{\rho v}.$$

The last term is just a constant, and therefore the robust problem reduces to the Markowitz model (1). The estimation errors do not affect the optimal portfolio!

Garlappi et al. (2007) consider a variation of model (1) where the risk is not a hard constraint $\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{x} \le v$ but instead a penalty in the objective, leading to the model $\max_{x\in\Delta} \boldsymbol{\mu}^\mathsf{T}\mathbf{x} - \gamma\, \mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{x}$ for a given penalty term $\gamma \in \mathbb{R}$. Robustifying as in (4), this model becomes $\max_{x\in\Delta} \boldsymbol{\mu}^\mathsf{T}\mathbf{x} - \gamma\, \mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{x} - \kappa\sqrt{\mathbf{x}^\mathsf{T}\boldsymbol{\Xi}\mathbf{x}}$. Garlappi et al. (2007) consider this model with $\boldsymbol{\Xi} = \boldsymbol{\Sigma}$. When $\kappa = 0$, this reduces to the Markowitz model and when $\kappa \to +\infty$, it reduces to the minimum variance portfolio. In their computational experiments, Garlappi et al. (2007) find that the minimum variance portfolio gives the best results in terms of measures such as the expected return or the Sharpe ratio.

In the remainder of this paper, we consider choices for $\boldsymbol{\Xi}$ that can outperform both the Markowitz portfolio and the minimum variance portfolio, using actual expected return as a measure.



(a) $\boldsymbol{\Xi} = \mathrm{diag}\big([1,1,1,1,1,1,1,\epsilon,1,1,1]\big)$ (b) $\boldsymbol{\Xi} = \mathrm{diag}\big([1,\epsilon,1,1,1,1,1,\epsilon,1,1,1]\big)$

Figure 2 Robust frontiers of several Ξ matrices.

To illustrate the potential of robust optimization, we design an experiment showing that good matrices $\boldsymbol{\Xi}$ exist, even among diagonal matrices with only two different diagonal entries. We construct $\boldsymbol{\Xi}$ based on partial information on the true expected return $\boldsymbol{\mu}$. In particular, in Figures $2a-2b$, we let some entries of the estimation-error matrix equal to $\epsilon = 0.001$ ($\kappa = 1$, sample size $=$ 24). In Figure 2a, the entry corresponding to the asset with the highest true return is set to ϵ, and in Figure 2b, the entries corresponding to the four assets with highest true return are set to ϵ.

The overall picture that we see in Figure 2 is that there is room for improvement if one is able to choose a good matrix $\boldsymbol{\Xi}$. It seems promising that the robust optimization can close the gap between the actual Markowitz frontier and the true frontier.

4. The class of diagonal estimation-error matrices

In this section, we focus on the class of diagonal estimation-error matrices. This choice is partly motivated by the study of Stubbs and Vance (2005). The authors argue that it is difficult to generate estimation-error matrices accurately, and they suggest that the use of simple diagonal matrices is beneficial. A natural question to ask is: Do we loose in performance by restricting our attention to diagonal matrices $\boldsymbol{\Xi}$?

The issue boils down to the following: Given an estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$, can we always find a diagonal matrix $\boldsymbol{\Xi}$ such that the resulting robust portfolio $\hat{\mathbf{x}}^R$ has an actual expected return very close to the true expected return? To this end, we start examining the optimality conditions of (4).

4.1. Analysis of the robust portfolio optimization problem

We first write (4) as a convex optimization problem of the following form.

$$\begin{aligned}
&\underset{\mathbf{x}}{\text{minimize}} && -\hat{\boldsymbol{\mu}}^{\mathsf{T}}\mathbf{x}+\sqrt{\mathbf{x}^{\mathsf{T}}\boldsymbol{\Xi}\mathbf{x}} && \text{(5a)}\\
&\text{subject to} && \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{x}\le v, && \text{(5b)}\\
& && \mathbf{1}^{\mathsf{T}}\mathbf{x}=1, && \text{(5c)}\\
& && -\mathbf{x}\le\mathbf{0}. && \text{(5d)}
\end{aligned}$$

We next make some observations about (5). The proof of these observations are deferred to the Appendix.

Proposition 1. *Suppose that $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$ are positive definite. Then, the following statements are true about* (5).

1. *It is a convex optimization problem with a strictly convex objective function. Therefore, the optimal solution $\hat{\mathbf{x}}^R$ is unique.*
2. *It satisfies Slater's condition.*
3. *It is sufficient to consider the optimality conditions for differentiable functions.*

We next derive the optimality conditions of (5). Let $\lambda^1 \in \mathbb{R}$, $\lambda^2 \in \mathbb{R}$, and $\boldsymbol{\lambda}^3 \in \mathbb{R}^n$ respectively denote the Lagrangian multipliers of the constraints of (5). The Lagrangian function associated with this problem is given below.

$$L\left(\mathbf{x}, \lambda^1, \lambda^2, \boldsymbol{\lambda}^3\right) = -\hat{\boldsymbol{\mu}}^\mathsf{T}\mathbf{x} + \sqrt{\mathbf{x}^\mathsf{T}\boldsymbol{\Xi}\mathbf{x}} + \lambda^1\left(\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{x} - v\right) + \lambda^2\left(\mathbf{1}^\mathsf{T}\mathbf{x} - 1\right) - \left(\boldsymbol{\lambda}^3\right)^\mathsf{T}\mathbf{x}$$

Using the Lagrangian function, we can write the optimality conditions of (5).

$$-\hat{\boldsymbol{\mu}} + \frac{\boldsymbol{\Xi}\mathbf{x}}{\sqrt{\mathbf{x}^\mathsf{T}\boldsymbol{\Xi}\mathbf{x}}} + 2\lambda^1\boldsymbol{\Sigma}\mathbf{x} + \lambda^2\mathbf{1} - \boldsymbol{\lambda}^3 = \mathbf{0} \tag{6a}$$

$$\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{x} \leq v \tag{6b}$$

$$\mathbf{1}^\mathsf{T}\mathbf{x} = 1 \tag{6c}$$

$$-\mathbf{x} \leq \mathbf{0} \tag{6d}$$

$$\lambda^1 \geq 0 \tag{6e}$$

$$\boldsymbol{\lambda}^3 \geq \mathbf{0} \tag{6f}$$

$$\lambda^1\left(\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{x} - v\right) + \left(\boldsymbol{\lambda}^3\right)^\mathsf{T}\mathbf{x} = 0 \tag{6g}$$

When $\boldsymbol{\Xi}$ is a diagonal matrix, (6a) can be written as

$$-\hat{\mu}_i + \frac{\xi_i x_i}{\sqrt{\sum_{j=1}^n \xi_j x_j^2}} + 2\lambda^1 \sum_{j=1}^n \sigma_{ij} x_j + \lambda^2 - \lambda_i^3 = 0, \quad i \in [1, n] \tag{7}$$

Here, $\boldsymbol{\xi}$ denotes the vector of diagonal entries of $\boldsymbol{\Xi}$ and $[1, n] := \{1, \ldots, n\}$. Now, we introduce additional variables and constraints to rewrite the square-root term in (7). Let

$$z_i = \frac{\xi_i x_i}{\sqrt{\sum_{j=1}^n \xi_j x_j^2}}, \quad i \in [1, n]. \tag{8}$$

Using this substitution, we rewrite (7) as

$$-\hat{\mu}_i + z_i + 2\lambda^1 \sum_{j=1}^n \sigma_{ij} x_j + \lambda^2 - \lambda_i^3 = 0, \quad i \in [1, n], \tag{9a}$$

$$\sum_{i=1}^n x_i z_i = \alpha, \tag{9b}$$

$$\xi_i x_i = z_i \alpha, \quad i \in [1, n], \tag{9c}$$

$$z_i \geq 0, \quad i \in [1,n], \tag{9d}$$

$$\xi_i > 0, \quad i \in [1,n], \tag{9e}$$

$$\alpha > 0 \tag{9f}$$

We have the following result, its proof is deferred to Appendix EC.2.2.

PROPOSITION 2. *The system* $\{(6b)-(6g),(9)\}$ *is a correct reformulation of the system* (6).

4.2. Analysis of the loss due to estimation error

Given a solution estimate $\hat{\mathbf{x}}$, we define the *loss* in objective value in the following way.

$$\text{loss}(\hat{\mathbf{x}}) = \boldsymbol{\mu}^\mathsf{T}\mathbf{x}^* - \boldsymbol{\mu}^\mathsf{T}\hat{\mathbf{x}}$$

A solution $\hat{\mathbf{x}}^R$ to the robust problem (4) depends on the expected return estimate $\hat{\boldsymbol{\mu}}$ and the estimation-error matrix $\boldsymbol{\Xi}$. In this case we write

$$\text{loss}(\hat{\boldsymbol{\mu}}, \boldsymbol{\Xi}) = \boldsymbol{\mu}^\mathsf{T}\mathbf{x}^* - \boldsymbol{\mu}^\mathsf{T}\hat{\mathbf{x}}^R$$

The first question we want to answer is the following: Given any $\hat{\boldsymbol{\mu}}$, does there exist a diagonal $\boldsymbol{\Xi}$ matrix such that $\text{loss}(\hat{\boldsymbol{\mu}}, \boldsymbol{\Xi})$ is equal to zero. The following example shows that it is not always possible to find such a diagonal matrix $\boldsymbol{\Xi}$.

EXAMPLE 1. Consider a two asset portfolio optimization problem. Let $\sigma_{11} < v$, $\sigma_{22} > v$. Let $\mu_1 > \mu_2$. Clearly $(x_1 = 1, x_2 = 0)$ is the true optimal solution. Suppose that $\hat{\mu}_2 > \hat{\mu}_1$.

It follows that we have $\lambda^1 = 0$ because the risk constraint is not tight. Furthermore, $\lambda_1^3 = 0$ because the non-negativity constraint of x_1 is not active. We see that the optimality conditions are satisfied except (6a). Using our reformulation idea, we must find a feasible solution to the following system.

$$-\hat{\mu}_1 + z_1 + \lambda^2 = 0$$

$$-\hat{\mu}_2 + z_2 + \lambda^2 - \lambda_2^3 = 0$$

$$x_1 z_1 + x_2 z_2 = \alpha$$

$$\xi_1 x_1 = z_1 \alpha$$

$$\xi_2 x_2 = z_2 \alpha$$

$$\xi_1,\ \xi_2 > 0$$

$$z_1,\ z_2 \geq 0$$

$$\alpha > 0$$

We see that because $x_2 = 0$ and $\alpha > 0$, it must be that $z_2 = 0$. Furthermore, we can rewrite the first two equations in the following way by treating z_1 and λ_2^3 as slack variables.

$$\lambda^2 \leq \hat{\mu}_1$$

$$\lambda^2 \geq \hat{\mu}_2$$

This is impossible since $\hat{\mu}_2 > \hat{\mu}_1$. Because the true optimal solution is unique, we conclude there does not exist $\boldsymbol{\xi}$ such that the robust portfolio optimization problem yields zero loss.

Our next result shows that even though it is not possible to reach zero loss, we can get arbitrarily close to it.

THEOREM 1. *Given $\epsilon > 0$, for every $\hat{\boldsymbol{\mu}}$ there exists a diagonal $\boldsymbol{\Xi}$ matrix such that loss$(\hat{\boldsymbol{\mu}}, \boldsymbol{\Xi}) \leq \epsilon$.*

Proof. Let $\mathbf{x}^{\text{mv}} := \arg\min_{\mathbf{x}} \{\mathbf{x}^\mathsf{T} \boldsymbol{\Sigma} \mathbf{x} : \mathbf{x} \in \Delta\}$ denote the minimum variance portfolio. Let $\mathbf{x}^{\text{eq}}$ denote the equal-weight portfolio, i.e. $x_i = 1/n,\ i \in [1, n]$. Pick $\delta \in (0,1)$ such that $\tilde{\mathbf{x}} = \delta \mathbf{x}^{\text{mv}} + (1 - \delta)\mathbf{x}^{\text{eq}}$ has a lower variance than the true optimal solution $\mathbf{x}^*$. Clearly, all components of $\tilde{\mathbf{x}}$ are strictly greater than zero. Let $\tilde{\mathbf{x}}^* = (1 - \bar{\epsilon})\mathbf{x}^* + \bar{\epsilon}\tilde{\mathbf{x}}$ for some $\bar{\epsilon} \in (0,1)$. Note that $\tilde{\mathbf{x}}^* \in \mathcal{X}$ and has a lower variance than $\mathbf{x}^*$, therefore it is a feasible solution to the portfolio problem.

Now, observe that the loss corresponding to $\tilde{\mathbf{x}}^*$ is given by

$$\begin{aligned}
\text{loss}(\tilde{\mathbf{x}}^*) &= \sum_{i=1}^{n} \mu_i \Big(x_i^* - \tilde{x}_i^* \Big) \\
&= \sum_{i=1}^{n} \mu_i \Big(x_i^* - \big[(1 - \bar{\epsilon}) x_i^* + \bar{\epsilon} \tilde{x}_i \big] \Big) \\
&= \bar{\epsilon} \sum_{i=1}^{n} \mu_i \Big(x_i^* - \tilde{x}_i \Big) \\
&= \bar{\epsilon} \Big(\boldsymbol{\mu}^\mathsf{T} \mathbf{x}^* - \boldsymbol{\mu}^\mathsf{T} \tilde{\mathbf{x}} \Big) > 0.
\end{aligned}$$

Therefore, given ϵ, we can pick $\bar{\epsilon} \in (0,1)$ such that

$$\bar{\epsilon} \leq \frac{\epsilon}{\left(\boldsymbol{\mu}^\mathsf{T}\mathbf{x}^* - \boldsymbol{\mu}^\mathsf{T}\tilde{\mathbf{x}}\right)}$$

This pick ensures that $\text{loss}(\tilde{\mathbf{x}}^*) \leq \epsilon$.

It suffices to find $\boldsymbol{\xi}$, λ^1, λ^2 and $\boldsymbol{\lambda}^3$ such that together with $\tilde{\mathbf{x}}^*$ we have a solution that satisfies the optimality conditions.

Let $\lambda^1 = 0$, and $\lambda_i^3 = 0$ for all $i \in [1,n]$. We see that all optimality conditions are satisfied except (6a). Using the reformulation idea, we see that finding such $\boldsymbol{\xi}$ is equivalent to finding a solution to the below system.

$$\begin{aligned}
-\hat{\mu}_i + z_i + \lambda^2 = 0, &\quad i \in [1,n] \\
\sum_{i=1}^{n} z_i \tilde{x}_i^* = \alpha & \\
\xi_i \tilde{x}_i^* = z_i \alpha, &\quad i \in [1,n] \\
\xi_i \geq 0, &\quad i \in [1,n] \\
z_i \geq 0, &\quad i \in [1,n] \\
\alpha \geq 0 &
\end{aligned}$$

Let $\lambda^2 < \min_i\{\hat{\mu}_i\}$. Let $z_i = \hat{\mu}_i - \lambda^2$ for all $i \in [1,n]$. Then $\alpha = \sum_{i=1}^n z_i \tilde{x}_i^*$. We let $\xi_i = \frac{z_i \alpha}{\tilde{x}_i^*}$ for all $i \in [1,n]$. Clearly, all ξ_i, z_i, α are greater than zero. Therefore, $\boldsymbol{\xi}$ satisfies the optimality conditions and this concludes the proof. □

Finally we conclude this section by stating a sufficient condition for obtaining zero loss.

Theorem 2. *If all the assets are active in the true optimal solution, then for every $\hat{\boldsymbol{\mu}}$ there exists a diagonal $\boldsymbol{\Xi}$ matrix such that loss$(\hat{\boldsymbol{\mu}}, \boldsymbol{\Xi}) = 0$.*

Proof. Without loss of generality, we assume that all elements of the given $\hat{\boldsymbol{\mu}}$ are greater than 0. It suffices to find $\boldsymbol{\xi}$, λ^1, λ^2 and $\boldsymbol{\lambda}^3$ such that together with $\mathbf{x}^*$ we have a solution that satisfies the optimality condition.

Let $\lambda^1 = 0$, $\lambda^2 = 0$, and $\lambda_i^3 = 0$ for all $i \in [1,n]$. Using the reformulation idea, we see that finding such $\boldsymbol{\xi}$ is equivalent to finding a solution to the below system.

$$
\begin{aligned}
-\hat{\mu}_i + z_i = 0, \quad i \in [1,n] \\
\sum_{i=1}^{n} z_i x_i^* = \alpha \\
\xi_i x_i^* = z_i \alpha, \quad i \in [1,n] \\
\xi_i > 0, \quad i \in [1,n] \\
z_i \geq 0, \quad i \in [1,n] \\
\alpha > 0
\end{aligned}
$$

Let $z_i = \hat{\mu}_i$ for all $i \in [1,n]$. Then we have that $\alpha = \sum_{i=1}^{n} \hat{\mu}_i x_i^*$. We let $\xi_i = \frac{z_i \alpha}{x_i^*}$ for all $i \in [1,n]$. Clearly, all ξ_i, z_i, α are greater than zero. Therefore, $\boldsymbol{\xi}$ satisfies the optimality conditions and this concludes the proof. $\square$

5. Finding the best estimation-error matrix

In this section, we generalize the results of the previous section to the situation where we have several estimates $\hat{\boldsymbol{\mu}}^1, \ldots, \hat{\boldsymbol{\mu}}^T$ of $\boldsymbol{\mu}$. We consider a setup in the same spirit with the computational experiments conducted to measure the performance of the robust optimization via an out-of-sample simulation in Scherer (2007). In particular, we investigate the ability of the robust model to produce portfolios that are close to the true optimal portfolio even when different estimated expected returns are used as an input to the robust model.

We show that we can always find a diagonal matrix $\boldsymbol{\Xi}$ such that the resulting robust portfolios $\hat{\mathbf{x}}^1, \ldots, \hat{\mathbf{x}}^T$ all have an actual expected return very close to the optimal expected return $\boldsymbol{\mu}^\mathsf{T} x^*$ where x^* is the optimal portfolio computed using the true $\boldsymbol{\mu}$. To this end, we write a mathematical program that outputs a matrix $\boldsymbol{\Xi}$ that achieves the minimum loss. We define the loss given a collection of return estimates $\{\hat{\boldsymbol{\mu}}^t\}_{t=1}^T$ and estimation-error matrix $\boldsymbol{\Xi}$ as

$$
\operatorname{loss}(\{\hat{\boldsymbol{\mu}}^t\}_{t=1}^T, \boldsymbol{\Xi}) = \sum_{t \in [1,T]} (\boldsymbol{\mu}^\mathsf{T} \mathbf{x}^* - \boldsymbol{\mu}^\mathsf{T} \hat{\mathbf{x}}^{R,t})
$$

where $\hat{\mathbf{x}}^{R,t}$ denotes the optimal robust portfolio for $t \in [1,T]$.

5.1. A bilevel programming formulation

In this section, we describe the bilevel programming formulation that computes the best estimation-error matrix $\mathbf{\Xi}$ for a given collection of return estimates $\{\hat{\boldsymbol{\mu}}^t\}_{t=1}^T$.

The following parameters are used in the description of the bilevel model.

- $\boldsymbol{\mu}$: True expected return vector.
- $\mathbf{\Sigma}$: True covariance matrix of the asset returns.
- $\hat{\boldsymbol{\mu}}^t$: Estimated return vector under trial $t \in [1, T]$.
- v : Risk threshold.
- $\mathbf{x}^*(v) = \arg\max_{\mathbf{x}\in\Delta^n} \left\{ \boldsymbol{\mu}^\mathsf{T}\mathbf{x} \ : \ \mathbf{x}^\mathsf{T}\mathbf{\Sigma}\mathbf{x} \le v \right\}$, The portfolio with maximum expected return for a given risk threshold value v.

Let $\hat{\mathbf{x}}^t(v, \mathbf{\Xi}, \hat{\boldsymbol{\mu}}^t)$ denote the optimal portfolio given that the estimated return vector is $\hat{\boldsymbol{\mu}}^t$, estimation-error matrix is $\mathbf{\Xi}$, and the risk threshold is v under trial $t \in [1, T]$. Note that this value is obtained by solving a convex optimization problem of the form (4). Therefore, we have the following bilevel program to compute the best $\mathbf{\Xi}$ that minimizes the loss.

$$\underset{\mathbf{\Xi}}{\text{minimize}} \quad \sum_{t=1}^{T} \left(\boldsymbol{\mu}^\mathsf{T}\mathbf{x}^*(v) - \boldsymbol{\mu}^\mathsf{T}\hat{\mathbf{x}}^t(v, \mathbf{\Xi}, \hat{\boldsymbol{\mu}}^t) \right) \tag{10a}$$

$$\text{subject to} \quad \hat{\mathbf{x}}^t(v, \mathbf{\Xi}, \hat{\boldsymbol{\mu}}^t) = \arg\max_{\mathbf{x}\in\Delta^n} \left\{ (\hat{\boldsymbol{\mu}}^t)^\mathsf{T}\mathbf{x} - \sqrt{\mathbf{x}^\mathsf{T}\mathbf{\Xi}\mathbf{x}} \ : \ \mathbf{x}^\mathsf{T}\mathbf{\Sigma}\mathbf{x} \le v \right\}, \quad t \in [1, T], \tag{10b}$$

$$\mathbf{\Xi} \in \mathcal{S}^n_{++}. \tag{10c}$$

In this formulation, $\mathbf{\Xi}$ is the upper-level decision variable, and each $\hat{\mathbf{x}}^t(v, \mathbf{\Xi}, \hat{\boldsymbol{\mu}}^t)$ is a lower-level decision variable. Note that we have as many lower-level problems as the number of trials we have.

The above bilevel program minimizes the sum of the difference between the value of the true frontier, and the value of the actual frontier under each trial. Alternatively, one can use other performance measures, for example minimizing the maximum of the differences. In this formulation, we assume the risk threshold v is large enough that the lower-level problem is feasible. Our next goal is to reformulate problem (10) as a single level optimization program.

5.2. Reformulating (10) as a single level program

We know that robust portfolio optimization problem is a convex optimization program, it satisfies Slater's condition, and it has a unique solution, see Proposition 1. Therefore, we can use the optimality conditions (6) to reformulate (10) as a single level program (Bard 1998).

$$
\begin{aligned}
\underset{\boldsymbol{\Xi}, \hat{\mathbf{x}}^{\mathsf{T}}, \lambda^{1,t}, \lambda^{2,t}, \boldsymbol{\lambda}^{3,t}}{\text{minimize}} \quad & \sum_{t=1}^{T} \left(\boldsymbol{\mu}^{\mathsf{T}} \mathbf{x}^{*} - \boldsymbol{\mu}^{\mathsf{T}} \mathbf{x}^{t} \right) && \text{(11a)} \\
\text{subject to} \quad & -\hat{\boldsymbol{\mu}}^{t} + \frac{\boldsymbol{\Xi}\mathbf{x}^{t}}{\sqrt{(\mathbf{x}^{t})^{\mathsf{T}} \boldsymbol{\Xi} \mathbf{x}}} + 2\lambda^{1,t} \boldsymbol{\Sigma} \mathbf{x}^{t} + \lambda^{2,t} \mathbf{1} - \boldsymbol{\lambda}^{3,t} = \mathbf{0}, \quad t \in [1,T], && \text{(11b)} \\
& (\mathbf{x}^{t})^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{x}^{t} \le v, \quad t \in [1,T], && \text{(11c)} \\
& \mathbf{1}^{\mathsf{T}} \mathbf{x}^{t} = 1, \quad t \in [1,T], && \text{(11d)} \\
& -\mathbf{x}^{t} \le \mathbf{0}, \quad t \in [1,T], && \text{(11e)} \\
& \lambda^{1,t} \ge 0, \quad t \in [1,T], && \text{(11f)} \\
& \boldsymbol{\lambda}^{3,t} \ge \mathbf{0}, \quad t \in [1,T], && \text{(11g)} \\
& \lambda^{1,t} \left((\mathbf{x}^{t})^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{x}^{t} - v \right) + \left(\boldsymbol{\lambda}^{3,t} \right)^{\mathsf{T}} \mathbf{x}^{t} = 0, \quad t \in [1,T], && \text{(11h)} \\
& \boldsymbol{\Xi} \in \mathcal{S}_{++}. && \text{(11i)}
\end{aligned}
$$

Note that in the above formulation, the optimality conditions are appended for each trial $t \in [1,T]$.

The bilevel model (11) is a non-convex optimization problem due to the stationary point equations (11b) and the complementary slackness conditions (11h). We next present a reformulation for (11) when $\boldsymbol{\Xi}$ is a positive definite diagonal matrix.

Using the earlier reformulation idea in Section 4, we can reformulate (11) so as to avoid the square-root terms.

$$
\begin{aligned}
\underset{\boldsymbol{\xi}, \mathbf{x}^{t}, \lambda^{1,t}, \lambda^{2,t}, \boldsymbol{\lambda}^{3,t}}{\text{minimize}} \quad & \text{(11a)} && \text{(12a)} \\
\text{subject to} \quad & \text{(11c)} - \text{(11h)} && \text{(12b)} \\
& -\hat{\mu}_i^{t} + z_i^{t} + 2\lambda^{1,t} \sum_{j=1}^{n} \sigma_{ij} x_j^{t} + \lambda^{2,t} - \lambda_i^{3,t} = 0, \quad i \in [1,n], \ t \in [T], && \text{(12c)}
\end{aligned}
$$

$$\sum_{i=1}^{n} x_i^t z_i^t = \alpha^t, \quad t \in [T], \tag{12d}$$

$$\xi_i x_i^t = z_i^t \alpha^t, \quad i \in [n],\ t \in [T], \tag{12e}$$

$$z_i^t \geq 0, \quad i \in [1,n],\ t \in [T], \tag{12f}$$

$$\alpha^t > 0, \quad t \in [T], \tag{12g}$$

$$\xi_i > 0. \tag{12h}$$

5.3. Analysis of the bilevel model

In this section, we show that the bilevel model (12) has always an optimal solution that yields arbitrarily small loss.

THEOREM 3. *For every $\epsilon > 0$ and $\{\hat{\boldsymbol{\mu}}^t\}_{t=1}^T$, there exists a diagonal matrix $\boldsymbol{\Xi}$ such that loss$(\{\hat{\boldsymbol{\mu}}\}_{t=1}^T, \boldsymbol{\Xi}) \leq \epsilon$.*

Proof. It suffices to find $\boldsymbol{\xi}$, $\lambda^{1,t}$, $\lambda^{2,t}$, $\boldsymbol{\lambda}^{3,t}$, and $\mathbf{x}^t$ such that they are feasible for (12), and all the vectors $\mathbf{x}^t$ are within a small neighborhood of $\mathbf{x}^*$.

Let $\lambda^{1,t} = 0$, and $\lambda_i^{3,t} = 0$ for all $i \in [1,n]$ and $t \in [1,t]$. Then (12c) can be written as

$$-\hat{\mu}_i^t + z_i^t + \lambda^{2,t} = 0, \quad i \in [1,n],\ t \in [T], \tag{13}$$

Using (12h), we substitute $x_i^t = z_i^t \alpha^t / \xi_i$ in (12g). We see that it suffices to solve the following system to complete the proof.

$$\sum_{i \in [1,n]} \frac{(\hat{\mu}_i^t - \lambda^{2,t})^2}{\xi_i} = 1, \quad t \in [1,T]. \tag{14a}$$

$$z_i^t \geq 0, \quad i \in [1,n],\ t \in [T], \tag{14b}$$

$$\alpha^t > 0, \quad t \in [T], \tag{14c}$$

$$\xi_i > 0. \tag{14d}$$

Let $\tilde{\mathbf{x}}^*$ be as in the proof of Theorem 1.

That is, $\tilde{\mathbf{x}}^*$ is a feasible solution of (1) that satisfies the constraints $\mathbf{x}^\mathsf{T} \boldsymbol{\Sigma} \mathbf{x} \leq v$ and $\mathbf{x} \geq \mathbf{0}$ strictly and is within a small ball of the true optimum $\mathbf{x}^*$.

Pick $\xi_i = M/\tilde{x}_i^*$ for some large M. With this choice, the equations in (14a) are written as

$$\sum_{i\in[1,n]} \frac{(\hat{\mu}_i^t - \lambda^{2,t})^2}{\tilde{x}_i^*} = M, \quad t \in [1,T]. \tag{15}$$

For each $t \in [1,T]$, start by letting $\lambda^{2,t} = \min_{i\in[1,n]}\{\hat{\mu}_i^t\}$ and decrease $\lambda^{2,t}$ until the equation in (15) is satisfied. This will happen, because the left hand side in (15) is continuous and monotonically increasing.

Now, let $z_i^t = \hat{\mu}_i^t - \lambda^{2,t}$. Let $\eta > 0$ be any positive real. Note that by choosing M sufficiently large, we can guarantee that, for each t, all z_i^t are within $1+\eta$ of the smallest. Namely, for all $i \in [1,n]$, $k_t \le z_i^t \le (1+\eta)k_t$, where $k_t = \min_j z_j^t$. This is because the $\hat{\mu}_i^t$ are fixed whereas $-\lambda^{2,t}$ increases as M increases.

For each $t \in [1,T]$, set $\alpha^t = \frac{1}{\sum_i \frac{z_i^t}{\xi_i}}$. Now let $x_i^t = \alpha^t \times \frac{z_i^t}{\xi_i}$. Note that we have $\sum_{i\in[1,n]} x_i^t = 1$ and $x_i^t > 0$. To show that $\mathbf{x}^t$ is feasible, we only need to show that it satisfies the variance constraint. This will follow from showing that $\mathbf{x}^t$ is in a small neighborhood of $\tilde{\mathbf{x}}^*$, which satisfies it strictly. We have $x_i^t = \alpha^t \times \frac{z_i^t}{\xi_i} = \alpha^t \times \frac{z_i^t \tilde{x}_i^*}{M}$. Using our bounds on z_i^t, we get

$$\frac{\alpha^t k_t}{M}\tilde{x}_i^* \le x_i^t \le (1+\eta)\frac{\alpha^t k_t}{M}\tilde{x}_i^*.$$

Adding over i we get $\frac{\alpha^t k_t}{M} \le 1 \le (1+\eta)\frac{\alpha^t k_t}{M}$. This implies

$$(1-\eta)\tilde{x}_i^* \le x_i^t \le (1+\eta)\tilde{x}_i^*.$$

This translates into a vanishing loss in the objective value as η goes to 0. And this concludes the proof. □

We conclude this section by noting that despite the exciting theoretical results of Sections 4 and 5, the non-convex non-linear bilevel model is intractable to solve using the current state-of-the-art software for global optimization. Developing an efficient method to solve the bilevel model presented in this section remains a future work.

6. Using the identity matrix as estimation-error matrix

The results of the previous sections demonstrate that one can focus on diagonal estimation-error matrices and consequently deal with calibrating only n parameters in constructing the uncertainty sets. In this section, we restrict our attention to the simplest diagonal estimation-error matrix, the identity matrix. In this case, we only need to calibrate one parameter(κ), which is advantageous for decision makers.

Our goal is to investigate whether using an identity matrix as the estimation-error matrix can improve the performance of the classical Markowitz model. To this end, we perform a simulation where we compare the out-of-sample performances of the two approaches. We generate samples from the distribution Normal($\boldsymbol{\mu}, \boldsymbol{\Sigma}/N$). The value of N determines the accuracy of the samples we generate (increasing N results in better estimates of the expected return). Then, we solve the classical Markowitz model and the robust portfolio model for varying κ values and compare their actual performances. When $\kappa = \infty$, the robust portfolio becomes the equal-weight portfolio.

		$\kappa \times N$										
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	∞
N	1	13.0	18.3	19.3	19.4	19.3	19.2	19.1	19.0	18.9	18.9	18.4
	3	7.8	13.7	16.3	17.2	17.4	17.4	17.2	17.0	16.8	16.7	15.1
	6	5.3	10.1	13.1	14.6	15.2	15.3	15.2	15.0	14.8	14.6	11.4
	12	3.6	7.0	9.6	11.3	12.2	12.5	12.6	12.4	12.2	11.9	5.8
	24	2.4	4.6	6.4	7.8	8.7	9.1	9.2	9.0	8.7	8.3	-3.4
	120	1.2	2.0	2.7	3.1	3.3	3.2	2.8	2.3	1.6	0.8	-59.2

Table 1 Percentage gap closed by the robust model when $\boldsymbol{\Xi} = \mathbf{I}$ ($v = 0.002$).

Table 1 gives the result of our simulation. Each cell in Table 1 contains the percentage gap closed by the robust solution compared to the Markowitz solution. Specifically, in each cell of Table 1, we report $(\bar{R} - \bar{M})/(T - \bar{M})$ where $\bar{R}$ denotes the actual performance of the robust solution estimates, $\bar{M}$ denotes the actual performance of the Markowitz solution estimates, and T denotes the true return. For example, the entry 19.4 for $N = 1$ and $\kappa \times N = 0.4$ is obtained by first computing the estimates $\bar{M} = 0.01262$ and $\bar{R} = 0.01314$. Similarly, for $N = 120$ and $\kappa \times N = 0.5$, we have that $\bar{M} = 0.01395$ and $\bar{R} = 0.01399$ which yields the entry 3.3. The true return T is equal to 0.01535. Note that for each cell, the values $\bar{R}$ and $\bar{M}$ are averages over 10000 trials; consequently, the standard error for each cell in Table 1 is less than 0.5.

We see from Table 1 that the robust model can significantly outperform the Markowitz model when κ lies in a wide range of values. This is interesting, because the identity matrix contains no information about the problem or the relationship between the assets. Furthermore, the equal-weight portfolio outperforms the Markowitz solution estimates up to $N = 12$. On the other hand, the Markowitz solution estimates are better than equal-weight portfolio when we have more accurate samples (i.e., $N = 24$ and $N = 120$). It is important to note that the robust solution estimates are better than the Markowitz solution estimates even when the samples are very accurate (i.e., $N = 120$).

We repeat the experiments on a different data set based on four international equity indices. We use the monthly returns for Canada, Switzerland, the United Kingdom and the United States between 2004 and 2022. The data is obtained from Morgan Stanley Capital International (MSCI).

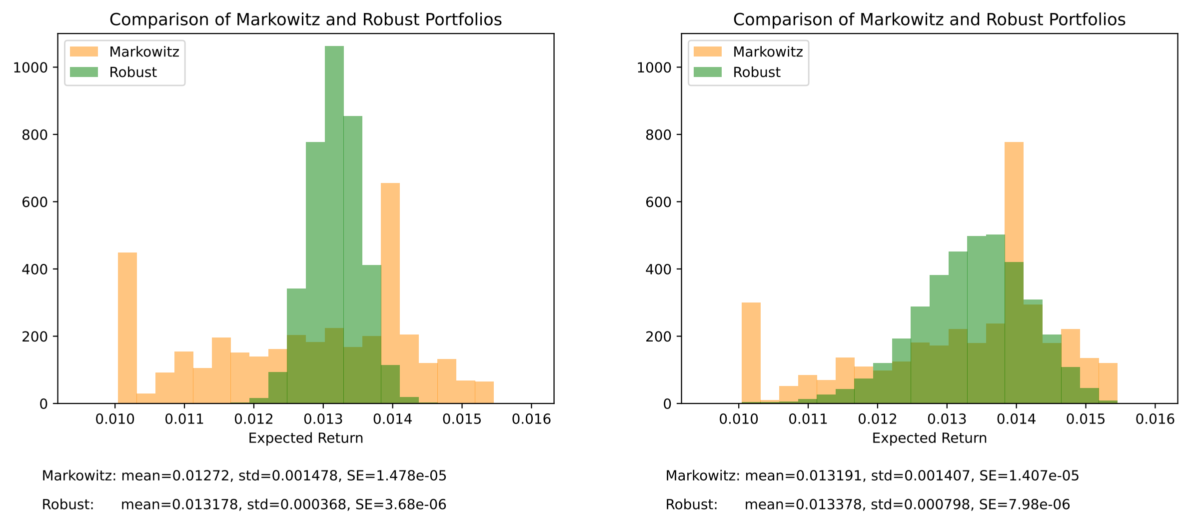
		$\kappa \times N$										
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	∞
N	1	32.9	31.5	30.8	30.5	30.3	30.2	30.1	30.0	30.0	29.9	29.6
	3	31.2	31.9	30.1	29.1	28.5	28.1	27.9	27.7	27.5	27.4	26.3
	6	28.2	32.8	30.6	28.6	27.4	26.5	26.0	25.5	25.2	24.9	22.6
	12	24.8	33.1	32.6	29.9	27.4	25.6	24.3	23.4	22.7	22.1	17.1
	24	21.5	32.0	34.4	32.8	29.7	26.6	23.9	21.8	20.1	18.8	7.4
	120	14.7	25.4	32.4	36.4	38.0	37.7	35.8	32.9	29.2	25.2	-63.3

Table 2 **Percentage gap closed by the robust model when $\Xi = \mathbf{I}$ ($v = 0.0013$).**

We see from Table 2 that the robust portfolios perform better than the portfolios obtained from the classical Markowitz model. In this data set, the expected returns of the minimum variance portfolio and of the equal-weight portfolio are 0.00108 and 0.00473, respectively. The true expected return T is equal to 0.00559. We note that the Markowitz model beats the equal-weight portfolio only when $N = 120$, whereas the robust model beats the equal-weight portfolio under all $N - \kappa \times N$ combinations where $\kappa \times N$ is finite in Table 2. Furthermore, the minimum variance portfolio is outperformed by both the Markowitz model and the robust model under all $N - \kappa \times N$ combinations.

The results in Table 1 and 2 illustrate that the $\kappa \times N$ values that perform best vary between 0.2 and 0.7. We observe that keeping $\kappa \times N$ constant when N varies works well for improving the actual performance of the portfolio. Not surprisingly, the need for robustification diminishes when we have better estimates as N increases.

In order to understand the results better, we focus on two particular distributions $N = 3$ and $N = 24$ when $\kappa \times N = 0.5$ in Table 1. We present the averages $\bar{M}$ and $\bar{R}$ as a histogram. Figure



(a) $n = 3$ (b) $n = 24$

Figure 3 Histograms of the actual performances of Robust and Markowitz models.

3 provides a clearer picture for the superior performance of the robust model. We see that the empirical distributions of the actual returns under the robust model have a smaller variance. On the other hand, the actual returns under the Markowtiz model features outliers that perform very poorly. As a result, the robust model outperforms the Markowitz model, and the difference is significant.

7. Conclusion

In this paper, we take a theoretical look at the robust portfolio optimization problem. Our main contribution is to show that the family of diagonal estimation-error matrices is sufficiently rich to achieve an arbitrarily small loss compared to the optimum expected return. We perform illustrative computational experiments to show that even using an identity matrix can improve on the actual performance of the Markowitz model when the size of the uncertainty set is correctly calibrated.

Our work is a step towards constructing better estimation-error matrices for robust portfolio optimization. Our bilevel model can be used to construct good estimation-error matrices using the available data. This will require developing a more efficient method to solve the bilevel model. Another research direction would be a theoretical analysis on the use of identity matrices as estimation-error matrices.

References

Ban G, El Karoui N, Lim A (2018) Machine learning and portfolio optimization. *Management Science* 64(3):1136–1154.

Bard JF (1998) *Practical Bilevel Optimization*, volume 30 of *Nonconvex Optimization and Its Applications* (Boston, MA: Springer US), ISBN 978-1-4419-4807-6 978-1-4757-2836-1, URL `http://dx.doi.org/10.1007/978-1-4757-2836-1`.

Ben-Tal A, Nemirovski A (1999) Robust solutions of uncertain linear programs. *Operations research letters* 25(1):1–13.

Best MJ, Grauer RR (1991) On the sensitivity of mean-variance efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies* 4(2):315–342.

Broadie M (1993) Computing efficient frontiers using estimated parameters. *Annals of Operations Research* 45(1):21–58, ISSN 0254-5330, 1572-9338, URL `http://dx.doi.org/10.1007/BF02282040`.

Ceria S, Stubbs RA (2006) Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management* 7(2):109–127.

Chopra V, Ziemba W (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management* 19(2):6–11.

Chopra VK (1993) Mean-variance revisited: Near-optimal portfolios and sensitivity to input variations. *Journal of Investing* .

DeMiguel V, Garlappi L, Nogales F, Uppal R (2009a) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55(5):798–812.

DeMiguel V, Garlappi L, Uppal R (2009b) Optimal versus naive diversification: How inneficient is the 1/n portfolio strategy? *Review of Financial Studies* 22(5):1915–1953.

Fabozzi FJ, Kolm PN, Pachamanova DA, Focardi SM (2007) *Robust portfolio optimization and management* (John Wiley & Sons).

Garlappi L, Uppal R, T W (2007) Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies* 20(1):41–81.

Goldfarb D, Iyengar G (2003) Robust Portfolio Selection Problems. *Mathematics of Operations Research* 28(1):1–38, ISSN 0364-765X, 1526-5471, URL `http://dx.doi.org/10.1287/moor.28.1.1.14260`.

Gotoh Jy, Takeda A (2011) On the role of norm constraints in portfolio selection. *Computational Management Science* 8(4):323–353, ISSN 1619-697X, 1619-6988, URL `http://dx.doi.org/10.1007/s10287-011-0130-2`.

Horst JRt, Roon FAd, Werker BJM (2001) Incorporating estimation risk in portfolio choice. *Technical report, Department of Finance, Tilburg University, Netherlands* .

Jagannathan R, Ma T (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* 58:1651–1684.

Kan R, Zhou G (2007) Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42(3):621–656.

Kocuk B, Cornuéjols G (2020) Incorporating black-litterman views in portfolio construction when stock returns are a mixture of normals. *Omega* 91:102008.

Lim A, Shanthikumar J, G-Y V (2012) Robust portfolio choice with learning in the framework of regret: Single-period case. *Management Science* 58(9):1732–1746.

Markowitz H (1952) Portfolio analysis. *Journal of Finance* 8:77–91.

Michaud R (2008) *Efficient asset management* (Oxford University Press).

Olivares-Nadal AV, DeMiguel V (2018) Technical Note—A Robust Perspective on Transaction Costs in Portfolio Optimization. *Operations Research* 66(3):733–739, ISSN 0030-364X, 1526-5463, URL `http://dx.doi.org/10.1287/opre.2017.1699`.

Scherer B (2007) Can robust portfolio optimisation help to build better portfolios? *Journal of Asset Management* 7(6):374–387, ISSN 1470-8272, 1479-179X, URL `http://dx.doi.org/10.1057/palgrave.jam.2250049`.

Stubbs RA, Vance P (2005) Computing return estimation error matrices for robust optimization. *Axioma Research Paper* .

Tutuncu R, Koenig M (2004) Robust asset allocation. *Annalas of Operations Research* 132:157–187.

E-Companion

EC.1. Data set

We use the data set provided in Kocuk and Cornuéjols (2020). The data set includes 360 monthly returns of 11 sectors based on the Global Industrial Classification Standard. We calculate the true return vector $\boldsymbol{\mu}$ and the true covariance matrix $\boldsymbol{\Sigma}$ as the sample average and the sample covariance matrix of these returns. We refer the reader to Kocuk and Cornuéjols (2020) for more details about the dataset.

EC.2. Additional proofs

In this section, we present the additional proofs.

EC.2.1. Proof of Proposition 1

1. We know that there exists a matrix $\mathbf{L}$ such that $\boldsymbol{\Xi} = \mathbf{L}\mathbf{L}^\intercal$, because $\boldsymbol{\Xi}$ is a positive definite matrix. Therefore, the objective function f of (5) can be written as

$$\begin{aligned} f(\mathbf{x}) &= -(\hat{\boldsymbol{\mu}})^\intercal \mathbf{x} + \sqrt{\mathbf{x}^\intercal \boldsymbol{\Xi} \mathbf{x}} \\ &= -(\hat{\boldsymbol{\mu}})^\intercal \mathbf{x} + \|\mathbf{x}\mathbf{L}\|_2 \end{aligned} \tag{EC.1}$$

To conclude the proof, it suffices to show that the norm function $\|\cdot\|_2$ is strictly convex on the feasible region of (5), because the composition with an affine mapping of a strictly convex function is strictly convex.

To this end, we first note that any two distinct point in the feasible region of (5) are linearly independent because of $\mathbf{1}^\intercal \mathbf{x} = 1$. Let $\lambda \in (0,1)$. We want to show that

$$\|\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\|_2 < \lambda \|\mathbf{x}_1\|_2 + (1-\lambda)\|\mathbf{x}_2\|_2$$

We take the square of both sides and observe that

$$\begin{aligned} \|\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\|_2^2 &= \big(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2,\ \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\big) \\ &= \lambda^2 \big(\mathbf{x}_1, \mathbf{x}_1\big) + (1-\lambda)^2 \big(\mathbf{x}_2, \mathbf{x}_2\big) + 2\lambda(1-\lambda)\big(\mathbf{x}_1, \mathbf{x}_2\big) \end{aligned}$$

$$= \lambda^2\|\mathbf{x}_1\|_2^2 + (1-\lambda)^2\|\mathbf{x}_2\|_2^2 + 2\lambda(1-\lambda)\big(\mathbf{x}_1, \mathbf{x}_2\big)$$

$$< \lambda^2\|\mathbf{x}_1\|_2^2 + (1-\lambda)^2\|\mathbf{x}_2\|_2^2 + 2\lambda(1-\lambda)\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2$$

$$= (\lambda\|\mathbf{x}_1\|_2 + (1-\lambda)\|\mathbf{x}_2\|_2)^2$$

The strict inequality above is true due to the fact that Cauchy-Schwarz inequality is strict when $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent.

2. It is clear that the robust portfolio optimization problem (5) is a convex program. It suffices to exhibit an interior point of the feasible region. We assume that v is strictly greater than the minimum variance portfolio. Let $\mathbf{x}^{\text{mv}}$ and $\mathbf{x}^{\text{eq}}$ denote the minimum variance and the equal-weight portfolios, respectively. There exists $\delta \in (0,1)$ such that the variance of $\tilde{\mathbf{x}} = \delta\mathbf{x}^{\text{mv}} + (1-\delta)\mathbf{x}^{\text{eq}}$ is less than v. Clearly, all components of $\tilde{\mathbf{x}}$ is greater than zero. Thus, $\tilde{\mathbf{x}}$ is an exterior point of the feasible region of (5).
3. The square-root function is non-differentiable only at the origin. We have that the objective function (EC.1) is always differentiable in the feasible region of (5), because $\boldsymbol{\Xi}$ is positive definite, and $\mathbf{1}^\mathsf{T}\mathbf{x} = 1$.

EC.2.2. Proof of Proposition 2

We want to show the equivalence of the system (6) and the system $\{(6\text{b}) - (6\text{g}), (9)\}$. We only need to show given $(\boldsymbol{\Xi}, \mathbf{x}, \lambda^1, \lambda^2, \boldsymbol{\lambda}^3)$ that satisfies (6) with $\boldsymbol{\Xi}$ is positive definite and diagonal, there exists $(\mathbf{z}, \alpha, \boldsymbol{\xi}, \mathbf{x}, \lambda^1, \lambda^2, \boldsymbol{\lambda}^3)$ that satisfies $\{(6\text{b}) - (6\text{g}), (9)\}$. The other direction is trivial since the $\boldsymbol{\xi}$ vector in any feasible solution to $\{(6\text{b}) - (6\text{g}), (9)\}$ is a diagonal estimation-error matrix that satisfies the optimality conditions.

Let $\boldsymbol{\xi}$ denote the diagonal entries of $\boldsymbol{\Xi}$. Because $\boldsymbol{\Xi}$ is positive definite, we know that all elements of $\boldsymbol{\xi}$ are strictly greater than zero. It suffices to find $(\mathbf{z}, \alpha)$ that satisfies (9), since $(\mathbf{x}, \lambda^1, \lambda^2, \boldsymbol{\lambda}^3)$ automatically satisfies $(6\text{b}) - (6\text{g})$.

Consider any given $\mathbf{x} >= 0$. Let $P := \{i \in [1,n] : x_i > 0\}$. Let $Z = [1,n] \setminus P$. Since $\mathbf{1}^\mathsf{T}\mathbf{x} = 1$, we know that P is non-empty. This implies that the right hand side of (9c) is positive for all $i \in P$. Therefore, α must be positive.

Now note that we can write (9b) as

$$\begin{aligned}\alpha &= \sum_{i\in P} x_i z_i + \sum_{i\in Z} x_i z_i \\ &= \sum_{i\in P} x_i z_i \\ &= \sum_{i\in P} x_i \frac{\xi_i x_i}{\alpha}\end{aligned}$$

Thus we have that $\alpha = \sqrt{\sum_{i\in P} x_i \xi_i x_i} > 0$. Accordingly, each z_i assumes the value $\frac{\xi_i x_i}{\alpha}$ for all $i \in P$. Letting each $z_i = 0$ for all $i \in Z$ concludes the proof.