Addressing Estimation Errors on Expected Asset Returns through Robust Portfolio Optimization

Gérard Cornuéjols, Özgün Elçi, Vrishabh Patil*
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA

It is well known that the performance of 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 parameters that specify this ellipsoid, namely the point estimate and the estimation-error matrix. We show that there exist diagonal estimation-error matrices that achieve an arbitrarily small loss in the expected portfolio return as compared to the optimum. We empirically investigate the sample size needed to compute the point estimate. We also conduct an empirical study of different estimation-error matrices and give a heuristic to choose the size of the uncertainty set. The results of our experiments show that robust portfolio models featuring a family of diagonal estimation-error matrices outperform benchmark portfolio models including the classical Markowitz model.

Key words: robust optimization, portfolio optimization, optimization under uncertainty

1. Introduction

Consider a portfolio optimization problem where we allocate capital across n assets to maximize the return on investment. If the return vector $\mathbf{r} \in \mathbb{R}^n$ is known, the problem can be formulated as $\max_{\mathbf{x}\in\mathbb{R}^n_+}\{\mathbf{r}^\top\mathbf{x} : \mathbf{1}^\top\mathbf{x} = 1\}$, where $\mathbf{x}$ denotes the fraction of investment in each asset assuming long positions only. In this case, the problem has a trivial optimal solution: invest only in the asset with the largest return.

In practice, however, investors must consider that the assets are risky and that the return vector $\mathbf{r} \sim \mathbb{D}$ belongs to some probability distribution. The classical mean-variance portfolio optimization problem introduced by Markowitz (1952) addresses this uncertainty by maximizing the *expected* return of the portfolio subject to a constraint on the risk modeled as the variance of the portfolio return. Let $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ denote the expectation vector and covariance matrix of the asset returns, respectively. Then the Markowitz model is formulated as

$$\max_{\mathbf{x}\in\mathbb{R}^n_+} \boldsymbol{\mu}^\top\mathbf{x} \tag{1}$$

* Corresponding author at: 4765 Forbes Ave, Pittsburgh, PA 15213, United States of America.
E-mail address: vmpatil@andrew.cmu.edu.

$$\text{subject to } \mathbf{x}^\top \boldsymbol{\Sigma} \mathbf{x} \leq v \tag{2}$$

$$\mathbf{1}^\top \mathbf{x} = 1. \tag{3}$$

Despite the theoretical promise of this mean-variance model, practitioners face an overwhelming challenge. The *true* expectation vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the random asset returns are unknown. Therefore, one can only optimize (1)-(3) with *estimated* parameters.

1.1. Literature Review

It has been observed that even small errors in estimating $\boldsymbol{\mu}$ produce large changes in the returns of portfolio holdings (see, for example, Best and Grauer 1991, Chopra and Ziemba 1993, Michaud and Michaud 2008). The issue of obtaining reliable estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ has been studied extensively, leading to a vast literature. Several papers advocate the use of shrinkage estimators, such as the James-Stein (James and Stein 1992), Jobson (Jobson 1979), Jorion (Jorion 1986), Frost-Savarino (Frost and Savarino 1986), and Ledoit-Wolf (Ledoit and Wolf 2003, 2004a,b, 2020) estimators. The James-Stein, Jobson, and Jorion procedures are estimators for the mean asset returns, the Frost-Savarino procedure is a joint estimator of the means and covariances, and the Ledoit-Wolf procedure is an estimator of the covariance. Shrinkage is often interpreted as a form of empirical Bayesian procedure, which assumes a prior to establish an exogenous structure on potential estimates. We refer the reader to Avramov and Zhou (2010) for a survey of Bayesian procedures used in portfolio selection. An interesting approach to addressing estimation errors is the Black-Litterman model (Black and Litterman 1990, 1992), which combines market information and investor views into the mean-variance optimization problem. Other works have incorporated diversification across market and estimation risk into the model (Jagannathan and Ma 2003, ter Horst et al. 2006, Kan and Zhou 2007). ter Horst et al. (2006) also suggest portfolio weight adjustments. DeMiguel et al. (2009), Brodie et al. (2009), Gotoh and Takeda (2011) additionally study the imposition of norm constraints to regularize the optimal portfolio against large errors. There have also been studies on the intersection of machine learning and portfolio optimization in the context of error mitigation (Lim et al. 2012, Ban et al. 2018). Finally, we name robust portfolio optimization. Robust optimization has been well-studied in the portfolio management literature Goldfarb and Iyengar (2003), Tütüncü and Koenig (2004), Natarajan et al. (2008), Calafiore and Monastero (2012), Bertsimas et al. (2018). This paper contributes to the stream of robust portfolio optimization literature, focusing on errors on the estimates for $\boldsymbol{\mu}$.

1.2. Background on Robust Optimization

We assume that the *true* expected return vector $\boldsymbol{\mu}$ that parametrizes the real-world returns distribution is unknown and belongs to an ellipsoidal uncertainty set given by

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

where $\hat{\boldsymbol{\mu}}$ is an *estimated* expected return vector and $\boldsymbol{\Xi}$ is a positive definite matrix referred to as the estimation-error matrix. The robust optimization problem hopes to find a portfolio that maximizes the expected returns of the investment in a worst-case scenario $\boldsymbol{\mu} \in \mathcal{U}$. That is, the robust optimization problem is formulated as

$$\max_{\mathbf{x}\in\mathcal{X}} \min_{\boldsymbol{\mu}\in\mathcal{U}} \left\{\boldsymbol{\mu}^\top \mathbf{x}\right\}$$

where $\mathcal{X}$ is the feasible region characterized by constraints (2) and (3) over the non-negative orthant. Estimation errors in the mean $\boldsymbol{\mu}$ have a greater impact on portfolio performance than those in the covariance matrix $\boldsymbol{\Sigma}$ or other parameters (Jagannathan and Ma 2003, DeMiguel et al. 2009, Gotoh and Takeda 2011). Indeed, Chopra and Ziemba (1993) show that errors in $\boldsymbol{\mu}$ are around twenty times as significant as errors in the covariances (see also Ulf and Raimond 2006). Additionally, there is a common perception that return variances and covariances are much easier to estimate from historical data, and that a few factors can capture the general covariance structure, making it more manageable to estimate compared to expected returns (Merton 1980, Nelson 1992, Chan et al. 1999). This paper focuses on the estimation of $\boldsymbol{\mu}$ and on how to deal with estimation errors in $\boldsymbol{\mu}$. For the remainder of the paper, we assume that the *true* covariance matrix $\boldsymbol{\Sigma}$ is known.

Following Ben-Tal and Nemirovski (1999), the robust optimization problem can be reformulated as a quadratically constrained convex programming problem

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

Here $\kappa > 0$ and the term $\sqrt{\mathbf{x}^\top \boldsymbol{\Xi}\mathbf{x}}$ can be interpreted as an estimation risk that must be considered in addition to the market risk $\mathbf{x}^\top \boldsymbol{\Sigma}\mathbf{x}$ (see Fabozzi et al. 2007, p. 371). While κ can be absorbed into $\boldsymbol{\Xi}$ in the reformulation, we maintain a distinction between the two terms as κ may be interpreted as the weight on the estimation risk relative to the expected return term in the objective.

In this paper we assume that we have access to historical data on asset returns over an extended period. A fascinating question in data science is to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Xi}$ from this data, especially when one can expect some degree of obsolescence in the older data.

1.3. Contributions

A common critique of robust portfolio models is that they often lack guidance on how to define the uncertainty set. Indeed, the literature on constructing $\mathcal{U}$ in portfolio management is scarce. Stubbs and Vance (2005) provide a comprehensive overview of the practical impacts of computing suitable estimation-error matrices. Additionally, there have been studies in which a scalar multiple of $\boldsymbol{\Sigma}$ is used as the estimation-error matrix (Scherer 2007, Garlappi et al. 2007). 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.

The primary contribution of this paper is a practical framework for constructing the uncertainty set $\mathcal{U}$ for the expected return vector $\boldsymbol{\mu}$. We focus on constructing $\mathcal{U}$ based on observed historical asset returns. We conduct theoretical and empirical analyses on the choice of the estimation-error matrix $\boldsymbol{\Xi}$ and the parameter κ, and provide empirically strong choices for the sample size of historical data used to estimate $\boldsymbol{\mu}$. We first give theoretical results on the class of *diagonal* estimation-error matrices for the choice of $\boldsymbol{\Xi}$ in Section 2. In particular, we show that one can achieve an arbitrarily small loss in the expected portfolio return as compared to the optimal portfolio. We accomplish this by showing that there exists a choice of parameters for the diagonal estimation-error matrix $\boldsymbol{\Xi}$ such that the resulting solution to (4) is arbitrarily close to the optimal portfolio. This is an existential result which, unfortunately, does not translate into actual portfolios in practice. In the following sections, we address constructive aspects of the uncertainty set. In Section 3, we evaluate different sample sizes for constructing the estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ from historical returns. Our results highlight the existence of a gap between the expected returns of the Markowitz mean-variance portfolio, constructed using the estimated $\hat{\boldsymbol{\mu}}$, and the optimal portfolio, constructed using the true $\boldsymbol{\mu}$, across all sample sizes. We observe that larger sample sizes do not produce superior portfolios; rather, an "intermediate" sample size achieves the best results. We explore choices for the estimation-error matrix $\boldsymbol{\Xi}$ through empirical experiments and present a heuristic to calibrate κ in Section 4. We perform computational experiments on simulated, synthetic data drawn from distributions with parameters constructed from historical asset returns and devise choices for $\boldsymbol{\Xi}$ and κ that lead to statistically significant improvements. Given that it is infeasible to meaningfully obtain valid and statistically significant results on real-world returns, we validate our findings on additional synthetic data with added temporal uncertainty in the asset returns. Our results demonstrate that a robust portfolio model featuring a family of diagonal estimation-error matrices and an appropriate choice for κ found using our proposed heuristic can significantly improve the performance of the Markowitz model. Section 5 contains some concluding remarks.

2. Diagonal Estimation-Error Matrices

In this section, we present two existential theorems, which demonstrate that one can focus on diagonal estimation-error matrices without compromising on the quality of the portfolio. The construction of these matrices in the proofs relies on knowledge of the true optimal portfolio, which is not available in practice. The purpose of these theoretical results is not to propose a directly implementable method, but rather to motivate the empirical investigation of diagonal estimation-error matrices by establishing that, in principle, such structures can achieve near-optimal performance. We explore practical, data-driven constructions of such matrices in later sections.

Diagonal estimation-error matrices were first studied by Stubbs and Vance (2005), where the authors argue that a simple diagonal estimation-error matrix is easier to generate than a dense

estimation-error matrix. However, a natural question arises: does there exist a trade-off in the performance of the robust portfolio if constructed with the simpler diagonal estimation-error matrix? We address this question by presenting our main theoretical result, which states that for any estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$, one can always choose a diagonal, positive definite matrix $\boldsymbol{\Xi}$ and a positive parameter κ such that the resulting robust portfolio has an expected return arbitrarily close to that of the optimal portfolio.

Given a feasible portfolio $\hat{\mathbf{x}}$ and an optimal portfolio $\mathbf{x}^*$, we define the *loss* in expected returns as

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

where $\boldsymbol{\mu}$ is the true expected return. A solution $\hat{\mathbf{x}}^R$ to the robust portfolio problem

$$\max_{\mathbf{x}} \ \hat{\boldsymbol{\mu}}^\top \mathbf{x} - \kappa\sqrt{\mathbf{x}^\top \boldsymbol{\Xi}\mathbf{x}} \tag{5}$$

$$\text{subject to } \mathbf{x}^\top \boldsymbol{\Sigma}\mathbf{x} \le v \tag{6}$$

$$\mathbf{1}^\top \mathbf{x} = 1 \tag{7}$$

$$\mathbf{x} \ge \mathbf{0} \tag{8}$$

depends on the expected return estimate $\hat{\boldsymbol{\mu}}$, the estimation-error matrix $\boldsymbol{\Xi}$, and the parameter κ. In this case, assuming feasibility, we write

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

THEOREM 1. *Given $\epsilon > 0$, for every $\hat{\boldsymbol{\mu}}$ there exists a diagonal, positive definite matrix $\boldsymbol{\Xi}$ and $\kappa > 0$ such that $\text{loss}(\hat{\boldsymbol{\mu}}, \boldsymbol{\Xi}, \kappa) < \epsilon$. Furthermore, for any optimal portfolio $\mathbf{x}^*$, we can choose $\boldsymbol{\Xi}$ and κ such that the robust problem (5)-(8) has a solution $\hat{\mathbf{x}}^R$ that is arbitrarily close to $\mathbf{x}^*$.*

Proof. Let $\mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n_+ : \ \mathbf{x}^\top \boldsymbol{\Sigma}\mathbf{x} \le v, \ \mathbf{1}^\top \mathbf{x} = 1\}$ be the set of feasible portfolios. If $\mathcal{X} = \emptyset$ or a single point the theorem trivially holds, so we assume that the dimension of $\mathcal{X}$ is at least 1. Let $I^0 := \{i \in \{1, 2, \ldots, n\} : \ x_i = 0 \text{ for all } \mathbf{x} \in \mathcal{X}\}$ and $I^+ := \{1, 2, \ldots, n\} \setminus I^0$. Note that I^0 may be nonempty, for example, when v is set equal to the variance of a minimum-variance portfolio: $v = \min_{\mathbf{x} \in \mathbb{R}^n_+} \{\mathbf{x}^\top \boldsymbol{\Sigma}\mathbf{x} : \mathbf{1}^\top \mathbf{x} = 1\}$. Pick $\tilde{\mathbf{x}} \in \text{rint}(\mathcal{X})$, where $\text{rint}(\mathcal{X})$ denotes the relative interior of the convex set $\mathcal{X}$. Note that $\tilde{x}_i = 0$ for $i \in I^0$ and $\tilde{x}_i > 0$ for $i \in I^+$. Let $\mathbf{x}^*$ be an optimal portfolio and let $\tilde{\mathbf{x}}^* = (1 - \epsilon')\mathbf{x}^* + \epsilon'\tilde{\mathbf{x}}$ for $\epsilon' \in (0, 1)$. Then $\tilde{\mathbf{x}}^* \in \text{rint}(\mathcal{X})$.

Observe that $\tilde{\mathbf{x}}^* \to \mathbf{x}^*$ as $\epsilon' \to 0$. This implies $\boldsymbol{\mu}^\top \tilde{\mathbf{x}}^* \to \boldsymbol{\mu}^\top \mathbf{x}^*$ as $\epsilon' \to 0$. Therefore, the loss corresponding to $\tilde{\mathbf{x}}^*$ is arbitrarily small. We now show that there exists a choice of $\boldsymbol{\Xi}$ and κ such that the solution to the robust portfolio problem is arbitrarily close to $\tilde{\mathbf{x}}^*$, thus proving the theorem.

First, suppose $\hat{\boldsymbol{\mu}} = \mathbf{0}$. Choose $\kappa > 0$, $\xi_i = \frac{1}{\tilde{x}_i^*}$ for $i \in I^+$ and $\xi_i = 1$ for $i \in I^0$. We claim that $\tilde{\mathbf{x}}^*$ solves the robust portfolio problem $\max_{\mathbf{x} \in \mathcal{X}} -\kappa\sqrt{\mathbf{x}^\top \boldsymbol{\Xi}\mathbf{x}} \equiv \min_{\mathbf{x} \in \mathcal{X}} \mathbf{x}^\top \boldsymbol{\Xi}\mathbf{x}$. Indeed, let $\mathcal{Y} := \{\mathbf{x} \in$

$\mathbb{R}^{I^+} \times \{0\}^{I^0} : \mathbf{1}^\top \mathbf{x} = 1\}$ be the relaxation of the set $\mathcal{X}$ where $x_i = 0$ for all $i \in I^0$ and the non-negativity and risk constraints have been dropped. We have $\min_{\mathbf{x}\in\mathcal{X}} \mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x} \geq \min_{\mathbf{x}\in\mathcal{Y}} \mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x} = \max_{\lambda\in\mathbb{R}} \min_{\mathbf{x}\in\mathbb{R}^{I^+}} \sum_{i\in I^+} \frac{x_i^2}{\tilde{x}_i^*} + \lambda(1 - \sum_{i\in I^+} x_i)$ where equality holds because $\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}$ is a convex function and $\mathcal{Y}$ is an affine space. Solving this Lagrangian problem, the partial derivatives are $\frac{2x_i}{\tilde{x}_i^*} - \lambda = 0$ for $i \in I^+$ and $1 - \sum_{i\in I^+} x_i = 0$, which yields the unique optimal solution $x_i = \tilde{x}_i^*$ for $i \in I^+$ and $\lambda = 2$. Since the optimal solution $\tilde{\mathbf{x}}^*$ to $\min_{\mathbf{x}\in\mathcal{Y}} \mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}$ is also a feasible solution in $\mathcal{X}$, it is the unique optimal solution to $\min_{\mathbf{x}\in\mathcal{X}} \mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}$ as well. This proves the claim.

For any given $\hat{\boldsymbol{\mu}}$, again let $\xi_i = \frac{1}{\tilde{x}_i^*}$ for $i \in I^+$ and $\xi_i = 1$ for $i \in I^0$. Consider the problem $\max_{\mathbf{x}\in\mathcal{X}} \hat{\boldsymbol{\mu}}^\top \mathbf{x} - \kappa\sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} \equiv \min_{\mathbf{x}\in\mathcal{X}} \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} - \eta \hat{\boldsymbol{\mu}}^\top \mathbf{x}$ where $\eta = \frac{1}{\kappa} > 0$. Let $g(\mathbf{x}, \eta) := \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} - \eta \hat{\boldsymbol{\mu}}^\top \mathbf{x}$. Let $g^*(\eta) := \min_{\mathbf{x}\in\mathcal{X}} g(\mathbf{x}, \eta)$ and let $\mathbf{x}^*(\eta)$ be an optimal solution. Define $M := \max_{\mathbf{x}\in\mathcal{X}} |\hat{\boldsymbol{\mu}}^\top \mathbf{x}|$. The existence of the maximum follows from the boundedness of $\mathcal{X}$. Let $h^*(\eta) := \min_{\mathbf{x}\in\mathcal{X}} \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} - \eta M = \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*} - \eta M$. Clearly, $h^*(\eta) \leq g^*(\eta) \leq \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*} - \eta \hat{\boldsymbol{\mu}}^\top \tilde{\mathbf{x}}^*$. Taking the limit as $\eta \to 0$, we get $\sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*} \leq \lim_{\eta\to 0} g^*(\eta) \leq \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*}$. Therefore $\lim_{\eta\to 0} g^*(\eta) = \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*}$. We claim that, in fact, $\lim_{\eta\to 0} \mathbf{x}^*(\eta) = \tilde{\mathbf{x}}^*$. To prove this claim, we will show that $\forall \epsilon'' > 0, \exists \eta_0 > 0$ such that $\forall\, 0 \leq \eta \leq \eta_0 \;\; ||\mathbf{x}^*(\eta) - \tilde{\mathbf{x}}^*||_2 \leq \epsilon''$.

Because $\sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}}$ is a convex function (the matrix $\boldsymbol{\Xi}$ is positive definite and therefore $\sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}}$ is a Mahalabonis norm), $\mathcal{X}$ is a convex set, and $\tilde{\mathbf{x}}^*$ is the unique optimal solution to $\min_{\mathbf{x}\in\mathcal{X}} \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}}$, it follows that (see, for example, Corollary 27.2.2 in Rockafellar (1970))

$$\forall \epsilon'' > 0 \;\; \exists \zeta > 0 \text{ such that } |\sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} - \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*}| \leq \zeta \text{ for } \mathbf{x} \in \mathcal{X}, \text{ implies } ||\mathbf{x} - \tilde{\mathbf{x}}^*||_2 \leq \epsilon''. \tag{9}$$

Given such a scalar ζ, the boundedness of $\mathcal{X}$ and $\lim_{\eta\to 0} g^*(\eta) = \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*}$ imply, respectively,

$$\exists \eta_1 > 0 \text{ such that } \forall\, 0 \leq \eta \leq \eta_1 \;\; |\eta \hat{\boldsymbol{\mu}}^\top \mathbf{x}| < \zeta/2 \text{ for all } \mathbf{x} \in \mathcal{X} \tag{10}$$

$$\exists \eta_2 > 0 \text{ such that } \forall\, 0 \leq \eta \leq \eta_2 \;\; |g^*(\eta) - \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*}| < \zeta/2 \tag{11}$$

Set $\eta_0 = \min\{\eta_1, \eta_2\}$ and let $0 \leq \eta \leq \eta_0$. Replacing $g^*(\eta)$ in (11), we get

$$|\sqrt{\mathbf{x}^*(\eta)^\top \boldsymbol{\Xi} \mathbf{x}^*(\eta)} - \eta \hat{\boldsymbol{\mu}}^\top \mathbf{x}^*(\eta) - \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*}| < \zeta/2$$

It now follows from (10) and the triangle inequality that

$$|\sqrt{\mathbf{x}^*(\eta)^\top \boldsymbol{\Xi} \mathbf{x}^*(\eta)} - \sqrt{\tilde{\mathbf{x}}^{*\top} \boldsymbol{\Xi} \tilde{\mathbf{x}}^*}| < \zeta.$$

Therefore, by (9),

$$\forall \epsilon'' > 0 \;\; \exists \eta_0 > 0 \;\; \forall\, 0 < \eta \leq \eta_0 \;\; ||\mathbf{x}^*(\eta) - \tilde{\mathbf{x}}^*||_2 \leq \epsilon''$$

proving the claim. □

We now give an example to show that it is not always possible to choose a diagonal positive definite matrix $\boldsymbol{\Xi}$ and a κ such that $\text{loss}(\hat{\boldsymbol{\mu}}, \boldsymbol{\Xi}, \kappa) = 0$.

EXAMPLE 1. Consider a two asset portfolio optimization problem. Let $\sigma_{11} < v$, $\sigma_{22} > v$, and $\mu_1 > \mu_2$. Suppose $\hat{\mu}_1 < \hat{\mu}_2$. Then there does not exist a diagonal positive definite matrix $\boldsymbol{\Xi}$ and a parameter κ such that $\text{loss}(\hat{\boldsymbol{\mu}}, \boldsymbol{\Xi}, \kappa) = 0$.

It is easy to see that the optimal portfolio to the problem (1)-(3) is $\mathbf{x}^* = (x_1 = 1, x_2 = 0)$ and that it is unique. Moreover, we observe that both constraint (6) and the constraint $x_1 \geq 0$ are not tight for $\mathbf{x}^*$. Then the Lagrangian dual of the robust portfolio problem can be written as

$$\min_{\lambda \in \mathbb{R},\ \tau \in \mathbb{R}_+} \ \max_{x_1, x_2} \ \hat{\mu}_1 x_1 + \hat{\mu}_2 x_2 - \kappa \sqrt{x_1^2 \xi_1 + x_2^2 \xi_2} + \lambda(1 - (x_1 + x_2)) + \tau x_2$$

with the optimality conditions

$$x_1 + x_2 = 1 \tag{12}$$

$$\hat{\mu}_1 - \frac{\kappa x_1 \xi_1}{\sqrt{x_1^2 \xi_1 + x_2^2 \xi_2}} - \lambda = 0 \tag{13}$$

$$\hat{\mu}_2 - \frac{\kappa x_2 \xi_2}{\sqrt{x_1^2 \xi_1 + x_2^2 \xi_2}} - \lambda + \tau = 0. \tag{14}$$

Since $\kappa \geq 0$, $\boldsymbol{\Xi} \succ 0$ and $\tau \geq 0$, the optimality conditions imply that $\hat{\mu}_1 = \lambda + \kappa\sqrt{\xi_1} \geq \lambda$ and $\hat{\mu}_2 = \lambda - \tau \leq \lambda$ for $\mathbf{x}^*$ for any choice of ξ_1, ξ_2. However, this is a contradiction to since $\hat{\mu}_1 < \hat{\mu}_2$.

Despite it not being possible to achieve zero loss generally, our next result states a sufficient condition for obtaining zero loss.

THEOREM 2. *If all the assets are active in the true optimal solution* $\mathbf{x}^*$*, then for every* $\hat{\boldsymbol{\mu}}$ *there exists a diagonal, positive definite matrix* $\boldsymbol{\Xi}$ *and* $\kappa > 0$ *such that* $\text{loss}(\hat{\boldsymbol{\mu}}, \boldsymbol{\Xi}, \kappa) = 0$.

Proof. Incorporating $\kappa > 0$ into the matrix $\boldsymbol{\Xi}$, the robust portfolio problem to solve is $\max_{\mathbf{x} \in \mathcal{X}} \hat{\boldsymbol{\mu}}^\top \mathbf{x} - \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}}$.

We will construct a diagonal matrix $\boldsymbol{\Xi}$ with diagonal entries ξ_i. Choose a scalar λ such that $\hat{\mu}_i + \lambda > 0$ for all $i = 1, \ldots, n$ and set

$$\xi_i = \frac{(\hat{\mu}_i + \lambda) \sum_{j=1}^n (\hat{\mu}_j + \lambda) x_j^*}{x_i^*}.$$

Note that $\xi_i > 0$ for $i = 1, \ldots, n$. So $\boldsymbol{\Xi}$ is a positive definite matrix.

Let $\mathcal{Y} = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{1}^\top \mathbf{x} = 1\}$ be the relaxation of $\mathcal{X}$ obtained by dropping the non-negativity and risk constraints. We have $\max_{\mathbf{x} \in \mathcal{X}} \hat{\boldsymbol{\mu}}^\top \mathbf{x} - \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} \leq \max_{\mathbf{x} \in \mathcal{Y}} \hat{\boldsymbol{\mu}}^\top \mathbf{x} - \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} = \min_{\lambda \in \mathbb{R}} \max_{\mathbf{x} \in \mathbb{R}^n} \hat{\boldsymbol{\mu}}^\top \mathbf{x} - \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} - \lambda(1 - \mathbf{1}^\top \mathbf{x})$.

Let $\alpha := \sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}} = \sqrt{\sum_{j=1}^n \xi_j x_j^2}$. Taking partial derivatives of the above Lagrangian function, we get the optimality conditions

$$\hat{\mu}_i - \frac{\xi_i x_i}{\alpha} + \lambda = 0 \qquad \text{for } i = 1, \ldots, n$$
$$\sum_{i=1}^n x_i = 1.$$

These optimality conditions are satisfied by the vector $\mathbf{x} = \mathbf{x}^*$. Indeed, when $\mathbf{x} = \mathbf{x}^*$, we have

$$\alpha = \sqrt{\sum_{i=1}^n \xi_i x_i^{*2}} = \sqrt{\sum_{i=1}^n (\hat{\mu}_i + \lambda)\big(\sum_{j=1}^n (\hat{\mu}_j + \lambda) x_j^*\big) x_i^*} = \sum_{j=1}^n (\hat{\mu}_j + \lambda) x_j^*$$

It follows that $\hat{\mu}_i - \frac{\xi_i x_i^*}{\alpha} + \lambda = 0$ for $i = 1, \ldots, n$. We also have $\sum_{i=1}^n x_i^* = 1$. Therefore $\mathbf{x}^*$ satisfies the optimality conditions of the Lagrangian problem. Since the solution $\mathbf{x} = \mathbf{x}^*$ belongs to $\mathcal{X}$, it is an optimal solution to the robust portfolio problem. □

3. Estimating the Expected Return Vector Using Observed Historical Asset Returns

In order to use model (4), a practitioner needs to provide Ξ and $\hat{\boldsymbol{\mu}}$. We discuss the choice of Ξ in Section 4. In this section, we address the estimation $\hat{\boldsymbol{\mu}}$ of the vector $\boldsymbol{\mu}$ of expected asset returns. We assume that we have access to historical data. If the historical returns are independent and identically distributed (i.i.d.) random variables, then certainly, by the law of large numbers, we obtain the best estimate for $\boldsymbol{\mu}$ by using all of the data. However, in reality, the historical returns may become obsolete over time. Therefore it may not be appropriate to assume time independence over long periods. Indeed, Ulf and Raimond (2006) present an empirical study where even the simple equal-weight portfolio outperforms portfolios constructed under the assumption that the returns are i.i.d. Our focus is on determining an appropriate sample size for estimating $\boldsymbol{\mu}$ from historical returns. Suppose $r^1, r^2, \ldots, r^H$ are the vectors of historical asset returns observed in the real world in some time horizon $1, 2, \ldots, H$, possibly over several decades. It is natural to estimate the expected asset returns as the average of the N most recent observations. But then we should ask: what choice of N gives the best estimates for $\boldsymbol{\mu}$? To this end, we investigate the performance of portfolios constructed using the Markovitz model with different sample sizes for estimating $\boldsymbol{\mu}$. To be able to repeat the experiments and obtain statistically significant results, we generate a set of synthetic, simulated asset returns, which we briefly describe below.

We generate simulated returns $\hat{r}^t \sim \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma})$, following a multivariate normal distribution with a "true" expected return vector $\boldsymbol{\mu}^t$ and covariance matrix $\boldsymbol{\Sigma}$. That is, the simulated returns $\hat{r}^t$ are independent random variables, drawn from different distributions for each $t = 1, \ldots, H$. We refer

the reader to Appendix A for our construction of the parameters $\boldsymbol{\mu}^t$ and $\boldsymbol{\Sigma}$. To summarize, $\boldsymbol{\mu}^t$ is computed from the real data r^i by averaging over T consecutive time periods, centered at t. Asset returns $\hat{r}^i$ are then generated using the distribution $\mathcal{N}(\boldsymbol{\mu}^i, \boldsymbol{\Sigma})$. A Markowitz portfolio $\hat{\mathbf{x}}^t$ is constructed at time t using the estimate $\hat{\boldsymbol{\mu}}^t = \frac{1}{N}\sum_{i=t-N+1}^{t} \hat{r}^i$ based on the asset returns $\hat{r}^i$ available in the N periods preceding t. This portfolio is then held until period $t+T$ and evaluated using $\boldsymbol{\mu}^{t+T}$. Note that, by construction, $\boldsymbol{\mu}^{t+T}$ is derived from future asset returns $\hat{r}^i$ not used in the estimate $\hat{\boldsymbol{\mu}}^t$ at time t. Clearly, if the data r^i were independent over time, the vector $\boldsymbol{\mu}^{t+T}$ would be independent of the values $\boldsymbol{\mu}^i$ for $i \leq t$. Therefore, estimates $\hat{\boldsymbol{\mu}}^t$ that are averaged over larger N would provide better estimates of $\boldsymbol{\mu}^{t+T}$ and therefore better Markowitz portfolios $\hat{\mathbf{x}}^t$. Our experiments below show that this is not the case.

In addition to the Markowitz portfolio $\hat{\mathbf{x}}^t := \max_{\mathbf{x}\in\mathcal{X}} \hat{\boldsymbol{\mu}}^{t\top}\mathbf{x}$, we construct the optimal portfolio at time t, namely $\mathbf{x}^{t*} := \max_{\mathbf{x}\in\mathcal{X}} \boldsymbol{\mu}^{t\top}\mathbf{x}$. We compute its *true expected return* $\boldsymbol{\mu}^{t+T^\top}\mathbf{x}^{t*}$, assuming we hold it for the next T periods.

For each choice of N, we compute the *estimated expected return* $\hat{\boldsymbol{\mu}}^{t^\top}\hat{\mathbf{x}}^t$ of the Markowitz portfolio $\hat{\mathbf{x}}^t$, and its *actual expected return* $\boldsymbol{\mu}^{t+T^\top}\hat{\mathbf{x}}^t$, all averaged over the simulated runs and over the time periods t. We evaluate the average performance of these portfolios over several simulated runs using *efficient frontiers* that plot the expected returns against different risk thresholds v. We achieve a standard error less than 0.05 for each reported value. A comprehensive overview of our experimental setup can be found in Appendix A.

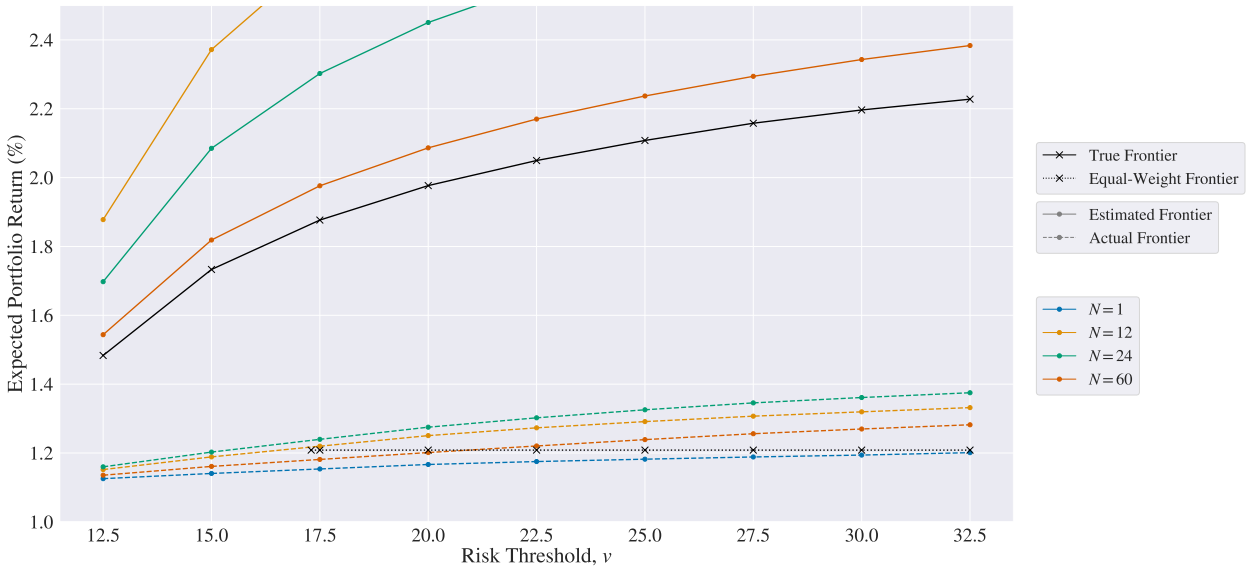

Figure 1 **Efficient frontiers of portfolios constructed on the 11-asset GICS monthly returns dataset.**

Figures 1 and 2 present the experimental results for an 11-asset dataset with monthly returns and a 12-asset dataset with daily returns. The average expected returns for the equal-weight portfolios

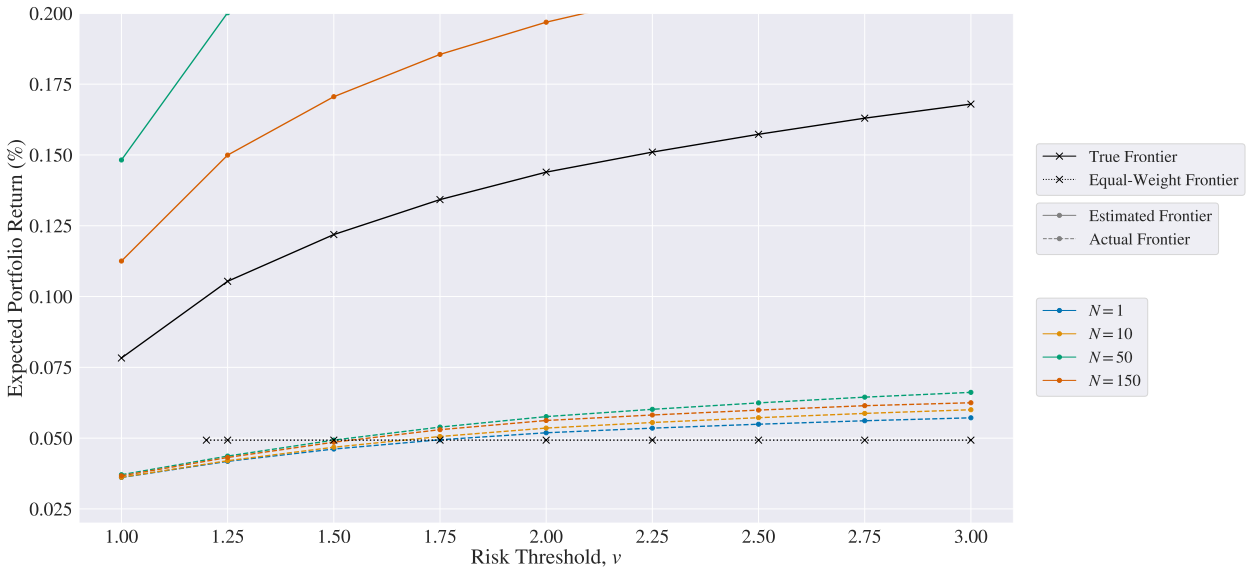

Figure 2 **Efficient frontiers of portfolios constructed on the 12-asset Fama-French daily returns dataset.**

were 1.21% per month and 0.049% per day, respectively, while the minimum variance portfolios yielded 1.23% and 0.034% on average, respectively. The corresponding minimum variances were 11.46 and 0.84. Figure 1 plots efficient frontiers for Markowitz portfolios obtained for different choices of N in the estimation of $\hat{\boldsymbol{\mu}}^t$. Note that for any $N > 1$, the Markowitz optimal portfolios outperform the equal-weight portfolio when the risk threshold is $v \geq 21.25$. This trend is also observed in Figure 2 for $v \geq 1.75$. Both figures reveal a significant gap between the true frontier and the actual frontier for all values of N. This gap quantifies the value of information: Knowing $\boldsymbol{\mu}$ improves the expected portfolio return from around 1.3% per month to about 2% per month for a risk level $v = 20$. Interestingly, the actual frontier initially improve with increasing N, but then deteriorate for larger N. The best results are achieved by choosing N = 24 months (Figure 1) and N = 50 days (Figure 2). Large values of N produce poor portfolios. Perhaps this is not surprising as the returns are not identically distributed over time. We note that the equal-weight portfolio has a fixed variance, which in these cases is 17.35 (Figure 1) and 1.18 (Figure 2). The equal-weight frontier is shown only for values of the risk threshold v above this level, since it is not feasible for the equal-weight portfolio to satisfy more stringent risk constraints. We repeated the experiments on other data sets summarized in Appendix C. The results are similar. The gap between true and actual frontiers is very significant. The quality of the Markowitz portfolios improves as N increases up to some value N^*, and then it deteriorates as N increases further. For monthly data, averaging over 5 data sets and 3 risk levels, we find $N^* \approx 28$ albeit with substantial variation; for daily data, averaging over 4 data sets and 3 risk levels, $N^* \approx 100$ again with much variation. In Section 4 we use $N^* = 24$ for monthly datasets and $N^* = 100$ for daily datasets. These values for N^* are not intended to be prescriptive across all datasets, but their existence demonstrates that one cannot

mitigate the effects of estimation errors on $\boldsymbol{\mu}$ by using a large N. This lends further support to the usefulness of a robust optimization framework.

While there exist several other methods for estimating the expected return vector $\boldsymbol{\mu}$, Jagannathan and Ma (2003), DeMiguel et al. (2009) show that estimation errors in the sample mean persist regardless of the estimation technique used. In our study, we limit our analyses to estimators obtained from observed historical asset returns.

The significant gap between the true and actual frontiers observed in Figures 1 and 2 highlights the potential for better portfolios. We show in the next section that robust optimization can indeed deliver superior portfolios.

4. Empirical Study of the Estimation-Error Matrix

In this section, we investigate choices for the estimation-error matrix $\boldsymbol{\Xi}$ in (4) and the associated parameter κ. While robust optimization techniques have been previously employed in the context of portfolio management, there is little guidance on constructing $\boldsymbol{\Xi}$. Following our theoretical results from Section 2, we explore the class of diagonal estimation-error matrices. Such a choice for $\boldsymbol{\Xi}$ requires practitioners to only calibrate $n+1$ parameters. We first conduct experiments on simulated streams of i.i.d. returns. In later parts of the section, we validate the results of these experiments with the setting introduced in Section 3, for which we consider returns that are not necessarily independently distributed.

Following the works of Scherer (2007) and Garlappi et al. (2007), we are initially motivated to examine candidate choices for a diagonal $\boldsymbol{\Xi}$ that involve the covariance matrix $\boldsymbol{\Sigma}$. Consider the family of matrices $\boldsymbol{\Xi}(k) = \text{diag}\left(\frac{1}{\sigma_i^k}\right)$, where $k \in \mathbb{R}$. We consider the following choices for $\boldsymbol{\Xi}(k)$.

1. $\boldsymbol{\Xi}(0) = I$: The identity matrix may be appealing to decision-makers because it requires calibrating only one parameter, κ. A large value of κ corresponds to choosing a solution close to the equally weighted portfolio. As κ decreases more emphasis is put on the estimates $\hat{\boldsymbol{\mu}}$. This trade-off makes the tuning of the parameter κ fairly intuitive.
2. $\boldsymbol{\Xi}(-2) = \text{diag}(\sigma_i^2)$: This choice of $\boldsymbol{\Xi}$ appeals to a fundamental intuition: assets with a higher variance tend to have a poorer estimate of their expected returns.
3. $\boldsymbol{\Xi}(2) = \text{diag}\left(\frac{1}{\sigma_i^2}\right)$: This choice of $\boldsymbol{\Xi}$ incorporates the notion of a "risk premium". This approach is particularly relevant when investors believe in the Capital Asset Pricing Model (CAPM), suggesting that assets with a greater variance in their return should also have a greater expected return.

Selecting an appropriate value for κ is critical. Given the objective function $\hat{\boldsymbol{\mu}}^\top \mathbf{x} - \kappa\sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}}$ in (4), we may consider κ to be the weight imposed on the penalty term $\sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}}$, relative to the expected return term $\hat{\boldsymbol{\mu}}^\top \mathbf{x}$. Therefore, we design a heuristic to calibrate κ such that a target ratio

r between these two terms $\frac{\hat{\boldsymbol{\mu}}^\top \mathbf{x}}{\sqrt{\mathbf{x}^\top \boldsymbol{\Xi} \mathbf{x}}}$ is achieved. We present the heuristic as Algorithm 1 in Appendix B. We evaluate three potential choices for the target r, namely $2, 3, 4$.

Our experimental framework is based on real-world stocks in the S&P 500 index and the Fama-French industry portfolio data library. We assume that the returns follow a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ computed as follows. Using the observed historical asset returns r^i for $i = 1, \ldots, H$, we set $\boldsymbol{\mu} = \frac{1}{H} \sum_{t=1}^{H} r^i$, and $\boldsymbol{\Sigma}$ to be the covariance matrix of the returns. The estimated vector $\hat{\boldsymbol{\mu}}$ is then the sample average of N random samples generated from the distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Based on the results previously presented in Section 3, we let $N = 24$ for our data on monthly returns, and $N = 100$ for our data on daily returns. Our goal is to investigate the various choices for $\boldsymbol{\Xi}$ and κ mentioned above. We construct portfolios over four evenly-spaced risk thresholds v for each dataset, denoted by Low, Medium, High, and Very High.

We compute the Markowitz portfolio to be $\mathbf{x}^M := \arg\max_{\mathbf{x} \in \mathbb{R}^n_+} \{\hat{\boldsymbol{\mu}} \mathbf{x} : \mathbf{x}^\top \boldsymbol{\Sigma} \mathbf{x} \leq v,\ \mathbf{1}^\top \mathbf{x} = 1\}$, and obtain the robust portfolio $\mathbf{x}^R$ by solving the robust portfolio problem (4) for a given choice of $\boldsymbol{\Xi}$ and κ. The expected return of the Markowitz portfolio is given by $\boldsymbol{\mu}^\top \mathbf{x}^M$, and that of the robust portfolio by $\boldsymbol{\mu}^\top \mathbf{x}^R$. We present the results of the experiments as the percentage gap closed by the robust portfolio relative to the Markowitz portfolio. Specifically, we report the ratio $(\bar{R} - \bar{M})/(T - \bar{M})$, where $\bar{R}$ denotes the average out-of-sample expected return of the robust portfolios, and $\bar{M}$ denotes the average out-of-sample expected returns of the Markowitz portfolio. The quantity T denotes the true optimal expected return obtained by solving problem (1)–(3) using the true mean vector. The values for $\bar{R}$ and $\bar{M}$ are averages of $10,000$ simulated runs. Consequently, the standard error of each reported percentage gap closed is less than 0.1. We further consolidate the results by presenting the average percentage gap closed across all datasets at four different risk thresholds in Table 1.

The optimal target ratio r was chosen by conducting a grid search for each $\boldsymbol{\Xi}(k)$, with the returns for the optimal parameters presented. The optimal target ratio was found to be $r = 4$ for the three rows corresponding to monthly data. For daily data, the ratio was $r = 4$ for $\boldsymbol{\Xi}(-2)$, and $r = 2$ for $\boldsymbol{\Xi}(0)$ and $\boldsymbol{\Xi}(2)$.

The largest percentage gaps closed for each risk threshold are highlighted in Table 1. Our simulations demonstrate that significant gains can be achieved over the Markowitz model with an appropriately chosen estimation-error matrix and parameter κ. Notably, even the simple choice of the identity matrix, $\boldsymbol{\Xi}(0)$, yields portfolios that outperform the Markowitz portfolio for Low, Medium, and High risk levels. On the other hand, we find that $\boldsymbol{\Xi}(-2)$ is not a strong choice for the estimation-error matrix, suggesting that errors in estimating the expected returns cannot be corrected based directly on the variance of the returns. Interestingly, robust portfolios constructed using $\boldsymbol{\Xi}(2)$ consistently outperform the Markowitz portfolio across all risk thresholds in datasets with both monthly and daily returns, closing the largest gaps.

	Risk Threshold, v	Low	Medium	High	Very High
Monthly	Gap Closed (%) with $\Xi(-2)$	0.5	-1.7	-4.6	-7.1
	Gap Closed (%) with $\Xi(0)$	2.8	2.7	1.0	-1.0
	Gap Closed (%) with $\Xi(2)$	**3.5**	**5.2**	**4.9**	**3.6**
	Risk Threshold, v	Low	Medium	High	Very High
Daily	Gap Closed (%) with $\Xi(-2)$	6.4	3.1	-0.1	-1.0
	Gap Closed (%) with $\Xi(0)$	10.8	6.7	2.5	1.0
	Gap Closed (%) with $\Xi(2)$	**11.7**	**8.6**	**5.6**	**5.2**

Table 1 Average percentage gap closed by the robust portfolio compared to the Markowitz portfolio for different choices of estimation-error matrices: $\Xi(-2) = \mathbf{diag}(\sigma_i^2)$, $\Xi(0) = I$, and $\Xi(2) = \mathbf{diag}\left(\frac{1}{\sigma_i^2}\right)$. The percentages are averages across 5 datasets with varying number of assets.

Overall, for the best choices of Ξ and κ, the gap between the expected returns of the Markowitz portfolios and the optimal portfolios is reduced by an average of 6.1% across all datasets and risk thresholds in our experiments.

To complement the tabular results, we also include a scatter plot that compares the out-of-sample expected returns of the robust portfolios with those of four benchmark strategies. Each point in the plot corresponds to one of 10,000 simulated runs, with the x-axis representing the expected return of the robust portfolio and the y-axis representing that of a benchmark portfolio. Benchmarks include the Markowitz portfolio, the true optimal portfolio (computed with $\boldsymbol{\mu}$), the equal-weight portfolio, and the minimum-variance portfolio. The true optimal, equal-weight, and minimum-variance portfolios do not depend on the sampled estimates of $\hat{\boldsymbol{\mu}}$, so their expected returns remain constant across simulations and appear as horizontal lines in the plot. A point lying below the identity line $x = y$ indicates that the robust portfolio outperforms the corresponding benchmark in that run. Since the choice $\Xi(2)$ yielded the most consistent improvement across all datasets and risk thresholds, we restrict the scatter plot analysis to this case. The plot corresponds to the medium risk threshold. This visualization, presented in Figure 3, offers a more granular view of performance and highlights the empirical advantage of the robust strategy.

As expected, the scatter plot reveals that robust optimization offers the greatest benefit when the Markowitz portfolio performs poorly, for example, when the equal-weight portfolio outperforms the Markowitz portfolio. In scenarios where the Markowitz portfolio achieves high returns, typically when the estimated mean vector $\hat{\boldsymbol{\mu}}$ is close to the true $\boldsymbol{\mu}$, the robust portfolio often yields similar or slightly lower returns. This aligns with the role of the robust approach, which is designed to guard against adverse estimation errors. When the input estimates are reliable, the added conservatism of robustness can become less advantageous. Conversely, when $\hat{\boldsymbol{\mu}}$ is inaccurate, the robust portfolio is more likely to outperform. We also observe meaningful variability across simulations, further underscoring the value of robust optimization in mitigating downside risk due to estimation error.

Figure 3 **Scatter plot comparing the out-of-sample expected returns of the robust portfolio and four benchmark portfolios over 10,000 simulations. Points below the identity line $x = y$ indicate that the robust portfolio outperformed the benchmark.**

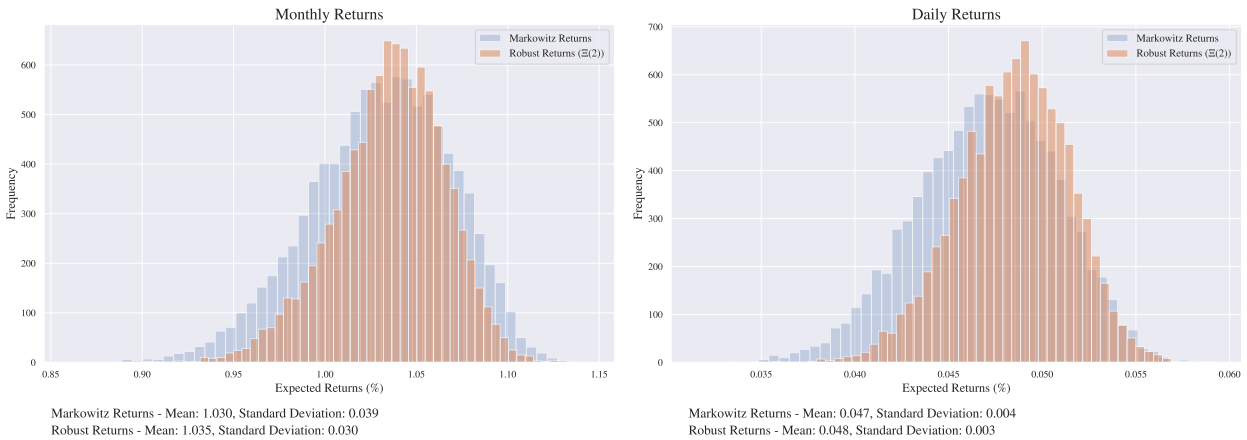

Figure 4 **Histogram of expected returns showing greater concentration for robust portfolios and wider variability for Markowitz portfolios across 10,000 simulations for $\Xi(2) = \mathbf{diag}\left(\frac{1}{\sigma_i^2}\right)$.**

To further explore the behavior of the robust and Markowitz portfolios, we present a histogram of their out-of-sample expected returns at the medium risk threshold across the 10,000 simulations in Figure 4. The distribution of robust portfolio returns is noticeably more concentrated around its mean, while the Markowitz portfolio exhibits a wider spread. While the Markowitz portfolio may occasionally outperform, its performance is more volatile and sensitive to estimation error.

Finally, we present the Sharpe ratios of the robust portfolios and benchmark strategies in Table 2. The risk-free rate was computed to be the average return of U.S. Treasury securities with 10-year

Risk Thresholds, v	Low	Medium	High	Very High
Robust Portfolio with $\boldsymbol{\Xi}(-2)$	0.155	0.146	0.142	0.140
Robust Portfolio with $\boldsymbol{\Xi}(0)$	0.155	0.144	0.139	0.135
Robust Portfolio with $\boldsymbol{\Xi}(2)$	0.155	0.143	0.136	0.132
Markowitz Portfolio	0.154	0.142	0.135	0.130
Optimal Portfolio	0.181	0.168	0.157	0.148
Equal Weight	0.146			
Minimum Variance	0.174			

Table 2 Sharpe ratios of robust and benchmark portfolios across four risk thresholds, averaged over five monthly-return datasets.

constant maturity over the same horizon as the dataset, yielding a monthly rate of 0.313% and a daily rate of 0.0063%. Among the robust and Markowitz portfolios, those constructed using $\boldsymbol{\Xi}(-2)$ yield the largest Sharpe ratios. This is consistent with the intuition that penalizing high-variance assets improves risk-adjusted performance. It is notable that the minimum-variance portfolio consistently outperforms all other methods with respect to the Sharpe ratio in the Medium to Very High risk regimes, including even the optimal portfolio, constructed using the true expected returns. We also note an interesting contrast between Tables 1 and 2 when comparing robust portfolios constructed with $\boldsymbol{\Xi}(-2)$, $\boldsymbol{\Xi}(0)$, and $\boldsymbol{\Xi}(2)$. Robust portfolios with $\boldsymbol{\Xi}(2)$ deliver higher expected returns than other choices (Table 1), while those with $\boldsymbol{\Xi}(-2)$ achieve higher Sharpe ratios (Table 2). This is explained by the difference in volatility. As an example, consider the 11-sector data set and a Medium risk threshold v. In our simulation, with 10,000 robust portfolios constructed for each of $\boldsymbol{\Xi}(-2)$, $\boldsymbol{\Xi}(0)$, and $\boldsymbol{\Xi}(2)$, the fraction that satisfied the risk constraint $\mathbf{x}^\top \boldsymbol{\Sigma} \mathbf{x} \leq v$ at equality was 58%, 72%, 82%, respectively; the average volatility was 4.36, 4.43, and 4.47, respectively (the Markowitz portfolio has an even higher average volatility of 4.5). Thus, $\boldsymbol{\Xi}(-2)$ is advantageous when risk-adjusted returns are prioritized. Importantly, the robust portfolios always outperform the Markowitz portfolio in terms of Sharpe ratio, underscoring the robustness benefits of our approach. Indeed, the Markowitz portfolios have lower returns and higher risk on average.

We also computed Sharpe ratios in the daily return setting. In this case, the robust model marginally outperforms the Markowitz portfolio. However, excluding the optimal portfolio, the equal-weight strategy consistently yields the highest Sharpe ratios overall, albeit by a small margin. We attribute this to the significantly higher standard deviation of the asset returns observed in the daily setting (see Tables 5–8 in Appendix C), which diminish the relative benefits of optimization-based strategies in favor of the diversification offered by equal weighting. These findings suggest that while our robust optimization framework improves expected returns and consistency, traditional benchmarks such as equal weighting remain competitive in high-volatility environments.

Our subsequent analysis explores the case of positive k, which aligns with the return-maximization perspective in Table 1. We conclude the section by validating our choice through additional experiments by incorporating a temporal component to the data. We utilize the same experimental setup as in Section 3 and described in Appendix A. We consider $k = 2, 4, 10$ for the estimation-error matrices $\mathbf{\Xi}(k)$. We let $N = 24$ for monthly returns and $N = 100$ for our daily returns, and consider an array of risk thresholds v for each dataset, with κ calibrated with the optimal target ratios found with a grid search. We present our results in Table 3 as the percentage gap closed by the robust portfolio compared to the Markowitz portfolio for each dataset, with a standard error less than 0.05 for each reported value.

Dataset	**Units**	**Number of Sectors**	Risk Threshold, v	Low	Medium	High	Very High
GICS	Monthly	11	Gap Closed (%) with $\mathbf{\Xi}(2)$	**4.1**	**3.6**	3.5	2.4
			Gap Closed (%) with $\mathbf{\Xi}(4)$	2.1	2.4	**3.7**	3.7
			Gap Closed (%) with $\mathbf{\Xi}(10)$	-2.1	-2.4	2.7	**6.3**
Fama-French	Monthly	5	Gap Closed (%) with $\mathbf{\Xi}(2)$	0.7	0.3	0.4	0.9
			Gap Closed (%) with $\mathbf{\Xi}(4)$	**1.9**	1.8	1.8	2.1
			Gap Closed (%) with $\mathbf{\Xi}(10)$	1.7	**2.1**	**2.4**	**3.8**
		10	Gap Closed (%) with $\mathbf{\Xi}(2)$	**4.0**	**5.8**	5.9	5.3
			Gap Closed (%) with $\mathbf{\Xi}(4)$	3.7	5.7	6.3	6.2
			Gap Closed (%) with $\mathbf{\Xi}(10)$	0.9	5.1	**9.9**	**11.1**
		12	Gap Closed (%) with $\mathbf{\Xi}(2)$	3.3	4.2	3.8	3.0
			Gap Closed (%) with $\mathbf{\Xi}(4)$	**3.9**	5.8	6.2	5.8
			Gap Closed (%) with $\mathbf{\Xi}(10)$	1.6	**6.3**	**10.2**	**10.7**
		17	Gap Closed (%) with $\mathbf{\Xi}(2)$	3.6	2.8	2.3	1.8
			Gap Closed (%) with $\mathbf{\Xi}(4)$	4.2	4.4	3.7	3.1
			Gap Closed (%) with $\mathbf{\Xi}(10)$	**4.5**	**5.0**	**7.7**	**6.5**
	Daily	5	Gap Closed (%) with $\mathbf{\Xi}(2)$	5.0	6.5	6.4	5.9
			Gap Closed (%) with $\mathbf{\Xi}(4)$	6.4	8.5	8.7	7.9
			Gap Closed (%) with $\mathbf{\Xi}(10)$	**6.7**	**10.0**	**11.9**	**12.7**
		10	Gap Closed (%) with $\mathbf{\Xi}(2)$	4.9	3.3	2.1	1.8
			Gap Closed (%) with $\mathbf{\Xi}(4)$	10.3	8.9	7.2	6.7
			Gap Closed (%) with $\mathbf{\Xi}(10)$	**12.7**	**14.6**	**14.3**	**13.6**
		12	Gap Closed (%) with $\mathbf{\Xi}(2)$	7.3	6.2	4.5	3.9
			Gap Closed (%) with $\mathbf{\Xi}(4)$	7.1	6.0	4.9	4.7
			Gap Closed (%) with $\mathbf{\Xi}(10)$	**9.6**	**11.5**	**12.2**	**12.2**
		17	Gap Closed (%) with $\mathbf{\Xi}(2)$	**3.4**	5.2	5.0	4.2
			Gap Closed (%) with $\mathbf{\Xi}(4)$	3.2	**6.2**	6.9	6.7
			Gap Closed (%) with $\mathbf{\Xi}(10)$	1.4	5.7	**9.8**	**11.5**

Table 3 Percentage gap closed by the robust portfolio with estimation-error matrices $\mathbf{\Xi}(2) = \mathbf{diag}\left(\frac{1}{\sigma_i^2}\right)$, $\mathbf{\Xi}(4) = \mathbf{diag}\left(\frac{1}{\sigma_i^4}\right)$, and $\mathbf{\Xi}(10) = \mathbf{diag}\left(\frac{1}{\sigma_i^{10}}\right)$ compared to the Markowitz portfolio, with temporal uncertainty in the data. Columns correspond to the portfolios constructed along various risk thresholds.

Our findings once again show that robust portfolio with the appropriate estimation-error matrix can improve upon the Markowitz portfolio. In particular, the choice of $\mathbf{\Xi}(4)$ results in robust portfolios that consistently outperform the Markowitz portfolios across all datasets on average for all risk thresholds, with $\mathbf{\Xi}(10)$ portfolios covering the largest gaps. For the best choices of $\mathbf{\Xi}$ and κ,

we observe that the gap between the expected returns of the Markowitz portfolios and the optimal portfolios is closed by an average of 8% across all datasets and risk thresholds in our experiments. We also note that portfolios with larger k tend to exhibit higher variance, with the majority of portfolios hitting the risk threshold v in our simulated runs. So their gains in expected return come with increased exposure to risk, as observed in Table 2. Overall, these experiments suggest that a robust portfolio approach can be extended and applied to settings where the historical returns are not identically distributed random variables.

5. Conclusion

In this paper, we offer a framework for constructing an ellipsoidal uncertainty set of the expected asset returns; this is needed in robust portfolio optimization. Our work addresses a gap in the literature by providing both theoretical and empirical insights into the selection of the estimation-error matrix $\mathbf{\Xi}$ and the weight κ assigned to the estimation risk. These choices are critical to the performance of robust portfolio optimization models. We prove an existential theorem showing that diagonal estimation-error matrices can yield robust portfolios with arbitrarily small loss in expected return compared to an optimum portfolio. Additionally, we challenge a conventional assumption often made for data-driven optimization methods that larger sample sizes from historical data invariably reduce estimation error. Instead, we demonstrate through empirical analysis that this is not always the case. Our empirical findings, based on synthetic data reflecting historical asset returns, support the practical usefulness of robust portfolio model as an improvement to the traditional Markowitz model. We find that the penalty term in the objective function can improve portfolio construction by reshaping how estimation risk interacts with asset volatility. Thus, the ellipsoid in the robust formulation provides a flexible mechanism that goes beyond modeling uncertainty and can be tuned to different investment priorities. Our work paves the way for further investigations in constructing estimation-error matrices for robust portfolio optimization, with implications for both academic research and practical financial decision-making.

Acknowledgements

This research was supported by the U.S. Office of Naval Research under award number N00014-22-1-2528. The authors would like to thank Matthias Köppe for his valuable contributions to an earlier version of this manuscript. We also thank Javier Peña, Karan Singh, and the referees for their helpful comments.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Avramov D, Zhou G (2010) Bayesian portfolio analysis. *Annual Review of Financial Economics* 2(1):25–47.

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

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

Bertsimas D, Gupta V, Kallus N (2018) Data-driven robust optimization. *Mathematical Programming* 167:235–292.

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

Black F, Litterman R (1990) Asset allocation: combining investor views with market equilibrium. *Goldman Sachs Fixed Income Research* 115(1):7–18.

Black F, Litterman R (1992) Global portfolio optimization. *Financial Analysts Journal* 48(5):28–43.

Brodie J, Daubechies I, De Mol C, Giannone D, Loris I (2009) Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences* 106(30):12267–12272.

Calafiore GC, Monastero B (2012) Data-driven asset allocation with guaranteed short-fall probability. *2012 American Control Conference (ACC)*, 3687–3692 (IEEE).

Chan LK, Karceski J, Lakonishok J (1999) On portfolio optimization: Forecasting covariances and choosing the risk model. *The Review of Financial Studies* 12(5):937–974.

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

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

Fabozzi FJ, Kolm PN, Pachamanova DA, Focardi SM (2007) *Robust Portfolio Optimization and Management* (John Wiley & Sons).

Frost PA, Savarino JE (1986) An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis* 21(3):293–305.

Garlappi L, Uppal R, Wang T (2007) Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The 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.

Gotoh Jy, Takeda A (2011) On the role of norm constraints in portfolio selection. *Computational Management Science* 8:323–353.

Gurobi Optimization L (2024) Gurobi optimizer reference manual. URL `https://www.gurobi.com`.

Jagannathan R, Ma T (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58(4):1651–1683.

James W, Stein C (1992) Estimation with quadratic loss. *Breakthroughs in statistics: Foundations and basic theory*, 443–460 (Springer).

Jobson J (1979) Improved estimation for markowitz portfolios using james-stein type estimators. *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, volume 71, 279–284.

Jorion P (1986) Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis* 21(3):279–292.

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.

Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10(5):603–621.

Ledoit O, Wolf M (2004a) Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management* 30(4).

Ledoit O, Wolf M (2004b) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2):365–411.

Ledoit O, Wolf M (2020) Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics* 48(5):3043–3065.

Lim AE, Shanthikumar JG, Vahn GY (2012) Robust portfolio choice with learning in the framework of regret: Single-period case. *Management Science* 58(9):1732–1746.

Markowitz H (1952) Portfolio selection. *The Journal of Finance* 7(1):77–91.

Merton RC (1980) On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8(4):323–361.

Michaud RO, Michaud RO (2008) *Efficient asset management: a practical guide to stock portfolio optimization and asset allocation* (Oxford University Press).

Natarajan K, Pachamanova D, Sim M (2008) Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science* 54(3):573–585.

Nelson DB (1992) Filtering and forecasting with misspecified arch models i: Getting the right variance with the wrong model. *Journal of Econometrics* 52(1-2):61–90.

Rockafellar RT (1970) *Convex analysis* (Princeton University Press).

Scherer B (2007) Can robust portfolio optimisation help to build better portfolios? *Journal of Asset Management* 7:374–387.

Stubbs R, Vance P (2005) Computing return estimation error matrices for robust optimization. *Axioma Research Papers* 1:1–9.

ter Horst J, De Roon F, Werker BJ (2006) Incorporating estimation risk in portfolio choice. *Advances in Corporate Finance and Asset Pricing. Elsevier, Amsterdam* 449–472.

Tütüncü RH, Koenig M (2004) Robust asset allocation. *Annals of Operations Research* 132:157–187.

Ulf H, Raimond M (2006) Portfolio choice and estimation risk. a comparison of bayesian to heuristic approaches. *ASTIN Bulletin: The Journal of the IAA* 36(1):135–160.

Appendix A: Detailed Experimental Setup

We conduct our experiments using simulated data modeled on stocks in the U.S. equity market. We use data on the historical returns of stocks in the S&P 500 index as consolidated by Kocuk and Cornuéjols (2020), classified into 11 sectors according to the Global Industrial Classification Standard (GICS). Table 4 gives the market-weighted averages and variances of their monthly returns over a 30-year period between January 1987 and December 2016. Additionally, we consider the 5-, 10-, 12-, and 17-sector datasets from the Fama-French data library. These datasets include monthly returns, for which we consider a 30-year period between March 1994 and February 2024, and daily returns, for which we consider a 10-year period between March 2014 and February 2024. The sectors are outlined in Tables 5-8 along with the market-weighted average and variance of their returns.

Observations of the financial market suggest that assumptions about structural stationarity are weak. Systemic changes to the market in recent history, such as the 1997 Asian financial crisis, the collapse of the dot-com bubble, the 2008 financial crisis, and the 2020 COVID-19 pandemic substantiate this claim. These events significantly impacted key industries across numerous sectors. Therefore, it seems more reasonable to consider multiple distributions that evolves over time rather than assuming one static distribution that describes the asset returns as a whole. Consequently, it is unlikely that the asset returns are i.i.d. random variables. Instead, we model the asset returns for each time period as random variables drawn from possibly different multivariate normal distributions.

A fundamental challenge is extracting the *true* expected returns that parameterize the distributions the real-world observed historical data are drawn from. To address this concern and to obtain statistically significant results, we use simulated data in our experiments. We assume that the returns are normally distributed over time with varying mean returns $\boldsymbol{\mu}^t$ and a fixed variance $\boldsymbol{\Sigma}$ across all time periods, Normal$(\boldsymbol{\mu}^t, \boldsymbol{\Sigma})$. Consider planning horizon for $t = 1, \ldots, H$. We compute the true expected returns as $\boldsymbol{\mu}^t = \frac{1}{T} \sum_{i=t-\frac{T}{2}+1}^{t+\frac{T}{2}} r^i$ for $t = \frac{T}{2}, \ldots, H - \frac{T}{2}$, where r^i is the realization of returns observed in the i^{th} time unit in the data and T is a prescribed parameter. We chose $T = 30$ and $T = 260$ for our experiments on the monthly and daily datasets, respectively. The parameter T was tuned in accordance with systemic changes in the market observed in the data. We let the true covariance matrix $\boldsymbol{\Sigma}$ be the covariance of the asset returns in the stock market data. We then sample $\hat{r}^t \sim \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma})$ to simulate additional streams of "observed" returns, and compute the

estimated expected return vectors as $\hat{\boldsymbol{\mu}}^t = \frac{1}{N}\sum_{i=t-N+1}^{t} \hat{r}^t$ for $t = \frac{T}{2} + N - 1, \ldots, H - \frac{T}{2} - T$, where N is a chosen parameter for the sample size of historical time-units to consider for the estimation. We simulated 50 runs for experiments on monthly returns where $H = 360$, and 10 runs for experiments on daily returns where $H = 2600$.

We define the optimal portfolios as $\mathbf{x}^{t*} := \max_{\mathbf{x}\in\mathcal{X}} \boldsymbol{\mu}^{t\top}\mathbf{x}$ for each $t = \frac{T}{2}, \ldots, H - \frac{T}{2}$. That is, for each time period, we find the portfolio that maximizes the true expected return $\boldsymbol{\mu}^t$. Similarly, we define the estimated Markowitz portfolios as $\hat{\mathbf{x}}^t := \max_{\mathbf{x}\in\mathcal{X}} \hat{\boldsymbol{\mu}}^{t\top}\mathbf{x}$ for each $t = \frac{T}{2} + N - 1, \ldots, H - \frac{T}{2} - T$. We utilize *efficient frontiers* to quantify the performance of the portfolios. An efficient frontier plots the maximum expected return of a portfolio of assets as a function of the risk thresholds (Markowitz 1952). The true frontier is computed using the true expected returns of the assets and the optimal portfolios. The estimated frontier is computed by using the estimated expected returns and the estimated Markowitz portfolios, which describes the expected return of the Markowitz portfolios should the estimated parameters be realized. The actual frontier plots the expected return one actually observes on the true expected returns when one invests in the portfolios constructed with the estimated expected returns. That is, for each choice of N, we compute the *true expected return* $\boldsymbol{\mu}^{t+T\top}\mathbf{x}^{t*}$, the *estimated expected return* $\hat{\boldsymbol{\mu}}^{t\top}\hat{\mathbf{x}}^t$, and the *actual expected return* $\boldsymbol{\mu}^{t+T\top}\hat{\mathbf{x}}^t$, all averaged over the simulated runs and over the time periods t. We evaluate the portfolios at time $t+T$ to negate any sampling bias in our set up. We illustrate our experimental setup in Figure 5.

All experiments were run on an Apple M3 processor with 8GB of memory, running Python version 3.12.4 and Gurobi Optimizer version 11.0.1 (Gurobi Optimization 2024).

Appendix B: A Heuristic to Calibrate κ

Algorithm 1 Heuristic to calibrate κ

Input: Estimates of the expected returns $\hat{\boldsymbol{\mu}} \in \mathbb{R}^n$, a diagonal positive definite estimation-error matrix $\boldsymbol{\Xi} \in \mathbb{R}^{n\times n}$, a lower bound $l \in \mathbb{R}$ and an upper bound $u \in \mathbb{R}$ for the ratio $\frac{\hat{\boldsymbol{\mu}}^\top\mathbf{x}}{\kappa\sqrt{\mathbf{x}^\top\boldsymbol{\Xi}\mathbf{x}}}$.

Output: $\kappa \in \mathbb{R}$.

Initialize $\bar{\mathbf{x}} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$, $\bar{\bar{\mathbf{x}}} = \left(\frac{\frac{1}{\xi_i}}{\sum_{i=1}^n \frac{1}{\xi_i}}, \ldots, \frac{\frac{1}{\xi_i}}{\sum_{i=1}^n \frac{1}{\xi_i}}\right)$, $r = \frac{u-l}{2}$.
Compute $\hat{\boldsymbol{\mu}}\bar{\mathbf{x}}$, $\sqrt{\bar{\bar{\mathbf{x}}}^\top\boldsymbol{\Xi}\bar{\bar{\mathbf{x}}}}$.
Choose $\kappa = \frac{\hat{\boldsymbol{\mu}}\bar{\mathbf{x}}}{r\sqrt{\bar{\bar{\mathbf{x}}}^\top\boldsymbol{\Xi}\bar{\bar{\mathbf{x}}}}}$.
while Stopping condition not met **do**
  Solve for $\mathbf{x}^R$ the solution to (4).
  if $l \le \frac{\hat{\boldsymbol{\mu}}\mathbf{x}^R}{\kappa\sqrt{\mathbf{x}^{R\top}\boldsymbol{\Xi}\mathbf{x}^R}} \le u$ **then**
    return κ.
  else
    $\kappa \leftarrow \frac{\hat{\boldsymbol{\mu}}\mathbf{x}^R}{r\sqrt{\mathbf{x}^{R\top}\boldsymbol{\Xi}\mathbf{x}^R}}$.
  end if
end while

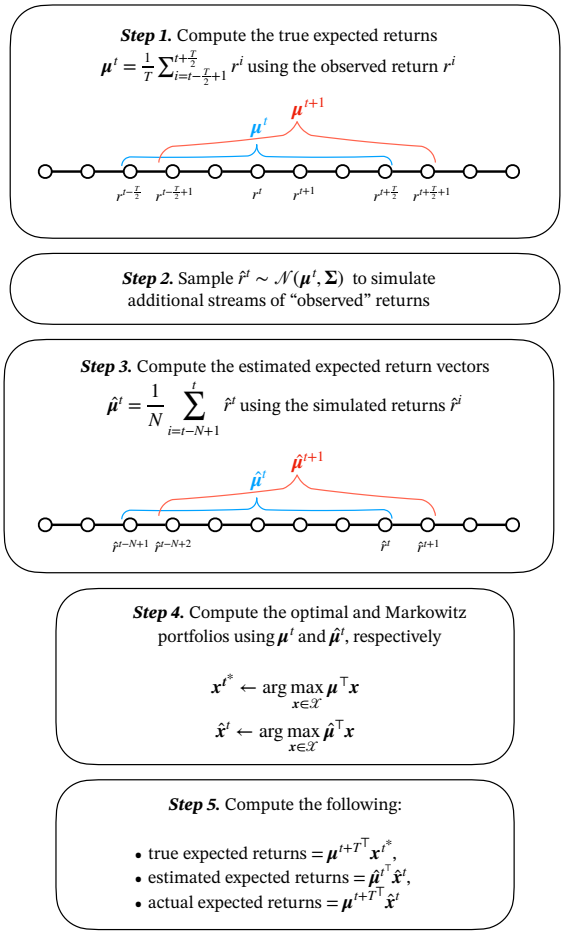

Figure 5 **A schematic summarizing our experimental setup**

Algorithm 1 first initializes the portfolios $\bar{\mathbf{x}}$ and $\bar{\bar{\mathbf{x}}}$. Here, $\bar{\mathbf{x}}$ is the equal-weight portfolio, $\bar{\bar{\mathbf{x}}}$ is a portfolio that is normalized relative to the inverse of ξ_i across all assets. The heuristic also initializes the target r for the ratio $\frac{\hat{\boldsymbol{\mu}}\bar{\mathbf{x}}}{\sqrt{\bar{\bar{\mathbf{x}}}^\top \boldsymbol{\Xi} \bar{\bar{\mathbf{x}}}}}$ to the middle point in a target range $[l, u]$. It then chooses the corresponding κ and solves for the associated robust portfolio $\mathbf{x}^R$. The heuristic returns κ if the ratio $\frac{\hat{\boldsymbol{\mu}}\mathbf{x}^R}{\kappa\sqrt{\mathbf{x}^{R\top} \boldsymbol{\Xi} \mathbf{x}^R}}$ falls in the target range $[l, u]$. If not, κ is re-calibrated and the ratio is checked again. We evaluate three potential choices for the range $[l, u]$, namely $[1, 3]$, $[2, 4]$, $[3, 5]$, conducting a grid search to determine the optimal values.

Appendix C: Data Summary

	Energy	Consumer Discre-tionary	Consumer Staples	Real Estate	Industrials	Financials	Telecomm--unications Services	Information Technology	Materials	Health Care	Utilities
μ	1.18	1.51	1.39	1.15	1.29	1.33	1.03	1.73	1.39	1.42	1.01
σ^2	39.5	28.3	17.2	52.5	26.5	39.5	30.0	50.5	32.4	21.6	18.3

Table 4 Sample averages and variances of historical returns of the GICS 11-sector dataset.

		Cnsmr	Manuf	HiTec	Hlth	Other
Daily	μ	0.0453	0.0457	0.0553	0.0500	0.0449
	σ^2	1.11	1.40	2.21	1.31	1.97
Monthly	μ	0.939	0.907	1.12	1.03	0.876
	σ^2	17.7	20.5	38.6	18.3	28.1

Table 5 Sample averages and variances of historical returns of the Fama-French 5-sector dataset.

		NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
Daily	μ	0.0410	0.0480	0.0494	0.0516	0.0630	0.0316	0.0489	0.0500	0.0384	0.0449
	σ^2	0.885	3.17	1.45	2.83	2.69	1.63	1.42	1.31	1.23	1.97
Monthly	μ	0.887	1.01	1.02	1.04	1.29	0.646	1.02	1.05	0.813	0.900
	σ^2	14.1	75.2	25.0	46.2	48.3	27.4	21.6	18.1	17.3	28.1

Table 6 Sample averages and variances of historical returns of the Fama-French 10-sector dataset.

		NoDur	Durbl	Manuf	Enrgy	Chems	BusEq	Telcm	Utils	Shops	Hlth	Money	Other
Daily	μ	0.0410	0.0480	0.0527	0.0516	0.0433	0.0632	0.0316	0.0384	0.0489	0.0500	0.0493	0.0364
	σ^2	0.885	3.17	1.75	2.83	1.25	2.70	1.63	1.23	1.42	1.31	2.45	1.59
Monthly	μ	0.887	1.01	1.09	1.04	0.877	1.29	0.646	0.813	1.02	1.05	0.969	0.739
	σ^2	14.1	75.2	31.6	46.2	18.5	48.4	27.4	17.3	21.6	18.1	32.5	25.8

Table 7 Sample averages and variances of historical returns of the Fama-French 12-sector dataset.

		Food	Mines	Oil	Clths	Durbl	Chems	Cnsum	Cnstr	Steel	FabPr	Machn	Cars	Trans	Utils	Rtail	Finan	Other
Daily	μ	0.0405	0.0534	0.0507	0.0419	0.0306	0.0445	0.0504	0.0574	0.0473	0.0504	0.0638	0.0528	0.0500	0.0384	0.0492	0.0493	0.0453
	σ^2	0.951	3.79	2.83	2.20	1.83	2.21	1.13	2.24	4.27	1.75	2.84	2.94	1.66	1.23	1.53	2.46	1.57
Monthly	μ	0.880	1.01	1.03	0.897	0.633	0.874	1.08	1.19	0.959	1.05	1.31	1.08	1.01	0.813	1.03	0.969	0.935
	σ^2	14.8	67.5	46.8	39.6	37.2	36.7	15.8	37.2	83.7	32.8	52.3	65.2	29.6	17.3	23.3	32.5	25.6

Table 8 Sample averages and variances of historical returns of the Fama-French 17-sector dataset.