Uncertainty Set Sizes, Sensitivity Analysis, in Robust Portfolio Optimization

by

Mingyu Yang

A research paper
presented to the University of Waterloo
in fulfillment of the
research paper requirement for the degree of
Master of Mathematics
in
Computational Mathematics

Waterloo, Ontario, Canada, 2019

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This research paper looks at portfolio optimization under uncertainty. One approach to dealing with uncertainty when distributions are unknown is applying robust optimization with uncertainty sets. The choice for uncertainty sets is often made to guarantee tractability of the robust counterpart. We consider the question of how to *better* determine the sizes of the uncertainty sets using sensitivity analysis as a guide.

We provide the first order derivative formulae for the perturbations with respect to the original parameters as well as with respect to the uncertainty set sizes. This results in a robustness measure. We use the L-shape in the curve of the robustness measure versus the size parameters under the global minimum CVaR problem. The L-curve is applied to the determination of size parameters. Therefore, the uncertainty set sizes are decided by both the data and the problems.

Acknowledgements

I would like to express my heartfelt thanks to my supervisors, Professor Henry Wolkowicz, for his guidance, support and encouragement throughout the program, and to my second reader Professor Yuying Li for taking her time to read my paper and providing insightful suggestions.

I would also like to thank all the friends and staffs in Computational Mathematics for all their help and the wonderful experience.

Table of Contents

List of Tables

List of Figures

Chapter 1

Background

1.1 Introduction

In this research report we study robust portfolio optimization with emphasis on the accompanying uncertainty sets. We introduce sensitivity analysis results to help the selection of uncertainty set sizes.

We look at portfolio optimization under uncertainty, including parameter uncertainty and distribution ambiguity. One approach to dealing with uncertainty, when the type of distribution as well as parameters are unknown, is applying robust optimization with uncertainty sets. The choice for uncertainty sets is often made to guarantee tractability of the robust counterpart. We consider the question of how to *better* determine the sizes of the uncertainty sets using sensitivity analysis as a guide.

We provide the first order derivative formulae for the perturbations with respect to the original parameters as well as with respect to the uncertainty set sizes. This results in a robustness measure. We use the maximum curvature point in the resulting L-shape in the curve of the robustness measure versus the size parameters under the global minimum CVaR problem. This so-called *L-curve* is applied to the determination of size parameters. Therefore, the uncertainty set sizes are decided by both the data and the properties of the problems.

1.1.1 Modern Portfolio Theory

Modern Portfolio Theory (MPT), put forward by Harry Markowitz in 1952 [12], is a method for asset allocation with the assumption that investors are risk-averse. Roughly speaking, it constructs a portfolio by minimizing risk while meeting a minimum on the investor's expected return. Variance or standard deviation is used as the risk measure.

We consider a portfolio with n assets and let the random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ be the unknown rates of return in the given time period. Let

$$\boldsymbol{\mu} = E(\mathbf{X}) = [E(X_1), E(X_2), \ldots, E(X_n)]^T$$

be the expected return vector. And let $\mathbf{w} = (w_1, w_2, \ldots, w_n)^T$ be the weighting component of each asset satisfying the budget constraint

$$\mathbf{1}_n^T \mathbf{w} = \sum_{i=1}^n w_i = 1,$$

where $\mathbf{1}_n$ is a n-dimensional vector of ones. Then $\boldsymbol{\mu}^T \mathbf{w}$ is the portfolio expected return.

Further, we let $\Sigma = \text{Cov}(\mathbf{X}) = (\text{Cov}(X_i, X_j)) \in \mathcal{S}_+^n$ be the *covariance matrix* of the returns on the assets in the portfolio so that $\mathbf{w}^T \Sigma \mathbf{w}$ is the variance of portfolio return. Here $\mathcal{S}^n$ denotes $n \times n$ symmetric matrices, and $\mathcal{S}_+^n$ is the cone of positive semidefinite matrices. Since the variance should always be nonnegative and $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i), \forall i, j$, the covariance matrix Σ is symmetric positive semidefinite by nature. It is positive definite unless one asset return is an exact linear function of the others. In asset allocation, the potential assets are generically not perfectly correlated. Therefore, we can assume $\Sigma \in \mathcal{S}_{++}^n$, without loss of generality.

The classic *Markowitz model* can be expressed as

$$\begin{aligned}
\min_{\mathbf{w} \in \mathbb{R}^n} \quad & \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} \quad && \text{(risk)} \\
\text{s.t.} \quad & \boldsymbol{\mu}^T \mathbf{w} = \mu_0 \quad && \text{(expected return)} \\
& \mathbf{1}_n^T \mathbf{w} = 1 \quad && \text{(budget)}
\end{aligned} \tag{1.1}$$

Remark 1.1.1. *Some researchers consider the expected return as an inequality constraint, i.e., $\boldsymbol{\mu}^T \mathbf{w} \geq \mu_0$. However, this inequality constraint is active (holds with equality) when μ_0 is in a specific set $[\mu_{\min}, \mu_{\max}]$. Here $\mu_{\min}$ is the expected return when the portfolio is at the global minimum risk; while $\mu_{\max}$ is the global maximum expected return. And for all $\mu_0 > \mu_{\max}$, the problem is infeasible.*

Other constraints, such as: *no-short-selling, diversification, box constraints, transaction costs, turnover*, can be added to the model when needed.

The Markowitz problem can be solved explicitly using Lagrange multipliers. The solution satisfies the following linear system:

$$\begin{bmatrix} \Sigma & -\boldsymbol{\mu} & -\mathbf{1}_n \\ \boldsymbol{\mu}^T & 0 & 0 \\ \mathbf{1}_n^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \lambda_1 \\ \sigma \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mu_0 \\ 1 \end{bmatrix}, \tag{1.2}$$

where λ_1 and σ are two Lagrangian multipliers. Moreover, this problem is equivalent to:

$$\min_{\mathbf{w} \in \mathbb{R}^n} \quad \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} - \lambda_2 \boldsymbol{\mu}^T \mathbf{w} \quad \text{(expected return adjusted risk)}$$
$$\text{s.t.} \quad \mathbf{1}_n^T \mathbf{w} = 1 \qquad \qquad \text{(budget)}$$

if we set $\lambda_2 = \lambda_1$, the optimal Lagrange multiplier from (1.2).

There is some criticism directed at the Markowitz theory. First, it is impossible to find the *true* mean and variance embedded behind each asset. A similar statement holds for the *true* correlations between any two assets. Therefore, the two important parameters: the *mean* and the *covariance matrix*, both have to be estimated by historical data or by personal judgment. This means that biases inevitably creep in. We note that [3] shows that the weight of an optimal mean-variance model is extremely sensitive to changes in the mean of assets' return; while [13] argues that the variance and covariance matrix can be estimated more accurately.

Second, variance is used as a risk measure in the model. It is a symmetric measure of the deviation of the portfolio return from its mean, treating the upside risk (gains) and downside risk (losses) in the same way. In practice, a risk-averse investor will be more concerned with the downside risk, than with the upside risk. If the portfolio return satisfies a Gaussian distribution, or any symmetric distribution, variance can be used as a downside risk measure because the downside risk is the same as the upside risk. However, much research and empirics show that the return should follow a *fat-tailed distribution* (with positive skewness). Variance, therefore, may not be a good risk measure in the model.

1.1.2 Coherent Risk Measures

We let $\Omega = \mathbb{R}^n$ be the finite dimensional set of states, $\mathcal{G}$ be the set of all risks (set of all real valued functions on $\mathbb{R}^n$), and we let random outcomes X_1 and X_2 be any elements

in $\mathcal{G}$. Then we set r to be the risk free return, if, in any state of nature, there exists an (risk-free) asset that has an initial value of 1 and a final price of r, in a fixed time period.

Definition 1.1.2 (Coherent measure [2]). *A risk measure, $\rho\colon \mathcal{G} \to \mathbb{R}$ is defined to be coherent if it satisfies the following properties:*

1. $\rho(X_1 + \alpha r) = \rho(X_1) - \alpha, \forall \alpha \in \mathbb{R}$ *(translation invariance);*

2. $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$ *(subadditivity);*

3. $\rho(\lambda X_1) = \lambda \rho(X_1), \forall \lambda \geq 0$ *(positive homogeneity);*

4. $\rho(X_1) \leq \rho(X_2)$ *if $X_1 \leq X_2$ (monotonicity).*

The coherent measures of risk are consistent with investors' intuition about risk. Besides, a subadditive and positive homogeneous risk measure is convex, which is an attractive property in optimization problems. Therefore, these assumptions have been well accepted since they first appeared in 1999.

In addition to the critical inherent symmetrical distribution assumption of variance as a risk measure, it is not a coherent risk measure because translation invariance and monotonicity properties cannot be satisfied:

- translation invariance: $\mathrm{Var}(X + \alpha r) = \mathrm{Var}(X) \neq \mathrm{Var}(X) - r, \forall \alpha \in \mathbb{R}$;

- monotonicity: by counterexample, let $X_1 = U(0,2)$ and $X_2 = U(2,3)$ be two random variables with uniform distributions. It is clear that $X_1 \leq X_2$. However, $\mathrm{Var}(X_1) > \mathrm{Var}(X_2)$ as $\mathrm{Var}(X_1) = \frac{1}{3}$ and $\mathrm{Var}(X_2) = \frac{1}{12}$.

1.1.3 Popular Risk Measures

To improve the model with regard to the second disadvantages of Markowitz model mentioned above, there are many risk measures that can be used as alternatives to the variance.

Some of the popular measure of the risk of loss are:

1. **Value-at-Risk** (VaR)

    VaR is a common quantile-based risk measure defined as:

    $$\mathrm{VaR}_{\beta,\mathrm{F}}(\mathbf{w}) := \min\left\{\alpha \in \mathbb{R} : \int_{l(\mathbf{X},\mathbf{w}) \leq \alpha} dF(\mathbf{X}) \geq \beta\right\}$$

for a given confidence level β (usually 99% or 95%) and a known cumulative distribution function $F(\mathbf{X})$ for $\mathbf{X}$. $l(\mathbf{X}, \mathbf{w})$ is the *loss function* with a decision vector $\mathbf{w}$ and a random vector $\mathbf{X}$. If we let $\mathbf{w}$ denote the weight of each asset and $\mathbf{X}$ be the random vector for the rates of return, $l(\mathbf{X}, \mathbf{w}) = -\mathbf{X}^T \mathbf{w}$ is the function of rate of loss for a given period.

VaR has been well accepted in financial risk management and reporting by investors and regulators to see how many assets are in need to cover the loss under most of possible future market conditions. However, it does not satisfy subadditivity and therefore not a coherent risk measure [2].

2. **Conditional Value-at-Risk** (CVaR)

CVaR, also called expected shortfall, is closely related to the definition of VaR. It is defined as the expected loss which exceeds VaR with a given confidence level β:

$$\text{CVaR}_{\beta,\text{F}}(\mathbf{w}) := \frac{1}{1-\beta} \int_{-\mathbf{X}^T\mathbf{w} \geq \text{VaR}_{\beta,\text{F}}(\mathbf{w})} -\mathbf{X}^T \mathbf{w} \, dF(\mathbf{X}).$$

CVaR is a coherent risk measure that has very appealing computational properties. Let

$$G_{\beta,F}(\mathbf{w}, \lambda) := \lambda + \frac{1}{1-\beta} \int_{\mathbf{X} \in \mathbb{R}^n} [-\mathbf{X}^T \mathbf{w} - \lambda]^+ dF(\mathbf{X}), \ [t]^+ := \max(t, 0). \qquad (1.3)$$

Rockafellar and Uryasev [15] propose that one can simultaneously compute VaR and CVaR by solving the following optimization problem in Theorem 1.1.3, (1.4).

Theorem 1.1.3 (From [15, Theorem 1]). *As a function of λ, $G_\beta(\mathbf{w}, \lambda)$ is convex and continuously differentiable. The* $\text{CVaR}_{\beta,\text{F}}$ *of the loss associated with any* $\mathbf{w} \in \mathcal{X}$ *can be determined from the formula* [1]

$$\text{CVaR}_{\beta,\text{F}}(\mathbf{w}) = \min_{\lambda \in \mathbb{R}} G_{\beta,F}(\mathbf{w}, \lambda) \qquad (1.4)$$

and

$$\text{VaR}_{\beta,\text{F}}(\mathbf{w}) \in \underset{\lambda \in \mathbb{R}}{\text{argmin}} \, G_{\beta,F}(\mathbf{w}, \lambda).$$

[1] $\mathcal{X}$, *the feasible set of* $\mathbf{w}$, either includes or excludes the target expected return constraint, depending on the context.

Theorem 1.1.4 (From [15, Theorem 2]). *The following two problems are equivalent in the sense that their optimal values are equal:*

$$p^* = \min_{\mathbf{w} \in \mathcal{X}} \text{CVaR}_{\beta,\mathbf{F}}(\mathbf{w})$$
$$= \min_{(\mathbf{w},\lambda) \in \mathcal{X} \times \mathbb{R}} G_\beta(\mathbf{w}, \lambda),$$

where the pair $(\mathbf{w}^*, \lambda^*)$ *attains the minimum for the second problem if, and only if,* $\mathbf{w}^*$ *attains the minimum for the first problem, and* $\lambda^* \in \text{argmin}_{\lambda \in \mathbb{R}} G_\beta(\mathbf{w}^*, \lambda)$. *Moreover,* $G_\beta(\mathbf{w}, \lambda)$ *is convex with respect to* $(\mathbf{w}, \lambda)$.

3. **Other Popular Risk Measures**

   - Entropic value at risk:

   $$\text{EVaR}(X) = \min_{t>0} t \ln M_X(\frac{1}{t}) - t \ln \alpha$$

   where $M_X(\cdot)$ is the moment-generating function. It is a coherent risk measure introduced in [1] in 2012.

   There are also many risk measures using the concept of lower partial moment:

   - Semi-deviation
   $$\text{SD} = \sqrt{E[([E(X) - X]^+)^2]}.$$

   This is equivalent to standard deviation for symmetric distributions so that it is not a coherent risk measure.

   - Target semi-deviation

   $$\text{TSD}(\tau) = \sqrt{E[([\tau - X]^+)^2]}.$$

   - Omega ratio
   $$\Omega(\tau) = \frac{E(X) - \tau}{E(\tau - X)^+} + 1,$$

   where τ is a threshold of desirable gain. This is not a coherent risk measure as it violates positive homogeneity. For counterexample, let $X = U(0,1)$ be a random variable with uniform distribution. Then $\Omega(2X) \neq 2\Omega(X)$ as $\Omega(X) = 0$ and $\Omega(2X) = 1 \neq 0$ when $\tau = 1$.

6

1.2 Robust Portfolio Optimization

In many optimization problems, the parameters, or the inputs of the models are unknown or are estimated inaccurately when the problems need to be solved. It can be a serious concern if the outputs (optimal values and objective function value) of the model are sensitive to the inputs. When it comes to portfolio optimization, as it has been discussed in the disadvantages of MPT, the measures of risk and return are usually calculated using historical data and are deemed to be wrong, which can lead to highly unreliable portfolios. Cornuéjols, Peña, Tütüncü, in [5], define robust optimization as the modelling of optimization problems with data uncertainty to obtain a solution that is guaranteed to be "good" for all or most possible realizations of the uncertain parameters. The possible realizations is described through *uncertainty sets,* $\mathcal{U}$.

According to the definition above, the robust optimization can be divided into two types: *absolute robust optimization* ("good" for all possible realizations) and *relative robust optimization* ("good" for most possible realizations). Absolute robust optimization can be formulated as:

$$
\begin{aligned}
\min_{x \in \mathcal{K}} \quad & f(x, p) \\
\text{s.t.} \quad & g(x, p) \geq B, \forall p \in \mathcal{U},
\end{aligned}
$$

where p is the uncertain parameters and $\mathcal{K}$ is a feasible set of x.

It can be reformulated as a min-max optimization:

$$
\begin{aligned}
\min_{x \in \mathcal{K}} \ \max_{p \in \mathcal{U}} \quad & f(x, p) \\
\text{s.t.} \quad \min_{p \in \mathcal{U}} \quad & g(x, p) \geq B.
\end{aligned}
$$

Relative robust optimization can be formulated as:

$$
\begin{aligned}
\min_{x \in \mathcal{K}} \ \max_{p \in \mathcal{U}} \quad & f(x, p) - f(x^*(p), p) \\
\text{s.t.} \quad \min_{p \in \mathcal{U}} \quad & g(x, p) \geq B,
\end{aligned}
$$

where $x^*(p) = \operatorname*{argmin}_{x \in \mathcal{K}} f(x, p)$.

Comparing to the fact that absolute robustness only cares for the optimization in the worst case, relative robustness measures the worst case relative to the best case under each scenario in the uncertainty set and is less conservative than the absolute robustness. It is more consistent with investment targets of many decision makers as the worst case might be very unlikely to happen. However, it introduces more computational complexity and can lose some good properties that the absolute robustness preserves.

The size and shape of uncertainty sets decide the desired robustness of the models. A consistent approach to choose the uncertainty sets does not exist. Some common types of uncertainty sets include finite scenario sets, convex hull sets, box uncertainty sets, ellipsoidal sets, etc. By implementing different types of uncertainty sets, the problems will be transformed into linear programs, semi-definite programs, second-order programs, conic programs, etc.

To determine the size of uncertainty set, Goldfarb and Iyengar proposed the confidence regions of a robust multivariate factor model as ellipsoidal uncertainty sets in [8]. Tütüncü and Koenig [16] use bootstrap method as well as moving averages to generate a box uncertainty set. The bigger the uncertainty set is, the more conservative the model will be.

1.2.1   Robust Mean-CVaR Portfolio Optimization

Recall the definition of $G_{\beta,F}$ in (1.3). The (non-robust) mean-CVaR problem (MC-general) can be formulated as:

$$\text{CVaR}_{\beta,\text{F}}^{*} = \min_{(\mathbf{w},\lambda)\in\mathcal{X}\times\mathbb{R}} \quad G_{\beta,F}(\mathbf{w},\lambda) \tag{MC-general}$$
$$\text{s.t.} \quad \boldsymbol{\mu}^{T}\mathbf{w} \geq \mu_0.$$

For most portfolio selection problems, decision makers only know the realized data, $\mathbf{X}_{[i]}, i = 1,\ldots,T$, rather than $\mathbf{X}$ itself. Therefore, the integral part of $G_{\beta,F}(\mathbf{w},\lambda)$ can be replaced by sample estimation:

$$\tilde{G}_{\beta}(\mathbf{w},\lambda) := \lambda + \frac{1}{1-\beta}\sum_{i=1}^{T} p_{[i]}[\mathbf{X}_{[i]}^{T}\mathbf{w} - \lambda]^{+},$$

where $p_{[i]}$ is the weight for $\mathbf{X}_{[i]}$ and $\sum_{i=1}^{T} p_{[i]} = 1$. Some investors will give greater weight/importance to more recent data. Usually, without other information, we set

$$p_{[i]} = \frac{1}{T}. \tag{1.5}$$

For time series, a common choice of $p_{[i]}$ is the *weighted moving averages*,

$$p_{[i]} = \frac{i}{\sum_{i=1}^{T} T} = \frac{2i}{T(T+1)}. \tag{1.6}$$

Similarly for $\boldsymbol{\mu}$, one can use the sample mean $\hat{\boldsymbol{\mu}} := \sum_{i=1}^{T} p_{[i]} \mathbf{X}_{[i]}.$[2] Note that $\tilde{G}_\beta(\mathbf{w}, \lambda)$ is convex and piecewise linear, with respect to λ, and is convex as a function of $\mathbf{w}$.

After a substitution using sample statistics, the *sample-based mean-CVaR problem* (SMC) is:

$$\begin{aligned} \min_{(\mathbf{w}, \lambda) \in \mathcal{X} \times \mathbb{R}} \quad & \tilde{G}_\beta(\mathbf{w}, \lambda) \\ \text{s.t.} \quad & \hat{\boldsymbol{\mu}}^T \mathbf{w} \geq \mu_0. \end{aligned} \qquad \text{(SMC)}$$

Although the mean-CVaR optimization model has gained popularity, Lim et al. [11] show that portfolios obtained by solving mean-CVaR and global minimum CVaR problems are unreliable. This is due to estimation errors of CVaR and/or the mean. Moreover these errors are magnified by optimization. Therefore, we need to introduce more robustness into the models.

Based on the discussion in the previous sections about CVaR and robust portfolio optimization, the absolute robust optimization model of mean-CVaR portfolio (RMC-general) selection problem is:

$$\begin{aligned} \text{WCVaR}_\beta^* = \min_{\mathbf{w} \in \mathcal{X}} \quad & \max_{F(\cdot) \in \mathcal{U}} \quad \text{CVaR}_{\beta, \mathrm{F}}(\mathbf{w}) \\ \text{s.t.} \quad & \min_{F(\cdot) \in \mathcal{U}} \boldsymbol{\mu}^T \mathbf{w} \geq \mu_0, \end{aligned} \qquad \text{(RMC-general)}$$

where $F(\cdot)$ is the cumulative distribution function of $\mathbf{X}$ and $\mathcal{U}$ is the uncertainty set of the possible distributions that $\mathbf{X}$ might follow.

In the situation that $F(\cdot)$ is ambiguous and characterized as a set $\mathcal{U}$, Zhu and Fukushima [18] define *worst case risk measures*,

$$\rho_w(X) = \max_{F(\cdot) \in \mathcal{U}} \rho(X),$$

and worst case CVaR as

$$\text{WCVaR}_\beta(\mathbf{w}) = \max_{F(\cdot) \in \mathcal{U}} \text{CVaR}_{\beta, \mathrm{F}}(\mathbf{w}).$$

They then prove that $\text{WCVaR}_\beta(\mathbf{w})$ is a coherent risk measure.

---

[2]Similar to sample weighted covariance matrix $\hat{\Sigma} := \sum_{i=1}^{T} p_{[i]} (\mathbf{X}_{[i]} - \boldsymbol{\mu})(\mathbf{X}_{[i]} - \boldsymbol{\mu})^T.$

Proposition 1.2.1 (From [18, Proposition 1]). *If ρ associated with determinate probability measure F is a coherent risk measure, then the corresponding ρ_w associated with ambiguous probability measure $\mathcal{U}$ remains a coherent risk measure.*

Huang et al. [9] propose the *relative robust CVaR model, $RCVaR_\beta$*, and claim that

$$\mathrm{RCVaR}_\beta(\mathbf{w}) = \max_{F(\cdot)\in\mathcal{U}} \left\{ \mathrm{CVaR}_{\beta,\mathrm{F}}(\mathbf{w}) - \mathrm{CVaR}_\alpha(\mathbf{w}^*) \right\},$$

where $\mathbf{w}^* = \underset{\mathbf{w}\in\mathcal{X}}{\operatorname{argmin}}\, \mathrm{CVaR}_{\beta,\mathrm{F}}(\mathbf{w})$, is not a coherent risk measure. As it is not the main focus of this research report, we do not discuss relative robustness further. We note that Delage and Ye [6] divide the uncertainty into distribution form as well as into moments form.

Much of the research of mean-CVaR portfolio selection focuses on the uncertainty of the first and/or second order moments. The concept of *distribution ambiguity*, referring to events for which the probabilities of the future outcomes are unknown so that it does not require the distribution of $\mathbf{X}$ to be specified, is integrated into some of them. For example, Quaranta and Zaffaroni [14] utilize the box uncertainty set of expected return

$$\mathcal{U} = \left\{ E(\mathbf{X}) = [E(X_1), \dots, E(X_n)]^T : |E(X_i) - \hat{\mu}_i| \le s_i, \forall i = 1, \dots, n \right\},$$

where $\hat{\mu}_i$ is the weighted average of asset i to amortize the *ghost effect*. And they then use *Soyster's approach* to build a linear program. Delage and Ye [6] use the set of ambiguous distribution with conic constraints of first and second order moments:

$$\mathcal{U} = \left\{ F(\cdot): \begin{array}{l} P(\mathbf{X} \in \Omega) = 1,\ P \in \mathcal{M}_+, \\ (E_F(\mathbf{X}) - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1}((E_F(\mathbf{X}) - \hat{\boldsymbol{\mu}}) \le \gamma_1, \\ E_F[(\mathbf{X} - \hat{\boldsymbol{\mu}})(\mathbf{X} - \hat{\boldsymbol{\mu}})^T] \le \gamma_2\hat{\Sigma}. \end{array} \right\},$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are empirical mean and covariance matrix, respectively and $\mathcal{M}_+$ is the set of all probability measures. Kang et al. [10] propose a simpler set of ambiguous distribution with ellipsoidal constraints,

$$\mathcal{U}_{(F,\gamma_1,\gamma_2)} = \left\{ F(\cdot): \begin{array}{l} P(\mathbf{X} \in \Omega) = 1,\ P \in \mathcal{M}_+, \\ (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \le \gamma_1, \\ \|\Sigma - \hat{\Sigma}\|_F \le \gamma_2, \\ \boldsymbol{\mu} = E_F(\mathbf{X}),\ \Sigma = \mathrm{Cov}_F(\mathbf{X}) \succ 0 \end{array} \right\}, \tag{1.7}$$

They also add a zero net adjustment constraint to reduce the conservativeness in the absolute robust model. We implement (1.7) in this work.

Other researchers take the uncertainty of the distribution form into consideration. Huang et al. [9] use a simple finite set of possible distributions. Zhu and Fukushima [18] investigate distributional uncertainty of mixture distribution uncertainty, box uncertainty, and ellipsoidal uncertainty.

For the mean-CVaR portfolio selection problem with moment uncertainty, the following Lemma 1.2.2 is useful:

Lemma 1.2.2 (From [4, Theorem 2.9]). *Let the random vector* $\mathbf{X} \in \mathbb{R}^n$ *have mean* $\boldsymbol{\mu}$, *covariance matrix* $\Sigma \succ 0$, *and an uncertain probability distribution function* $F(\cdot)$ *that follows a family of distributions*

$$\mathcal{F}_{(\boldsymbol{\mu},\Sigma)} = \left\{ F(\cdot) : P(\mathbf{X} \in \Omega) = 1, E_F(\mathbf{X}) = \boldsymbol{\mu}, \mathrm{Cov}_F(\mathbf{X}) = \Sigma \right\},$$

with $\Omega = \mathbb{R}^n$. *Then there exists a tight upper bound for the worst case CVaR with respect to the uncertainty set* $\mathcal{F}_{(\boldsymbol{\mu},\Sigma)}$:

$$\max_{F(\cdot) \in \mathcal{F}_{(\boldsymbol{\mu},\Sigma)}} \mathrm{CVaR}_{\beta,\mathrm{F}}(\mathbf{w}) = -\boldsymbol{\mu}^T \mathbf{w} + \kappa \sqrt{\mathbf{w}^T \Sigma \mathbf{w}}, \tag{1.8}$$

where $\kappa = \sqrt{\frac{\beta}{1-\beta}}$.

The lemma indicates that the robust mean-CVaR model is an extension of the mean-variance model but magnifies the error of the bias from estimation of $\boldsymbol{\mu}$ and Σ.

Kang et al. [10] propose the following Proposition 1.2.3 but without a detailed proof. We include the details below for completeness.

Proposition 1.2.3 (From [10, Proposition 2.2]). *Let* $\mathbf{w} \in \mathbb{R}^n$ *be given. Then*

$$\max_{F(\cdot) \in \mathcal{U}_{(F,\gamma_1,\gamma_2)}} \mathrm{CVaR}_{\beta,\mathrm{F}}(\mathbf{w}) = \max_{(\boldsymbol{\mu},\Sigma) \in \mathcal{U}_{(\hat{\boldsymbol{\mu}},\hat{\Sigma},\gamma_1,\gamma_2)}} \max_{F(\cdot) \in \mathcal{F}_{(\boldsymbol{\mu},\Sigma)}} \mathrm{CVaR}_{\beta,\mathrm{F}}(\mathbf{w}), \tag{1.9}$$

where

$$\mathcal{U}_{(\hat{\boldsymbol{\mu}},\hat{\Sigma},\gamma_1,\gamma_2)} = \left\{ (\boldsymbol{\mu}, \Sigma) \in \mathbb{R}^n \times \mathcal{S}_{++}^n : (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \gamma_1, \|\Sigma - \hat{\Sigma}\|_F \leq \gamma_2 \right\}. \tag{1.10}$$

Proof. We need to prove that the feasible sets for both problems in (1.9) are the same.

1. $(\subseteq)$: $\forall F(\cdot) \in \mathcal{U}_{(F,\gamma_1,\gamma_2)}$,

$$(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \gamma_1, \; \|\Sigma - \hat{\Sigma}\|_F \leq \gamma_2, \; \Sigma \in \mathcal{S}^n_{++} \Rightarrow (\boldsymbol{\mu}, \Sigma) \in \mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1, \gamma_2)},$$

where $E_F(\mathbf{X}) = \boldsymbol{\mu}$ and $\mathrm{Cov}_F(\mathbf{X}) = \Sigma$ are as defined. So $F(\cdot) \in \mathcal{F}_{(\boldsymbol{\mu}, \Sigma)}$ follows.

2. $(\supseteq)$: $\forall F(\cdot) \in \mathcal{F}_{(\boldsymbol{\mu}, \Sigma)}$,

$$E_F(\mathbf{X}) = \boldsymbol{\mu} \; \text{and} \; \mathrm{Cov}_F(\mathbf{X}) = \Sigma,$$

and since $(\boldsymbol{\mu}, \Sigma) \in \mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1, \gamma_2)}$,

$$\Sigma \in \mathcal{S}^n_{++}, \; (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \gamma_1, \; \|\Sigma - \hat{\Sigma}\|_F \leq \gamma_2 \Rightarrow F(\cdot) \in \mathcal{U}_{(F,\gamma_1,\gamma_2)}.$$

$\square$

The proof for Proposition 1.2.3 also works for

$$\min_{F(\cdot) \in \mathcal{U}_{(F,\gamma_1,\gamma_2)}} \mathrm{CVaR}_{\beta,\mathrm{F}}(\mathbf{w}) = \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1, \gamma_2)}} \min_{F(\cdot) \in \mathcal{F}_{(\boldsymbol{\mu}, \Sigma)}} \mathrm{CVaR}_{\beta,\mathrm{F}}(\mathbf{w}).$$

We can see that the problem depends on the *size parameters* (γ_1, γ_2) of the uncertainty set, and (RMC-general) can be formulated as:

$$\mathrm{WCVaR}^*_\beta(\gamma_1, \gamma_2) = \min_{\mathbf{w} \in \mathcal{X}} \max_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1, \gamma_2)}} \max_{F(\cdot) \in \mathcal{F}_{(\boldsymbol{\mu}, \Sigma)}} \mathrm{CVaR}_{\beta,\mathrm{F}}(\mathbf{w})$$
$$\text{s.t.} \quad \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1, \gamma_2)}} \min_{F(\cdot) \in \mathcal{F}_{(\boldsymbol{\mu}, \Sigma)}} \boldsymbol{\mu}^T \mathbf{w} \geq \mu_0.$$

By Lemma 1.2.2, the problem becomes:

$$\mathrm{WCVaR}^*_\beta(\gamma_1, \gamma_2) = \min_{\mathbf{w} \in \mathcal{X}} \max_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1, \gamma_2)}} -\boldsymbol{\mu}^T \mathbf{w} + \kappa \sqrt{\mathbf{w}^T \Sigma \mathbf{w}}$$
$$\text{s.t.} \quad \min_{\boldsymbol{\mu} \in \mathcal{U}^1_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1)}} \boldsymbol{\mu}^T \mathbf{w} \geq \mu_0. \qquad \qquad (\text{RMC-}\mathcal{U}_\gamma)$$

Note that $\mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1, \gamma_2)}$ can be separate as the union of two sets:

$$\mathcal{U}^1_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1)} = \left\{ \boldsymbol{\mu} \in \mathbb{R}^n : (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \gamma_1 \right\}$$
$$\text{and} \mathcal{U}^2_{(\hat{\Sigma}, \gamma_2)} = \left\{ \Sigma \in \mathcal{S}^n_{++} : \|\Sigma - \hat{\Sigma}\|_F \leq \gamma_2 \right\}.$$

Accordingly, the robust *global minimum risk problem*, which is the original problem without the target expected return constraint, is,

$$\text{WCVaR}_\beta^*(\gamma_1, \gamma_2) = \min_{\mathbf{w} \in \mathcal{X}} \max_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1, \gamma_2)}} -\boldsymbol{\mu}^T \mathbf{w} + \kappa \sqrt{\mathbf{w}^T \Sigma \mathbf{w}}. \tag{1.11}$$

The *global maximum expected return problem*, which instead maximize the *worst case expected return, WR* without considering risk, is,

$$\text{WR}^*(\gamma_1) = \max_{\mathbf{w} \in \mathcal{X}} \min_{\boldsymbol{\mu} \in \mathcal{U}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \gamma_1)}^1} \boldsymbol{\mu}^T \mathbf{w}. \tag{1.12}$$

1.3 Sensitivity Analysis

As discussed in Section 1.2, it is a nontrivial problem to determine the *best* size of an uncertainty set. For the uncertainty set $\mathcal{U}_{(F, \gamma_1, \gamma_2)}$, γ_1 and γ_2 are the *size parameters* for the set. Sensitivity analysis reflects the impact of small perturbations in the inputs to the outputs. It can be a complementary alternative to robust optimization. We continue now to combine the sensitivity analysis into the determination for the size of $\mathcal{U}_{(F, \gamma_1, \gamma_2)}$.

We consider a general nonlinear program and general *perturbations* $\boldsymbol{\epsilon} \in \mathbb{R}^k$ appearing in the objective function and in the constraints:

$$\begin{aligned} \min_{\mathbf{w}} \quad & f(\mathbf{w}, \boldsymbol{\epsilon}) \\ \text{s.t.} \quad & g(\mathbf{w}, \boldsymbol{\epsilon}) \geq 0 \\ & h(\mathbf{w}, \boldsymbol{\epsilon}) = 0. \end{aligned} \qquad P_g(\boldsymbol{\epsilon})$$

Here $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$, and $h : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^p$. The Lagrangian of $P_g(\boldsymbol{\epsilon})$ is

$$L(\mathbf{w}, \mathbf{u}, \mathbf{v}, \boldsymbol{\epsilon}) = f(\mathbf{w}, \boldsymbol{\epsilon}) - \sum_{i=1}^m u_i g_i(\mathbf{w}, \boldsymbol{\epsilon}) + \sum_{j=1}^p v_j h_j(\mathbf{w}, \boldsymbol{\epsilon}),$$

where $\mathbf{u} = (u_1, \dots, u_m)$ and $\mathbf{v} = (v_1, \dots, v_p)$ are vectors of Lagrangian multipliers.

From Fiacco [7], the following results are useful.

Theorem 1.3.1 (From [7, Theorem 3.2.2, Corollary 3.2.4]). *Let the following four assumptions in Items I to IV hold:*

   I the functions defining $P_g(\boldsymbol{\epsilon})$ are twice continuously differentiable in $\mathbf{w}$ and their gradients with respect to $\mathbf{w}$ and the constraints are once continuously differentiable in $\boldsymbol{\epsilon}$ in a neighborhood of $(\mathbf{w}^*, 0)$;

II the second-order sufficient conditions for a local minimum of $P(0)$ hold at $\mathbf{w}^*$, with associated Lagrange multipliers $\mathbf{u}^*$ and $\mathbf{v}^*$;

III the gradients $\nabla g_i(\mathbf{w}^*, 0)$ (for i such that $g_i(\mathbf{w}^*, 0) = 0$) and $\nabla h_j(\mathbf{w}^*, 0)$ (all j) are linearly independent (linear independence constraint qualification, LICQ);

IV $u_i^* > 0$ when $g_i(\mathbf{w}^*, 0) = 0$ ($i = 1, \ldots, m$), i.e., strict complementary slackness holds.

Then:

(a) $\mathbf{w}^*$ is a local isolated minimizing point of $P(0)$, and the associated Lagrange multipliers $\mathbf{u}^*$ and $\mathbf{v}^*$ are unique;

(b) for $\boldsymbol{\epsilon}$ in a neighborhood of 0, there exists a unique, once continuously differentiable vector function $\mathbf{y}^*(\boldsymbol{\epsilon}) = [\mathbf{w}^*(\boldsymbol{\epsilon}), \mathbf{u}^*(\boldsymbol{\epsilon}), \mathbf{v}^*(\boldsymbol{\epsilon})]^T$ satisfying the second-order sufficient conditions for a local minimum of problem $P_g(\boldsymbol{\epsilon})$ such that $\mathbf{y}^*(0) = [\mathbf{w}^*, \mathbf{u}^*, \mathbf{v}^*]^T = \mathbf{y}^*$, and hence $\mathbf{w}^*(\boldsymbol{\epsilon})$ is a locally unique local minimum of problem $P_g(\boldsymbol{\epsilon})$ with associated Lagrange multipliers $\mathbf{u}^*(\boldsymbol{\epsilon})$ and $\mathbf{v}^*(\boldsymbol{\epsilon})$;

(c) for $\boldsymbol{\epsilon}$ near 0, the set of binding inequalities is unchanged, strict complementary slackness holds, and the binding constraint gradients are linearly independent at $\mathbf{w}^*(\boldsymbol{\epsilon})$;

(d) a first-order approximation of $[\mathbf{w}^*(\boldsymbol{\epsilon}), \mathbf{u}^*(\boldsymbol{\epsilon}), \mathbf{v}^*(\boldsymbol{\epsilon})]$ in a neighborhood of $\boldsymbol{\epsilon} = 0$ is given by

$$\begin{bmatrix} \mathbf{w}^*(\boldsymbol{\epsilon}) \\ \mathbf{u}^*(\boldsymbol{\epsilon}) \\ \mathbf{v}^*(\boldsymbol{\epsilon}) \end{bmatrix} = \begin{bmatrix} \mathbf{w}^* \\ \mathbf{u}^* \\ \mathbf{v}^* \end{bmatrix} + (M^*)^{-1} N^* \boldsymbol{\epsilon} + o(\|\boldsymbol{\epsilon}\|), \tag{1.13}$$

where $\phi(\boldsymbol{\epsilon}) = o(\|\boldsymbol{\epsilon}\|)$ means $\lim_{\boldsymbol{\epsilon} \to 0} \phi(\boldsymbol{\epsilon})/\|\boldsymbol{\epsilon}\| = 0$. $M^* = M(0)$ and $N^* = N(0)$, where $M(\boldsymbol{\epsilon})$, $N(\boldsymbol{\epsilon})$ are defined below:

$$M(\boldsymbol{\epsilon}) = \left[\begin{array}{c|ccc|ccc} \nabla_{\mathbf{ww}}^2 L & \cdots & -\nabla_{\mathbf{w}} g_i & \cdots & \cdots & -\nabla_{\mathbf{w}} h_j & \cdots \\ \vdots & \ddots & & & & & \\ u_i \nabla_{\mathbf{w}} g_i^T & & g_i & & & 0 & \\ \vdots & & & \ddots & & & \\ \nabla_{\mathbf{w}} h_j^T & & 0 & & & 0 & \end{array} \right] \in \mathbb{R}^{(n+m+p) \times (n+m+p)},$$

$$N(\boldsymbol{\epsilon}) = \left[-\nabla_{\boldsymbol{\epsilon}\mathbf{w}}^2 L, \cdots, -u_i \nabla_{\boldsymbol{\epsilon}} g_i, \cdots, \nabla_{\boldsymbol{\epsilon}} h_j, \cdots \right]^T \in \mathbb{R}^{(n+m+p) \times k}.$$

$$\tag{1.14}$$

Theorem 1.3.2 (From [7, Theorem 3.4.1]). *Let the local optimal value function be $f^*(\boldsymbol{\epsilon}) = f(\mathbf{w}^*(\boldsymbol{\epsilon}), \boldsymbol{\epsilon})$, and the optimal value Lagrangian be $L^*(\boldsymbol{\epsilon}) = L(\mathbf{w}^*(\boldsymbol{\epsilon}), \mathbf{u}^*(\boldsymbol{\epsilon}), \mathbf{v}^*(\boldsymbol{\epsilon}), \boldsymbol{\epsilon})$. If the conditions of Theorem 1.3.1 hold for $P_g(\boldsymbol{\epsilon})$, and if the problem functions are twice continuously differentiable in $(\mathbf{w}^*(\boldsymbol{\epsilon}), \boldsymbol{\epsilon})$ near $(\mathbf{w}^*, 0)$ then, in a neighborhood of $\boldsymbol{\epsilon} = 0$,*

$$f^*(\boldsymbol{\epsilon}) = L^*(\boldsymbol{\epsilon}). \tag{1.15}$$

From (1.13), we denote the first order approximation as

$$\begin{bmatrix} \hat{\mathbf{w}}(\boldsymbol{\epsilon}) \\ \hat{\mathbf{u}}(\boldsymbol{\epsilon}) \\ \hat{\mathbf{v}}(\boldsymbol{\epsilon}) \end{bmatrix} = \begin{bmatrix} \mathbf{w}^* \\ \mathbf{u}^* \\ \mathbf{v}^* \end{bmatrix} + (M^*)^{-1} N^* \boldsymbol{\epsilon} \tag{1.16}$$

By Theorem 1.3.1(b), $\mathbf{w}^*(\boldsymbol{\epsilon})$ is the unique local minimum of problem $P_g(\boldsymbol{\epsilon})$. For a convex optimization problem, $\mathbf{w}^*(\boldsymbol{\epsilon})$ is the unique global minimum.

1.3.1 Perturbation Towards the Original Parameters

Let's consider the direct perturbation towards the original parameters.

First, assume that the population statistics are estimated exactly by the sample statistics, which means $\gamma_1 = \gamma_2 = 0$. RMC-$\mathcal{U}_\gamma$ becomes RMC-$\mathcal{U}_0$:

$$\text{WCVaR}_\beta^*(0,0) = \min_{\mathbf{w} \in \mathcal{X}} \quad -\hat{\boldsymbol{\mu}}^T \mathbf{w} + \kappa \sqrt{\mathbf{w}^T \hat{\Sigma} \mathbf{w}} \qquad \text{(RMC-}\mathcal{U}_0\text{)}$$
$$\text{s.t.} \quad \hat{\boldsymbol{\mu}}^T \mathbf{w} \geq \mu_0.$$

We now introduce a *perturbation,* $\boldsymbol{\epsilon}$, to the sample statistics, and denote the statistics under the perturbations as $\hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}$ and $\hat{\Sigma}_{\boldsymbol{\epsilon}}$. We use sensitivity analysis to test how much the output $\mathbf{w}^*$ and the expected return $\hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}^T \mathbf{w}^*$ are sensitive to $\boldsymbol{\epsilon}$. That is we need to test: if we have sample estimation that is wrong, then how serious is the influence on the final decision? The intuition is that, if either the $\mathbf{w}^*$ or $\hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}^T \mathbf{w}^*$ is sensitive to error, then RMC-$\mathcal{U}_0$ will behave badly once the sample statistics are not equal to the population statistics. Therefore γ_1 and/or γ_2 should be large.

Definition 1.3.3. *Let* chol$(\cdot)$ *denote the* Cholesky decomposition *of a positive definite matrix, i.e., for $A \in \mathcal{S}_{++}^n$, we get $L =$ chol(A), where $A = LL^T$, and L is a lower triangular matrix.*

Define the *triangular number,* $t(n) = n(n+1)/2$.

Definition 1.3.4. *Let* $\mathrm{LMat}(\cdot) : \mathbb{R}^{t(n)} \to \mathbb{R}^{n \times n}$ *be the transformation that generates a lower triangular matrix from a vector.*

For example,

$$\mathrm{LMat}(\mathbf{a}) = \begin{pmatrix} a_1 & 0 \\ a_2 & a_3 \end{pmatrix}$$

reshapes $\mathbf{a} \in \mathbb{R}^3$ into a 2-by-2 lower triangular matrix. It is clear that $\mathrm{LMat}(\cdot)$ is a linear transformation.

Now let

$$\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2) \in \mathbb{R}^n \times \mathbb{R}^{t(n)}, \quad \hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}} = \hat{\boldsymbol{\mu}} + \boldsymbol{\epsilon}_1, \quad \hat{\Sigma}_{\boldsymbol{\epsilon}} = (L + \mathrm{LMat}(\boldsymbol{\epsilon}_2))(L + \mathrm{LMat}(\boldsymbol{\epsilon}_2))^T \in \mathcal{S}_+^n,$$

where $L = \mathrm{chol}(\hat{\Sigma})$. Note that we obtain positive definiteness of $\hat{\Sigma}_{\boldsymbol{\epsilon}}$ if $\boldsymbol{\epsilon}_2$ is *small*. Besides, $\boldsymbol{\epsilon}_1 = [I_n, \mathbf{0}_{n \times t(n)}]\boldsymbol{\epsilon}$, $\boldsymbol{\epsilon}_2 = [\mathbf{0}_{t(n) \times n}, I_{t(n)}]\boldsymbol{\epsilon}$.

Recall that the constraint set $\mathcal{X} = \left\{ \mathbf{w} \in \mathbb{R}^n : \mathbf{w} \geq \mathbf{0}_n, \mathbf{1}_n^T \mathbf{w} = 1 \right\}$. We need to write RMC-$\mathcal{U}_0$ with the perturbation explicitly:

$$\begin{aligned} \mathrm{WCVaR}^*_{\beta, \boldsymbol{\epsilon}}(0, 0) = \quad &\min_{\mathbf{w} \in \mathbb{R}^n} \quad -\hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}^T \mathbf{w} + \kappa \sqrt{\mathbf{w}^T \hat{\Sigma}_{\boldsymbol{\epsilon}} \mathbf{w}} \\ &\text{s.t.} \quad \hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}^T \mathbf{w} \geq \mu_0, \\ &\qquad \mathbf{1}_n^T \mathbf{w} = 1, \ \mathbf{w} \geq \mathbf{0}_n. \end{aligned} \qquad \text{(RMC-}\mathcal{U}_{\boldsymbol{\epsilon}})$$

The Lagrangian for (RMC-$\mathcal{U}_{\boldsymbol{\epsilon}}$) is:

$$\begin{aligned} L(\mathbf{w}, \mathbf{u}, \mathbf{v}, \boldsymbol{\epsilon}) &= f(\mathbf{w}, \boldsymbol{\epsilon}) - \sum_{i=0}^{n} u_i g_i(\mathbf{w}, \boldsymbol{\epsilon}) + v h(\mathbf{w}, \boldsymbol{\epsilon}) \\ &= -\hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}^T \mathbf{w} + \kappa \sqrt{\mathbf{w}^T \hat{\Sigma}_{\boldsymbol{\epsilon}} \mathbf{w}} - u_0(\hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}^T \mathbf{w} - \mu_0) - \sum_{i=1}^{n} u_i \mathbf{e}_i^T \mathbf{w} + v(\mathbf{1}_n^T \mathbf{w} - 1) \\ &= -(1 + u_0)\boldsymbol{\epsilon}_1^T \mathbf{w} + \kappa \left\| (L + \mathrm{LMat}(\boldsymbol{\epsilon}_2))^T \mathbf{w} \right\|_2 + C(\mathbf{w}), \end{aligned}$$

where $\mathbf{e}_i$ denotes the i-th unit vector. Note that

$$\begin{aligned} g_0(\mathbf{w}, \boldsymbol{\epsilon}) &= \hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}^T \mathbf{w} - \mu_0, \\ g_i(\mathbf{w}, \boldsymbol{\epsilon}) &= \mathbf{e}_i^T \mathbf{w}, \ i = 1, \dots, n, \\ h(\mathbf{w}, \boldsymbol{\epsilon}) &= \mathbf{1}_n^T \mathbf{w} - 1. \end{aligned}$$

16

And

$$C(\mathbf{w}) = -(1 + u_0)\hat{\boldsymbol{\mu}}^T\mathbf{w} - u_0\mu_0 - \sum_{i=1}^{n} u_i\mathbf{e}_i^T\mathbf{w} + v(\mathbf{1}_n^T\mathbf{w} - 1)$$

is a function independent of $\boldsymbol{\epsilon}$.

Then, by (1.14), for elements in $M(\boldsymbol{\epsilon}) \in \mathbb{R}^{(2n+2)\times(2n+2)}$,

$$\nabla_{\mathbf{ww}}^2 L = \kappa \left(\mathbf{w}^T\hat{\Sigma}_{\boldsymbol{\epsilon}}\mathbf{w}\right)^{-\frac{3}{2}} \left[\left(\mathbf{w}^T\hat{\Sigma}_{\boldsymbol{\epsilon}}\mathbf{w}\right)\hat{\Sigma}_{\boldsymbol{\epsilon}} - \left(\hat{\Sigma}_{\boldsymbol{\epsilon}}\mathbf{w}\right)\left(\hat{\Sigma}_{\boldsymbol{\epsilon}}\mathbf{w}\right)^T\right],$$

$$\nabla_{\mathbf{w}}g_0 = \hat{\boldsymbol{\mu}}_{\boldsymbol{\epsilon}}, \tag{1.17}$$

$$\nabla_{\mathbf{w}}g_i = \mathbf{e}_i, i = 1, \ldots, n,$$

$$\nabla_{\mathbf{w}}h = \mathbf{1}_n,$$

and for elements in $N(\boldsymbol{\epsilon}) \in \mathbb{R}^{(2n+2)\times(n+t(n))}$,

$$\nabla_{\boldsymbol{\epsilon}\mathbf{w}}^2 L = -(1 + u_0)\begin{bmatrix} I_n & \mathbf{0}_{n\times t(n)} \end{bmatrix}^T + \kappa\nabla_{\boldsymbol{\epsilon}\mathbf{w}}^2 \left\|(L + \mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\right\|_2,$$

$$\nabla_{\boldsymbol{\epsilon}}g_0 = \begin{bmatrix} I_n & \mathbf{0}_{n\times t(n)} \end{bmatrix}^T \mathbf{w}, \tag{1.18}$$

$$\nabla_{\boldsymbol{\epsilon}}g_i = \mathbf{0}_{n+t(n)}, i = 1, \ldots, n,$$

$$\nabla_{\boldsymbol{\epsilon}}h = \mathbf{0}_{n+t(n)}.$$

The detailed calculation of $\nabla_{\boldsymbol{\epsilon}\mathbf{w}}^2 L$ can be found in appendix B, see (B.9).

With the first order approximation in (1.16), the perturbed optimal CVaR and expected return can be approximated as well. This sensitivity results towards the original parameters can be used to determine both the sizes and the shape of uncertainty sets. However, when the portfolio size n is large, the analysis could become very complicated. This application could be an interesting topic for the future work.

1.3.2   Perturbation Towards Size Parameters of Uncertainty Set

If a size of an uncertainty set is given, we can use sensitivity analysis to analyze whether the given size is appropriate. From Kang et al. [10], for the given γ_1, γ_2, the robust counterpart (RMC-$\mathcal{U}_\gamma$) becomes[3]:

$$\begin{aligned} \mathrm{WCVaR}_\beta^*(\gamma_1, \gamma_2) = \min_{\mathbf{w}\in\mathbb{R}^n} \quad & -\hat{\boldsymbol{\mu}}^T\mathbf{w} + \kappa\left\|(\hat{\Sigma} + \gamma_2 I_n)^{1/2}\mathbf{w}\right\|_2 + \sqrt{\gamma_1}\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2 \\ \text{s.t.} \quad & \sqrt{\gamma_1}\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2 \leq \hat{\boldsymbol{\mu}}^T\mathbf{w} - \mu_0, \\ & \mathbf{1}_n^T\mathbf{w} = 1, \mathbf{w} \geq \mathbf{0}. \end{aligned} \tag{P_0}$$

---

[3]It can be reformulated as a second-order cone program with auxiliary variables.

By separating $\displaystyle\max_{(\boldsymbol{\mu},\Sigma)\in\mathcal{U}_{(\hat{\boldsymbol{\mu}},\hat{\Sigma},\gamma_1,\gamma_2)}} -\boldsymbol{\mu}^T\mathbf{w}+\kappa\sqrt{\mathbf{w}^T\Sigma\mathbf{w}}$ from the (RMC-$\mathcal{U}_\gamma$) into two subproblems, the worst case is attained when

$$
\begin{aligned}
\boldsymbol{\mu}_{wc} &= \operatorname*{argmax}_{(\boldsymbol{\mu},\Sigma)\in\mathcal{U}_{(\hat{\boldsymbol{\mu}},\hat{\Sigma},\gamma_1,\gamma_2)}} -\boldsymbol{\mu}^T\mathbf{w} = \hat{\boldsymbol{\mu}} - \frac{\sqrt{\gamma_1}\hat{\Sigma}\mathbf{w}}{\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2}, \\
\text{and} \quad \Sigma_{wc} &= \operatorname*{argmax}_{(\boldsymbol{\mu},\Sigma)\in\mathcal{U}_{(\hat{\boldsymbol{\mu}},\hat{\Sigma},\gamma_1,\gamma_2)}} \kappa\sqrt{\mathbf{w}^T\Sigma\mathbf{w}} = \hat{\Sigma} + \frac{\gamma_2\mathbf{w}\mathbf{w}^T}{\mathbf{w}^T\mathbf{w}}.
\end{aligned}
\tag{1.19}
$$

Now let
$$
\boldsymbol{\epsilon} = (\epsilon_1,\epsilon_2) \in \mathbb{R}\times\mathbb{R},\ \gamma_1^{\boldsymbol{\epsilon}} = \gamma_1 + \epsilon_1,\ \gamma_2^{\boldsymbol{\epsilon}} = \gamma_2 + \epsilon_2.
$$
Substitute (γ_1,γ_2) in (P_0) by $(\gamma_1^{\boldsymbol{\epsilon}},\gamma_2^{\boldsymbol{\epsilon}})$:

$$
\begin{aligned}
\mathrm{WCVaR}_\beta^*(\gamma_1^{\boldsymbol{\epsilon}},\gamma_2^{\boldsymbol{\epsilon}}) = \ \min_{\mathbf{w}\in\mathbb{R}^n} \ & -\hat{\boldsymbol{\mu}}^T\mathbf{w} + \kappa\left\|(\hat{\Sigma}+\gamma_2^{\boldsymbol{\epsilon}}I_n)^{1/2}\mathbf{w}\right\|_2 + \sqrt{\gamma_1^{\boldsymbol{\epsilon}}}\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2 \\
\text{s.t.} \quad & \sqrt{\gamma_1^{\boldsymbol{\epsilon}}}\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2 \le \hat{\boldsymbol{\mu}}^T\mathbf{w} - \mu_0, \\
& \mathbf{1}_n^T\mathbf{w} = 1,\ \mathbf{w}\ge\mathbf{0}.
\end{aligned}
\tag{$P_{\boldsymbol{\epsilon}}$}
$$

The Lagrangian for $(P_{\boldsymbol{\epsilon}})$ is:

$$
\begin{aligned}
L(\mathbf{w},\mathbf{u},\mathbf{v},\boldsymbol{\epsilon}) &= f(\mathbf{w},\boldsymbol{\epsilon}) - \sum_{i=0}^{n} u_i g_i(\mathbf{w},\boldsymbol{\epsilon}) + vh(\mathbf{w},\boldsymbol{\epsilon}) \\
&= -\hat{\boldsymbol{\mu}}^T\mathbf{w} + \kappa\left\|(\hat{\Sigma}+\gamma_2^{\boldsymbol{\epsilon}}I_n)^{1/2}\mathbf{w}\right\|_2 + \sqrt{\gamma_1^{\boldsymbol{\epsilon}}}\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2 \\
&\quad -u_0\left(-\sqrt{\gamma_1^{\boldsymbol{\epsilon}}}\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2 + \hat{\boldsymbol{\mu}}^T\mathbf{w} - \mu_0\right) - \sum_{i=1}^{n} u_i\mathbf{e}_i^T\mathbf{w} + v(\mathbf{1}_n^T\mathbf{w}-1) \\
&= \kappa\left\|(\hat{\Sigma}+\gamma_2^{\boldsymbol{\epsilon}}I_n)^{1/2}\mathbf{w}\right\|_2 + \sqrt{\gamma_1^{\boldsymbol{\epsilon}}}(1+u_0)\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2 - (1+u_0)\hat{\boldsymbol{\mu}}^T\mathbf{w} \\
&\quad +u_0\mu_0 - \sum_{i=1}^{n} u_i\mathbf{e}_i^T\mathbf{w} + v(\mathbf{1}_n^T\mathbf{w}-1).
\end{aligned}
$$

Note that
$$
\begin{aligned}
g_0(\mathbf{w},\boldsymbol{\epsilon}) &= -\sqrt{\gamma_1^{\boldsymbol{\epsilon}}}\left\|\hat{\Sigma}^{1/2}\mathbf{w}\right\|_2 + \hat{\boldsymbol{\mu}}^T\mathbf{w} - \mu_0, \\
g_i(\mathbf{w},\boldsymbol{\epsilon}) &= \mathbf{e}_i^T\mathbf{w},\ i=1,\ldots,n, \\
h(\mathbf{w},\boldsymbol{\epsilon}) &= \mathbf{1}_n^T\mathbf{w} - 1.
\end{aligned}
$$

Then, for elements in $M(\boldsymbol{\epsilon}) \in \mathbb{R}^{(2n+2)\times(2n+2)}$,

$$\nabla^2_{\mathbf{ww}} L = \kappa \left\| \left(\hat{\Sigma} + \gamma_2^{\boldsymbol{\epsilon}} I_n\right)^{1/2} \mathbf{w} \right\|_2^{-3}$$

$$\left[\left\| \left(\hat{\Sigma} + \gamma_2^{\boldsymbol{\epsilon}} I_n\right)^{1/2} \mathbf{w} \right\|_2^2 \left(\hat{\Sigma} + \gamma_2^{\boldsymbol{\epsilon}} I_n\right) - \left(\left(\hat{\Sigma} + \gamma_2^{\boldsymbol{\epsilon}} I_n\right) \mathbf{w}\right) \left(\left(\hat{\Sigma} + \gamma_2^{\boldsymbol{\epsilon}} I_n\right) \mathbf{w}\right)^T \right]$$

$$+ \sqrt{\gamma_1^{\boldsymbol{\epsilon}}} \left(1 + u_0\right) \left\| \hat{\Sigma}^{1/2} \mathbf{w} \right\|_2^{-3} \left[\left\| \hat{\Sigma}^{1/2} \mathbf{w} \right\|_2^2 \hat{\Sigma} - \left(\hat{\Sigma} \mathbf{w}\right) \left(\hat{\Sigma} \mathbf{w}\right)^T \right]$$

$$\nabla_{\mathbf{w}} g_0 = -\sqrt{\gamma_1^{\boldsymbol{\epsilon}}} \frac{\hat{\Sigma} \mathbf{w}}{\left\| \hat{\Sigma}^{1/2} \mathbf{w} \right\|_2} + \hat{\boldsymbol{\mu}},$$

$$\nabla_{\mathbf{w}} g_i = \mathbf{e}_i, i = 1, \ldots, n,$$

$$\nabla_{\mathbf{w}} h = \mathbf{1}_n,$$

(1.20)

and for elements in $N(\boldsymbol{\epsilon}) \in \mathbb{R}^{(2n+2)\times 2}$,

$$\nabla^2_{\boldsymbol{\epsilon}\mathbf{w}} L = \left[\frac{1 + u_0}{2\sqrt{\gamma_1^{\boldsymbol{\epsilon}}}} \cdot \frac{\hat{\Sigma} \mathbf{w}}{\left\| \hat{\Sigma}^{1/2} \mathbf{w} \right\|_2} \quad \frac{2 \left\| \left(\hat{\Sigma} + \gamma_2^{\boldsymbol{\epsilon}} I_n\right)^{1/2} \mathbf{w} \right\|_2^2 I_n - \mathbf{w}^T \mathbf{w} \left(\hat{\Sigma} + \gamma_2^{\boldsymbol{\epsilon}} I_n\right)}{2 \left\| \left(\hat{\Sigma} + \gamma_2^{\boldsymbol{\epsilon}} I_n\right)^{1/2} \mathbf{w} \right\|_2^3} \mathbf{w} \right]^T$$

$$\nabla_{\boldsymbol{\epsilon}} g_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T \frac{1}{2\sqrt{\gamma_1^{\boldsymbol{\epsilon}}}} \left\| \hat{\Sigma}^{1/2} \mathbf{w} \right\|_2,$$

$$\nabla_{\boldsymbol{\epsilon}} g_i = \mathbf{0}_2, i = 1, \ldots, n,$$

$$\nabla_{\boldsymbol{\epsilon}} h = \mathbf{0}_2.$$

(1.21)

The first order approximation of CVaR and $\mathbf{w}$ is as (1.15) and (1.16). Besides, the first n rows of $(M^*)^{-1} N^*$ is the first order derivative $\frac{d\mathbf{w}}{d\boldsymbol{\epsilon}}$, which can be seen as the sensitivity of $\mathbf{w}$ to $\boldsymbol{\epsilon}$. Note that $\gamma_1 \neq 0$ is implied by N^*.

Chapter 2

Methodology and Simulation Analysis

In this chapter, we will illustrate the sensitivity analysis results in section 1.3.2 to show how to pick appropriate uncertainty sizes. We will use the example below to generate random samples for a simulation analysis. First, we will compare the true size parameters with the bootstrapping parameters in order to see how the bootstrap method could be improved. Then, the analysis is done separately for the global minimum risk problem and for the mean-CVaR problem. We will show how the first order approximation in the sensitivity analysis from section 1.3.2 performs. For the global minimum risk problem, we will propose three methods to find the size parameters by finding the elbow points and/or by controlling the robustness measure. Finally, we will compare the different methods by the efficient frontiers and portfolio compositions.

Example 2.0.1. *In the following and continuing below, we use an example of four underlying assets in Yam et al. [17] whose true mean vector and covariance matrix are given by:*

$$\boldsymbol{\mu} = \begin{bmatrix} 0.061166 \\ 0.109547 \\ 0.090358 \\ 0.040923 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0.018632 & 0.020056 & 0.020646 & 0.015213 \\ 0.020056 & 0.034507 & 0.027412 & 0.020652 \\ 0.020646 & 0.027412 & 0.048680 & 0.021663 \\ 0.015213 & 0.020652 & 0.021663 & 0.018791 \end{bmatrix}.$$

We assume that the returns of the four assets follows a multivariate Gaussian distribution Gaussian(μ, Σ) if without specific note.

Notation 2.0.2. *As in (1.13) and (1.16), we continue denoting the optimal value of $(P_{\boldsymbol{\epsilon}})$*

as $\mathbf{w}^*(\boldsymbol{\epsilon})$, and its first order approximation as $\hat{\mathbf{w}}(\boldsymbol{\epsilon})$. Similarly, we denote $WCVaR^*(\boldsymbol{\epsilon}) = WCVaR_\beta^*(\gamma_1^\epsilon, \gamma_2^\epsilon)$ when (γ_1, γ_2) is clear.

2.1 Bootstrap Method

As we discussed in the previous chapter, the bootstrap method has been used to determine the size parameters γ_1 and γ_2. Suppose we have the realized data of length T: $\mathbf{X}_{[i]}$, $i = 1, \ldots, T$ with mean $\hat{\boldsymbol{\mu}}$ and covariance matrix $\hat{\Sigma}$. Recall that we use (1.7) as the uncertainty set. We follow the bootstrapping procedure in Kang et al. [10]:

Step 1. Construct B bootstrap samples $\mathbf{X}^b = \{\mathbf{X}_{[i]}^b\}_{i=1}^T$, $b = 1, \ldots, B$ by drawing random samples with replacement from the available observations. The length of each bootstrap samples is T.

Step 2. For each bootstrap sample $\mathbf{X}^b$, compute the corresponding mean $\hat{\boldsymbol{\mu}}_b$ and covariance matrix $\hat{\Sigma}_b$, and then generate a sample
$$C = \left\{ (\hat{\boldsymbol{\mu}}_b, \hat{\Sigma}_b), \ b = 1, \ldots, B \right\}.$$

Step 3. For sample C, define data set C_{γ_1} and C_{γ_2} as
$$C_{\gamma_1} = \left\{ \gamma_{1b} : \gamma_{1b} = (\hat{\boldsymbol{\mu}}_b - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1} (\hat{\boldsymbol{\mu}}_b - \hat{\boldsymbol{\mu}}), \ b = 1, \ldots, B \right\},$$
$$C_{\gamma_2} = \left\{ \gamma_{2b} : \gamma_{2b} = \left\| \hat{\Sigma}_b - \hat{\Sigma} \right\|_F, \ b = 1, \ldots, B \right\}.$$

Step 4. Pick η as the confidence region. Compute the upper quantile of the corresponding data set:
$$\gamma_1 = \text{quantile}_\eta(C_{\gamma_1}), \ \gamma_2 = \text{quantile}_\eta(C_{\gamma_2}).$$

In the following, we will always take 95% as the confidence region.

2.1.1 Performance of Bootstrap Method to Pick Size Parameters

In the simulation analysis with the *true* mean and covariance matrix given, the parameter uncertainties due to the estimation error can be computed by:

$$\text{true } \gamma_1 = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \hat{\Sigma}^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}), \ \text{true } \gamma_2 = \left\| \hat{\Sigma} - \Sigma \right\|_F. \tag{2.1}$$

We want to compare bootstrapping values with the true values of γ_i by taking random samples (30 in multivariate Gaussian distribution and 30 in multivariate t distribution, using the parameters in Example 2.0.1), and then calculating their true γ_i and bootstrapping γ_i with $B = 10000$ resamples. We can see in figure 2.1 that for most cases, the bootstrapping values are larger than the true values, for both γ_1 and γ_2. Therefore, we can treat the bootstrapping values as a reference of upper bounds for each γ_i.
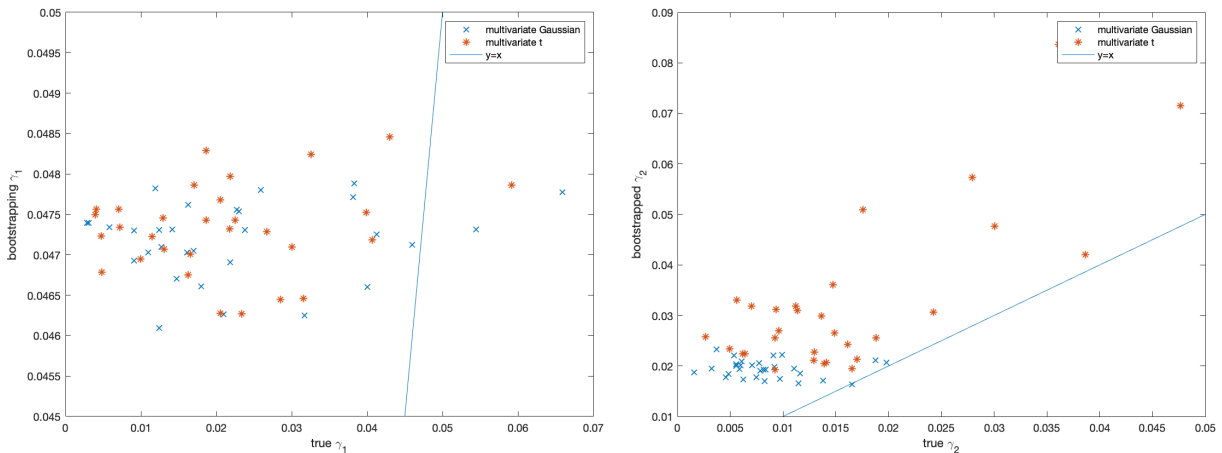
Figure 2.1: Bootstrapping values tend to overestimate true γ_i

2.2 Robust Global Minimum Risk Problem

Recall that the global minimum risk problem in (1.11) can be reformulated as the robust counterpart:

$$\text{WCVaR}_\beta^*(\gamma_1, \gamma_2) = \min_{\mathbf{w} \in \mathbb{R}^n} \quad -\hat{\boldsymbol{\mu}}^T\mathbf{w} + \kappa \left\| (\hat{\Sigma} + \gamma_2 I_n)^{1/2}\mathbf{w} \right\|_2 + \sqrt{\gamma_1} \left\| \hat{\Sigma}^{1/2}\mathbf{w} \right\|_2$$
$$\text{s.t.} \quad \mathbf{1}_n^T\mathbf{w} = 1, \ \mathbf{w} \geq \mathbf{0}.$$

In this section, we analyze the sensitivity analysis results for the global minimum risk problem. We also propose three methods to pick and analyze the size parameters for the uncertainty set. We have to notice that the above problem is dependent on the pair of parameters (γ_1, γ_2). We set $\beta = 0.95$.[1]

---

[1]Due to the complementary slackness, $u_0 = 0$ is set to calculate M^* and N^*; or it is equivalent to set μ_0 to a very small nagative number in (P_0) so that the target expected return constraint is inactive.

2.2.1 Performance of the First Order Approximation

We generate 20 random samples with length $T = 200$ and compare their first order approximation results to $\mathbf{w}^*(\boldsymbol{\epsilon})$ and $\text{WCVaR}^*(\boldsymbol{\epsilon})$ of $(P_{\boldsymbol{\epsilon}})$ with the exact solutions in figures 2.2 and 2.3 on pages 24, 25, respectively. The plots on the left show the first order approximations of $\text{WCVaR}^*(\boldsymbol{\epsilon}) = \text{WCVaR}^*_{\beta}(\gamma_1^{\boldsymbol{\epsilon}}, \gamma_2^{\boldsymbol{\epsilon}})$ versus the exact function values. When the points are almost on the identity line $y = x$, the first order approximations are good to approximate the optimal function value. The plots on the right show the difference $\hat{\mathbf{w}}(\boldsymbol{\epsilon}) - \mathbf{w}^*(\boldsymbol{\epsilon})$ for 4 assets. When the differences for all 4 assets are small, the first order approximations are good.

We can see that the plots indicate the first order approximations of $\mathbf{w}^*(\boldsymbol{\epsilon})$ and $\text{WCVaR}^*(\boldsymbol{\epsilon})$ after perturbation are good when $|\epsilon_1| = 1/10\gamma_1$ (the scale for the $\mathbf{w}^*(\boldsymbol{\epsilon}) - \hat{\mathbf{w}}(\boldsymbol{\epsilon})$ is 10^{-5}) but not as good even with the small change $|\epsilon_2| = 1/50\gamma_2$ (the scale for the $\mathbf{w}^*(\boldsymbol{\epsilon}) - \hat{\mathbf{w}}(\boldsymbol{\epsilon})$ is 10^{-3}). Note that in this analysis (γ_1, γ_2) are taken as the bootstrapping values for each sample (20 pairs of bootstrapping (γ_1, γ_2) with means $0.0474, 0.0199$, variances $2.58 \times 10^{-7}, 3.82 \times 10^{-6}$, respectively) and $\mu_0 = 0$. But we can also take other initial values for the γ_i.
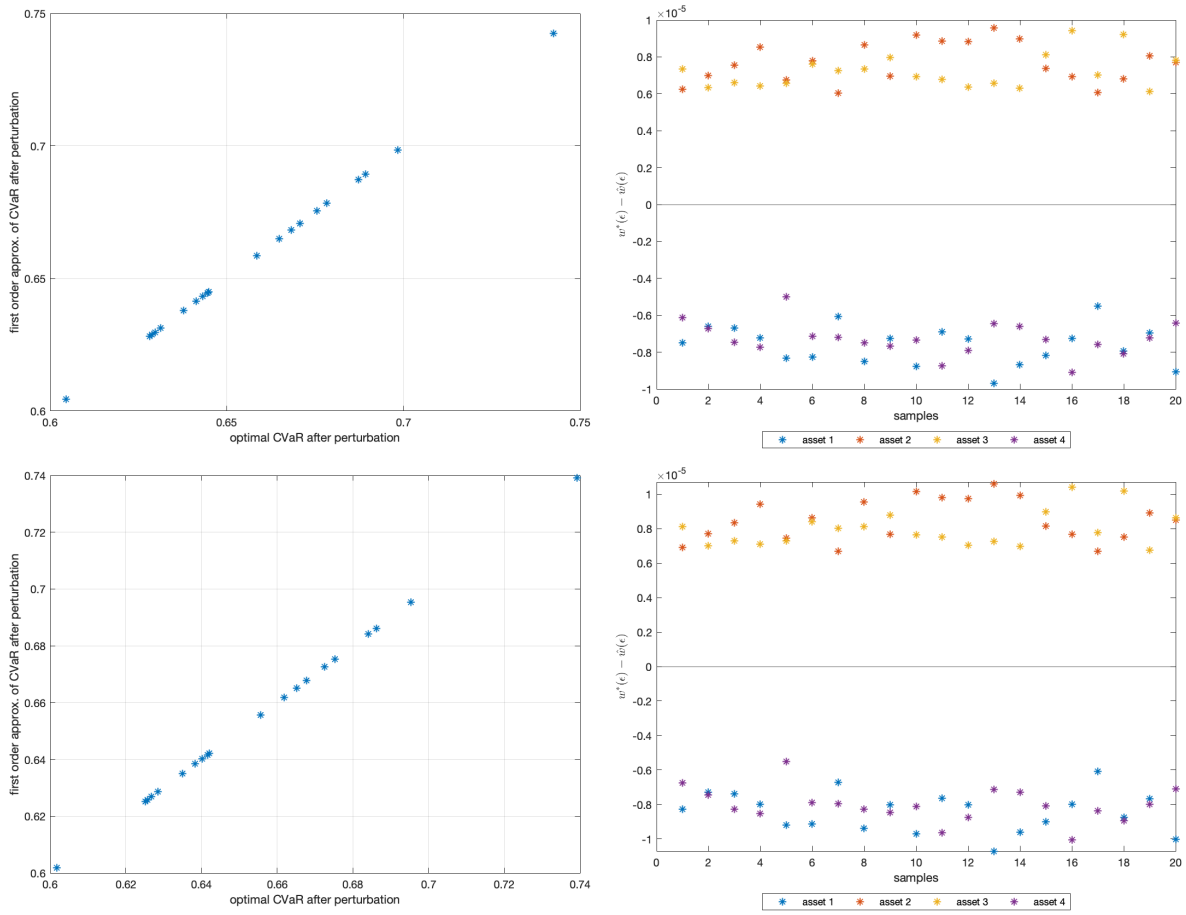
Figure 2.2: first order approximation is good when $\epsilon_1 = \pm 1/10\gamma_1$ and $\epsilon_2 = 0$ (upper $\epsilon_1 = 1/10\gamma_1$, lower $\epsilon_1 = -1/10\gamma_1$)
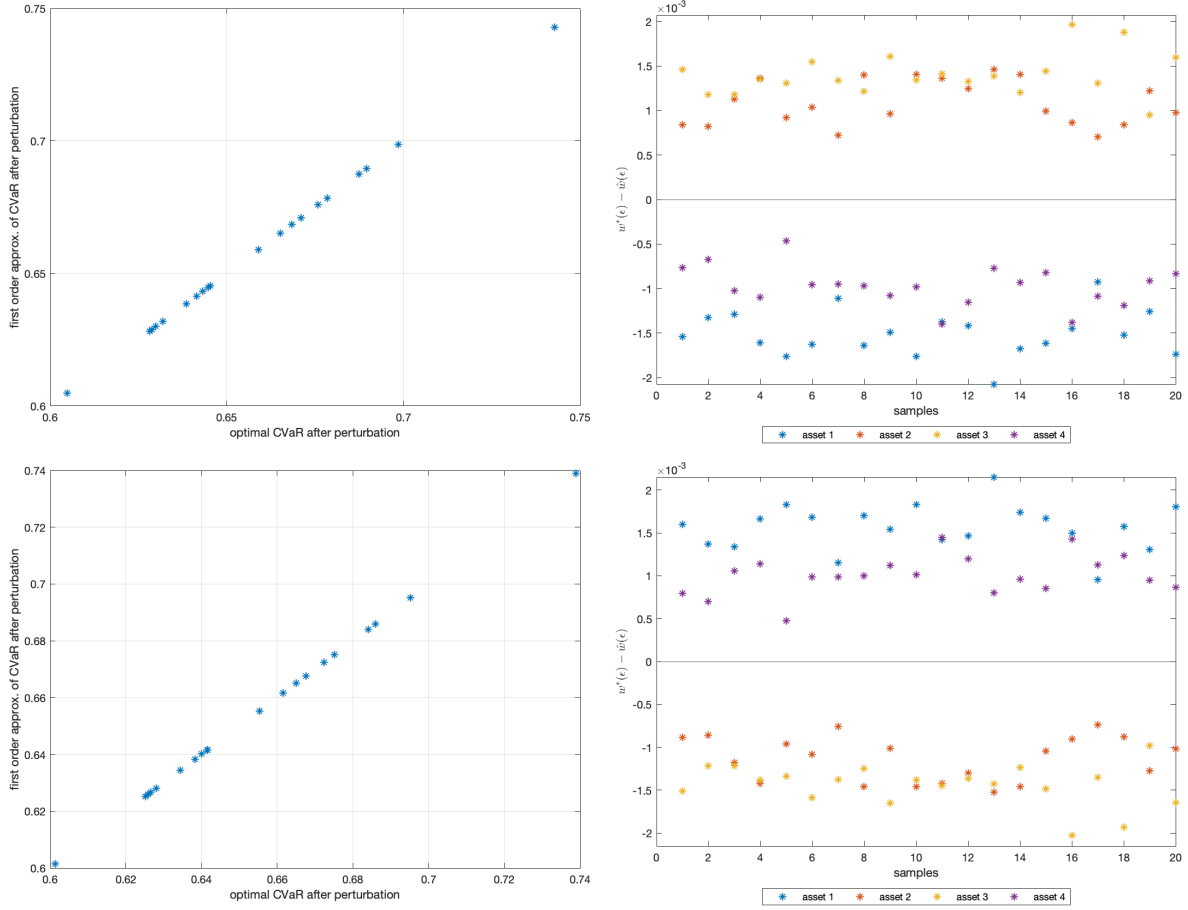
Figure 2.3: first order approximation is not as good when $\epsilon_2 = \pm 1/50\gamma_2$ and $\epsilon_1 = 0$ (upper $\epsilon_2 = 1/50\gamma_2$, lower $\epsilon_2 = -1/50\gamma_2$)

Therefore, in this case, we can trust the first order approximation of $\mathbf{w}^*(\boldsymbol{\epsilon})$ and WCVaR$^*(\boldsymbol{\epsilon})$ within $\pm 10\%$ of γ_1 (we need to find a smaller "trust region" for the first order approximation or have to investigate the higher order approximation for γ_2) and compute the relative change of optimal $\mathbf{w}$,

$$\Delta_1 := \frac{\|\mathbf{w}^*(\boldsymbol{\epsilon}) - \mathbf{w}^*\|_1}{\|\mathbf{w}^*\|_1} = \|\mathbf{w}^*(\boldsymbol{\epsilon}) - \mathbf{w}^*\|_1 \,,^2 \tag{2.2}$$

---

[2]Recall the budget constraint $\|\mathbf{w}^*\|_1 = 1$.

or that of the optimal worst case CVaR,

$$\Delta_2 := \frac{\text{WCVaR}^*(\boldsymbol{\epsilon}) - \text{WCVaR}^*}{\text{WCVaR}^*} \tag{2.3}$$

to see if the output is sensitive to the uncertainty parameters. We will focus on the sensitivity of $\mathbf{w}$ and take (2.2) as a *robustness measure*. Note that Δ_1 can also be approximated by

$$\tilde{\Delta}_1 := \|\hat{\mathbf{w}}(\epsilon) - \mathbf{w}^*\|_1 \approx \Delta_1.$$

2.2.2 Three Methods to Determine Size of Uncertainty Set

Method 1: Find Elbow Point for the L-Curve

By fixing one of (γ_1, γ_2) and changing the other, we can plot the curve of cumulative $\mathbf{w}$ and Δ_1 by solving (P_0) multiple times, see figures 2.4 and 2.5. Note that although the grid of γ_i (i.e. x-axis) is evenly-spaced in the examples, Δ_1 will be normalized. When the grid is not evenly spaced, the uneven space will influence Δ_1 so that normalization $\frac{\Delta_1}{\epsilon_i}$ is necessary. Furthermore, we need to test the consistency about the normalized Δ_1, which means different grids generate the curve with similar shape and elbow point (see an example in figure 2.6).

In statistics as well as ill-conditioned problems in optimization, a curve with a negative first derivative and a unimodal second derivative is often termed an *L-curve* as it typically has an L-shaped graph. The point with maximum *absolute curvature*

$$k = \frac{|y''|}{(1 + y'^2)^{\frac{3}{2}}}$$

is the *elbow point* of the curve. In general, the elbow point is the point with most information preserved in a small space. Figures 2.4b and 2.5b are very similar to an L-curve with some random noise. In fact, the small change around $\gamma_2 = 0.025$ in figure 2.5b is from the asset 3 being added to the portfolio.
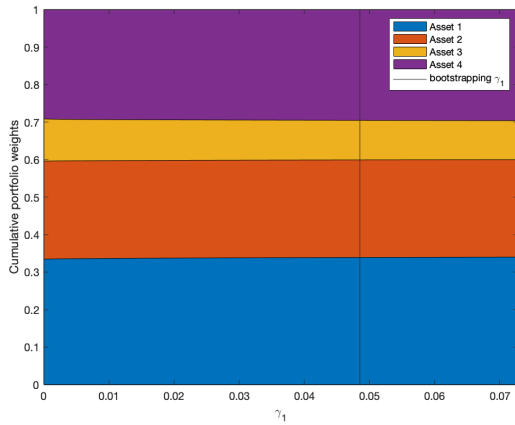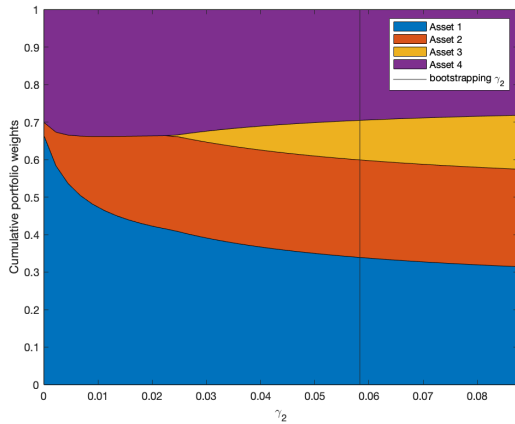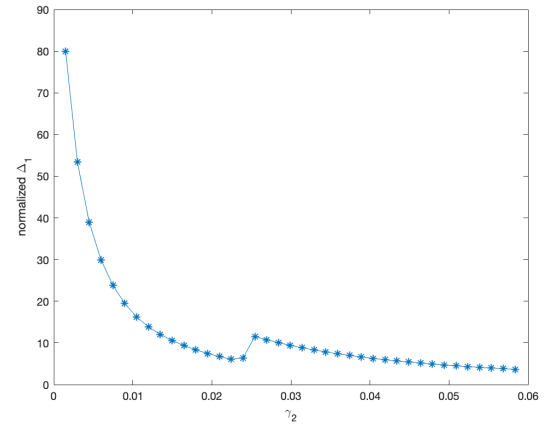
(a) cumulative $\mathbf{w}$ v.s. γ_1
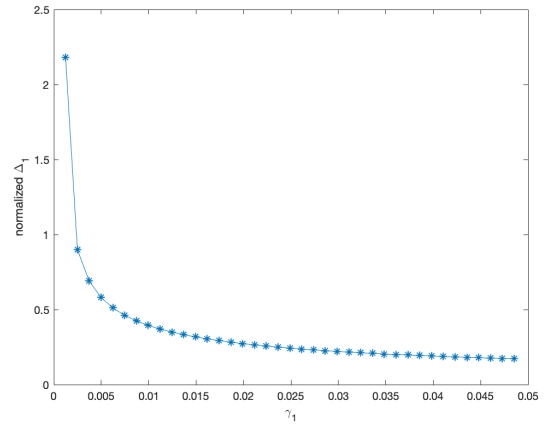
(b) Δ_1/ϵ_1 v.s. γ_1

Figure 2.4: $\mathbf{w}$ and normalized Δ_1 versus γ_1 by solving P_0 40 times, with $\gamma_2 = 0.0584$

(a) cumulative $\mathbf{w}$ v.s. γ_2

(b) Δ_1/ϵ_2 v.s. γ_2

Figure 2.5: $\mathbf{w}$ and normalized Δ_1 versus γ_2 by solving P_0 40 times, with $\gamma_1 = 0.0485$

Therefore, we can choose the elbow points as γ_i because they are conservative enough.[3] The steps for this method are:

---

[3]There are some risk in the calculation of κ: when there are significant difference between the scale of x and y, κ cannot be found properly. So in our algorithm to find the elbow point, we will first adjust the axis: $y_{adj} = y/\max(y) \times \max(x)$.
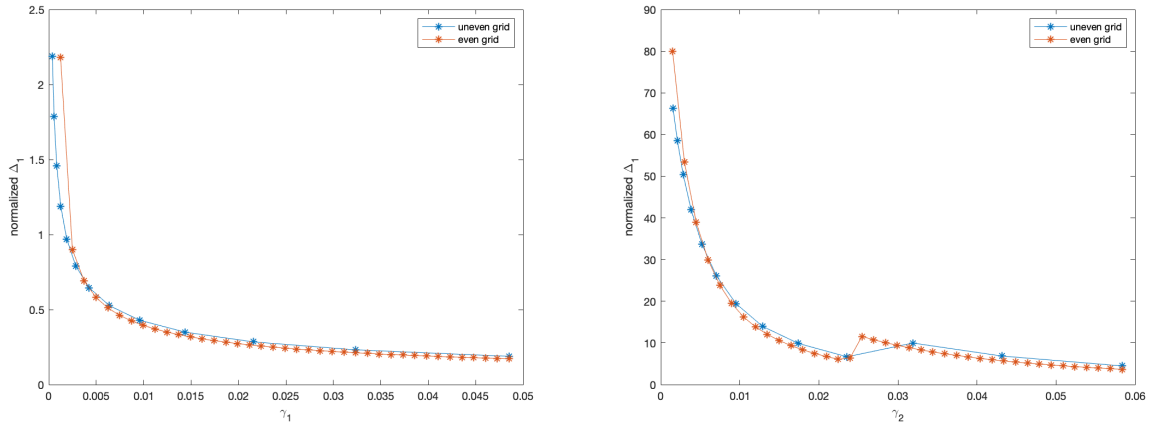
27

Figure 2.6: even and uneven grids will generate similar L-curves

Step 1. Pick an initial pair of $(\gamma_1^{(0)}, \gamma_2^{(0)})$. It can be from the bootstrap method[4] or by subjective judgment.

Step 2. Find grids[5] for γ_i.

Step 3. Repeat the following steps till convergence (i.e. $(\gamma_1^{(r)}, \gamma_2^{(r)}) = (\gamma_1^{(r-1)}, \gamma_2^{(r-1)})$[6]):

  a. Fix $\gamma_2 = \gamma_2^{(r-1)}$. Solve P_0 for each γ_1 on the grid and compute the normalized Δ_1.

  b. Find $\gamma_1^{(r)}$ as the elbow point of the normalized Δ_1 versus γ_1.

  c. Fix $\gamma_1 = \gamma_1^{(r)}$. Solve P_0 for each γ_2 and compute the normalized Δ_1.

  d. Find $\gamma_2^{(r)}$ as the elbow point of the normalized Δ_1 versus γ_1.

Step 4. Take $(\gamma_1^{(r)}, \gamma_2^{(r)})$ as the final pair.

Back to the simulation example, step 3 ends after two loops. Start from the bootstrapping pair $(\gamma_1^{(0)}, \gamma_2^{(0)}) = (0.0485, 0.0584)$ with the grids equally spaced between 0 and $\gamma_i^{(0)}$ by 40 points, the final pair $(\gamma_1^{(2)}, \gamma_2^{(2)}) = (0.0062, 0.0209)$.

---

[4]There is no need for many bootstrapping resamples because we just want to get the sense about how large it should be.

[5]The grids can be either evenly-spaced or unevenly-spaced between 0 and $k\gamma_i^{(0)}$ with $k \geq 1$.

[6]Because we have a discrete grid for γ_i, the equality can be satisfied.

Method 2: Control the Sensitivity with a Threshold

Let $D = (M^*)^{-1}N^*$. We can observe from (1.16) that

$$w_j^*(\boldsymbol{\epsilon}) = w_j^* + D(j,:)\boldsymbol{\epsilon} + o(\|\boldsymbol{\epsilon}\|), \ \ j = 1, \ldots, n.$$

Therefore,

$$D(j,i) = \frac{dw_j}{d\epsilon_i}, \ \ j = 1, \ldots, n; \ \ i = 1, 2.$$

In the meantime,

$$\lim_{\epsilon_j \to 0^+} \frac{\Delta_1}{\epsilon_j} = \lim_{\epsilon_i \to 0^+} \frac{\|\mathbf{w}^*(\boldsymbol{\epsilon}) - \mathbf{w}^*\|_1}{\epsilon_i} = \lim_{\epsilon_i \to 0^+} \frac{1}{\epsilon_i} \sum_{j=1}^{n} |w_j(\boldsymbol{\epsilon}) - w_j|$$

$$= \sum_{j=1}^{n} \lim_{\epsilon_i \to 0^+} \frac{|w_j(\boldsymbol{\epsilon}) - w_j|}{\epsilon_i} = \sum_{j=1}^{n} \left|\frac{dw_j}{d\epsilon_i}\right|, \ \ i = 1, 2.$$

So,

$$\|D(1:n,i)\|_1 = \lim_{\epsilon_i \to 0^+} \frac{\Delta_1}{\epsilon_i}, \ \ i = 1, 2. \tag{2.4}$$

Let $d_1 = \|D(1:n,1)\|_1$ and $d_2 = \|D(1:n,2)\|_1$ If, as we observed from figures 2.4b and 2.5b that the normalized Δ_1 has a decreasing trend, it will be safe to use the following procedure to pick (γ_1, γ_2):

Step 1. Decide a termination condition. For example, $d_i < L$, $i = 1, 2$ for some small threshold L.

Step 2. Decide a *magnification factor, f* > 1.

Step 3. Start from a small pair of $(\gamma_1^{(0)}, \gamma_2^{(0)})$.

Step 4. Repeat the following steps till the termination condition is satisfied:

    a. Solve P_0 with the pair $(\gamma_1^{(r)}, \gamma_2^{(r)})$ and compute d_1 and d_2.

    b. For $i = 1, 2$, if $d_i \geq L$, update $\gamma_i^{(r+1)} = f \times \gamma_i^{(r)}$. Otherwise, keep $\gamma_i^{(r+1)} = \gamma_i^{(r)}$.

Compared with Method 1, this method doesn't need to compute all the optimal solution on the grid but take L as good threshold and believe that $\mathbf{w}^*$ is stable enough. In the algorithm, it updates γ_1 and γ_2 at the same time.

The biggest challenge for this method is how to pick L and f. Smaller L and f will increase the computation. L can also have an important impact on the search route as well as on the final pair. In intuition, d_i will tend to 0 (L can be very small) when the uncertainty set gets larger (i.e. both γ_i are larger). This is because when the uncertainty set is large enough, the worst case $\mathbf{w}$ will be so bad[7] that, further enlarging the uncertainty set will not make the result worse much. In addition, when the number of assets increase, L should be smaller.

Figures 2.7 and 2.8 in pages 31, 32 show the surfaces of d_1 and d_2 versus γ_1 and γ_2. Both d_i tend to decrease when γ_i are larger. We also show the search routes from $0.1\gamma_i^{(0)}$ with $f = 1.25$ and $L = 0.5, 0.8$ on the surface. $(\gamma_1^{(0)}, \gamma_2^{(0)}) = (0.0485, 0.0584)$ is the bootstrapping pair. The final pairs are $(0.0095, 0.0849)$ for $L = 0.5$ and $(0.0049, 0.0679)$ for $L = 0.8$.

---

[7]The worst of all cases is to put all the money in the asset with lowest risk and expected return, as long as it can satisfy the expected return constraint.
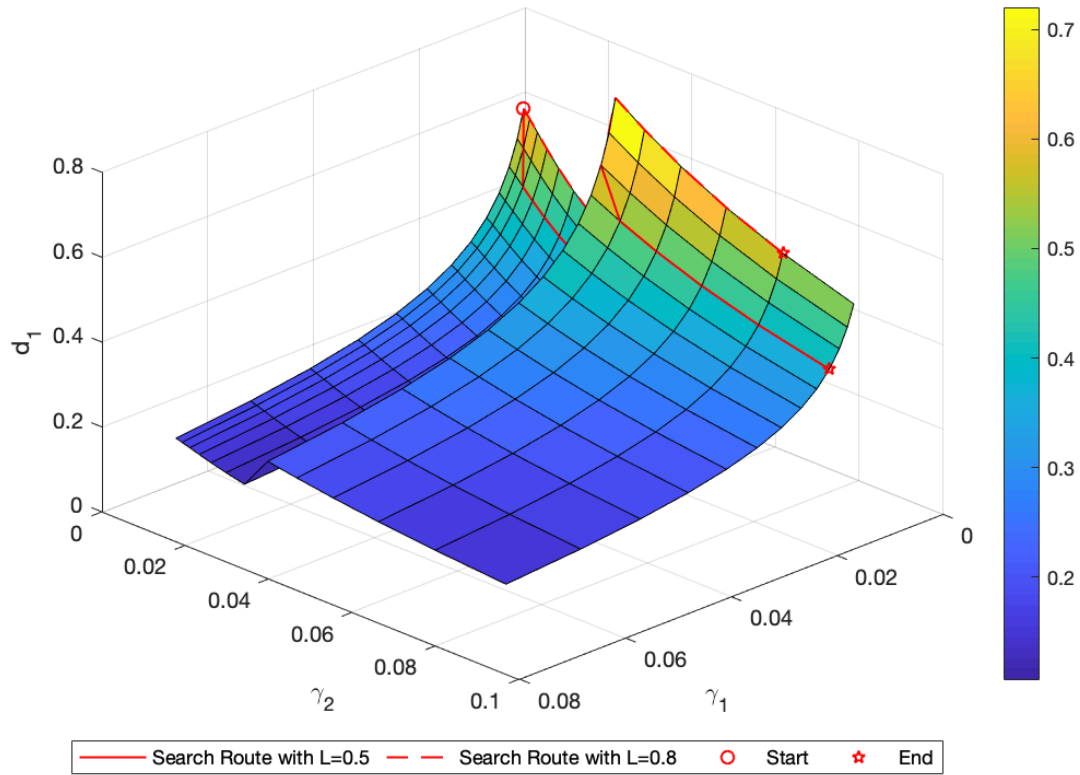
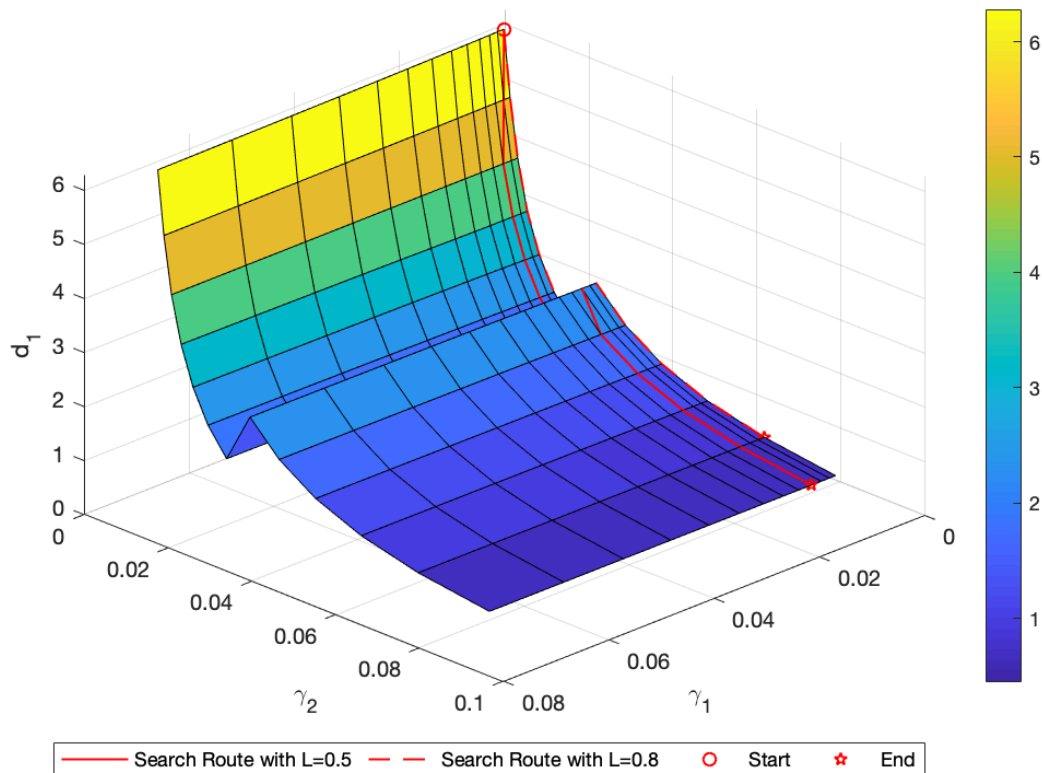Figure 2.7: d_1 versus γ_1 and γ_2 with search routes L=0.5 & L=0.8

Figure 2.8: d_2 versus γ_1 and γ_2 with search routes L=0.5 & L=0.8

Method 3: Find Elbow Point with the First Order Derivative

With the relationship of the first order derivative and normalized Δ_1 in (2.4), mixing with the idea of elbow points and L-curves in Method 1, we can alternatively find the elbow points of d_1 and d_2. The procedure is as follows:

Step 1. Pick an initial pair of $(\gamma_1^{(0)}, \gamma_2^{(0)})$.

Step 2. Find grids for γ_i.

Step 3. Repeat the following steps till convergence (i.e. $(\gamma_1^{(r)}, \gamma_2^{(r)}) = (\gamma_1^{(r-1)}, \gamma_2^{(r-1)})$):

a. Fix $\gamma_2 = \gamma_2^{(r-1)}$. Solve P_0 for each γ_1 on the grid and compute d_1.

b. Find $\gamma_1^{(r)}$ as the elbow point of d_1 versus γ_1.

c. Fix $\gamma_1 = \gamma_1^{(r)}$. Solve P_0 for each γ_2 and compute d_2.

d. Find $\gamma_2^{(r)}$ as the elbow point of d_2 versus γ_1.

Step 4. Take $(\gamma_1^{(r)}, \gamma_2^{(r)})$ as the final pair.

In our example, step 3 ends after two loops. Start from the bootstrapping pair $(\gamma_1^{(0)}, \gamma_2^{(0)}) = (0.0485, 0.0584)$ with the grids equally spaced between 0 and $\gamma_i^{(0)}$ by 40 points, the final pair $(\gamma_1^{(2)}, \gamma_2^{(2)}) = (0.0087, 0.0224)$.

This method is much more compute-intensive than method 2. It is also slightly more compute-intensive than method 1 because in addition to solve P_0 multiple times till convergence, it needs to calculate M^* and N^* for d_i.

2.3 Robust Mean-CVaR Problem

Recall the robust mean-CVaR problem in (P_0). In this section, we will analyze the sensitivity analysis results for the robust mean-CVaR problem. $\mu_0 = 0.05$ and $\beta = 0.95$ are set.

2.3.1 Performance of the First Order Approximation

Using the same method and scale of $\epsilon_1 = 1/10\gamma_1$, $\epsilon_2 = 1/50\gamma_2$ in section 2.2.1, we plot the figures 2.9 and 2.10. Note that the perturbed problem is infeasible for some samples (e.g. sample 5, 11, 16, 20 when $\epsilon_1 = 1/10\gamma_1$ and $\epsilon_2 = 0$).

The performance of the first order approximation in the mean-CVaR problem is not so good as that of global minimum risk problem in the last section. The additional target expected return constraint introduces a more complicated shape to solution space for the problem.
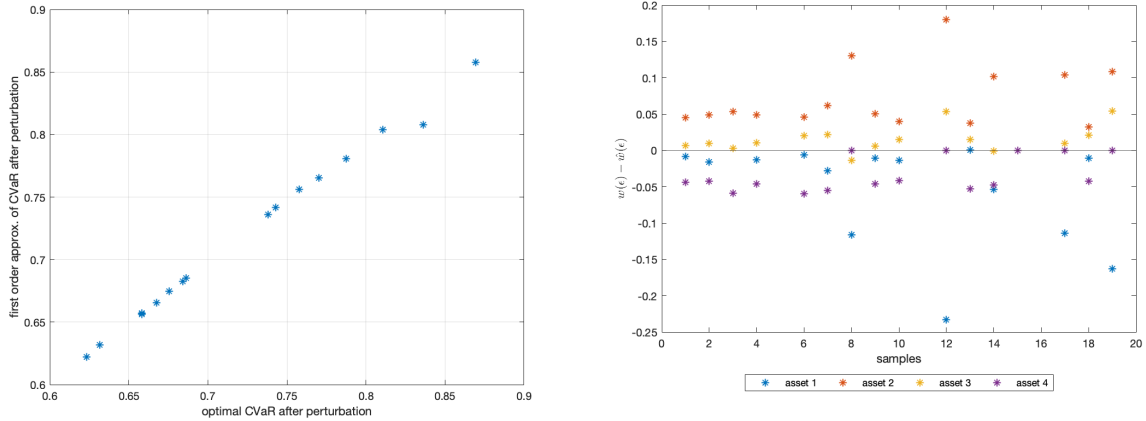
Figure 2.9: first order approximation is bad when $\epsilon_1 = 1/10\gamma_1$ and $\epsilon_2 = 0$
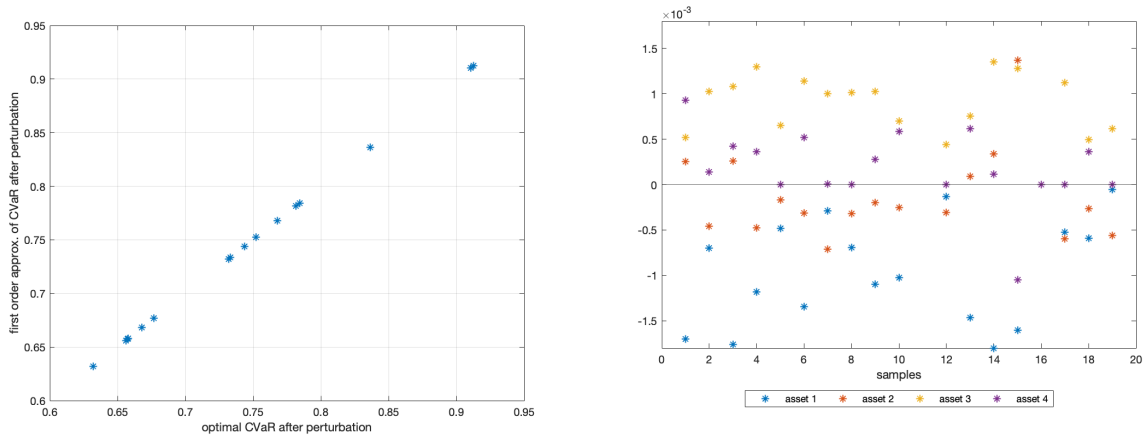(Similar result when $\epsilon_1 = -1/10\gamma_1$)

Figure 2.10: first order approximation is not good when $\epsilon_2 = 1/50\gamma_2$ and $\epsilon_1 = 0$
(Similar result when $\epsilon_2 = -1/50\gamma_2$)

2.3.2 No L-curve for Mean-CVaR Problem

In section 2.2.2, we have observed that the normalized Δ_1 (or d_i) versus γ_i is similar to an L-curve. Based on the observation, we have proposed three methods to pick γ_i. Therefore, we need to test whether the L-shape preserves with the target expected return constraint.

In figure 2.11 in page 35, we plot Δ_1 versus one of γ_i with the other γ_i fixed. We can

34

see that when γ_i is small, there are several small points, indicating the similarity of the optimal strategy, $\mathbf{w}^*$. The target expected return constraint is inactive for the first two γ_1 while the constraint is always active for the γ_2 in the plot. In figure 2.12 in page 36, we plot d_i versus the pair of (γ_1, γ_2) and show each three-dimension plot in two directions. The shape of the surfaces are not as smooth as in figures 2.7 and 2.8 in pages 31, 32. The range of d_i (i.e. d_1 as large as 60) also shows the instability of the strategy, $\mathbf{w}^*$.

Both figures indicate the failure of L-shape for the mean-CVaR problem so the methods we proposed in section 2.2.2 are no longer appropriate. We will leave the reason for the observation as an open question.
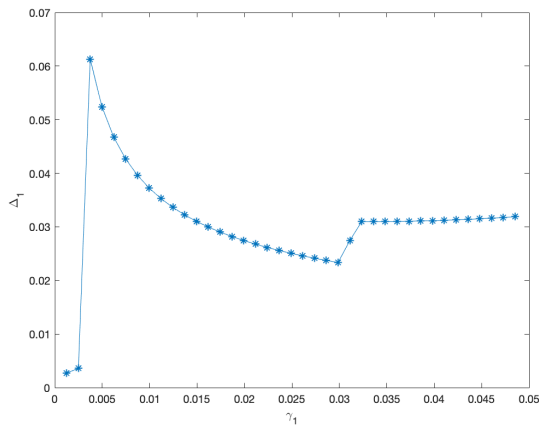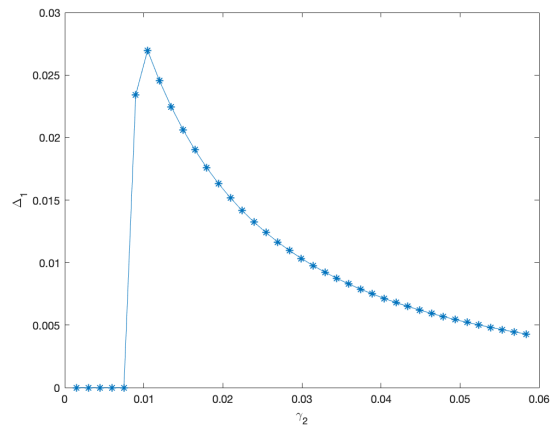
(a) Δ_1 versus γ_1 with $\gamma_2 = 0.0584$ (b) Δ_1 versus γ_2 with $\gamma_1 = 0.0485$

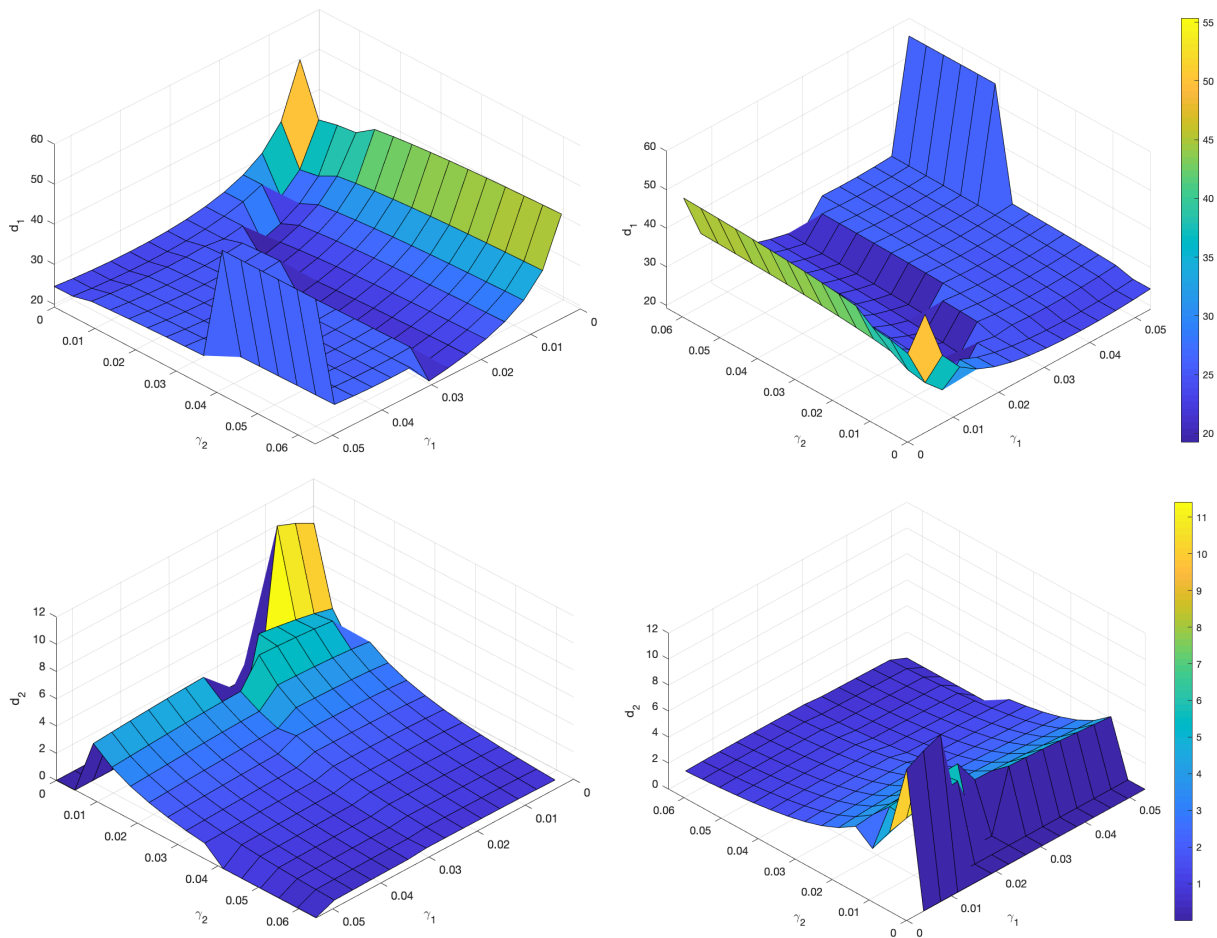Figure 2.11: There is no L-curve for mean-CVaR problem

35

Figure 2.12: The surfaces of d_1 and d_2 are not L-shape

Although we don't observe the L-curve for the robustness measure, the sensitivity analysis result can still be applied to decide whether a given pair of (γ_1, γ_2) is robust enough. Recall the interpretation of d_i. It shows the relative change of $\mathbf{w}^*$ measured by 1-norm when γ_i is perturbed by a unit of small noise. In our example with given $(\gamma_1, \gamma_2) = (0.0485, 0.0584)$, $d_1 = 26.4814$, $d_2 = 0.6456$. As d_2 is smaller than 1, $\gamma_2 = 0.0584$ seems to be appropriate. However, the large d_1 indicates the inappropriateness of γ_1. One possible re-selection is $\gamma_1 = 0.0243$ where d_1 is the smallest when $\gamma_2 = 0.0584$ is fixed.

2.4 Comparison of Efficient Frontiers and Portfolio Composition

Below is a summary of γ_i selected by multiple methods for the global minimum risk problem in section 2.2. The last column shows the true values as in (2.1).

γ_i	Bootstrapping	M1: Elbow	M2: $L = 0.5$	M2: $L = 0.8$	M3: Elbow	True
γ_1	0.0485	0.0062	0.0095	0.0049	0.0087	0.0238
γ_2	0.0584	0.0209	0.0849	0.0679	0.0224	0.0400

Table 1: Summary of γ_i in the simulation analysis for different methods

Although the pairs of (γ_1, γ_2) are selected for the global minimum risk problem, it is interesting to use them to solve the mean-CVaR problem and compare their performance.

The portfolio cumulative compositions (cumulative composition versus expected return) under each pair of γ_i are shown in figure 2.13 in page 38.

For a given $\mathbf{w}$ and a given pair of (γ_1, γ_2), the worst case $\boldsymbol{\mu}_{wc}$ and Σ_{wc} are in (1.19). Accordingly, the worst case CVaR and expected return (WR) can be computed by:

$$\text{WCVaR}_{\mathbf{w}} = -\boldsymbol{\mu}_{wc}^T \mathbf{w} + \kappa \sqrt{\mathbf{w}\Sigma_{wc}\mathbf{w}}, \ \text{WR}_{\mathbf{w}} = \boldsymbol{\mu}_{wc}^T \mathbf{w}.$$

Therefore, taking the true γ_i in table 1, as well as all the $\mathbf{w}$ on each efficient frontier (with different γ_i), we can compare the frontiers in figure 2.14 in page 39. We can see that each frontier has an approximately identical shape to the efficient frontier (blue in figure 2.14), which supports our use of γ_i by each method (for the global minimum risk problem and even for mean-CVaR problem). The frontiers of portfolios got from (SMC) and (RMC-$\mathcal{U}_0$) are also graphed as a reference.
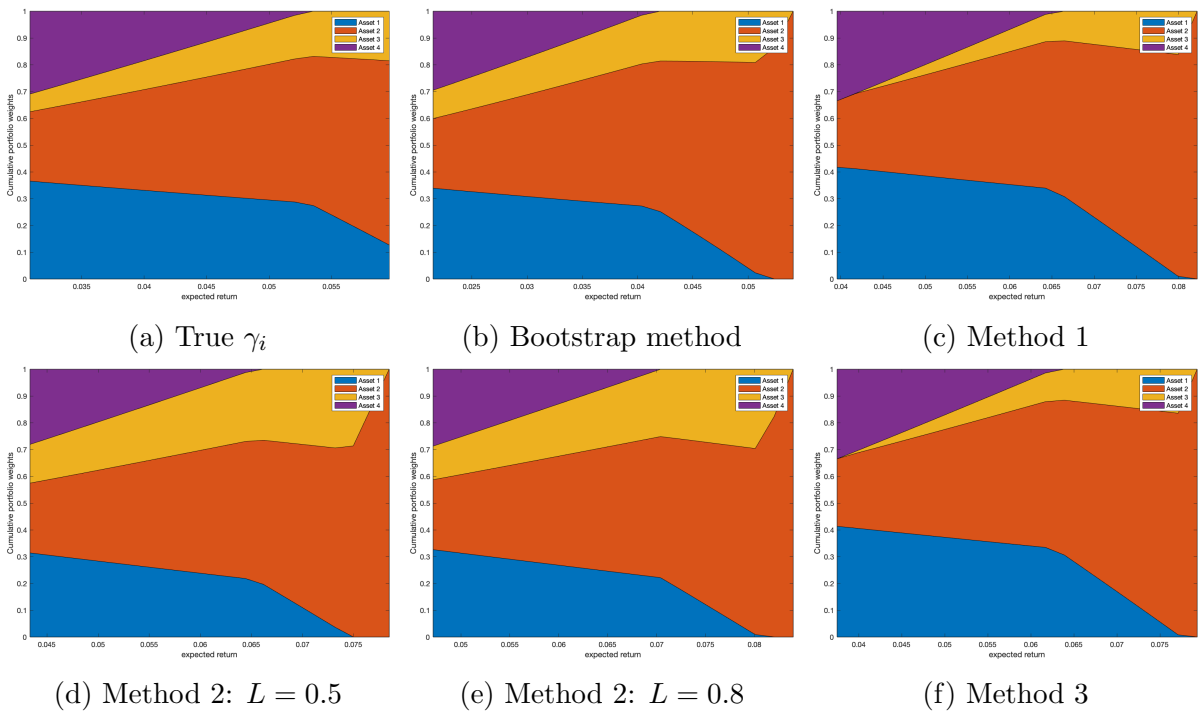
(a) True γ_i (b) Bootstrap method   (c) Method 1

(d) Method 2: $L = 0.5$ (e) Method 2: $L = 0.8$ (f) Method 3

Figure 2.13: Composition of portfolios on the efficient frontiers under each (γ_1, γ_2)
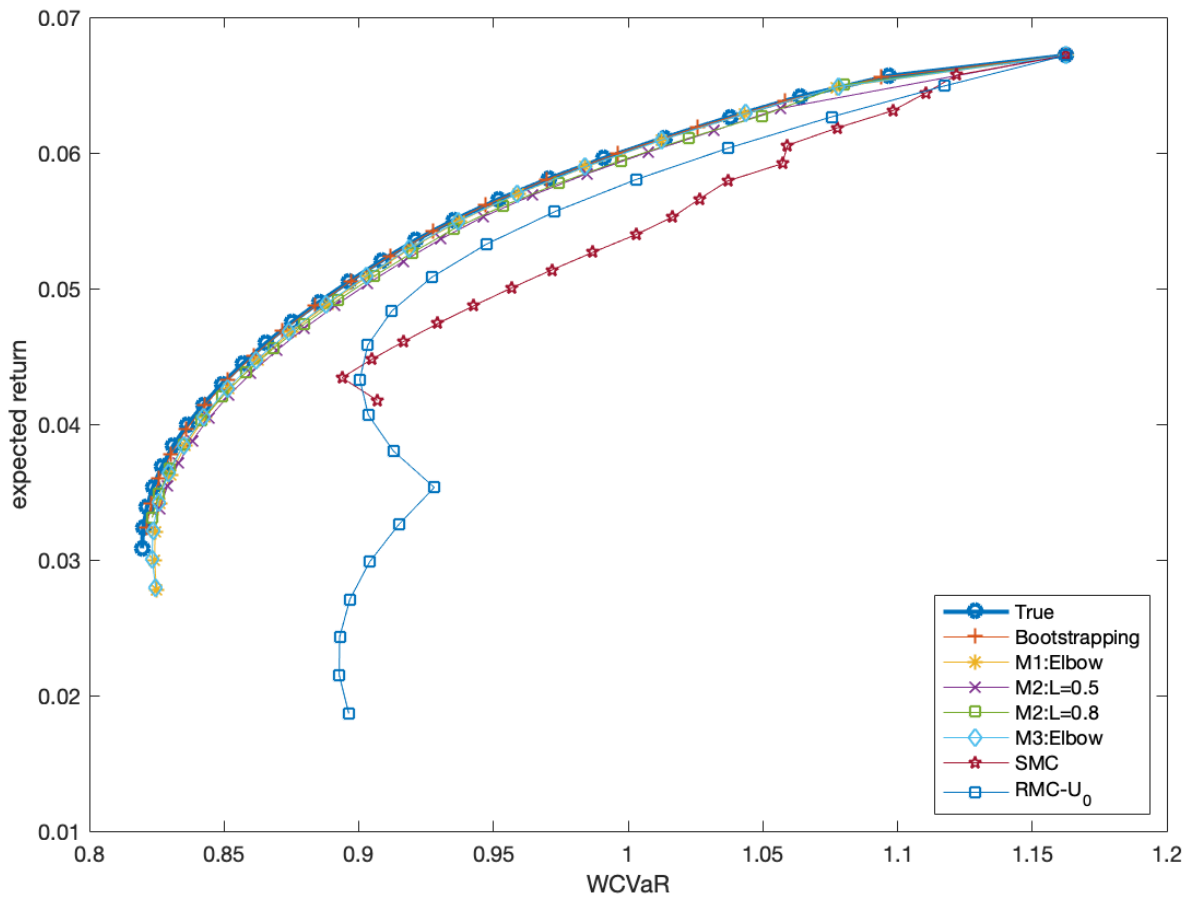
Figure 2.14: Mean-CVaR Frontiers under true $(\gamma_1, \gamma_2) = (0.0238, 0.0400)$

Chapter 3

Empirical Analysis

This chapter reports the empirical study of the robust portfolios based on real data. We use the moving window analysis to evaluate the *out-of-sample performance* of different inputs in the model.

3.1 Data and Methodology

The real data used here is the Credit Suisse Hedge Fund Index comprised by 12 subsets in table 1. The sample period is from January 2000 to October 2019. The *return on revenue, ROR* is given on a monthly basis. Figure 3.1 shows a histogram of Credit Suisse Convertible Arbitrage Hedge Fund Index ROR during this period. The ROR observations below -10% are in September and October 2008 during the financial crisis period.

The analysis picks the period of January 2000 to December 2014 as the initial training set with 180 observations and rebalances the portfolio once a month. We use a *moving window* of 180 observations, which means, when the data is realized after a month, it will be added into the training data set and the oldest one will be removed. Accordingly, the *investment window* is from January 2015 to October 2019. We have to point out that the moving training window always contains the extremely high loss period during the 2008 financial crisis. Therefore, the problem is how to make a robust investment decision while some extreme data has been included in the training dataset. In other words, the investment strategy should avoid being too conservative.

To calculate the sample mean and covariance matrix, we use the time-weighted estimates with $p_{[i]}$ in (1.6).

1	Credit Suisse Convertible Arbitrage Hedge Fund Index
2	Credit Suisse Emerging Markets Hedge Fund Index
3	Credit Suisse Equity Market Neutral Hedge Fund Index
4	Credit Suisse Event Driven Hedge Fund Index
5	Credit Suisse Event Driven Distressed Hedge Fund Index
6	Credit Suisse Event Driven Multi-Strategy Hedge Fund Index
7	Credit Suisse Event Driven Risk Arbitrage Hedge Fund Index
8	Credit Suisse Fixed Income Arbitrage Hedge Fund Index
9	Credit Suisse Global Macro Hedge Fund Index
10	Credit Suisse Long/Short Equity Hedge Fund Index
11	Credit Suisse Managed Futures Hedge Fund Index
12	Credit Suisse Multi-Strategy Hedge Fund Index

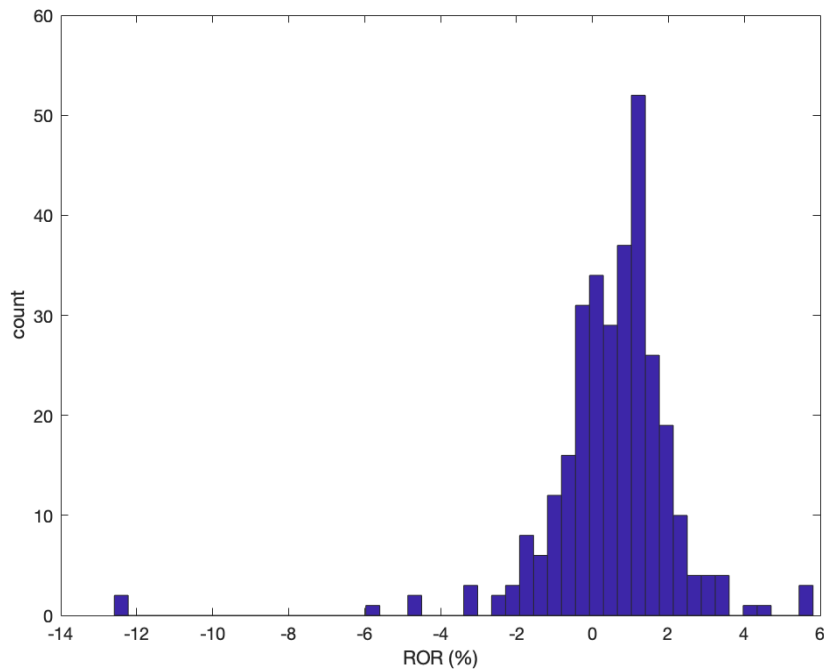Table 1: Subsets of Credit Suisse Hedge Fund Index

Figure 3.1: Histogram of Credit Suisse Convertible Arbitrage Hedge Fund Index monthly ROR during January 2000 to October 2019

To mostly take advantage of our methods to choose different (γ_1, γ_2) but save the computation, we re-calculate it once a year.

We use bootstrapping method with 5000 resamples, Method 1, Method 2 with $L = 0.1$ and $L = 0.3$ to calculate γ_i. For Method 1, 2 and 3, $(\gamma_1^{(0)}, \gamma_2^{(0)})$ is taken as bootstrapping γ_i from each moving window for re-calculation. The grids for both Method 1 and Method 3 are evenly spaced between 0 and $1.1\gamma_i^{(0)}$ by 20 points. We decide the magnification factor $f = 1.25$ for Method 2. In addition to the robust problem, we also solve the non-robust sampled base problem as in (SMC) as a reference. We use the time-weighted estimates with $p_{[i]}$ in (1.6) for both the objective function and the expected return.

For each method, we compute the *accumulative wealth*, W_i in the investment window with initial wealth $W_0 = 1$ by the equation:

$$W_{i+1} = W_i \times [(\mathbf{1}_n + \text{ROR}_{i+1})^T \mathbf{w}_{i+1}^*],$$

where ROR_{i+1} is the vector of recently realized return on revenue for according assets. We also compute the *actual ROR of the portfolio, $ROR_{P,i+1}$* by:

$$\text{ROR}_{\text{P},i+1} = W_{i+1}/W_i - 1 = \text{ROR}_{i+1}{}^T \mathbf{w}_{i+1}^*.$$

The variance of $\{\text{ROR}_{\text{P},i}\}$ will be reported as an indicator for the fluctuation of actual ROR during the investment period.

3.2   Global Minimum Risk Portfolio

In this section, we solve for global minimum risk portfolios using different methods. Recall the global minimum risk problem is (P_0) without the target expected return constraint:

$$\text{WCVaR}_{\beta}^*(\gamma_1, \gamma_2) = \min_{\mathbf{w} \in \mathbb{R}^n} \quad -\hat{\boldsymbol{\mu}}^T \mathbf{w} + \kappa \left\| (\hat{\Sigma} + \gamma_2 I_n)^{1/2} \mathbf{w} \right\|_2 + \sqrt{\gamma_1} \left\| \hat{\Sigma}^{1/2} \mathbf{w} \right\|_2$$
$$\text{s.t.} \quad \mathbf{1}_n^T \mathbf{w} = 1, \, \mathbf{w} \geq \mathbf{0}.$$

The pairs of (γ_1, γ_2) from different methods, re-calculated once a year, are shown in table 2. It is shown that for γ_1, the method 2 with $L = 0.01$ generates extremely large values compared to the other methods. The sensitivity of γ_i to L indicates the importance as well as the difficulty to choose an appropriate L. For γ_2, the bootstrapping values are the largest.

Figure 3.2 shows the accumulative wealth by different methods. In addition to the robust problems (P_0), we also solve (SMC) without the expected return constraint. We

γ_i	Method	Year				
		2015	2016	2017	2018	2019
γ_1	Bootstrapping	0.3329	0.2973	0.2036	0.2223	0.2074
	M1: Elbow	0.0964	0.0861	0.0589	0.0644	0.0600
	M2: $L = 0.01$	8.9642	12.5079	10.7072	11.6931	10.9090
	M2: $L = 0.05$	0.6160	0.5501	0.5886	0.4114	0.4798
	M3: Elbow	0.0964	0.3098	0.1179	0.0644	0.0600
γ_2	Bootstrapping	19.2412	19.8333	21.1493	22.7042	24.7905
	M1: Elbow	2.2279	9.1860	7.3466	6.5723	5.7410
	M2: $L = 0.01$	9.3332	9.6204	10.2587	8.8104	9.6200
	M2: $L = 0.05$	1.5659	1.6140	1.3769	1.4781	1.6140
	M3: Elbow	2.2279	2.2965	3.6733	6.5723	5.7410

Table 2: 5 year γ_i from different methods

can see that the method 2 with $L = 0.05$ has the overall best performance. Note that it also has the smallest γ_2 but the second largest γ_1 of all years. Specifically, in figure 3.3 we plot the actual ROR of each portfolio for the bootstrap method and method 2 with $L = 0.05$, which has the overall best and worst performance concerning the final accumulative wealth in figure 3.2. In addition, we also plot the actual ROR from the method 2 with $L = 0.01$ as it has the largest γ_1 and from (SMC).
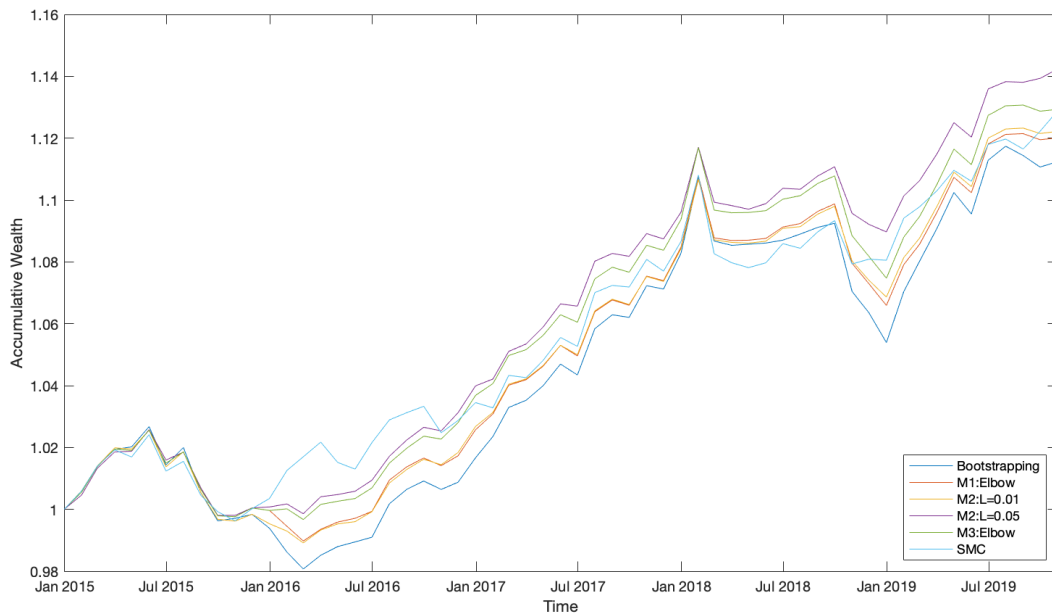
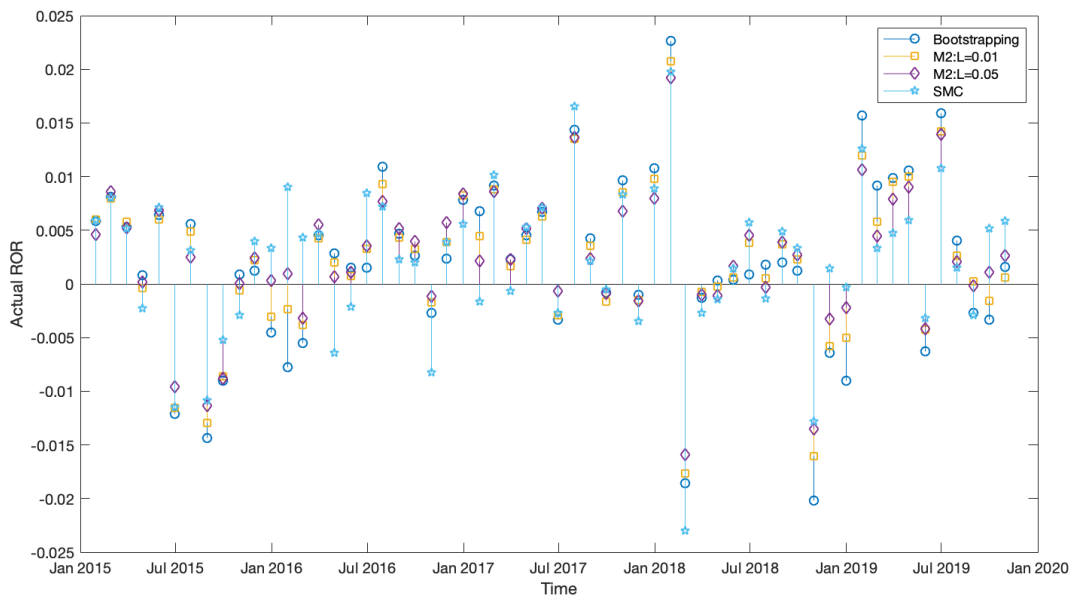Figure 3.2: Accumulative wealth by different methods

Figure 3.3: Actual ROR by different methods

44

In figure 3.3, we can see that method 2 with $L = 0.05$, which has the overall highest accumulative wealth, is not the one with the highest $\text{ROR}_{\text{P},i}$ but the one with the most stable $\text{ROR}_{\text{P},i}$. The variance of its actual ROR is 3.9737×10^{-5}. The most fluctuate ROR is by the bootstrap method with variance of 6.5163×10^{-5}. The variance of the ROR is reported in table 3.

Method	Bootstrapping	M1: Elbow	M2: $L = 0.01$	M2: $L = 0.05$	M3: Elbow	SMC
Variance $(\times 10^{-5})$	6.5163	5.1847	4.9617	3.9737	4.9548	5.0626

Table 3: Variance of actual ROR by different methods

3.3 Robust Mean-CVaR Portfolio

Recall the global maximum expected return problem in (1.12). It can be reformulated as:

$$\text{WR}^*(\gamma_1) = \max_{\mathbf{w} \in \mathbb{R}^n} \quad \hat{\boldsymbol{\mu}}^T \mathbf{w} - \sqrt{\gamma_1} \left\| \hat{\Sigma}^{1/2} \mathbf{w} \right\|_2$$
$$\text{s.t.} \quad \mathbf{1}_n^T \mathbf{w} = 1, \ \mathbf{w} \geq \mathbf{0}.$$

By solving the global maximum expected return portfolio, we get an upper bound of feasible μ_0 for (P_0) with bootstrapping γ_i. We decide a target expected return, $\mu_0 = -0.2\%$, which means, under the worst case characterized by (γ_1, γ_2) in (P_0), the expected loss is less than 0.2%. Although there is no L-curve for the mean-CVaR problem, the frontiers in figure 2.14 in page 39 support γ_i picked for the global minimum risk problem to be used in the mean-CVaR problem. Therefore, we use the γ_i in table 2. As in the last section, the result from the non-robust problem (SMC) will be presented as a reference.
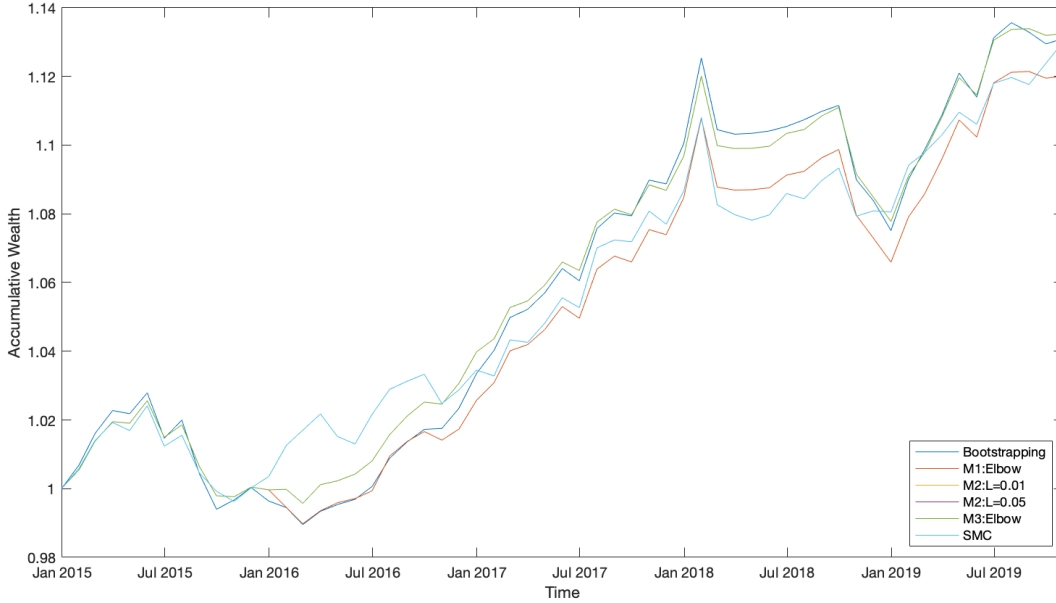
Figure 3.4: Accumulative wealth by different methods with $\mu = -0.2\%$

In figure 3.4, the accumulative wealth with different γ_i is shown. Method 2 with both L is infeasible with the expected return constraint. Method 1 with $\mu = -0.2\%$ has the expected return constraint always inactive, which means they are the same as the global minimum risk portfolio. The bootstrap method has a trend very similar to method 3, with the final accumulative wealth 1.1311 and 1.1325, respectively. The variance of actual ROR is 6.2693×10^{-5}, 4.9861×10^{-5}, respectively, indicating a more stable investment strategy by method 3. The non-robust problem has a very different pattern compared with the other lines in the plot. It outperforms the robust problem significantly during the first half year of 2016 but the performance gets worse till the end of year 2018. Its final accumulative wealth is 1.1303. The actual ROR is shown in figure 3.5. The variance of the ROR is reported in table 4.
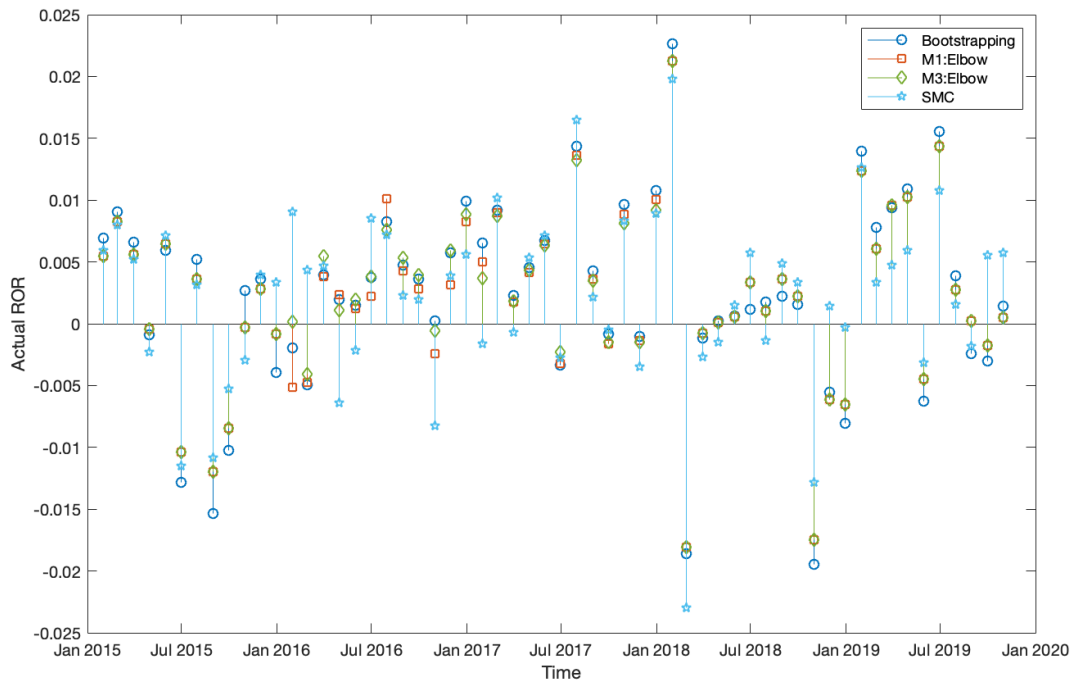
Figure 3.5: Actual ROR by different methods with $\mu = -0.2\%$

	Bootstrapping	M1: Elbow	M3: Elbow	SMC
Variance($\times 10^{-5}$)	6.2693	5.1847	4.9861	5.0491

Table 4: Variance of actual ROR by different methods with $\mu = -0.2\%$

Chapter 4

Conclusion and Future Work

4.1 Conclusion

In this report, we illustrated sensitivity analysis results in the robust mean-CVaR optimization problem in order to improve the size parameters for uncertainty sets in the prospect of the underlying problem but not only the data. The robustness measures are proposed to quantify the robustness of the results directly. They have a closed relationship with the first order derivative resulting from the sensitivity analysis.

Especially, we utilized the result of perturbation result towards size parameters of uncertainty set and observed the L-shape under the global minimum CVaR problem. As a result of the L-curve, we proposed three methods to select size parameters. The three methods include finding the elbow point for the L-curve with respect to the computed robustness measure and to the first order derivative of the optimal solution, as well as controlling the sensitivity with a threshold.

4.2 Future Work

One future work which naturally follows from here is to understand the L-shape we observed for the global minimum risk problem but not for the mean-CVaR problem. Another direction is to better determine the size parameter by utilizing the sensitivity analysis towards the original parameters as in section 1.3.1 and the direct inputs to the problem.

APPENDICES

Appendix A

Additional Notation

Notation A.0.1. *For symbolic consistency, bold lower case letters (e.g. $\mathbf{w}$) are (non-random) vectors, bold upper case letters (e.g. $\mathbf{X}$) are random vectors, bold upper case letters with square bracket subscripts (e.g. $\mathbf{X}_{[i]}$) are realized data of random vectors, non-bold upper case letters (e.g. X_1) are random variables or matrices without further notations. I_n is the $n \times n$ identity matrix. $\mathbf{0}_n$ is the n-dimensional zero vector. $\mathbf{0}_{m \times n}$ is the $m \times n$ zero matrix. $\mathcal{S}^n$ is the space of $n \times n$ symmetric metrics and $\mathcal{S}_+^n$ ($\mathcal{S}_{++}^n$) is the cone of $n \times n$ symmetric positive semidefinite (positive definite, respectively) matrices, For two matrices $A, B \in \mathcal{S}^n$, we let $A \preceq B$ or $B - A \succeq 0$ refer to $B - A \in \mathcal{S}_+^n$ (similarly for $\succ 0$ and $\mathcal{S}_{++}^n$).*

Notation A.0.2. *The submatrix notation in the following will be in consistence with MAT-LAB.*

Appendix B

Calculation of the Second Order Derivative $\nabla^2_{\boldsymbol{\epsilon}\mathbf{w}}\|(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2$

First, by differentiating the quadratic form in $\mathbf{w} \in \mathbb{R}^n$, we note that the gradient is

$$\nabla_{\mathbf{w}}\|(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2 = \frac{(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}}{\|(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2}. \tag{B.1}$$

Now for $\boldsymbol{\epsilon} \in \mathbb{R}^{n+t(n)}$, since this is only a function of $\boldsymbol{\epsilon}_2$, we see that the gradient is

$$\begin{aligned}
\nabla_{\boldsymbol{\epsilon}_2}\|(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2 &= \nabla_{\boldsymbol{\epsilon}_2}\|L^T\mathbf{w}+\mathrm{LMat}(\boldsymbol{\epsilon}_2)^T\mathbf{w}\|_2 \\
&= \nabla_{\boldsymbol{\epsilon}_2}\|\mathbf{b_w}+\mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2)\|_2 \\
&= \frac{1}{2\|\mathbf{b_w}+\mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2)\|_2}\nabla_{\boldsymbol{\epsilon}_2}\langle \mathbf{b_w}+\mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2), \mathbf{b_w}+\mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2)\rangle \\
&= \frac{1}{2\|\mathbf{b_w}+\mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2)\|_2}\nabla_{\boldsymbol{\epsilon}_2}(\langle \mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2), \mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2)\rangle + 2\langle \mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2), \mathbf{b_w}\rangle),
\end{aligned} \tag{B.2}$$

where $\mathbf{b_w} = L^T\mathbf{w}$ and $\mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2) = \mathrm{LMat}(\boldsymbol{\epsilon}_2)^T\mathbf{w} : \mathbb{R}^{t(n)} \to \mathbb{R}^n$ is a linear transformation as $\mathrm{LMat}(\boldsymbol{\epsilon}_2))$ is a linear transformation. We note that the partial with respect to $\boldsymbol{\epsilon}_1$ is $\mathbf{0}_n$, so that:

$$\begin{aligned}
\nabla_{\boldsymbol{\epsilon}}\|(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2 &= \begin{bmatrix} \mathbf{0}_n \\ \nabla_{\boldsymbol{\epsilon}_2}\|(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0}_{t(n)\times n} & I_{t(n)} \end{bmatrix}^T \nabla_{\boldsymbol{\epsilon}_2}\|(L+\mathrm{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2
\end{aligned} \tag{B.3}$$

Recall that the *adjoint* $\mathcal{L}^*$ of a linear transformation $\mathcal{L} : V \to W$ is defined by

$$\langle w, \mathcal{L}(v)\rangle = \langle \mathcal{L}^*(w), v\rangle, \quad \forall v \in V, \forall w \in W.$$

Proposition B.0.1. *Let $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with adjoint $\mathcal{L}^*$. Then the gradients satisfy*

$$\nabla_{\boldsymbol{\epsilon}} \langle \mathcal{L}(\boldsymbol{\epsilon}), \mathcal{L}(\boldsymbol{\epsilon}) \rangle = 2\mathcal{L}^*(\mathcal{L}(\boldsymbol{\epsilon})) \tag{B.4}$$

$$\nabla_{\boldsymbol{\epsilon}} \langle \mathcal{L}(\boldsymbol{\epsilon}), \mathbf{b} \rangle = \mathcal{L}^*(\mathbf{b}) \tag{B.5}$$

Proof. Let $f(\boldsymbol{\epsilon}) = \langle \mathcal{L}(\boldsymbol{\epsilon}), \mathcal{L}(\boldsymbol{\epsilon}) \rangle$ and $g(\boldsymbol{\epsilon}) = \langle \mathcal{L}(\boldsymbol{\epsilon}), \mathbf{b} \rangle$. Use expansion with $\boldsymbol{\epsilon_\delta} = \boldsymbol{\epsilon} + \boldsymbol{\delta}$:

$$
\begin{aligned}
f(\boldsymbol{\epsilon_\delta}) &= \langle \mathcal{L}(\boldsymbol{\epsilon} + \boldsymbol{\delta}), \mathcal{L}(\boldsymbol{\epsilon} + \boldsymbol{\delta}) \rangle \\
&= \langle \mathcal{L}(\boldsymbol{\epsilon}), \mathcal{L}(\boldsymbol{\epsilon}) \rangle + 2 \langle \mathcal{L}(\boldsymbol{\epsilon}), \mathcal{L}(\boldsymbol{\delta}) \rangle + \langle \mathcal{L}(\boldsymbol{\delta}), \mathcal{L}(\boldsymbol{\delta}) \rangle \\
&= f(\boldsymbol{\epsilon}) + \langle 2\mathcal{L}^*(\mathcal{L}(\boldsymbol{\epsilon})), \boldsymbol{\delta} \rangle + \tfrac{1}{2} \langle 2\mathcal{L}^*(\mathcal{L}(\boldsymbol{\delta})), \boldsymbol{\delta} \rangle \\
g(\boldsymbol{\epsilon_\delta}) &= \langle \mathcal{L}(\boldsymbol{\epsilon} + \boldsymbol{\delta}), \mathbf{b} \rangle \\
&= \langle \mathcal{L}(\boldsymbol{\epsilon}), \mathbf{b} \rangle + \langle \mathcal{L}(\boldsymbol{\delta}), \mathbf{b} \rangle \\
&= g(\boldsymbol{\epsilon}) + \langle \mathcal{L}^*(\mathbf{b}), \boldsymbol{\delta} \rangle .
\end{aligned}
$$

Therefore, by the definition of gradients, $\nabla_{\boldsymbol{\epsilon}} f(\boldsymbol{\epsilon}) = 2\mathcal{L}^*(\mathcal{L}(\boldsymbol{\epsilon}))$ proves (B.4), and $\nabla_{\boldsymbol{\epsilon}} g(\boldsymbol{\epsilon}) = \mathcal{L}^*(\mathbf{b})$ proves (B.5). $\qquad\square$

Definition B.0.2. *We denote* $\mathrm{vec}(\cdot)$ *to be the* vectorization *of a matrix, columnwise, i.e.,*

$$A \in \mathbb{R}^{m \times n}, \quad \mathrm{vec}(A) = (a_{11}, \ldots, a_{m1}, a_{12}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{mn})^T \in \mathbb{R}^{mn}.$$

Now we need to find the adjoint $\mathcal{L}_{\mathbf{w}}^*(\mathbf{u})$. Note that

$$
\begin{aligned}
\langle \mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2), \mathbf{u} \rangle &= \langle \mathrm{LMat}(\boldsymbol{\epsilon}_2)^T \mathbf{w}, \mathbf{u} \rangle \\
&= \mathbf{w}^T \mathrm{LMat}(\boldsymbol{\epsilon}_2) \mathbf{u} \\
&= \mathrm{trace}(\mathbf{w}^T \mathrm{LMat}(\boldsymbol{\epsilon}_2) \mathbf{u}) \\
&= \mathrm{trace}(\mathbf{u}\mathbf{w}^T \mathrm{LMat}(\boldsymbol{\epsilon}_2)) \\
&= \mathrm{trace}((\mathbf{w}\mathbf{u}^T)^T \mathrm{LMat}(\boldsymbol{\epsilon}_2)) \\
&= \mathrm{vec}(\mathbf{w}\mathbf{u}^T)^T \mathrm{vec}(\mathrm{LMat}(\boldsymbol{\epsilon}_2)).
\end{aligned} \tag{B.6}
$$

Definition B.0.3. *We denote* $\mathrm{lvec}(\cdot)$ *to be the* lower vectorization *of a matrix, i.e.,*

$$A \in \mathbb{R}^{n \times n}, \quad \mathrm{lvec}(A) = (a_{11}, \ldots, a_{n1}, a_{22}, \ldots, a_{n2}, \ldots, a_{n-1,n-1}, a_{n,n-1}, a_{nn})^T \in \mathbb{R}^{t(n)}.$$

$\mathrm{lvec}(\cdot)$ is a linear transformation. For a vector $\mathbf{x} \in \mathbb{R}^{t(n)}$, $\mathrm{lvec}(\mathrm{LMat}(\mathbf{x})) = \mathbf{x}$. For a lower triangular matrix $L \in \mathbb{R}^{n \times n}$, $\mathrm{LMat}(\mathrm{lvec}(L)) = L$.

Therefore, continuing from (B.6), we find the desired adjoint from

$$
\begin{aligned}
\langle \mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2), \mathbf{u} \rangle &= \operatorname{vec}(\mathbf{w}\mathbf{u}^T)^T \operatorname{vec}(\operatorname{LMat}(\boldsymbol{\epsilon}_2)) \\
&= \operatorname{lvec}(\mathbf{w}\mathbf{u}^T)^T \boldsymbol{\epsilon}_2 \\
&= \langle \operatorname{lvec}(\mathbf{w}\mathbf{u}^T), \boldsymbol{\epsilon}_2 \rangle \\
\implies \mathcal{L}_{\mathbf{w}}^*(\mathbf{u}) &= \operatorname{lvec}(\mathbf{w}\mathbf{u}^T).
\end{aligned}
$$

We now continue on finding the gradients and parts of Hessians. By Proposition B.0.1,

$$
\nabla_{\boldsymbol{\epsilon}_2} \langle \mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2), \mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2) \rangle = 2\mathcal{L}_{\mathbf{w}}^*(\mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2)) = 2\operatorname{lvec}(\mathbf{w}\mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2)^T) = 2\operatorname{lvec}(\mathbf{w}\mathbf{w}^T \operatorname{LMat}(\boldsymbol{\epsilon}_2))
$$
$$
\nabla_{\boldsymbol{\epsilon}_2} \langle \mathcal{L}_{\mathbf{w}}(\boldsymbol{\epsilon}_2), b_{\mathbf{w}} \rangle = \mathcal{L}_{\mathbf{w}}^*(b_{\mathbf{w}}) = \operatorname{lvec}(\mathbf{w}\mathbf{w}^T L)
$$

Therefore, continuing from (B.2), the first partial derivative with respect to $\boldsymbol{\epsilon}_2$ is:

$$
\begin{aligned}
\nabla_{\boldsymbol{\epsilon}_2} \|(L + \operatorname{LMat}(\boldsymbol{\epsilon}_2))^T \mathbf{w}\|_2 &= \frac{1}{\|(L+\operatorname{LMat}(\boldsymbol{\epsilon}_2))^T \mathbf{w}\|_2} (\operatorname{lvec}(\mathbf{w}\mathbf{w}^T \operatorname{LMat}(\boldsymbol{\epsilon}_2)) + \operatorname{lvec}(\mathbf{w}\mathbf{w}^T L)) \\
&= \frac{1}{\|(L+\operatorname{LMat}(\boldsymbol{\epsilon}_2))^T \mathbf{w}\|_2} \operatorname{lvec}\left(\mathbf{w}\mathbf{w}^T \left(L + \operatorname{LMat}(\boldsymbol{\epsilon}_2)\right)\right)
\end{aligned}
$$
$$(B.7)$$

For the second derivative with respect to $\mathbf{w}$, use expansion with $\mathbf{w}_{\boldsymbol{\delta}} = \mathbf{w} + \boldsymbol{\delta}$:

$$
\operatorname{lvec}((\mathbf{w} + \boldsymbol{\delta})(\mathbf{w} + \boldsymbol{\delta})^T L) = \operatorname{lvec}(\mathbf{w}\mathbf{w}^T L) + \operatorname{lvec}(\mathbf{w}\boldsymbol{\delta}^T L) + \operatorname{lvec}(\boldsymbol{\delta}\mathbf{w}^T L) + \operatorname{lvec}(\boldsymbol{\delta}\boldsymbol{\delta}^T L).
$$

If we can rewrite $\operatorname{lvec}(\mathbf{w}\boldsymbol{\delta}^T L) + \operatorname{lvec}(\boldsymbol{\delta}\mathbf{w}^T L)$ as an inner product of $\langle A_{\mathbf{w}}, \boldsymbol{\delta} \rangle$, then $\nabla_{\mathbf{w}} \operatorname{lvec}(\mathbf{w}\mathbf{w}^T L) = A_{\mathbf{w}}$.

Proposition B.0.4.
$$
\operatorname{lvec}(\mathbf{w}\boldsymbol{\delta}^T L) = \langle L F_{\mathbf{w}}, \boldsymbol{\delta} \rangle
$$

where $F_{\mathbf{w}} \in \mathbb{R}^{n \times t(n)}$ is as below:

$$
F_{\mathbf{w}} = \begin{bmatrix}
w_1 & w_2 & \cdots & w_n & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & w_2 & \cdots & w_n & 0 & \cdots & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & w_3 & \cdots & w_n & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & w_n
\end{bmatrix}
$$

For the i-th row, w_i in the $(2n + 2 - i)(i - 1)/2 + 1$-th column with $w_{i+1}, \ldots, w_n$ in the following columns and 0s in the other columns.

The readers can verify it by running the following code by Symbolic Math Toolbox$^{\text{TM}}$ in MATLAB:

53

```
Require: real number n
tn=(n+1)*n/2;
w=sym('w',[n 1],'real'); delta=sym('d',[n 1],'real');
L=sym('l',[n n],'real'); L=tril(L);
vec1=lvec(w*delta'*L);

F = sym('f',[n tn],'real'); F(:,:)=0; % construct the matrix F
for i=1:n
    begin_index=(2*n+2-i)*(i-1)/2+1; end_index=(2*n+2-i)*(i-1)/2+1+n-i;
    F(i,begin_index:end_index)=w(i:end);
end
vec2=(L*F)'*delta;
isequaln(vec1,vec2)
```

where lvec is consistent to Definition B.0.3:

```
function y=lvec(X)
y=[];
n=size(X,1);
for i=1:n
    y=[y;X(i:n,i)];
end
end
```

Proposition B.0.5.
$$\text{lvec}(\boldsymbol{\delta}\mathbf{w}^T L) = \langle G_{L,\mathbf{w}}, \boldsymbol{\delta} \rangle$$

where $G_{\mathbf{w}} \in \mathbb{R}^{n \times t(n)}$ is a block matrix as below:

Let the coefficient vector $\mathbf{c} \in \mathbb{R}^n$ be $\mathbf{c} = L^T \mathbf{w} = (c_1, \ldots, c_n)^T$. Then,

$$G_{L,\mathbf{w}} = \begin{pmatrix} c_1 I_n & & & & \\ & c_2 I_{n-1} & & \text{\Large 0} & \\ & & \ddots & & \\ & & & c_{n-1} I_2 & \\ & & & & c_n I_1 \end{pmatrix}$$

The code for verification is as followed:

```
Require: real number n
tn=(n+1)*n/2;
w=sym('w',[n 1],'real'); delta=sym('d',[n 1],'real');
L=sym('l',[n n],'real'); L=tril(L);
vec1=lvec(delta*w'*L);

c=L'*w;
G=[]; % construct the matrix G
for i=1:n
    temp=[zeros(i-1,n-i+1);c(i)*eye(n-i+1)]; G=[G,temp];
end
vec2=expand(G'*delta);
isequaln(vec1,vec2)
```

By Propositions B.0.4 and B.0.5,

$$\nabla_{\mathbf{w}} \operatorname{lvec}\left(\mathbf{w}\mathbf{w}^T\left(L + \operatorname{LMat}\left(\boldsymbol{\epsilon}_2\right)\right)\right) = (L + \operatorname{LMat}(\boldsymbol{\epsilon}_2))F_{\mathbf{w}} + G_{(L+\operatorname{LMat}(\boldsymbol{\epsilon}_2)),\mathbf{w}} \tag{B.8}$$

Therefore, from (B.3) and (B.7) and by the Quotient Rule as well as (B.1) and (B.8), we get that

$$
\begin{aligned}
\nabla^2_{\boldsymbol{\epsilon}\mathbf{w}}\|(L + \operatorname{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2 = \quad & \left(\begin{bmatrix}\mathbf{0}_{t(n)\times n} & I_{t(n)}\end{bmatrix}^T \frac{1}{\|(L+\operatorname{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2^2}\right) \\
& \left(\left(\nabla_{\mathbf{w}}\operatorname{lvec}\left(\mathbf{w}\mathbf{w}^T\left(L + \operatorname{LMat}\left(\boldsymbol{\epsilon}_2\right)\right)\right)\right)^T \|(L+\operatorname{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2\right. \\
& \left. - \left(\operatorname{lvec}\left(\mathbf{w}\mathbf{w}^T\left(L + \operatorname{LMat}\left(\boldsymbol{\epsilon}_2\right)\right)\right)\right)\left(\nabla_{\mathbf{w}}\|(L+\operatorname{LMat}(\boldsymbol{\epsilon}_2))^T\mathbf{w}\|_2^T\right)\right)
\end{aligned}
\tag{B.9}
$$

Index

Bibliography

[1] A. Ahmadi-Javid. Entropic value-at-risk: a new coherent risk measure. *J. Optim. Theory Appl.*, 155(3):1105–1123, 2012.

[2] P. Artzner, F. Delbaen, J-M Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

[3] M.J. Best and R.R. Grauer. 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, 1991.

[4] Li Chen, Simai He, and Shuzhong Zhang. Tight bounds for some risk measures, with applications to robust portfolio selection. *Oper. Res.*, 59(4):847–865, 2011.

[5] G. Cornuejols and R. Tütüncü. *Optimization methods in finance*. Mathematics, Finance and Risk. Cambridge University Press, Cambridge, 2007.

[6] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.*, 58(3):595–612, 2010.

[7] Anthony V. Fiacco. *Introduction to sensitivity and stability analysis in nonlinear programming*, volume 165 of *Mathematics in Science and Engineering*. Academic Press, Inc., Orlando, FL, 1983.

[8] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Math. Oper. Res.*, 28(1):1–38, 2003.

[9] D. Huang, S. Zhu, F.J. Fabozzi, and M. Fukushima. Portfolio selection under distributional uncertainty: A relative robust cvar approach. *European Journal of Operational Research*, 203(1):185–194, 2010.

[10] Z. Kang, X. Li, Z. Li, and S. Zhu. Data-driven robust mean-CVaR portfolio selection under distribution ambiguity. *Quant. Finance*, 19(1):105–121, 2019.

[11] A. E. B. Lim, J. G. Shanthikumar, and G-Y Vahn. Conditional value-at-risk in portfolio optimization: coherent but fragile. *Oper. Res. Lett.*, 39(3):163–171, 2011.

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

[13] R.C. Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323 – 361, 1980.

[14] A.G. Quaranta and A. Zaffaroni. Robust optimization of conditional value at risk and portfolio selection. *Journal of Banking & Finance*, 32(10):2046–2056, 2008.

[15] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

[16] R. H. Tütüncü and M. Koenig. Robust asset allocation. *Ann. Oper. Res.*, 132:157–187, 2004.

[17] S. C. P. Yam, H. Yang, and F. L. Yuen. Optimal asset allocation: risk and information uncertainty. *European J. Oper. Res.*, 251(2):554–561, 2016.

[18] S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Oper. Res.*, 57(5):1155–1168, 2009.