Robust Portfolio Optimization

PAR

Zheng Wei YAP

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

Résumé

Depuis la crise financière mondiale de 2008, le marché financier est devenu plus imprévisible que jamais, et il semble qu'il le restera dans un avenir prévisible. Cela signifie qu'un investisseur est confronté à des risques sans précédent, d'où le besoin croissant d'optimisation de portefeuille robuste pour les protéger contre l'incertitude, qui est potentiellement dévastateur si non supervisé mais ignoré dans le modèle classique de Markowitz, dont une autre carence est l'absence de moments d'ordre élevé dans son hypothèse de la distribution des rendements des actifs. Nous établissons une équivalence entre le modèle de Markowitz et le problème d'optimisation de la valeur à risque du portefeuille sous la normalité multivariée des rendements des actifs, de sorte que nous pouvons ajouter ces caractéristiques exclues dans le premier implicitement en les incorporant dans le second. Nous proposons également une méthode d'approximation de spline probabiliste à lissage et un modèle déterministe dans le cadre de la localisation-échelle sous la distribution elliptique des rendements des actifs pour résoudre le problème robuste d'optimisation de la valeur au risque du rendement du portefeuille. En particulier pour le modèle déterministe, nous introduisons un nouvel ensemble d'incertitude qui vit dans l'espace défini positif pour la matrice d'échelle sans compromettre la complexité et le conservatisme du problème d'optimisation, inventons une méthode pour déterminer la taille des ensembles, le testons sur des données réelles, et explorons ses propriétés de diversification. Bien que la valeur à risque soit la mesure de risque standard adoptée par le secteur bancaire et de l'assurance depuis le début des années nonante, elle a depuis suscité de nombreuses critiques, notamment de McNeil et al. (2005) et le Comité de Bâle sur le contrôle bancaire en 2012, également connu sous le nom de Bâle 3.5 [21, 23]. Bâle 4 [22] suggère même de passer de la «valeur à risque» à la mesure de la «valeur à risque conditionnelle». Nous verrons que la première peut Ãłtre remplacÃľ par la dernière ou mÃłme d'autres mesures de risque dans nos formulations facilement.

Abstract

Since the 2008 Global Financial Crisis, the financial market has become more unpredictable than ever before, and it seems set to remain so in the forseeable future. This means an investor faces unprecedented risks, hence the increasing need for robust portfolio optimization to protect them against uncertainty, which is potentially devastating if unattended yet ignored in the classical Markowitz model, whose another deficiency is the absence of higher moments in its assumption of the distribution of asset returns. We establish an equivalence between the Markowitz model and the portfolio return value-at-risk optimization problem under multivariate normality of asset returns, so that we can add these excluded features into the former implicitly by incorporating them into the latter. We also provide a probabilistic smoothing spline approximation method and a deterministic model within the location-scale framework under elliptical distribution of the asset returns to solve the robust portfolio return value-at-risk optimization problem. In particular for the deterministic model, we introduce a novel eigendecomposition uncertainty set which lives in the positive definite space for the scale matrix without compromising on the computational complexity and conservativeness of the optimization problem, invent a method to determine the size of the involved uncertainty sets, test it out on real data, and explore its diversification properties. Although the value-at-risk has been the standard risk measure adopted by the banking and insurance industry since the early nineties, it has since attracted many criticisms, in particular from McNeil et al. (2005) and the Basel Committee on Banking Supervision in 2012, also known as Basel 3.5 [21, 23]. Basel 4 [22] even suggests a move away from the "what" value-at-risk to the "what-if" conditional value-at-risk' measure. We shall see that the former may be replaced with the latter or even other risk measures in our formulations easily.

Acknowledgments

I would like to thank all friends and family members who have helped me one way or another in the course of my dissertation, and especially express my gratitude towards Professor Morgenthaler for taking me under his tutelage and giving me the freedom to pursue my academic interests. Soli Deo gloria.

Contents

Chapter 1

Preliminaries

1.1   Introduction

Since the 2008 Global Financial Crisis, there has been a general lack of confidence in the global financial system and the world economy. The European Sovereign Debt Crisis which resulted in Greece almost leaving the European Union and thus jeopardizing the single currency project did not help matters. Whether the European model will continue remains to be seen, although some observers already see it begin to unravel with the unpegging of the Swiss Franc (CHF) against the Euro (EUR) since January 2015, as well as the Quantitative Easing (QE) introduced by the European Central Bank (ECB) shortly afterwards in March the same year that was to last for at least sixteen months and worth no less than 1.1 trillion EUR. Theresa May will soon trigger Article 50 to start the two-year countdown to Brexit, which adds uncertainty to a world already in a state of flux.

Crossing over to the Asia-Pacific, China is in the process of restructuring into a more consumption based economy from one that is driven by massive state investment, and whether it can navigate through this transition period and emerge out of these murky waters successfully is still an unknown. Its slower growth as a result also means that Australia, whose economy depends heavily on iron ore exports to China, is adversely affected. Dealing with serious corruption and pollution issues will also be crucial in its ability to retain and attract talent, maintain social stability and continue its trajectory of growth. Moreover, China is seeking to expand its hegemony in the region, the evidence of which lies in its recent disputes with other Southeast Asian nations over the South China Sea, including a rejection of the Hague Tribunal's ruling in favor of the Philippines, with Japan over the Diaoyu or Senkaku Islands, with Taiwan over its independence, and even with Hong Kong over the freedom to elect its own Chief Executive. The simmering geopolitical tensions in the region and the threat of their escalating into war and violence is real.

In the United States, Donald Trump was elected in a freak election. What will the world become under him? Will there be a deregulation of banks causing yet another financial crisis, an anti-globalization and protectionist stance that further dampens sluggish world trade, a disengagement with and retreat of American military presence in Asia leading to an imbalance of power tilted towards China, resulting in a "might is right" instead of a rules-based order in the region? How about his foreign policies in the Middle East and their

implications?

Add to the mix that we are living in a disruptive age, at the cusp of the so-called "fourth industrial revolution", where machine learning, robotics and blockchain technology among others are all set to displace millions of jobs worldwide, changing the way we live and work, the only certainty for an investor is uncertainty, hence the Basel Committee's call to move away from the value-at-risk to the conditional value-at-risk measure, a need for robust portfolio optimization and the necessity to take into account the possible occurrence of a black swan event.

1.2   Markowitz Model

[134, Chapter 1] In modern portfolio theory, one is almost certainly reminded of the Markowitz model, where the investor has terminal wealth w_T at the end of the trading period $[0, T]$ with utility function $\mathcal{U}(w_T)$ such that $\mathcal{U}'(w_T) > 0$ and $\mathcal{U}''(w_T) < 0$. The positivity of the first derivative means that the higher the terminal wealth, the "happier" the investor is (non-satiation property). The negativity of the second derivative can be interpreted as the investor having decreasing added "happiness" with increasing wealth (risk-averse property). Furthermore, the utility function is of the quadratic form

$$\mathcal{U}(w_T) = \varphi w_T - \psi w_T^2$$

where $w_T = w_0(1 + r_{\mathcal{P}})$ such that w_0 is the initial wealth and $r_{\mathcal{P}}$ is the portfolio return over the period $[0, T]$, so that

$$\begin{aligned}
\mathcal{U}(w_T) &= \varphi w_0(1 + r_{\mathcal{P}}) - \psi w_0^2(1 + r_{\mathcal{P}})^2 \\
&= (\varphi w_0 - \psi w_0^2) + (\varphi w_0 - 2\psi w_0^2)r_{\mathcal{P}} - (\psi w_0^2)r_{\mathcal{P}}^2 \\
&= a + b r_{\mathcal{P}} - c r_{\mathcal{P}}^2,
\end{aligned}$$

where $a = \varphi w_0 - \psi w_0^2$, $b = \varphi w_0 - 2\psi w_0^2$ and $c = \psi w_0^2$, and the expected utility is

$$\begin{aligned}
\mathbb{E}(\mathcal{U}(w_T)) &= \mathbb{E}(a + b r_{\mathcal{P}} - c r_{\mathcal{P}}^2) \\
&= a + b\mathbb{E}(r_{\mathcal{P}}) - c\mathbb{E}(r_{\mathcal{P}}^2) \\
&= a + (b - c)\mathbb{E}(r_{\mathcal{P}}) - c\mathbb{V}(r_{\mathcal{P}}),
\end{aligned}$$

which the investor would like to maximize. Therefore, the quadratic utility function does not capture aversion to higher-order moments directly like, for example, the Constant Relative Risk Aversion (CRRA) utility function does. However, Kacperczyk and Damien (2011) show that the magnitude of this direct effect is negligible qualitatively.

In addition, the market is frictionless, that is, without taxes, transaction costs or short sales. Simply put, the investor minimizes the portfolio variance subject to a targeted portfolio expected return:

$$\min_{\mathbf{w} \in \mathbb{R}^n} \left\{ \frac{1}{2}\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w} : \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} = r, \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1 \right\},$$

where $\Sigma \in \mathbb{R}^{n \times n}$ is a positive definite covariance matrix, $\boldsymbol{\mu} \in \mathbb{R}^n$ is the expected return vector, $r \leq \max\{\mu_1, \ldots, \mu_n\}$ is the targeted expected return over $[0, T]$, $\mathbf{w} \in \mathbb{R}^n$ is the weight

vector and μ is linearly independent of $\mathbb{1} \in \mathbb{R}^n$ to avoid a degnerate scenario, in which the constraints $\mathbf{w}^T \mathbb{1} = 1$ and $\mathbf{w}^T \mu = r$ contradict each other unless $nr = \mu^T \mathbb{1}$. The 1/2 is inserted for notational convenience when considering the first-order conditions.

Theorem 1.1 (Black, 1972)
The global optimal solution of

$$\min_{\mathbf{w} \in \mathbb{R}^n} \left\{ \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} : \mathbf{w}^T \mu = r, \mathbf{w}^T \mathbb{1} = 1 \right\} \tag{1.1}$$

where $\Sigma \in \mathbb{R}^{n \times n}$ is positive definite, $r \in \mathbb{R}$ and $\mu \in \mathbb{R}^n$ is linearly independent of $\mathbb{1} \in \mathbb{R}^n$, is

$$\mathbf{w}_* = \frac{(Ar - C)\Sigma^{-1}\mu + (B - Cr)\Sigma^{-1}\mathbb{1}}{D} \tag{1.2}$$

where $A = \mathbb{1}^T \Sigma^{-1} \mathbb{1}$, $B = \mu^T \Sigma^{-1} \mu$, $C = \mathbb{1}^T \Sigma^{-1} \mu$ and $D = AB - C^2$.

Proof: Notice that the first order conditions of the equivalent Lagrangian problem

$$\min_{\mathbf{w} \in \mathbb{R}^n, v_1, v_2 \in \mathbb{R}} \left\{ \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} - v_1(\mathbf{w}^T \mathbb{1} - 1) - v_2(\mathbf{w}^T \mu - r) : v_1, v_2 > 0 \right\}$$

are

$$\mathbf{0} = \Sigma \mathbf{w} - v_1 \mathbb{1} - v_2 \mu, \tag{1.3}$$
$$1 = \mathbf{w}^T \mathbb{1}, \tag{1.4}$$
$$r = \mathbf{w}^T \mu. \tag{1.5}$$

Left-multiplying (1.3) with Σ^{-1} yields

$$\mathbf{w} = v_1 \Sigma^{-1} \mathbb{1} + v_2 \Sigma^{-1} \mu. \tag{1.6}$$

Then, left-multiplying (1.6) with $\mathbb{1}^T$ and μ^T, and using (1.4) and (1.5) obtains

$$1 = v_1 A + v_2 C \text{ and } r = v_1 C + v_2 B$$

respectively, which in matrix notation is

$$[1, r]^T = \begin{bmatrix} A & C \\ C & B \end{bmatrix} [v_1, v_2]^T.$$

Note that $\begin{bmatrix} A & C \\ C & B \end{bmatrix}$ is invertible since its determinant

$$D = AB - C^2 = \|\mathbf{P}^{-1}\mathbb{1}\|^2 \|\mathbf{P}^{-1}\mu\|^2 - \langle \mathbf{P}^{-1}\mathbb{1}, \mathbf{P}^{-1}\mu \rangle^2$$

is positive, where $\Sigma = \mathbf{P}\mathbf{P}^T$, $\|\cdot\|$ is the Euclidean norm, and $\langle \cdot, \cdot \rangle$ is the inner dot product, due to the linear indepdence of $\mathbf{P}^{-1}\mathbb{1}$ and $\mathbf{P}^{-1}\mu$ and the strict Cauchy-Schwarz inequality $\langle x, y \rangle < \|x\|^2 \|y\|^2$ when x and y are linearly independent. Now

$$[v_1, v_2]^T = \begin{bmatrix} A & C \\ C & B \end{bmatrix}^{-1} [1, r]^T$$

$$= [Cr - B, Ar - C]^{\mathrm{T}}/D$$

and substituting it into (1.6) obtains the desired result. $\qquad\square$

The standard deviation of the portfolio is

$$\sigma(r) = \sqrt{\frac{Ar^2 - 2Cr + B}{D}}$$

which is a hyperbola function.

Definition 1.1

If two portfolios have the same expected return, the one with lower standard deviation is said to dominate the other with higher standard deviation. On the other hand, if two portfolios have the same standard deviation, the one with higher expected return is said to dominate the other with lower expected return. A portfolio that is not dominated is called efficient.

It does not make sense to set $r < r_{\min}$ as there will always be another portfolio with the same standard deviation but a higher expected return. The optimal portfolios with $r \geq r_{\min}$ are effcient and the arc that represents them in the $(\sigma(r), r)$-space is called the efficient frontier. To eliminate the possibility of choosing $r < r_{\min}$, the portfolio expected return is written as an affine function of the portfolio standard deviation with intercept k and a non-negative gradient h:

$$\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} = k + h\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}},$$

so that by rearranging the intercept is

$$k = \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - h\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}},$$

which is maximized over the feasible domain. That is to say, we solve

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - h\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1\right\}, \tag{1.7}$$

to obtain an optimal portfolio with the chosen risk parameter $h \in [0, \infty)$, which can be interpreted as the marginal risk premium, or more precisely, the expected additional portolio return with a unit increase in portfolio standard deviation, corresponding to an $r \geq r_{\min}$. The higher h is, the more risk averse is the investor. Note that (1.7) is equivalent to

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{1/(1+h)\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - h/(1+h)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1\right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w}\in\mathbb{R}^n}\left\{\theta\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - (1-\theta)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1\right\}$$

where $\theta = \frac{1}{1+h} \in [0, 1]$.

1.2.1 With Riskless Asset

Now, suppose there is a riskless asset return $\mu_0 \in \mathbb{R}$ less than the targeted expected return r and denote the weight placed on it as $\eta \in \mathbb{R}$. The Markowitz model with riskless asset then becomes

$$\min_{\mathbf{w} \in \mathbb{R}^n, \eta \in \mathbb{R}} \left\{ \frac{1}{2} \mathbf{w}^\mathrm{T} \Sigma \mathbf{w} : \mathbf{w}^\mathrm{T} \boldsymbol{\mu} = r - \eta \mu_0, \mathbf{w}^\mathrm{T} \mathbb{1} = 1 - \eta \right\}.$$

Theorem 1.2
The optimal solution of

$$\min_{\mathbf{w} \in \mathbb{R}^n, \eta \in \mathbb{R}} \left\{ \frac{1}{2} \mathbf{w}^\mathrm{T} \Sigma \mathbf{w} : \mathbf{w}^\mathrm{T} \boldsymbol{\mu} = r - \eta \mu_0, \mathbf{w}^\mathrm{T} \mathbb{1} = 1 - \eta \right\} \tag{1.8}$$

where $\Sigma \in \mathbb{R}^{n \times n}$ is positive definite, $r, \mu_0 \in \mathbb{R}$ such that $r \geq \mu_0$ and $\boldsymbol{\mu} \in \mathbb{R}^n$ is linearly independent of $\mathbb{1} \in \mathbb{R}^n$ is

$$\mathbf{w}_* = \frac{r - \mu_0}{A\mu_0^2 - 2C\mu_0 + B} \Sigma^{-1}(\boldsymbol{\mu} - \mu_0 \mathbb{1}) \quad \text{and} \quad \eta_* = 1 - \mathbf{w}_*^\mathrm{T} \mathbb{1}. \tag{1.9}$$

Proof: The Lagrangian is

$$\min_{\mathbf{w} \in \mathbb{R}^n, \eta, v_1, v_2 \in \mathbb{R}} \left\{ \frac{1}{2} \mathbf{w}^\mathrm{T} \Sigma \mathbf{w} - v_1(\mathbf{w}^\mathrm{T} \mathbb{1} + w_0 - 1) - v_2(\mathbf{w}^\mathrm{T} \boldsymbol{\mu} + \eta \mu_0 - r) : v_1, v_2 > 0 \right\}.$$

which has first order conditions

$$\mathbf{0} = \Sigma \mathbf{w} - v_1 \mathbb{1} - v_2 \boldsymbol{\mu}, \tag{1.10}$$

$$v_1 = -v_2 \mu_0, \tag{1.11}$$

$$1 = \mathbf{w}^\mathrm{T} \mathbb{1} + \eta, \tag{1.12}$$

$$r = \mathbf{w}^\mathrm{T} \boldsymbol{\mu} + \eta \mu_0. \tag{1.13}$$

Substituting (1.11) into (1.10) and left-multiplying the result by Σ^{-1} obtains

$$\mathbf{w} = v_2 \Sigma^{-1}(\boldsymbol{\mu} - \mu_0 \mathbb{1}). \tag{1.14}$$

Furthermore, if we left-multiply (1.14) by $\mathbb{1}^\mathrm{T}$ and $\boldsymbol{\mu}^\mathrm{T}$ and make use of the relations (1.12) and (1.13), we get

$$1 - \eta = v_2(C - Ar) \text{ and } r - \eta \mu_0 = v_2(B - C\mu_0)$$

respectively. Solving for η in the former, before substituting it in the latter and rearranging obtains

$$v_2 = \frac{r - \mu_0}{A\mu_0^2 - 2C\mu_0 + B}$$

which, when plugged into (1.14), yields the desired optimal solution. It remains to compute η_*. $\qquad \square$

The standard deviation of the optimal portfolio given by (1.9) is

$$\sigma(r) = \sqrt{\frac{(r-\mu_0)^2(\boldsymbol{\mu}-\mu_0\mathbb{1})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{\mu}-\mu_0\mathbb{1})}{(A\mu_0^2 - 2C\mu_0 + B)^2}}$$

$$= \sqrt{\frac{(r-\mu_0)^2(A\mu_0^2 - 2C\mu_0 + B)}{(A\mu_0^2 - 2C\mu_0 + B)^2}}$$

$$= \frac{r-\mu_0}{\sqrt{A\mu_0^2 - 2C\mu_0 + B}},$$

which draws out the efficient line in the $(\sigma(r), r)$-space.

Theorem 1.3

The efficient line

$$\sigma(r) = \frac{r-\mu_0}{\sqrt{A\mu_0^2 - 2C\mu_0 + B}} \tag{1.15}$$

of the Markowitz model with riskless asset (1.8) *[p. 5] is a tangent to the efficient frontier*

$$\sigma(r) = \sqrt{(Ar^2 - 2Cr + B)/D}. \tag{1.16}$$

of the Markowitz model without riskless asset (1.1) *[p. 3].*

Proof: If we invest nothing into the riskless asset, weights on the risky assets add up to one and left-multiplying $\mathbf{w}_*$ in (1.9) by $\mathbb{1}^{\mathrm{T}}$ yields

$$1 = \frac{r-\mu_0}{A\mu_0^2 - 2C\mu_0 + B}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{\mu}-\mu_0\mathbb{1}) \Rightarrow \frac{r-\mu_0}{A\mu_0^2 - 2C\mu_0 + B} = \frac{1}{C - A\mu_0}$$

and when it is substituted back into (1.9), gets

$$\mathbf{w}_{\mathrm{tan}} = \frac{1}{C - A\mu_0}\Sigma^{-1}(\boldsymbol{\mu}-\mu_0\mathbb{1}), \tag{1.17}$$

which we shall call (1.17) the tangency portfolio. This tangency portfolio is efficient in the Markowitz model without riskless asset. To prove it, note that its expected return is

$$r = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}_{\mathrm{tan}} = \frac{B - C\mu_0}{C - A\mu_0}$$

so that inserting it into (1.2) obtains

$$\mathbf{w}_{\mathrm{tan}} = \frac{\left(A\left(\frac{B-C\mu_0}{C-A\mu_0}\right) - C\right)\Sigma^{-1}\boldsymbol{\mu} + \left(B - C\left(\frac{B-C\mu_0}{C-A\mu_0}\right)\right)\Sigma^{-1}\mathbb{1}}{D}.$$

Finally, note that the first derivative of the efficient frontier (1.16) with respect to r evaluated at the point corresponding to the tangency portfolio reads

$$\left.\frac{d\sigma(r)}{dr}\right|_{r=\frac{B-C\mu_0}{C-A\mu_0}} = \frac{A\left(\frac{B-C\mu_0}{C-A\mu_0}\right) - C}{\sqrt{D\left(A\left(\frac{B-C\mu_0}{C-A\mu_0}\right)^2 - 2C\left(\frac{B-C\mu_0}{C-A\mu_0}\right) + B\right)}} = \sqrt{A\mu_0^2 - 2C\mu_0 + B},$$

which is exactly the gradient of the efficient line (1.15). $\qquad\square$

1.3 Value-At-Risk Optimization

Definition 1.2

The ϵ-level value-at-risk of a random variable $X \in \mathbb{R}$

$$\text{V@R}_\epsilon(X) := \max_{a \in \mathbb{R}} \{a : \mathbb{P}(X < a) \leq \epsilon\}$$

is the largest value of $a \in \mathbb{R}$ such that the probability of X being less than a is not greater than ϵ.

Under multivariate normality of returns, the problem of maximizing the portfolio return value-at-risk

$$\max_{\mathbf{w} \in \mathbb{R}^n} \left\{ \text{V@R}_\epsilon(\mathbf{w}^{\text{T}}\mathbf{R}) : \mathbf{w}^{\text{T}}\mathbb{1} = 1, \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \right\} \tag{1.18}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : t < \text{V@R}_\epsilon(\mathbf{w}^{\text{T}}\mathbf{R}), \mathbf{w}^{\text{T}}\mathbb{1} = 1, \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \right\}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : \mathbb{P}_{\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)} \{\mathbf{w}^{\text{T}}\mathbf{R} < t\} \leq \epsilon, \mathbf{w}^{\text{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : \mathbb{P}_{Z \sim \mathcal{N}(0,1)} \left\{ Z < (t - \mathbf{w}^{\text{T}}\boldsymbol{\mu})/\sqrt{\mathbf{w}^{\text{T}}\Sigma\mathbf{w}} \right\} \leq \epsilon, \mathbf{w}^{\text{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : (t - \mathbf{w}^{\text{T}}\boldsymbol{\mu})/\sqrt{\mathbf{w}^{\text{T}}\Sigma\mathbf{w}} \leq -z_{1-\epsilon}, \mathbf{w}^{\text{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : t \leq \mathbf{w}^{\text{T}}\boldsymbol{\mu} - z_{1-\epsilon}\sqrt{\mathbf{w}^{\text{T}}\Sigma\mathbf{w}}, \mathbf{w}^{\text{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n} \left\{ \mathbf{w}^{\text{T}}\boldsymbol{\mu} - z_{1-\epsilon}\sqrt{\mathbf{w}^{\text{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\text{T}}\mathbb{1} = 1 \right\} \tag{1.19}$$

where $z_{1-\epsilon}$ is the $100(1-\epsilon)$th quantile of the standard normal distribution. Note that (1.18) is maximizing the quantile of the portfolio return and (1.19), which first appeared in Roy (1952) and is also called the downside-risk problem, can be interpreted as minimizing the portfolio standard deviation subject to a targeted expected return if $\epsilon < 0.5$, and is thus equivalent to the Markowitz model.

1.3.1 With Riskless Asset

Now assume that the weight η_* on the riskless asset return μ_0 is determined exogeneously to be a value less than one, then the portfolio return V@R_ϵ optimization problem under multivariate normality of returns becomes

$$\max_{\mathbf{w} \in \mathbb{R}^n} \left\{ \text{V@R}_\epsilon((1 - \eta_*)\mathbf{w}^{\text{T}}\mathbf{R} + \eta_*\mu_0) : (1 - \eta_*)\mathbf{w}^{\text{T}}\mathbb{1} + \eta_* = 1, \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \right\} \tag{1.20}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : t < \text{V@R}_\epsilon((1 - \eta_*)\mathbf{w}^{\text{T}}\mathbf{R} + \eta_*\mu_0), \mathbf{w}^{\text{T}}\mathbb{1} = 1, \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \right\}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : \mathbb{P}_{\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)} \{(1 - \eta_*)\mathbf{w}^{\text{T}}\mathbf{R} + \eta_*\mu_0 < t\} \leq \epsilon, \mathbf{w}^{\text{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : \mathbb{P}_{Z \sim \mathcal{N}(0,1)} \left\{ Z < \frac{t - (1 - \eta_*)\mathbf{w}^{\text{T}}\boldsymbol{\mu} - \eta_*\mu_0}{(1 - \eta_*)\sqrt{\mathbf{w}^{\text{T}}\Sigma\mathbf{w}}} \right\} \leq \epsilon, \mathbf{w}^{\text{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w}\in\mathbb{R}^n, t\in\mathbb{R}} \left\{ t : t \le (1-\eta_*)\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + \eta_*\mu_0 - (1-\eta_*)z_{1-\epsilon}\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}, \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w}\in\mathbb{R}^n} \left\{ (1-\eta_*)\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + \eta_*\mu_0 - (1-\eta_*)z_{1-\epsilon}\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w}\in\mathbb{R}^n} \left\{ (1-\eta_*)\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + \eta_*\mu_0 - (1-\eta_*)z_{1-\epsilon}\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1 \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w}\in\mathbb{R}^n} \left\{ \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - z_{1-\epsilon}\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1 \right\},$$

which is equivalent to the Markowitz model. The optimal portfolio return value-at-risk of problem (1.20) is

$$\mathrm{V@R}_\epsilon((1-\eta_*)\mathbf{w}_*^{\mathrm{T}}\mathbf{R} + \eta_*\mu_0) = (1-\eta_*)\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu} + \eta\mu_0 - z_{1-\epsilon}(1-\eta_*)\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}$$

$$= \eta_*\mu_0 + (1-\eta_*)\left(\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu} - z_{1-\epsilon}\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*} \right) \tag{1.21}$$

where $\mathbf{w}_*$ is the optimal solution of (1.19). Note that (1.21) suggests that if η_* is to be determined endogeneously, then $\eta_* = 1$ if

$$\mu_0 > \mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu} - z_{1-\epsilon}\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}$$

$$\Leftrightarrow \quad \mu_0 - \mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu} + z_{1-\epsilon}\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*} > 0$$

$$\Leftrightarrow \quad z_{1-\epsilon} > \frac{\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu} - \mu_0}{\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}}$$

$$\Leftrightarrow \quad 1 - \epsilon > \Phi\left(\frac{\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu} - \mu_0}{\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}} \right)$$

$$\Leftrightarrow \quad \epsilon < \Phi\left(\frac{\mu_0 - \mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu}}{\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}} \right) \tag{1.22}$$

and $\eta_* = -\infty$ if

$$\epsilon > \Phi\left(\frac{\mu_0 - \mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu}}{\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}} \right).$$

However, if $\mu_0 = \mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu} - z_{1-\epsilon}\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}$, then η_* has to be determined exogeneously since its taking of any value less than or equal to one results in the same optimal portfolio return value-at-risk.

1.4 Short-Selling Constraints

It does not make sense to allow infinite short-selling or borrowing. Thus, we let $\eta \ge \ell_\eta \in \mathbb{R}_-$ and $\mathbf{w} \ge \ell \in \mathbb{R}_-^n$ so as to restrict borrowing and short-selling respectively.

1.4.1 Markowitz Model

Jagannathan and Ma (2003) show that the Markowitz model with short-selling constraints

$$\min_{\mathbf{w}\in\mathbb{R}^n}\left\{\frac{1}{2}\mathbf{w}^\mathrm{T}\Sigma\mathbf{w} : \mathbf{w}^\mathrm{T}\boldsymbol{\mu} = r, \mathbf{w}^\mathrm{T}\mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell}\right\} \tag{1.23}$$

is equivalent to

$$\min_{\mathbf{w}\in\mathbb{R}^n}\left\{\frac{1}{2}\mathbf{w}^\mathrm{T}\Sigma\mathbf{w} : \mathbf{w}^\mathrm{T}(\boldsymbol{\mu} + \boldsymbol{\upsilon}/\nu) = r, \mathbf{w}^\mathrm{T}\mathbb{1} = 1\right\}$$

where $\boldsymbol{\ell} = [\ell_1,\ldots,\ell_n]^\mathrm{T}$, $\boldsymbol{\upsilon}$ is the vector of Lagrange multipliers associated with the constraint $\mathbf{w} \geq \boldsymbol{\ell}$ and ν is the Lagrange multiplier associated with $\mathbf{w}^\mathrm{T}\mathbb{1} = 1$. In other words, if we add $\frac{\boldsymbol{\upsilon}}{\nu}$ to $\boldsymbol{\mu}$ in (1.23), the short-selling constraint can be removed without any consequences and the results in Section 1.2 can be applied directly.

The Markowitz model with short-selling constraints and a riskless asset

$$\min_{\mathbf{w}\in\mathbb{R}^n,\eta\in\mathbb{R}}\left\{\frac{1}{2}\mathbf{w}^\mathrm{T}\Sigma\mathbf{w} : \mathbf{w}^\mathrm{T}\boldsymbol{\mu} = r - \eta\mu_0, \mathbf{w}^\mathrm{T}\mathbb{1} = 1 - \eta, \mathbf{w} \geq \boldsymbol{\ell}\right\}, \tag{1.24}$$

has targeted expected return

$$r = \eta_*\mu_0 + (1-\eta_*)\left(\frac{\mathbf{w}_*}{1-\eta_*}\right)^\mathrm{T}\boldsymbol{\mu} \tag{1.25}$$

where $(\mathbf{w}_*, \eta_*)$ is the optimal solution of (1.24), and standard deviation

$$\sigma(r) = (1-\eta_*)\sqrt{\left(\frac{\mathbf{w}_*}{1-\eta_*}\right)^\mathrm{T}\Sigma\left(\frac{\mathbf{w}_*}{1-\eta_*}\right)}. \tag{1.26}$$

From (1.26), we have

$$\eta_* = 1 - \frac{\sigma(r)}{\sqrt{\left(\frac{\mathbf{w}_*}{1-\eta_*}\right)^\mathrm{T}\Sigma\left(\frac{\mathbf{w}_*}{1-\eta_*}\right)}}$$

and substituting into (1.25) obtains the efficient line

$$r = \mu_0 + \frac{\left(\frac{\mathbf{w}_*}{1-\eta_*}\right)^\mathrm{T}\boldsymbol{\mu} - \mu_0}{\sqrt{\left(\frac{\mathbf{w}_*}{1-\eta_*}\right)^\mathrm{T}\Sigma\left(\frac{\mathbf{w}_*}{1-\eta_*}\right)}}\sigma(r)$$

in the $(\sigma(r), r)$-space where its radient is the famous Sharpe ratio [180].

1.4.2 Value-At-Risk Optimization

The portfolio return value-at-risk optimization problem with short-selling constraints

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{\mathrm{V@R}_\epsilon(\mathbf{w}^\mathrm{T}\mathbf{R}) : \mathbf{w}^\mathrm{T}\mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell}, \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)\right\}$$

can be written as

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-z_{1-\epsilon}\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}:\mathbf{w}^{\mathrm{T}}\mathbb{1}=1,\mathbf{w}\geq\boldsymbol{\ell}\right\} \tag{1.27}$$

following the same previous arguments. Analogous to Section 1.3.1, if the weight η_* on the riskless asset is determined exogeneously to be a value less than one, then we are interested in solving

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{\mathrm{V@R}_\epsilon((1-\eta_*)\mathbf{w}^{\mathrm{T}}\mathbf{R}+\eta_*\mu_0):(1-\eta_*)\mathbf{w}^{\mathrm{T}}\mathbb{1}+\eta_*=1,\mathbf{w}\geq\boldsymbol{\ell},\mathbf{R}\sim\mathcal{N}(\boldsymbol{\mu},\Sigma)\right\},$$

which is equivalent to (1.27) and whose optimal portfolio return value-at-risk is

$$\mathrm{V@R}_\epsilon((1-\eta_*)\mathbf{w}_*^{\mathrm{T}}\mathbf{R}+\eta_*\mu_0)=\eta_*\mu_0+(1-\eta_*)\left(\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu}-z_{1-\epsilon}\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}\right), \tag{1.28}$$

where $\mathbf{w}_*$ is the optimal solution of (1.27). Using reasoning similar to that in Section 1.4.1, the efficient line in the $(\sigma(r),r)$-space is represented by

$$r=\mu_0+\frac{\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu}-\mu_0}{\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}}\sigma(r).$$

If η_* is to be determined endogeneously, then (1.28) suggests that

$$\eta_*=\begin{cases}1 & \text{if } \epsilon<\Phi\left(\frac{\mu_0-\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu}}{\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}}\right),\\[2ex]0 & \text{if } \epsilon>\Phi\left(\frac{\mu_0-\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu}}{\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}}\right).\end{cases}$$

If $\mu_0=\mathbf{w}_*^{\mathrm{T}}\boldsymbol{\mu}-z_{1-\epsilon}\sqrt{\mathbf{w}_*^{\mathrm{T}}\Sigma\mathbf{w}_*}$, then η_* has to be determined exogeneously since its taking of any value in the interval $[\ell_\eta,1]$ results in the same optimal return value-at-risk.
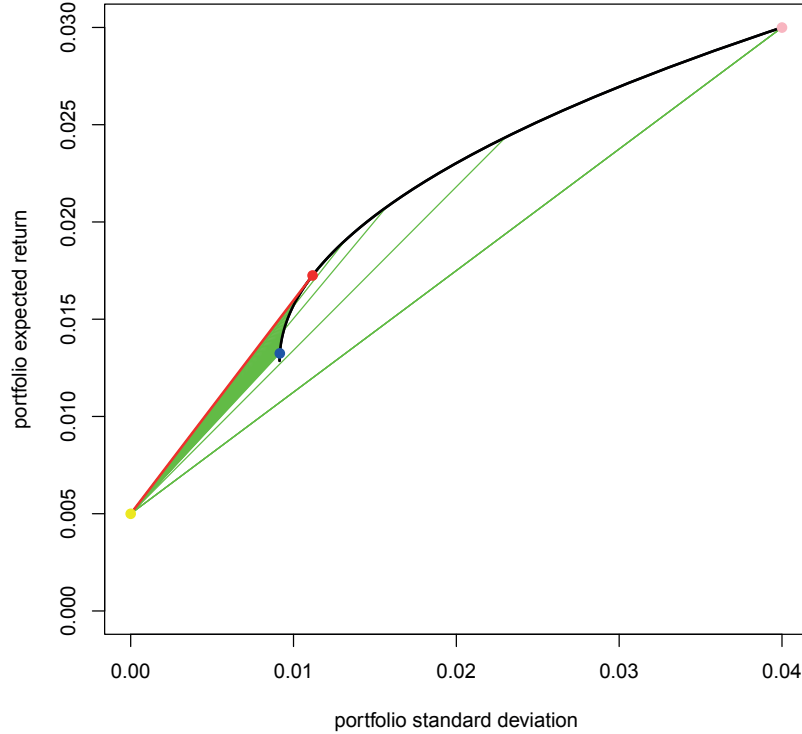
Figure 1.1: black line represents efficient frontier of Markowitz model without riskless asset, red line represents efficient frontier of Markowitz model with riskless asset of return $\mu_0 = 0.005$, and green lines represent optimal solutions of portfolio return value-at-risk optimization problem with weight on riskless asset determined exogeneously for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where $[r_1, r_2]^T \sim \mathcal{N}\left([\mu_1, \mu_2]^T, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}\right)$ such that $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$, $\rho = 0.2$.

1.5 Numerical Example

Assume there are two bivariate normally distributed risky assets with expected returns $\mu_1 = 0.01$, $\mu_2 = 0.03$, standard deviations $\sigma_1 = 0.1$, $\sigma_2 = 0.2$, correlation $\rho = 0.2$, a riskless asset $\mu_0 = 0.005$, and no short-selling. Unless otherwise stated, all programs hereafter are solved by the R package 'Alabama' by Ravi (2015).

Figure 1.1 shows the efficient frontiers of the Markowitz model with and without riskless asset (red and black lines respectively), as well as those which represent the portfolio return value-at-risk optimization problems (green lines) for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where the amount invested in the riskless asset is determined exogeneously. Note that the green line joins the yellow and blue dots at $\epsilon = 0.01$. As ϵ increases, the gradient of the green line increases then decreases, eventually joining the yellow and pink dots.

1.6   Higher Moments

It is an empirical fact that the distribution of returns often exhibit properties related to higher moments. Of course, the portfolio return (which is a convex combination of the individual assets) is asymptotically normal as the number of assets increases due to the Central Limit Theorem if they are independent, but obviously this is not realistic. Tsiang (1972), Francis (1975), Friend and Westerfield (1980), and Scott and Horvath (1980) take into account higher moments in their models, but they lead to serious issues in terms of solvability and complexity. Another way of including higher moments is through the use of dynamics, where the asset price process is modeled by variants of the stochastic differential equation. These models are highly demanding and usually used in derivative hedging involving only one risky and one riskless asset, and are seldom used in cases where more assets are included. Heuristic optimization methods, like in Maringer (2005), can be used to overcome the non-convexity and computational challenges that come with introducing higher moments, but things such as numerical instability and inaccuracy arise (since the optimal solution and convergence rate are different each time the optimization is performed, we are never sure of the global optimality of the resulting portfolio). Therefore, the inclusion of higher moments into the Markowitz model poses formidable problems.

1.7   Safe Convex Approximation

Fortunately, the equivalence between the Markowitz model and the portfolio return value-at-risk optimization problem under multivariate normality of asset returns means that higher moments can be added into the Markowitz model implicitly by incorporating them into a distribution Q, and considering

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{\text{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}}\mathbf{R}) : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell}, \mathbf{R} \sim Q\right\}. \tag{1.29}$$

Nevertheless, the value-at-risk objective function in (1.29) is generally non-concave. An approach to overcome this problem is to first notice that (1.29) is equivalent to

$$\max_{\mathbf{w}\in\mathbb{R}^n, t\in\mathbb{R}}\left\{t : t < \text{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}}\mathbf{R}), \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell}, \mathbf{R} \sim Q\right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w}\in\mathbb{R}^n, t\in\mathbb{R}}\left\{t : -\text{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}}\mathbf{R} - t) < 0, \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell}, \mathbf{R} \sim Q\right\}, \tag{1.30}$$

and then replace $-\text{V@R}_\epsilon(t - \mathbf{w}^{\mathrm{T}}\mathbf{R})$ with a convex upper bound to obtain a "safe" convex approximation of the associated constraint [25, p. 91]. We next introduce the concept of a coherent utility measure and then provide a couple of examples of it whose negation are safe convex upper bounds to the negated value-at-risk.

1.7.1   Coherent Utility Measures

We define coherent utility measures as an analogous counterpart to coherent risk measures, which are first introduced by Artzner et al. (1999). Let $Z : \Omega \to \mathbb{R}$ be a random function of

the space $\mathcal{Z}$ of all $\mathcal{F}$-measurable functions, defined on the measure space $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a sigma algebra. For $Z, Z' \in \mathcal{Z}$, let the relation $Z \geq Z'$ imply that

$$Z(\omega) \geq Z'(\omega) \text{ for amost every } \omega \in \Omega.$$

A utility function $\rho(Z)$ which maps Z into the extended real line $\mathbb{R} \cup \{+\infty\}$ is said to be coherent if it satisfies the following axioms:

(i) **Concavity**: $\rho(tZ+(1-t)Z') \geq t\rho(Z)+(1-t)\rho(Z')$ for all $Z, Z' \in \mathcal{Z}$ and for all $t \in [0,1]$.

(ii) **Monotonicity**: If $Z, Z' \in \mathcal{Z}$ and $Z \geq Z'$, then $\rho(Z) \geq \rho(Z')$.

(iii) **Translation Invariance**: If $d \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z+d) = \rho(Z)+d$.

(iv) **Positive Homogeneity**: If $t > 0$ and $Z \in \mathcal{Z}$, then $\rho(tZ) = t\rho(Z)$.

1.7.2 Conditional Value-At-Risk

[25, p. 93-96] The conditional value-at-risk is a coherent utility measure defined as

$$\text{CV@R}_\epsilon := \sup_{\nu \in \mathbb{R}} \left\{ \frac{1}{\epsilon} \mathbb{E}_{\mathbf{R} \sim Q} \left(\min\left(-Z + \nu, 0\right)\right) - \nu \right\}.$$

It is sometimes called the average value-at-risk

$$\text{AV@R}_\epsilon(Z) := \frac{1}{\epsilon} \int_0^\epsilon \text{V@R}_\nu(Z) d\nu$$

or the expected shortfall

$$\text{ES}_\epsilon := \mathbb{E}(Z | Z < \text{V@R}_\epsilon(Z)),$$

which can easily be obtained by substituting $\text{V@R}_\nu(Z) = F_Z^{-1}(\nu) = t$ into the integral of the AV@R_ϵ definition, where $F_Z(\cdot)$ represents the distribution function of the variable Z. The ES_ϵ is also shown to be equivalent to the CV@R_ϵ in Rockafellar and Uryasev (1999).

Replacing $-\text{V@R}_\epsilon$ with $-\text{CV@R}_\epsilon$, the Basel Committee's recommended risk measure, in (1.30) obtains a safe concave approximation, provided that $\mathbb{E}_{\mathbf{R} \sim Q}\{\|\mathbf{R}\|_2\}$ is bounded. To see this, note that the constraint

$$-\text{V@R}_\epsilon(\mathbf{w}^\mathrm{T}\mathbf{R} - t) < 0$$
$$\Leftrightarrow \quad \mathbb{P}_{\mathbf{R} \sim Q}\{\mathbf{w}^\mathrm{T}\mathbf{R} - t \leq 0\} < \epsilon$$
$$\Leftrightarrow \quad \mathbb{P}_{\mathbf{R} \sim Q}\{t - \mathbf{w}^\mathrm{T}\mathbf{R} > 0\} < \epsilon$$
$$\Leftrightarrow \quad \int \chi\left(t - \mathbf{w}^\mathrm{T}\mathbf{R}\right) dQ(\mathbf{R}) := p(\mathbf{w}, t) < \epsilon \qquad (1.31)$$

where $\chi(s)$ is the characteristic function

$$\chi(s) = \begin{cases} 0, & s < 0 \\ 1, & s \geq 0. \end{cases}$$

Due to the fact that $\chi(\cdot)$ is not a convex function, (1.31) is not necessarily a convex constraint. However, if we let $g(\cdot)$ be a convex function that is at least as large as $\chi(\cdot)$ everywhere, then

$$\int g\left(t - \mathbf{w}^{\mathrm{T}}\mathbf{R}\right)dQ(\mathbf{R}) := \Psi(\mathbf{w}, t) < \epsilon \tag{1.32}$$

is a safe approximation to (1.31), since

$$p(\mathbf{w}, t) \leq \Psi(\mathbf{w}, t) < \epsilon$$

so that whenever (1.32) is satisfied, so is (1.31). (1.32) is also a convex constraint because

$$\begin{aligned}
\Psi\left(\tau\mathbf{u} + (1-\tau)\mathbf{w}, \tau s + (1-\tau)t\right) &= \int g\left(\tau s + (1-\tau)t - (\tau\mathbf{u} + (1-\tau)\mathbf{w})^{\mathrm{T}}\mathbf{R}\right)dQ(\mathbf{R}) \\
&= \int g\left(\tau\left(s - \mathbf{u}^{\mathrm{T}}\mathbf{R}\right) + (1-\tau)\left(t - \mathbf{w}^{\mathrm{T}}\mathbf{R}\right)\right)dQ(\mathbf{R}) \\
&\leq \int \tau g\left(s - \mathbf{u}^{\mathrm{T}}\mathbf{R}\right) + (1-\tau)g\left(t - \mathbf{w}^{\mathrm{T}}\mathbf{R}\right)dQ(\mathbf{R}) \\
&= \tau\Psi(\mathbf{u}, s) + (1-\tau)\Psi(\mathbf{w}, t).
\end{aligned}$$

for all $(\mathbf{u}, s), (\mathbf{w}, t) \in \mathbb{R}^{n+1}$ and $\tau \in [0, 1]$.

Notice the inequality

$$v\Psi\left(v^{-1}(\mathbf{w}, t)\right) - v\epsilon < 0,$$

where $v > 0$ is a variable, is also a safe convex approximation to (1.31) since

$$g\left(v^{-1}s\right) \geq \chi\left(v^{-1}s\right) = \chi(s) \quad \forall v \in \mathbb{R}_{++}$$

so that

$$p(\mathbf{w}, t) \leq \Psi(v^{-1}(\mathbf{w}, t)),$$

and $v\Psi\left(v^{-1}(\mathbf{w}, t)\right)$ is the perspective function of the convex function $\Psi(\mathbf{w}, t)$.[1]

This implies

$$\exists v > 0 : v\Psi(v^{-1}(\mathbf{w}, t)) - v\epsilon < 0$$

is a safe convex approximation of (1.31). Now assume $g(\cdot)$ is a generator (a nonnegative increasing function such that $g(0) \geq 1$ and $\lim_{s \to -\infty} g(s) \to 0$), then it can be shown by using the lower semicontinuity of $\Psi(\cdot)$ and the construction of $G(\mathbf{w}, t)$ that the weaker condition

$$\inf_{v>0}\left\{v\Psi(v^{-1}(\mathbf{w}, t)) - v\epsilon\right\} := G(\mathbf{w}, t) < 0 \tag{1.33}$$

is also a safe convex approximation of (1.31). If we let $g_*(s) = \max\left[1 + s, 0\right]$ and $\Psi_*(\mathbf{w}, t) = \mathbb{E}_{\mathbf{R} \sim Q}\left\{g_*\left(t - \mathbf{w}^{\mathrm{T}}\mathbf{R}\right)\right\}$ be the generator and safe convex upper bound for $p(\mathbf{w}, t)$ respectively, then

$$G_*(\mathbf{w}, t) < 0$$

---

[1] The perspective of an $f : \mathbb{R}^n \to \mathbb{R}$ is the function $g(x, t) = tf(x/t)$ with domain $\{(x, t)|x/t \in \mathbf{dom}\, f, t > 0\}$. It is a well known fact that the perspective operation preserves convexity: if f is convex, then g is also convex.

$$\Leftrightarrow \quad \inf_{v>0}\left\{v\mathbb{E}_{\mathbf{R}\sim Q}\left(\max\left\{1+v^{-1}\left(t-\mathbf{w}^{\mathrm{T}}\mathbf{R}\right),0\right\}\right)-v\epsilon\right\}<0$$

$$\Leftrightarrow \quad \inf_{v>0}\left\{\mathbb{E}_{\mathbf{R}\sim Q}\left(\max\left\{v+t-\mathbf{w}^{\mathrm{T}}\mathbf{R},0\right\}\right)-v\epsilon\right\}<0$$

$$\Leftrightarrow \quad \inf_{v>0}\left\{\frac{1}{\epsilon}\mathbb{E}_{\mathbf{R}\sim Q}\left(\max\left\{v+t-\mathbf{w}^{\mathrm{T}}\mathbf{R},0\right\}\right)-v\right\}<0$$

$$\Leftrightarrow \quad \inf_{v<0}\left\{\frac{1}{\epsilon}\mathbb{E}_{\mathbf{R}\sim Q}\left(\max\left\{t-\mathbf{w}^{\mathrm{T}}\mathbf{R}-v,0\right\}\right)+v\right\}<0$$

$$\Leftrightarrow \quad \inf_{v\in\mathbb{R}}\left\{\frac{1}{\epsilon}\mathbb{E}_{\mathbf{R}\sim Q}\left(\max\left\{t-\mathbf{w}^{\mathrm{T}}\mathbf{R}-v,0\right\}\right)+v\right\}<0$$

$$\Leftrightarrow \quad -\sup_{v\in\mathbb{R}}\left\{\frac{1}{\epsilon}\mathbb{E}_{\mathbf{R}\sim Q}\left(\min\left\{\mathbf{w}^{\mathrm{T}}\mathbf{R}-t+v,0\right\}\right)-v\right\}<0$$

$$\Leftrightarrow \quad -\mathrm{CV@R}_{\epsilon}(\mathbf{w}^{\mathrm{T}}\mathbf{R}-t)<0,$$

which proves our case. Note that $g_*(s)$ and $\Psi_*(\mathbf{w},t)$ are the least conservative generator and safe convex upper bound for $p(\mathbf{w},t)$ respectively, and the $\mathrm{CV@R}_{\epsilon}$ constraint is serendipitously the best known safe convex approximation to the $\mathrm{V@R}_{\epsilon}$ constraint. However, calculating the $\mathrm{CV@R}_{\epsilon}$ requires multi-dimensional integration, which is normally intractable. The only practical way to compute $\mathrm{CV@R}_{\epsilon}$ is via Monte Carlo simulation, which is also time consuming, especially when ϵ is small. The only generic case in which the $\mathrm{CV@R}_{\epsilon}$ calculation is tractable is when the support of Q is a finite set $\{\mathbf{R}^1,\ldots,\mathbf{R}^N\}$ so that

$$\Psi_*(\mathbf{w}) = \sum_{i=1}^{N} \pi_i\max\left\{0,1+t-\mathbf{w}^{\mathrm{T}}\mathbf{R}^i\right\},$$

where N is a moderate positive integer and $\pi_i = \mathbb{P}\{\mathbf{R}=\mathbf{R}^i\}$.

1.7.3 Entropic Value-At-Risk

The entropic value-at-risk

$$\mathrm{EV@R}_{\epsilon}(Z) = \sup_{v>0}\left\{-\frac{1}{v}\log(M_Z(-v)/\epsilon)\right\}$$

is a recently introduced coherent utility measure [3] whose negation is also a safe convex upper bound to the negated value-at-risk. To see this, first note that the general Chernoff bound is

$$\mathbb{P}(Z\leq k) \leq \exp\{vk\}M_Z(-v), v>0$$

where $M_Z(\cdot)$ is the moment-generating function of Z. Solving $\exp\{vk\}M_Z(-v) = \epsilon$ for k yields

$$k_Z(v) = -\frac{1}{v}\log(M_Z(-v)/\epsilon)$$

so that

$$\mathbb{P}\left(Z\leq -\frac{1}{v}\log(M_Z(-v)/\epsilon)\right)\leq \epsilon.$$

This implies that $-\frac{1}{\nu}\log(M_Z(-\nu)/\epsilon)$ is a lower bound to $\text{V@R}_\epsilon(Z)$ for all values of $\nu > 0$, so that by the definition of the entropic value-at-risk we have $-\text{V@R}_\epsilon \leq -\text{EV@R}_\epsilon$. In fact, if we let $g(s) = \exp(s)$ and $\Psi(\mathbf{w}, t) = \mathbb{E}_{\mathbf{R} \sim Q}\{\exp(t - \mathbf{w}^\mathsf{T}\mathbf{R})\}$, then

$$
\begin{aligned}
& G(\mathbf{w}, t) < 0 \\
\Leftrightarrow\quad & \inf_{\nu > 0}\left\{\nu\mathbb{E}_{\mathbf{R} \sim Q}\left(\exp\left\{\nu^{-1}\left(t - \mathbf{w}^\mathsf{T}\mathbf{R}\right)\right\}\right) - \nu\epsilon\right\} < 0 \\
\Leftrightarrow\quad & \inf_{\nu > 0}\left\{\mathbb{E}_{\mathbf{R} \sim Q}\left(\exp\left\{\nu^{-1}\left(t - \mathbf{w}^\mathsf{T}\mathbf{R}\right)\right\}\right)\right\} < \epsilon \\
\Leftrightarrow\quad & \mathbb{E}_{\mathbf{R} \sim Q}\left(\exp\left\{\nu_*^{-1}\left(t - \mathbf{w}^\mathsf{T}\mathbf{R}\right)\right\}\right) < \epsilon \\
\Leftrightarrow\quad & \log\mathbb{E}_{\mathbf{R} \sim Q}\left(\exp\left\{\nu_*^{-1}\left(t - \mathbf{w}^\mathsf{T}\mathbf{R}\right)\right\}\right) < \log\epsilon \\
\Leftrightarrow\quad & \inf_{\nu > 0}\left\{\log\mathbb{E}_{\mathbf{R} \sim Q}\left(\exp\left\{\nu^{-1}\left(t - \mathbf{w}^\mathsf{T}\mathbf{R}\right)\right\}\right)\right\} < \log\epsilon \\
\Leftrightarrow\quad & \inf_{\nu > 0}\left\{\log(M_{t - \mathbf{w}^\mathsf{T}\mathbf{R}}(\nu^{-1})/\epsilon)\right\} < 0 \\
\Leftrightarrow\quad & \inf_{\nu^{-1} > 0}\left\{\nu\log(M_{t - \mathbf{w}^\mathsf{T}\mathbf{R}}(\nu^{-1})/\epsilon)\right\} < 0 \\
\Leftrightarrow\quad & \inf_{\nu > 0}\left\{\frac{1}{\nu}\log(M_{t - \mathbf{w}^\mathsf{T}\mathbf{R}}(\nu)/\epsilon)\right\} < 0 \\
\Leftrightarrow\quad & -\sup_{\nu > 0}\left\{-\frac{1}{\nu}\log(M_{\mathbf{w}^\mathsf{T}\mathbf{R} - t}(-\nu)/\epsilon)\right\} < 0 \\
\Leftrightarrow\quad & -\text{EV@R}_\epsilon(\mathbf{w}^\mathsf{T}\mathbf{R} - t) < 0,
\end{aligned}
$$

where ν_* is the value which achieves the infimum for $\mathbb{E}_{\mathbf{R} \sim Q}\left(\exp\left\{\nu^{-1}\left(t - \mathbf{w}^\mathsf{T}\mathbf{R}\right)\right\}\right)$ and its logarithm. Therefore, the EV@R_ϵ constraint is also a safe convex approximation to the V@R_ϵ constraint. Note that since $\exp(s) \geq \max[1 + s, 0] \geq \chi(s)$, we have

$$
-\text{V@R}_\epsilon \leq -\text{CV@R}_\epsilon \leq -\text{EV@R}_\epsilon.
$$

Other than the need for the existence and knowledge of the moment-generating function of the portfolio return, which is not always the case, $-\text{EV@R}_\epsilon$ is a very conservative upper bound of the $-\text{V@R}_\epsilon$, due to the exponential generator being used.

1.8   Robust Optimization

[25, Chapters 1- 4] Apart from the absence of higher moments is the absence of model uncertainty in the Markowitz model. Garlappi et al. (2007), Amarov and Zhou (2010), and Harvey et al. (2011) treat this issue with a Bayesian flavor. Their recurring theme is to choose a prior for the distribution of returns, find the posterior, and then calculate and maximize the expected utility function using Bayesian methods. We are more interested in the concept of robust optimization, or some call it data-driven optimization, where model uncertainty is taken into account in the optimization process.

In classical optimization, model uncertainty is usually ignored, and the problem is solved under the assumption that there is perfect information, in the hope that this will not affect

the feasibility and optimality of the solutions significantly, and that minor adjustments of the nominal solution would suffice. However, these hopes are not always justified, and even a little model uncertainty may deserve our attention. For examples to illustrate this point, refer to Ben-Tal et al. (2009).

1.8.1 Uncertain Linear Optimization

A linear optimization problem is defined as

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \mathbf{c}^\mathsf{T} \mathbf{x} + d : \mathbf{A}\mathbf{x} \leq \mathbf{b} \right\} \tag{1.34}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$ and $d \in \mathbb{R}$. The data of the problem is the collection $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$ and the structure of (1.34) is determined by the number of constraints m and the number of variables n.

An uncertain linear optimization problem is defined as

$$\left\{ \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \mathbf{c}^\mathsf{T} \mathbf{x} + d : \mathbf{A}\mathbf{x} \leq \mathbf{b} \right\} : (\mathbf{A}, \mathbf{b}, \mathbf{c}, d) \in \mathcal{U} \right\}, \tag{1.35}$$

a collection of linear optimization problems with the same number of constraints and variables where

$$\mathcal{U} = \left\{ \left[\begin{array}{c|c} \mathbf{c}^\mathsf{T} & d \\ \hline \mathbf{A} & \mathbf{b} \end{array} \right] = \left[\begin{array}{c|c} \mathbf{c}_0^\mathsf{T} & d_0 \\ \hline \mathbf{A}_0 & \mathbf{b}_0 \end{array} \right] + \sum_{\ell=1}^{L} \zeta_\ell \left[\begin{array}{c|c} \mathbf{c}_\ell^\mathsf{T} & \mathbf{d}_\ell \\ \hline \mathbf{A}_\ell & \mathbf{b}_\ell \end{array} \right] : \boldsymbol{\zeta} \in \mathcal{Z} \subset \mathbb{R}^L \right\}$$

is an uncertainty set parameterized in an affine fashion by a vector $\boldsymbol{\zeta}$ varying in a given perturbation set $\mathcal{Z}$ [25, p. 7]. A vector $\mathbf{x} \in \mathbb{R}^n$ is called a robust feasible solution to (1.35) if it satisfies all realizations of the constraints from the uncertainty set, that is, $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ for all $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d) \in \mathcal{U}$. The robust value of the objective function in (1.35) at a robust feasible solution $\mathbf{x}$ is the largest value of $\mathbf{c}^\mathsf{T} \mathbf{x} + d$ over all realizations of the data in $\mathcal{U}$. The robust counterpart of (1.35) is the problem of minimizing the robust value over all the robust feasible solutions of (1.35), and can be written as

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \max_{(\mathbf{A}, \mathbf{b}, \mathbf{c}, d) \in \mathcal{U}} [\mathbf{c}^\mathsf{T} \mathbf{x} + d] : \mathbf{A}\mathbf{x} \leq \mathbf{b} \; \forall (\mathbf{A}, \mathbf{b}, \mathbf{c}, d) \in \mathcal{U} \right\}. \tag{1.36}$$

An optimal solution and value of (1.36) are called a robust optimal solution and value of (1.35) respectively [25, p. 9]. Essentially, we want to obtain the best of the worst objective functions, each calculated at a solution that remains feasible in the worst possible scenario. Note that (1.36) can be rewritten as

$$\min_{\mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : \begin{array}{c} \mathbf{c}^\mathsf{T} \mathbf{x} - t \leq -d \\ \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{array} \right\} \forall (\mathbf{A}, \mathbf{b}, \mathbf{c}, d) \in \mathcal{U} \right\},$$

where the uncertain objective is pushed into the constraints. Therefore, we lose nothing if we restrict ourselves to uncertain linear optimzation programs with a certain objective, and write the robust counterpart as

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{ \mathbf{c}^\mathsf{T} \mathbf{x} + d : \mathbf{A}\mathbf{x} \leq \mathbf{b} \; \forall (\mathbf{A}, \mathbf{b}) \in \mathcal{U} \}. \tag{1.37}$$

Note that by definition, (1.37) remains intact when the original uncertainty set $\mathcal{U}$ is extended to the direct product

$$\hat{\mathcal{U}} = \mathcal{U}_1 \times \ldots \times \mathcal{U}_m$$

where

$$\mathcal{U}_i = \{[\mathbf{a}_i; b_i] : [\mathbf{A}, \mathbf{b}] \in \mathcal{U}\}$$

is the projection of $\mathcal{U}$ onto the data space of the i-th constraint. Therefore, the constraints in (1.37) can be replaced with

$$\mathbf{a}_i^\mathrm{T} \mathbf{x} \le b_i \; \forall [\mathbf{a}_i, b_i] \in \mathcal{U}_i \tag{1.38}$$

where $\mathbf{a}_i^\mathrm{T}$ is the i-th row of $\mathbf{A}$, for $i = 1, \ldots, m$. If $\mathbf{x}$ is a robust feasible solution of (1.38), then $\mathbf{x}$ remains robust feasible when we extend the uncertainty set to its convex hull $\mathrm{Conv}(\mathcal{U}_i)$. To see this, note that if $[\bar{\mathbf{a}}_i; \bar{b}_i] \in \mathrm{Conv}(\mathcal{U}_i)$, then

$$[\bar{\mathbf{a}}_i; \bar{b}_i] = \sum_{j=1}^{J} \tau_j [\mathbf{a}_i^j; b_i^j]$$

with appropriately chosen $[\mathbf{a}_i^j; b_i^j] \in \mathcal{U}_i$ and $\tau_j \ge 0$ such that $\sum_j \tau_j = 1$. We now have

$$\bar{\mathbf{a}}_i^\mathrm{T} \mathbf{x} = \sum_{j=1}^{J} \tau_j [\mathbf{a}_i^j]^\mathrm{T} x \le \sum_{j=1}^{J} \tau_j b_i^j = \bar{b}_i, \tag{1.39}$$

where the inequality is due to the fact that $\mathbf{x}$ is feasible for (1.38) and $[\mathbf{a}_i^j; b_i^j] \in \mathcal{U}_i$. Using similar arguments, the set of robust feasible solutions to (1.38) remains intact when we extend $\mathcal{U}_i$ to its closure. Combining the observations above, we conclude that nothing is lost if, right from the beginning, $\mathcal{U}$ is replaced by the direct product $\hat{\mathcal{V}} = \mathcal{V}_1 \times \ldots \times \mathcal{V}_m$ where $\mathcal{V}_i$ is the closed convex hull of $\mathcal{U}_i$ [25, p. 10-13]. Skipping all details which the reader is referred to the first chapter of Ben-Tal et al. (2009), for some uncertainty structures we are then able to express each uncertain linear constraint as a finite set of explicit convex constraints, and reformulate the robust counterpart (1.37) as a computationally tractable[2] convex program, which unfortunately is very conservative and often leads to uninformative and impractical results where there might not even be a single feasible solution! The answer to this problem is to consider stochastic perturbations of the data and replace the ith uncertain linear inequality (1.38) with the chance constraint [25, p. 29-30]

$$\mathbb{P}_{\zeta \sim Q} \left\{ \zeta : [\mathbf{a}^0]^\mathrm{T} \mathbf{x} + \sum_{\ell=1}^{L} \zeta_\ell [\mathbf{a}^\ell]^\mathrm{T} \mathbf{x} \le b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell \right\} \ge 1 - \epsilon \tag{1.40}$$

where ζ is a random vector with probability distribution Q and $\epsilon \in (0,1)$ is a small tolerance level.

---

[2]Refer to Appendix A of [25] for a mathematical treatment of tractability.

To deal with uncertainty, we may assume that only partial information is known about the distribution of ζ so that $Q \in \mathcal{Q}$, where $\mathcal{Q}$ is a set of probability measures. In this case, we require

$$\forall Q \in \mathcal{Q} : \mathbb{P}_{\zeta \sim Q} \left\{ \zeta : [\mathbf{a}^0]^{\mathrm{T}} \mathbf{x} + \sum_{\ell=1}^{L} \zeta_\ell [\mathbf{a}^\ell]^{\mathrm{T}} \mathbf{x} \leq b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell \right\} \geq 1 - \epsilon \tag{1.41}$$

which we shall call the ambiguous chance constraint.

1.8.2 Bernstein Approximation

[25, p. 83-86] The equivalence between the Markowitz model and the portfolio return value-at-risk optimization problem means that we can not only add higher moments, but also model uncertainty into the former implicitly by incorporating them into the distribution of returns of the latter. We then end up with the collection of problems

$$\left\{ \max_{\mathbf{w} \in \mathbb{R}^n} \left\{ \mathrm{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}} \mathbf{R}) : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \ell, \mathbf{R} \sim Q \right\} : Q \in \mathcal{Q} \right\}$$

whose robust counterpart is

$$\max_{\mathbf{w} \in \mathbb{R}^n} \left\{ \min_{Q \in \mathcal{Q}} \left\{ \mathrm{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}} \mathbf{R}) : \mathbf{R} \sim Q \right\} : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \ell \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \ell, t \leq \min_{Q \in \mathcal{Q}} \left\{ \mathrm{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}} \mathbf{R}) : Q \in \mathcal{Q} \right\} \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \ell, t \leq \left\{ \mathrm{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}} \mathbf{R}) : Q \in \mathcal{Q} \right\} \ \forall Q \in \mathcal{Q} \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w} \in \mathbb{R}^n, t \in \mathbb{R}} \left\{ t : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \ell, \mathbb{P}_{\mathbf{R} \sim Q} \{ t - \mathbf{w}^{\mathrm{T}} \mathbf{R} > 0 \} \leq \epsilon \ \forall Q \in \mathcal{Q} \right\}. \tag{1.42}$$

If we let

$$\Psi^+(\mathbf{w}, t) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_{\mathbf{R} \sim Q} \left\{ g_* \left(t - \mathbf{w}^{\mathrm{T}} \mathbf{R} \right) \right\}$$

and assume $\mathbb{E}_{\mathbf{R} \sim Q} \{ \|\mathbf{R}\|_2 \}$ is uniformly bounded in $\mathcal{Q}$, then

$$\forall Q \in \mathcal{Q} : G(\mathbf{w}, t) \leq 0 \Leftrightarrow \mathrm{CV@R}_\epsilon(t - \mathbf{w}^{\mathrm{T}} \mathbf{R}) \leq 0$$

is the least conservative safe convex approximation to the ambiguous chance constraint of (1.42). Although convex, a problem with such infinitely many non-linear constraints is NP-hard. Fortunately, we are able to obtain a safe, convex and tractable approximation of (1.42). Assume

(i) $g(s) = \exp\{s\}$,

(ii) $\mathcal{Q}$ comprises of all product-type probability distributions $Q = Q_1 \times \ldots \times Q_n$ with marginals $Q_i \in \mathcal{Q}_i$ running independently in their respective families $\mathcal{Q}_i$, where $\mathcal{Q}_i$ is a given family of probability distributions on $\mathbb{R}$, $i = 1, \ldots, n$,

(iii) the functions $\Phi_i^*(s) := \sup_{Q_i \in \mathcal{Q}_i} \log \mathbb{E}_{x \sim Q_i}\{\exp\{xs\}\}$ are convex, lower semicontinuous such that $0 \in \mathrm{intDom}\ \Phi_i^*$, and

(iv) efficiently computable lower semicontinuous convex functions $\Phi_i^+(\cdot) \geq \Phi_i^*(\cdot)$ such that $0 \in \mathrm{intDom}\ \Phi_i^+$ are available.

Under these assumptions,

$$\mathbb{E}_{\mathbf{R} \sim Q}\left\{\exp\left\{t - \mathbf{w}^\mathsf{T}\mathbf{R}\right\}\right\} = \exp\{t\} \prod_{i=1}^{n} \mathbb{E}_{r_i \sim Q_i}\{\exp\{-w_i r_i\}\} \leq \exp\{t\} \prod_{i=1}^{n} \exp\left\{\Phi_i^*(-w_i)\right\}$$

and by setting

$$\Psi^+(\mathbf{w}, t) = \exp\{t\} \prod_{i=1}^{n} \exp\left\{\Phi_i^+(-w_i)\right\},$$

it is easily seen that the condition

$$\exists v > 0 : \Psi^+(v^{-1}(\mathbf{w}, t)) \leq \epsilon$$

is sufficient to satisfy (1.42) and can be rewritten as

$$\exists v > 0 : \log \Psi^+\left(v^{-1}(\mathbf{w}, t)\right) \leq \log \epsilon$$

$$\Leftrightarrow \quad \exists v > 0 : t + v \sum_{i=1}^{n} \Phi_i^+(-v^{-1} w_i) + v \log(1/\epsilon) \leq 0,$$

which can be weakened to

$$t + \inf_{v > 0}\left\{v \sum_{i=1}^{n} \Phi_i^+(-v^{-1} w_i) + v \log(1/\epsilon)\right\} \leq 0, \tag{1.43}$$

also known as the Bernstein approximation.

Remark 1.1

Notice that if each $\mathcal{Q}_i$ is a singleton, then the left-hand side of (1.43) is exactly $-\mathrm{EV@R}_\epsilon$. However, the $\mathrm{EV@R}_\epsilon$ constraint is not a special case of the Bernstein approximation, since the latter requires independence amongst the asset returns while the former does not.

Although tractable, this approximation is also very conservative due to the exponential generator chosen. To reduce conservativeness, we could artifically increase the tolerance level ϵ, but this is somewhat arbitrary. Efforts have been made to bridge the gap between the $\mathrm{CV@R}_\epsilon$ and Bernstein approximations [25, p. 97-100], but these methods are rather difficult to implement. The Lagrangian approximation from Bertsimas et al. (2000), and Bertsimas and Popescu (2005) can be used to include correlations, but nonetheless suffers from the same drawback of being too conservative, albeit computationally tractable.

1.9 Thesis Outlook

We will treat the absence of higher moments and model uncertainty in the Markowitz model by solving

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{Q\in\mathcal{Q}} \left\{ \mathrm{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}}\mathbf{R}) : \mathbf{R}\sim Q \right\} \tag{1.44}$$

where $\left\{\mathbf{w}\in\mathbb{R}^n : \mathbf{w}^{\mathrm{T}}\mathbb{1}=1, \mathbf{w}\geq\ell\right\}$, and for which the solutions shown thus far are either too conservative or intractable. Since the methods we develop can easily be extended to the conditional value-at-risk, in line with the Basel Committee's advice, or even the entropic value-at-risk, all our numerical implementations will only consider the value-at-risk for the sake of convenience. Chapter 2 proposes a spline approximation method where the smoothed minimum (over a finite set of distributions of the asset returns) portfolio return sample value-at-risk is maximized over the feasible domain. In Chapter 3, we work with elliptical distributions such that (1.44) possesses a location-scale form. Lobo and Boyd (2000), El Ghaoui et al. (2003), Goldfarb and Iyengar (2003), Lobo et al. (2007), Natarajan et al. (2008 and 2010), Ye et al. (2012), Zymler et al. (2013) and Rujeerapaiboon et al. (2015) all do some related work, but we introduce a novel uncertainty set for the scale matrix in the positive definite space where the eigenvalues vary in a box uncertainty set and the eigenvectors each varies in a cone uncertainty set with orthogonality preserved among them, so that the robust counterpart of the location-scale problem can be converted into a semi-definite program (SDP) which is solvable in polynomial time. In Chapter 4, we invent a method to determine the size of the uncertainty sets we use in the robust location-scale problem and perform numerical experiments on some real data. Chapter 5 includes trading costs and integer constraints into the robust location-scale problem and converts the resulting model into a mixed-integer program. Chapter 6 talks about portfolio diversification properties and Chapter 7 concludes.

Chapter 2

Spline Approximation

In this chapter, we solve the maxmin $V@R_\epsilon$ problem (1.44) using an additive spline approximation to the minimal $V@R_\epsilon$. The minimal $V@R_\epsilon$ over a finite set of scenarios is found by simulating quantile values at each point of the discretized set of the feasible domain (see below) and then fitting the minimal quantile with an additive second degree spline, which can then be maximized by any optimizer.

Gaivoronski and Pflug (2005) investigate a related method for finding the portfolio by maximizing the $V@R_\epsilon$. In their case, the V@R is approximated by a weighted sum of simulated or observed portfolio returns, with weights that depend on a smoothing parameter chosen by the user. The goal is to filter out the local noise and to be left with the global component. They do not take into account model uncertainty.

We assume that $\mathcal{Q}$ is a finite set containing M distributions of the return vector. Then, we estimate the minimum value-at-risk function

$$\text{MV@R}_\epsilon(\mathbf{w}) = \min_{Q \in \mathcal{Q}} \text{V@R}_\epsilon(\mathbf{w}^{\mathsf{T}}\mathbf{R})$$

by the minimum sample value-at-risk function

$$\text{MSV@R}_\epsilon(\mathbf{w}) = \min_{Q \in \mathcal{Q}} \min^{\lfloor K\epsilon \rfloor + 1} \left\{ \mathbf{w}^{\mathsf{T}}\mathbf{r}_1(Q), \ldots, \mathbf{w}^{\mathsf{T}}\mathbf{r}_K(Q) \right\} \tag{2.1}$$

based on the simulated return vectors $r_1(Q), \ldots, r_K(Q)$ for each $Q \in \mathcal{Q}$. In the above formula, $min^u(s_1, \ldots, s_n)$ denotes the uth smallest value among $s_1, \ldots, s_n$. Finally, we compute an additive spline that approximates $\text{MSV@R}_\epsilon(\mathbf{w})$ in the feasible domain $\mathcal{W}_\ell$, before maximizing it.

2.1 Additive Spline Approximation

We use univariate quadratic B-splines to approximate the function $\text{MSV@R}_\epsilon(\mathbf{w})$. First, the feasible domain $\mathcal{W}_\ell$ is discretized into

$$\mathbb{W}_\ell = \left\{ \mathbf{w} : w_i \in G_i, i = 1, \ldots, n-1, \text{ and } \mathbf{w}^{\mathsf{T}}\mathbb{1} = 1 \right\}$$

via evenly spaced grids along the axes, $G_i = \{\ell_i, \ell_i + \Delta, \ell_i + 2\Delta, \ldots, 1 - \sum_{j \neq i} \ell_i\}$, where $\Delta = (1 - \ell_1 - \cdots - \ell_n)/(d-1)$ is the spacing between adjacent nodes in each direction such that d is the number of nodes in each direction, freely chosen by the user. Note that the length of the interval in which each asset weight may vary is the same at $1 - \sum_{i=1}^{n} \ell_i$. We may build $\mathbb{W}_\ell$ using the following algorithm, for $n \geq 2$:

Algorithm 2.1

1. Let $\mathbb{W}_\ell = G_1$.

2. If $n = 2$, stop algorithm.

3. Otherwise, repeat n-2 times:

i. Let $\mathbb{W} = \mathbb{W}_\ell$.

ii. For each element e in $\mathbb{W}$:

$$Let\ \mathbb{W}_\ell = (\mathbb{W}_\ell \backslash e) \cup \left\{ \begin{bmatrix} e \\ \ell_{i+1} \end{bmatrix}, \begin{bmatrix} e \\ \ell_{i+1} + \Delta \end{bmatrix}, \begin{bmatrix} e \\ \ell_{i+1} + 2\Delta \end{bmatrix}, \ldots, \begin{bmatrix} e \\ 1 - e^{\mathrm{T}} \mathbb{1} \end{bmatrix} \right\}.$$

The cardinality of $\mathbb{W}_\ell$ can be computed with the help of Pascal's triangle. It is the sum of the first d elements of the $(n-1)$th diagonal parallel to the triangle's edge and including the first d values, that is,

$$\sum_{j=0}^{d-1} \binom{n+j-2}{j}.$$

This can be shown to be of order $\mathcal{O}(n^{d-1})$ and thus only grows polynomially with the number of assets.

Let $\mathbf{w}_j = \left[w_{j1}, \ldots, w_{j,n-1}, 1 - \sum_{i=1}^{n-1} w_{ji} \right]^{\mathrm{T}}$ denote the jth node. Then, $\mathrm{MSV@R}_\epsilon(\mathbf{w}_j)$ is approximated by the additive model

$$\sum_{i=1}^{n-1} f_i(w_{ji})$$

for $j = 1, \ldots, |\mathbb{W}_\ell|$ where

$$f_i(w_{ji}) = \sum_{k=1}^{q} \beta_{ik} B_k^1(w_{ji})$$

is the q-parameter quadratic B-spline. To fix the B-spline basis, we have to choose $q-3$ internal knots and the two endpoints of the feasible domain. The kth basis function of order o is defined recursively ([57] and [122]) as

$$B_k^{o+1}(w_{ji}) = \frac{w_{ji} - x_{ik}^*}{x_{i,k+2}^* - x_{ik}^*} B_k^o(w_{ji}) + \frac{x_{i,k+3}^* - w_{ji}}{x_{i,k+3}^* - x_{i,k+1}^*} B_{k+1}^o(w_{ji}),$$

with

$$B_k^{o=-1}(w_{ji}) = \begin{cases} 1, & \text{if } x_{ik}^* \leq w_{ji} < x_{i,k+1}^*, \\ 0, & \text{otherwise}, \end{cases}$$

23

where the knots $x_{i1}^* < x_{i2}^* < \ldots < x_{i,q-1}^*$ in the ith direction are evenly spaced and for the end intervals articificial knots outside of the feasible domain have to be added. Subsequently, we fit the B-spline coefficients such that the least squares distance between $\mathbf{y} = \left[\text{MSV@R}_\epsilon(\mathbf{w}_1), \ldots, \text{MSV@R}_\epsilon(\mathbf{w}_{|\mathbb{W}_\ell|}) \right]^{\text{T}}$ and $\mathbf{W}\boldsymbol{\beta}$ is minimized. The j-th row of $\mathbf{W}$ is

$$\mathbf{W}_j = \left[B_1^1(w_{j1}), \ldots, B_q^1(w_{j1}), \ldots, B_1^1(w_{j,n-1}), \ldots, B_q^1(w_{j,n-1}) \right]$$

and

$$\boldsymbol{\beta} = \left[\beta_{11}, \ldots, \beta_{1q}, \ldots, \beta_{n-1,1}, \ldots, \beta_{n-1,q} \right]^{\text{T}}.$$

2.2 Further Comments

The smoothness of the fit is controlled by the choice of q. No additional penalty term is needed. In order to avoid numerical problems due to ill-conditioned matrices, $q(n-1)$ should be smaller than the cardinality of $\mathbb{W}_\ell$. The choice of q is an art rather than a science, and always has an associated risk of under-smoothing or over-smoothing. We find that q between 5 and 10 gives satisfactory results. Although it is rather restrictive to use the additive model $\sum_{i=1}^{n-1} f_i(w_i)$ instead of the general model $f(w_1, \ldots, w_{n-1})$, empirical evidence by Gaivoronski and Pflug (2005) shows that the global component of a portfolio value-at-risk function only has a few extrema, which suggests that the former might be sufficient. Moreover, using an additive model means that the number of basis parameters only grows linearly with the number of assets, thus reducing the number of parameters to be estimated greatly. To avoid simulating a huge number of asset return vectors from each distribution so that an accurate spline approximation can be obtained, we can employ an iterative method with successively more simulations, where after each iteration a smaller space around the maximum found for the current spline approximation is used for the next iteration. Roughness such as kinks and discontinuities cannot be captured by the spline approximation and is a source of potential inaccuracy. Finally, note that other risk measures including the CV@R$_\epsilon$ and EV@R$_\epsilon$ can be used instead of the V@R$_\epsilon$.

2.3 Numerical Examples

The purpose of this section is to illustrate the effects of model uncertainty on the max V@R$_\epsilon$ problem by solving its robust counterpart using the spline approximation method described above, and assess its strengths and weaknesses. Thus, we only involve two risky assets, which are further assumed to be bivariate normally distributed in order to compare results obtained from the method with the theoretical solutions. In each of Figures 2.1 - 2.3 and Figures B.1 - B.14 in Appendix B, short-selling is disallowed; subfigure (a) shows the optimal weight on the second asset/difference in optimal weight between the two assets and subfigure (b) the negative portfolio value-at-risk against the tolerance level ϵ; the blue line represents the case where returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; the red line represents the robust counterpart with the associated perturbation set of the parameter vector; the grey vertical line indicates the position of

$\epsilon = 0.05$ along the horizontal axis; the cyan line in subfigure (a) represents the difference between red and blue lines; the black line in subfigure (b) is calculated based on the solution without uncertainty and the parameters that give the highest possible portfolio negated value-at-risk at that solution. In each application of the spline approximation method, five million asset return vectors of each parameter in the corresponding perturbation set are simulated, the 100ϵth percent sample quantiles are calculated at $w_1 = 0, 0.05, \ldots, 1$, and a univariate second degree B-spline function with twelve parameters and knot locations $0 < 0.1 < \ldots < 1$ is used. Since the spline is fitted over a single dimension, the maximum of its values computed at $w_1 = 0, 0.001, \ldots, 1$ is taken. The approximation results are very close to the theoretical values and thus omitted in the figures. Note that all wealth is invested into the risky assets if the negated (robust) optimal value-at-risk is less than the negated risk-free interest rate, otherwise all wealth is kept in the riskless asset.

In general, the standard deviation uncertainty has a larger effect than mean uncertainty for smaller tolerance levels, which should not be surprising given that the significance of the standard deviation term diminishes in the normal quantile function compared to the mean term as the tolerance level increases. The inverse happens happens for tolerance levels nearer to 0.5. This is also not surprising, because when the tolerance is exactly 0.5, the mean term alone determines the optimal investment. The influence of the correlation on the optimal portfolio is relatively small. In any case, model uncertainty has great ramifications potentially, and we should in no way ignore them, otherwise it might give us a false sense of security by making the negative portfolio value-at-risk seem lower than it really is, as illustrated by the black line being higher than the blue line in each subfigure (b) of Figures 2.1 - 2.3 and Figures B.1 - B.14 in Appendix B.

In each of Figures 2.4 - 2.6, subfigure (a) shows the optimal weight on the second asset and subfigure (b) the negative portfolio value-at-risk against the tolerance level; the blue line represents the case in which returns follow the independent bivariate normal distribution where $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$; the cyan line represents the case in which returns follow an independent bivariate Cauchy distribution where either the fifth, fifteenth or twenty fifth percent quantile of each of its marginals coincides with that of the corresponding marginal of the aforementioned bivariate normal distribution; the red line represents the case in which the optimal weights are distributionally robust against the previous two distributions just mentioned; the green line represents the spline approximation of the previous case; the grey vertical line indicates the tolerance level of $\epsilon = 0.05$. The distributionally robust optimal solution is the same as that of the bivariate Cauchy distribution for small tolerance levels, before transiting to that of the bivariate normal distribution as the tolerance level increases. The larger the quantile being matched, the later the onset of and slower the transition. In Figures 2.5 and 2.6, inaccuracies in the approximated solution start to appear in the transition period since the splines are unable to capture the kinks in the functions to be smoothed. In each application of the spline approximation method, five million asset return vectors are simulated each from the independent bivariate normal distribution and the associated independent bivariate Cauchy distribution, the 100ϵth percent sample quantiles are calculated at $w_1 = 0, 0.05, \ldots, 1$, and a univariate second degree B-spline function with twelve parameters and knot locations $0 < 0.1 < \ldots < 1$ is used. Since the spline is fitted over a single dimension, the maximum of its values computed at $w_1 = 0, 0.001, \ldots, 1$ is taken.

Figures B.15 - B.34 in Appendix B are analogous to those mentioned above where all things remain constant except that short-selling is allowed up to a maximum of one-fifth the total wealth for each asset, and the cyan line in each subfigure (a) and the black line in each subfigure (b) are omitted. In each application of the spline approximation method, five million asset return vectors are simulated for each parameter vector in the corresponding perturbation set, the 100ϵth percent sample quantiles are calculated at $w_1 = -0.2, -0.1, \ldots, 1.2$, and a univariate second degree B-spline function with sixteen parameters and knot locations $-0.2 < -0.1 < \ldots < 1.2$ is used. Since the spline is fitted over a single dimension, the maximum of its values computed at $w_1 = -0.2, -0.199, \ldots, 1.2$ is taken. They suggest parallel observations, albeit with slightly more inaccuracies in the approximated solutions when short-selling is involved due to more kinks in the functions to be smoothed, especially at large tolerance levels.
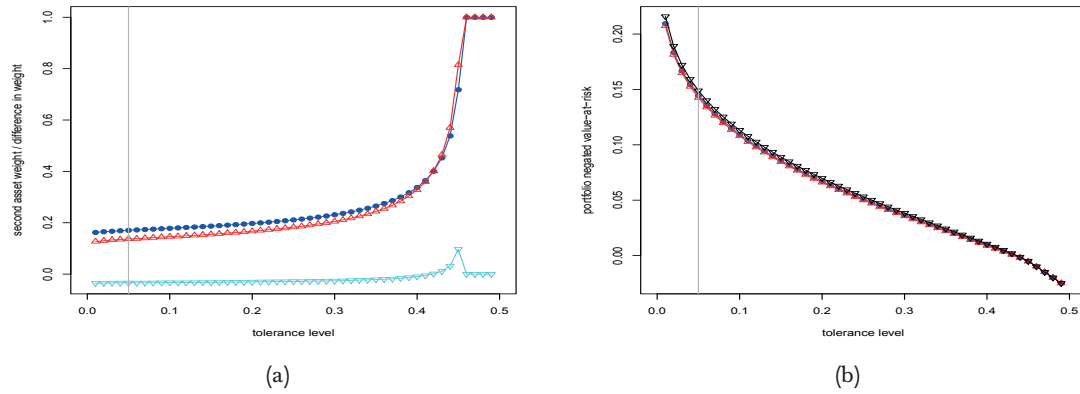
(a)   (b)
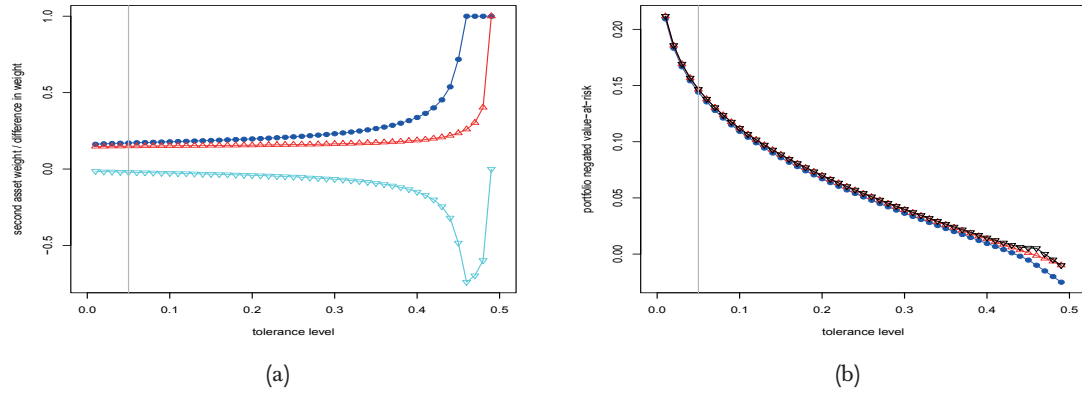
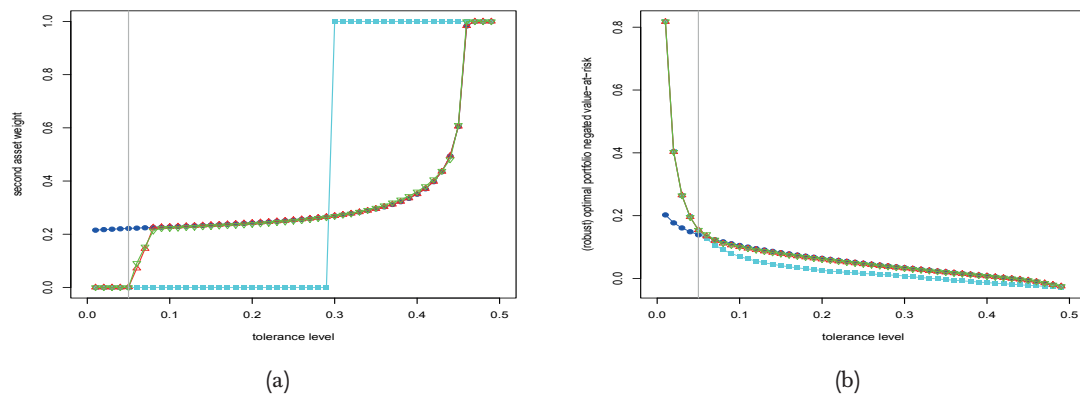Figure 2.1: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2] \in \{0.01, 0.015, 0.02, 0.025\} \times \{0.015, 0.02, 0.025, 0.03\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and corresponding parameter in the perturbation set that gives the highest possible portfolio negated value-at-risk.

(a)   (b)
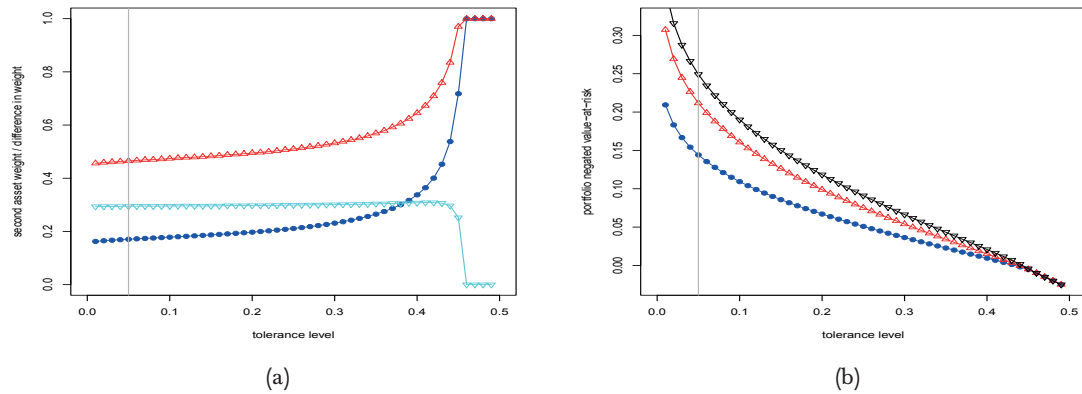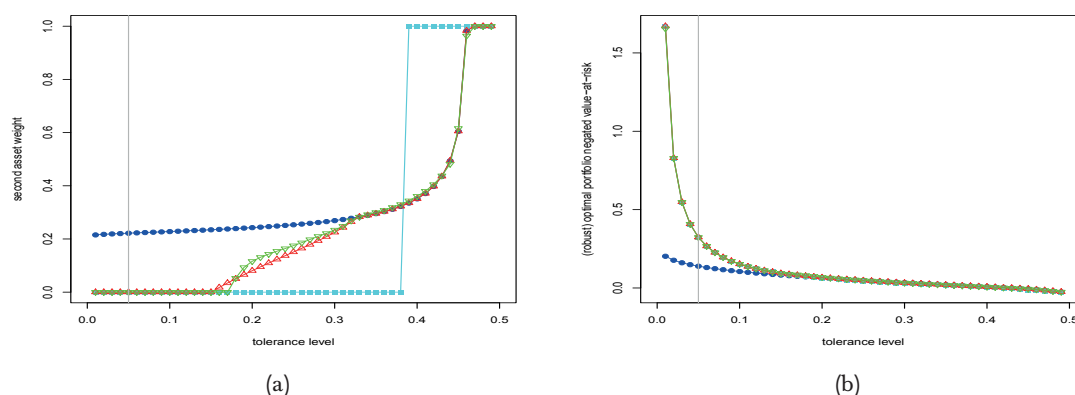
Figure 2.2: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\sigma_1, \sigma_2] \in \{0.1, 0.12, \ldots, 0.18\} \times \{0.12, 0.14, \ldots, 0.2\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and corresponding parameter in the perturbation set that gives the highest possible portfolio negated value-at-risk.

(a)   (b)

Figure 2.3: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$ where short-selling is disallowed; red - robust counterpart where $\rho \in \{0.1, 0.2, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and corresponding parameter in the perturbation set that gives the highest possible portfolio negated value-at-risk.

(a)   (b)

Figure 2.4: (a) optimal weight on second asset and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow an independent bivariate normal distribution where $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$; cyan - returns follow an independent bivariate Cauchy distribution where $m_1 = 0.01$, $m_2 = 0.03$, $\gamma_1 = 0.026$ and $\gamma_2 = 0.052$ such that the **fifth** percent quantile of each of its marginals coincides with that of the corresponding marginal of the aforementioned bivariate normal distribution; red - robust against both bivariate distributions; green - spline approximation of red case.
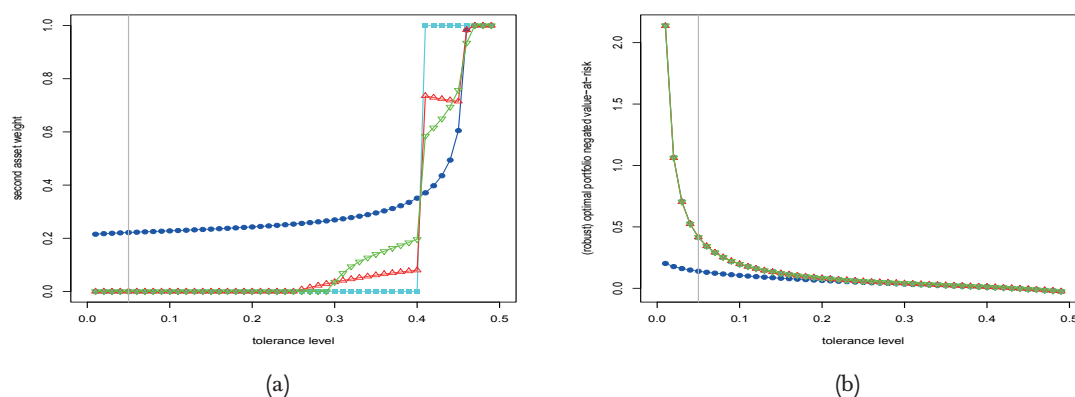
Figure 2.5: (a) optimal weight on second asset and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow an independent bivariate normal distribution where $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$; cyan - returns follow an independent bivariate Cauchy distribution where $m_1 = 0.01$, $m_2 = 0.03$, $\gamma_1 = \mathbf{0.053}$ and $\gamma_2 = \mathbf{0.106}$ such that the **fifteenth** percent quantile of each of its marginals coincides with that of the corresponding marginal of the aforementioned bivariate normal distribution; red - robust against both bivariate distributions; green - spline approximation of red case.

Figure 2.6: (a) optimal weight on second asset and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow an independent bivariate normal distribution where $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$; cyan - returns follow an independent bivariate Cauchy distribution where $m_1 = 0.01$, $m_2 = 0.03$, $\gamma_1 = \mathbf{0.067}$ and $\gamma_2 = \mathbf{0.135}$ such that the **twenty-fifth** percent quantile of each of its marginals coincides with that of the corresponding marginal of the aforementioned bivariate normal distribution; red - robust against both bivariate distributions; green - spline approximation of red case.

Chapter 3

Robust V@R$_\epsilon$ Optimization For Elliptical Distributions

The robust optimization of the V@R$_\epsilon$, CV@R$_\epsilon$ or EV@R$_\epsilon$ for elliptical distributions is of a location-scale form, which we will explore in this chapter. In particular, the problem in this general context is reformulated into a semi-definite program (SDP) under a novel uncertainty set for the scale matrix.

3.1 Multivariate Normal Distribution

Assuming that the asset return vector follows a multivariate normal distribution, the robust optimization of the V@R$_\epsilon$ becomes

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}} \left\{ \text{V@R}_\epsilon(\mathbf{w}^\mathrm{T}\mathbf{R}) : \mathbf{R}\sim\mathcal{N}(\boldsymbol{\mu},\Sigma) \right\} \tag{3.1}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}} \left\{ \mathbf{w}^\mathrm{T}\boldsymbol{\mu} - z_{1-\epsilon}\sqrt{\mathbf{w}^\mathrm{T}\Sigma\mathbf{w}} \right\}.$$

3.1.1 CV@R$_\epsilon$ Optimization

With V@R$_\epsilon$ replaced by CV@R$_\epsilon$, (3.1) is

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}} \left\{ \frac{1}{\epsilon}\int_0^\epsilon \text{V@R}_\nu(\mathbf{w}^\mathrm{T}\mathbf{R})d\nu : \mathbf{R}\sim\mathcal{N}(\boldsymbol{\mu},\Sigma) \right\}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}} \left\{ \frac{1}{\epsilon}\int_0^\epsilon \mathbf{w}^\mathrm{T}\boldsymbol{\mu} + z_\nu\sqrt{\mathbf{w}^\mathrm{T}\Sigma\mathbf{w}}d\nu : \mathbf{R}\sim\mathcal{N}(\boldsymbol{\mu},\Sigma) \right\}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}} \left\{ \mathbf{w}^\mathrm{T}\boldsymbol{\mu} + \frac{\sqrt{\mathbf{w}^\mathrm{T}\Sigma\mathbf{w}}}{\sqrt{2\pi\epsilon}}\int_{-\infty}^{z_\epsilon} x\exp\{-x^2/2\}dx : \mathbf{R}\sim\mathcal{N}(\boldsymbol{\mu},\Sigma) \right\} \quad \text{(let } z_\gamma = x)$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}} \left\{ \mathbf{w}^\mathrm{T}\boldsymbol{\mu} - \left(\exp\{-z_\epsilon^2/2\}/\sqrt{2\pi}\epsilon\right)\sqrt{\mathbf{w}^\mathrm{T}\Sigma\mathbf{w}} : \mathbf{R}\sim\mathcal{N}(\boldsymbol{\mu},\Sigma) \right\}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}} \left\{ \mathbf{w}^\mathrm{T}\boldsymbol{\mu} - (\phi(z_\epsilon)/\epsilon)\sqrt{\mathbf{w}^\mathrm{T}\Sigma\mathbf{w}} : \mathbf{R}\sim\mathcal{N}(\boldsymbol{\mu},\Sigma) \right\},$$

where $\phi(\cdot)$ is the density function of the standard normal distribution.

3.1.2 EV@R$_\epsilon$ Optimization

With V@R$_\epsilon$ replaced by EV@R$_\epsilon$, (3.1) is

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}}\left\{\max_{\nu>0}\left\{-\log\left(M_{\mathbf{w}^{\mathrm{T}}\mathbf{R}}(-\nu)/\epsilon\right)/\nu:\mathbf{R}\sim\mathcal{N}(\boldsymbol{\mu},\Sigma)\right\}\right\}$$

$$\Leftrightarrow\quad\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}}\left\{\max_{\nu>0}\left\{-\log\left(\exp\left\{-\nu\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}+\nu^2\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}/2\right\}\big/\epsilon\right)\big/\nu\right\}\right\}$$

$$\Leftrightarrow\quad\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}}\left\{\max_{\nu>0}\left\{\left(\nu\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\nu^2\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}/2+\log\epsilon\right)\big/\nu\right\}\right\}$$

$$\Leftrightarrow\quad\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}}\left\{\max_{\nu>0}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\nu\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}/2+(\log\epsilon)/\nu\right\}\right\}$$

$$\Leftrightarrow\quad\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\sqrt{-2\log\epsilon}\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}$$

$$\Leftrightarrow\quad\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\sqrt{-2\log\epsilon}\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\},$$

where the second last line is obtained by setting the first-order condition of the inner minimization problem to zero. Note that the entropic value-at-risk is not well-defined when returns follow an elliptically contoured α-stable distribution, since its moment-generating function does not exist.

3.2 Elliptically Contoured α-Stable Distribution

There is much evidence that asset returns display heavy-tailed properties, starting from the pioneering work of Mandelbrot (1963a, 1963b, 1967a, 1967b) and Fama (1963, 1965a, 1965b) among others which come later, including So (1987), Embrechts et al. (1997), Loretan and Phillips (1994), Rachev and Mittnik (2000), and Meerschaert and Scheffler (2003). However, some of these heavy-tailed distributions are criticized for their infinite variance, but as Nolan (2005) says, "…, bounded data are routinely modeled by normal distributions which have infinite support. The only justification for this is that the normal distribution gives a usable description of the shape of the distribution, even though it is clearly inappropriate for the tails for any problem with naturally bounded data. The same justification can be used for stable models … The variance is one measure of spread; the scale in a stable model is another."

Another reason against utilizing such distributions is the high computational complexity involved in the density calculation, but with modern computing power and the development of algorithms, this is no longer an issue.

In view of all these, portfolio optimization problems need to be robust against heavy-tailed distributions. To achieve that, we can let the returns $\mathbf{R}\sim\mathrm{EStable}(\alpha,\boldsymbol{\mu},\Sigma)$ follow an elliptically contoured α-stable distribution, which is essentially a scaled mixture of multivariate normal distributions [175] with joint characteristic function

$$\mathbb{E}(\exp\{\mathfrak{i}\mathbf{t}^{\mathrm{T}}\mathbf{R}\})=\exp\{-(\mathbf{t}^{\mathrm{T}}\Sigma\mathbf{t})^{\alpha/2}+\mathfrak{i}\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu}\},\tag{3.2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the location vector and $\Sigma \in \mathbb{S}_+^n$ is the shape matrix. This results in the portfolio return possessing a univariate α-stable distribution, because a random variable X has an α-stable distribution with shape $\alpha \in (0,2]$, skew $\beta \in [-1,1]$, scale $\gamma \in \mathbb{R}_{++}$ and location $\delta \in \mathbb{R}$ parameters, in which case we write $X \sim \text{Stable}(\alpha, \beta, \gamma, \delta)$, if it has characteristic function[1]

$$\mathbb{E}(\exp\{iuX\}) = \begin{cases} \exp\{-\gamma^\alpha |u|^\alpha [1 - i\beta(\tan\frac{\pi\alpha}{2}(\text{sign } u)] + i\delta u\}, & \alpha \neq 1, \\ \exp\{-\gamma|u|[1 + i\beta\frac{2}{\pi}(\text{sign } u)\log|u|] + i\delta u\}, & \alpha = 1, \end{cases} \tag{3.3}$$

and substituting $\mathbf{t} = u\mathbf{w}$ into (3.2) yields the characteristic function of $\mathbf{w}^{\mathrm{T}}\mathbf{R}$ in the form of (3.3) with $\beta = 0$, $\delta = \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}$ and $\gamma = \sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}$. Note that

$$\mathbf{w}^{\mathrm{T}}\mathbf{R} \stackrel{d}{=} \sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}Z_\alpha + \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu},$$

where $Z_\alpha \sim \text{Stable}(\alpha, 0, 1, 0)$ has characteristic function $\mathbb{E}(\exp\{iuZ_\alpha\}) = \exp\{-|u|^\alpha\}$, follows a $\mathcal{N}(0,2)$ distribution if $\alpha = 2$, a heavy-tailed distribution[2] with infinite variance if $\alpha < 2$, and a Cauchy$(1,0)$ distribution if $\alpha = 1$. If we assume $\alpha \in [1,2]$, then the maxmin V@R$_\epsilon$ problem can be written as

$$\max_{\mathbf{w} \in \mathcal{W}_\ell} \min_{(\alpha, \boldsymbol{\mu}, \Sigma) \in [1,2] \times \mathcal{M} \times \mathcal{S}} \left\{ \text{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}}\mathbf{R}) : \mathbf{R} \sim \text{EStable}(\alpha, \boldsymbol{\mu}, \Sigma) \right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},t) \in \mathcal{W}_\ell \times \mathbb{R}} \left\{ t : t \leq \min_{(\alpha, \boldsymbol{\mu}, \Sigma) \in [1,2] \times \mathcal{M} \times \mathcal{S}} \left\{ \text{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}}\mathbf{R}) : \mathbf{R} \sim \text{EStable}(\alpha, \boldsymbol{\mu}, \Sigma) \right\} \right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},t) \in \mathcal{W}_\ell \times \mathbb{R}} \left\{ t : t \leq \left\{ \text{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}}\mathbf{R}) : \mathbf{R} \sim \text{EStable}(\alpha, \boldsymbol{\mu}, \Sigma) \right\} \, \forall (\alpha, \boldsymbol{\mu}, \Sigma) \in [1,2] \times \mathcal{M} \times \mathcal{S} \right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},t) \in \mathcal{W}_\ell \times \mathbb{R}} \left\{ t : \mathbb{P}_{\mathbf{R} \sim \text{EStable}(\alpha, \boldsymbol{\mu}, \Sigma)} \{\mathbf{w}^{\mathrm{T}}\mathbf{R} < t\} \leq \epsilon \, \forall (\alpha, \boldsymbol{\mu}, \Sigma) \in [1,2] \times \mathcal{M} \times \mathcal{S} \right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},t) \in \mathcal{W}_\ell \times \mathbb{R}} \left\{ t : \mathbb{P}_{Z_\alpha} \left\{ Z_\alpha < (t - \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu})/\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} \right\} \leq \epsilon \, \forall (\alpha, \boldsymbol{\mu}, \Sigma) \in [1,2] \times \mathcal{M} \times \mathcal{S} \right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},t) \in \mathcal{W}_\ell \times \mathbb{R}} \left\{ t : (t - \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu})/\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} \leq -F_{Z_\alpha}^{-1}(1 - \epsilon) \, \forall (\alpha, \boldsymbol{\mu}, \Sigma) \in [1,2] \times \mathcal{M} \times \mathcal{S} \right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},t) \in \mathcal{W}_\ell \times \mathbb{R}} \left\{ t : t \leq \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - F_{Z_\alpha}^{-1}(1 - \epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} \, \forall (\alpha, \boldsymbol{\mu}, \Sigma) \in [1,2] \times \mathcal{M} \times \mathcal{S} \right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},t) \in \mathcal{W}_\ell \times \mathbb{R}} \left\{ t : t \leq \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{M} \times \mathcal{S}} \left\{ \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \max_{\alpha \in [1,2]} F_{Z_\alpha}^{-1}(1 - \epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} \right\} \right\}$$

$$\Leftrightarrow \max_{\mathbf{w} \in \mathcal{W}_\ell} \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{M} \times \mathcal{S}} \left\{ \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \max_{\alpha \in [1,2]} F_{Z_\alpha}^{-1}(1 - \epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} \right\}, \tag{3.4}$$

where $F_{Z_\alpha}^{-1}(1-\epsilon)$ is the $100(1-\epsilon)$th quantile of Z_α. To enable comparison with the Markowitz model, we should scale $F_{Z_\alpha}(1 - \epsilon)$ in (3.4) by a factor of $F_{Z_2}^{-1}(1 - \epsilon_*)/F_{Z_\alpha}^{-1}(1 - \epsilon_*)$ so that it is

---

[1]There are many parameterizations of the α-stable distribution; we use the 1-parameterization in Nolan (2015)

[2]If $X \sim \text{Stable}(\alpha, \beta, \gamma, \mu)$ where $0 < \alpha < 2$, the pth absolute moment $\mathbb{E}(|X|^p) = \int |x|^p f(x)dx$ is finite if and only if $p < \alpha$.

exactly $F_{Z_2}^{-1}(1 - \epsilon_*)$ when $\epsilon = \epsilon_*$, to obtain

$$\max_{\mathbf{w} \in \mathcal{W}_\ell} \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{M} \times \mathcal{S}} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} - \max_{\alpha \in [1,2]} \left\{ F_{Z_2}^{-1}(1 - \epsilon_*) F_{Z_\alpha}^{-1}(1 - \epsilon) / F_{Z_\alpha}^{-1}(1 - \epsilon_*) \right\} \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} \right\}. \qquad (3.5)$$

Remark 3.2
The support of $X \sim \mathrm{Stable}(\alpha, \beta, \gamma, \delta)$ is

$$supp(X) = \begin{cases} (-\infty, \delta] & \text{if } \alpha < 1, \beta = -1, \\ [\delta, \infty) & \text{if } \alpha < 1, \beta = 1, \\ (-\infty, \infty) & \text{otherwise.} \end{cases}$$

For our case in particular, the portfolio return $\mathbf{w}^{\mathrm{T}} \mathbf{R}$ is supported on the real line since $\beta = 0$, which certainly makes sense.

3.2.1 CV@R$_\epsilon$ Optimization

Note that (3.5) can be written as

$$\max_{\mathbf{w} \in \mathcal{W}_\ell} \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{M} \times \mathcal{S}} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} - (F_{Z_{\alpha_*}}^{-1}(1 - \epsilon) z_{1 - \epsilon_*} / F_{Z_{\alpha_*}}(1 - \epsilon_*)) \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w} \in \mathcal{W}_\ell} \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{M} \times \mathcal{S}} \left\{ \mathrm{V@R}_\epsilon(\mathbf{w}^{\mathrm{T}} \mathbf{R}) : \mathbf{R} \sim \mathrm{EStable}(\alpha_*, \boldsymbol{\mu}, c(\alpha_*, \epsilon_*) \Sigma) \right\} \qquad (3.6)$$

where $c(\alpha_*, \epsilon_*) = \sqrt{z_{1 - \epsilon_*} / F_{Z_{\alpha_*}}(1 - \epsilon_*)}$ such that α_* is an argument that maximizes $F_{Z_\alpha}^{-1}(1 - \epsilon) z_{1 - \epsilon_*} / F_{Z_\alpha}(1 - \epsilon_*)$ over $\alpha \in [1, 2]$. Replacing V@R$_\epsilon$ with CV@R$_\epsilon$ in (3.6) yields

$$\max_{\mathbf{w} \in \mathcal{W}_\ell} \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{M} \times \mathcal{S}} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} + \left(z_{1 - \epsilon_*} \int_0^\epsilon F_{Z_{\alpha_*}}^{-1}(\nu) d\nu / \left(\epsilon F_{Z_{\alpha_*}}(1 - \epsilon_*) \right) \right) \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} \right\},$$

where the integral $\int_0^\epsilon F_{Z_{\alpha_*}}^{-1}(\nu) d\nu$ can be calculated using, for example, methods in Stoyanov et al. (2006).

3.3 Distributions With Known Mean and Covariance

El-Ghaoui et al. (2003) show that if $\mathcal{Q}$ is the set of all probability distributions with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, then the maxmin V@R$_\epsilon$ problem is

$$\max_{\mathbf{w} \in \mathcal{W}_\ell} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} - \sqrt{\epsilon/(1 - \epsilon)} \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} \right\}$$

so that with added parameter uncertainty it becomes

$$\max_{\mathbf{w} \in \mathcal{W}_\ell} \min_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{M} \times \mathcal{S}} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} - \sqrt{\epsilon/(1 - \epsilon)} \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} \right\}.$$

Notice that the problems discussed thus far in this chapter are of the location-scale form

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}} \left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}, \tag{3.7}$$

where $\kappa(\epsilon) > 0$ is a decreasing function of the tolerance level ϵ. We shall focus on (3.7) hereafter. We next look at uncertainty sets for the location vector and scale matrix.

3.4 Location Uncertainty Sets

We first consider location uncertainty sets, which are well-studied (see for example [125, section 3]), so that (3.7) reduces to

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{\boldsymbol{\mu}\in\mathcal{M}} \left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}. \tag{3.8}$$

3.4.1 Box Uncertainty Set

[125, p. 7] The location box uncertainty set is $\mathcal{M}_{\mathrm{box}} = \left\{\boldsymbol{\mu} : \boldsymbol{\mu} \in [\underline{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}}]\right\}$, where $\underline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\mu}}$ contain the lower and upper bounds of $\boldsymbol{\mu}$ entry-wise respectively. Without short-selling, it is obvious that since each entry of $\mathbf{w}$ is non-negative, (3.8) with $\mathcal{M} = \mathcal{M}_{\mathrm{box}}$ can be written as

$$\max_{\mathbf{w}\in\mathcal{W}_0} \left\{\mathbf{w}^{\mathrm{T}}\underline{\boldsymbol{\mu}} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}$$

where $\mathcal{W}_0 = \{\mathbf{w} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1, \mathbf{w} \geq \mathbf{0}\}$, which is equivalent to a Second-Order Cone Program (SOCP). On the other hand, with short-selling (3.8) with $\mathcal{M} = \mathcal{M}_{\mathrm{box}}$ can be written as

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \left\{(\mathbf{w}^+)^{\mathrm{T}}\underline{\boldsymbol{\mu}} - (\mathbf{w}^-)^{\mathrm{T}}\overline{\boldsymbol{\mu}} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\} \tag{3.9}$$

where $\mathbf{w}^+$ has ith entry $w_i^+ = \max\{w_i, 0\}$ and $\mathbf{w}^-$ has ith entry $w_i^- = \max\{-w_i, 0\}$. Unless $\overline{\boldsymbol{\mu}} \geq \mathbf{0}$ and $\underline{\boldsymbol{\mu}} \leq \mathbf{0}$, in which case the objective function is a sum of concave functions and thus concave, (3.9) is not a convex optimization problem in general. However, (3.8) with $\mathcal{M} = \mathcal{M}_{\mathrm{box}}$ can always be converted into a convex problem which is equivalent to an SOCP:

Theorem 3.4
Let $\mathbf{p} = (\overline{\boldsymbol{\mu}} + \underline{\boldsymbol{\mu}})/2$ and $\mathbf{q} = (\overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}})/2$. Then, (3.8) with $\mathcal{M} = \mathcal{M}_{box}$ is equivalent to

$$\max_{(\mathbf{w},\mathbf{x})\in\mathcal{W}_\ell\times\mathbb{R}^n} \left\{\mathbf{w}^{\mathrm{T}}\mathbf{p} - \mathbf{x}^{\mathrm{T}}\mathbf{q} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : -\mathbf{x} \leq \mathbf{w} \leq \mathbf{x}\right\}.$$

Proof: Note that

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{\boldsymbol{\mu}\in\mathcal{M}_{\mathrm{box}}} \left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \left\{\max_{s\in\mathbb{R}} \left\{s : s \leq \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} \ \forall \boldsymbol{\mu} \in [\underline{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}}]\right\} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},s)\in\mathcal{W}_\ell\times\mathbb{R}} \left\{s - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : s \le \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} \ \forall\boldsymbol{\mu}\in[\underline{\boldsymbol{\mu}},\overline{\boldsymbol{\mu}}]\right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},s)\in\mathcal{W}_\ell\times\mathbb{R}} \left\{s - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : s \le \min_{\boldsymbol{\mu}\in[\underline{\boldsymbol{\mu}},\overline{\boldsymbol{\mu}}]} \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}\right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},s)\in\mathcal{W}_\ell\times\mathbb{R}} \left\{s - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : s \le \mathbf{w}^{\mathrm{T}}\mathbf{p} + \sum_{i=1}^{n}\min_{\zeta_i\in[-1,1]} \zeta_i w_i q_i\right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},s)\in\mathcal{W}_\ell\times\mathbb{R}} \left\{s - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : s \le \mathbf{w}^{\mathrm{T}}\mathbf{p} - \sum_{i=1}^{n}|w_i q_i|\right\}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \left\{\mathbf{w}^{\mathrm{T}}\mathbf{p} - \sum_{i=1}^{n}|w_i|q_i - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},\mathbf{x})\in\mathcal{W}_\ell\times\mathbb{R}^n} \left\{\mathbf{w}^{\mathrm{T}}\mathbf{p} - \sum_{i=1}^{n}x_i q_i - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : |w_i| \le x_i, i=1,\dots,n\right\}$$

$$\Leftrightarrow \max_{(\mathbf{w},\mathbf{x})\in\mathcal{W}_\ell\times\mathbb{R}^n} \left\{\mathbf{w}^{\mathrm{T}}\mathbf{p} - \mathbf{x}\mathbf{q} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : -\mathbf{x} \le \mathbf{w} \le \mathbf{x}\right\}.$$

$\square$

3.4.2 Ellipsoidal Uncertainty Set

[125, p. 7-8] The ellipsoidal uncertainty set is

$$\mathcal{M}_{\text{ellipsoid}} = \{\boldsymbol{\mu} : (\boldsymbol{\mu}-\boldsymbol{\mu}_0)^{\mathrm{T}}\mathbf{S}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) \le 1\}.$$

We have that

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{\boldsymbol{\mu}\in\mathcal{M}_{\text{ellipsoid}}} \left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{\boldsymbol{\mu}_1\in\mathbb{R}^n} \left\{\mathbf{w}^{\mathrm{T}}(\boldsymbol{\mu}_0+\boldsymbol{\mu}_1) - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : \|\mathbf{S}^{-1/2}\boldsymbol{\mu}_1\| \le 1\right\}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}_0 + \min_{\|\mathbf{x}\|\le 1} \mathbf{w}^{\mathrm{T}}\mathbf{S}^{1/2}\mathbf{x} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}$$

$$\Leftrightarrow \max_{\mathbf{w}\in\mathcal{W}_\ell} \left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}_0 - \|\mathbf{S}^{1/2}\mathbf{w}\| - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\},$$

where the second line is obtained by letting $\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1$, the third line by setting $\mathbf{S}^{-1/2}\boldsymbol{\mu}_1$ as $\mathbf{x}$, and the last line by noting that the optimal solution to the inner minimization problem is

$$\mathbf{x}_* = -\frac{\mathbf{S}^{1/2}\mathbf{w}}{\|\mathbf{S}^{1/2}\mathbf{w}\|},$$

the vector opposite in direction to $\mathbf{S}^{1/2}\mathbf{w}$ with the maximum possible length.

3.5   Scale Uncertainty Sets

We now consider the problem

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{\Sigma\in\mathcal{S}} \left\{ \mathbf{w}^\mathrm{T}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^\mathrm{T}\Sigma\mathbf{w}} \right\} \tag{3.10}$$

and look at some uncertainty sets for the scale matrix Σ.

3.5.1   Box Uncertainty Set

[125, p. 10] The box uncertainty set is

$$\mathcal{S} = \left\{ \Sigma : \underline{\sigma}_{ij} \leq \sigma_{ij} \leq \overline{\sigma}_{ij}, i = 1,\dots,n, \forall j \leq i, \Sigma \geq 0 \right\},$$

where $\underline{\sigma}_{ij}$ and $\overline{\sigma}_{ij}$ are the lower and upper bounds of σ_{ij} respectively and $\Sigma \geq 0$ ensures Σ is positive semi-definite.

3.5.2   Ellipsoidal Uncertainty Set

[125, p. 11] Denoting $\hat{\mathbf{s}}$ as the estimated mean vector and $\hat{\mathbf{V}}$ as the covariance matrix of the vector $\mathbf{s}$ of the upper triangular entries of the estimated scale matrix, we define the ellipsoidal uncertainty set as

$$\mathcal{S} = \left\{ \Sigma : (\mathbf{s} - \hat{\mathbf{s}})^\mathrm{T}\hat{\mathbf{V}}(\mathbf{s} - \hat{\mathbf{s}}) \leq c, \Sigma \geq 0 \right\},$$

where $c > 0$ is such that the higher its value, the larger the confidence region for $\mathbf{s}$. One drawback of this uncertainty set is the huge computational effort needed to calculate the entries of $\hat{\mathbf{V}}$, which is of order $\mathcal{O}(n^4)$.

3.5.3   Correlation Coefficient Uncertainty Set

[125, p. 11-12] We define the correlation coefficient uncertainty set as

$$\begin{aligned}
\mathcal{S} &= \left\{ \Sigma : \underline{\rho}_{ij} \leq \rho_{ij} \leq \overline{\rho}_{ij}, \Sigma \geq 0 \right\} \\
&= \left\{ \Sigma : \underline{\rho}_{ij}\sigma_i\sigma_j \leq \sigma_{ij} \leq \overline{\rho}_{ij}\sigma_i\sigma_j, \Sigma \geq 0 \right\},
\end{aligned}$$

where $\underline{\rho}_{ij}$ and $\overline{\rho}_{ij}$ are the lower and upper bounds of the correlation coefficient ρ_{ij} respectively. Note that the inequalities are generally non-convex in $(\sigma_i, \sigma_j, \sigma_{ij})$ unless $\underline{\rho}_{ij} \leq 0$ and $\overline{\rho}_{ij} \geq 0$, in which case we can introduce an auxiliary variable $t \in \mathbb{R}$ and write the uncertainty set as

$$\begin{aligned}
\mathcal{S} &= \left\{ \Sigma : \underline{\rho}_{ij}t \leq \sigma_{ij} \leq \overline{\rho}_{ij}t, t^2 \leq \sigma_i^2\sigma_j^2, \Sigma \geq 0 \right\} \\
&= \left\{ \Sigma : \underline{\rho}_{ij}t \leq \sigma_{ij} \leq \overline{\rho}_{ij}t, \left\| \begin{bmatrix} 2t \\ \sigma_i^2 - \sigma_j^2 \end{bmatrix} \right\| \leq \sigma_i^2 + \sigma_j^2, \Sigma \geq 0 \right\},
\end{aligned}$$

where the constraint $t^2 \leq \sigma_i^2 \sigma_j^2$ is equivalent to the second-order cone constraint $\left\| \begin{bmatrix} 2t \\ \sigma_i^2 - \sigma_j^2 \end{bmatrix} \right\| \leq \sigma_i^2 + \sigma_j^2$ due to Theorem A.20.

3.5.4 Specific Portfolio Scale Uncertainty Set

[125, p. 12] By specifying certain portfolios, albeit arbitrarily, and imposing constraints on their scales, we define the specific portfolio scale uncertainty set as

$$S = \left\{ \Sigma : l_i \leq \mathbf{w}_i^\mathsf{T} \Sigma \mathbf{w}_i \leq u_i, i = 1, \ldots, p, \Sigma \geq 0 \right\}$$

where $\mathbf{w}_i$ is the ith chosen portfolio with l_i and u_i being the lower and upper bounds of its squared scale respectively.

3.5.5 One-Factor Model Uncertainty Set

[125, p. 13] The one-factor model uncertainty set is generally non-convex and defined as

$$S = \left\{ \Sigma : \Sigma = \mathrm{diag}(\mathbf{g}) + \mathbf{h}\mathbf{h}^\mathsf{T}, (\mathbf{g}, \mathbf{h}) \in \mathcal{U} \right\}$$

where $\mathrm{diag}(\mathbf{g}) + \mathbf{h}\mathbf{h}^\mathsf{T}$ is the one-factor decomposition of Σ and $\mathcal{U}$ is a convex set.

Remark 3.3
Methods to solve (3.10) *under the uncertainty sets for* Σ *introduced in this section can be found in [125, Sections 5-7] and the references therein.*

3.6 Eigendecomposition Uncertainty Set

We next introduce a novel eigendecomposition uncertainty set for the scale matrix. First, we write Σ in the eigendecomposition form $\sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$. Then, we allow the positive eigenvalues $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]^\mathsf{T}$ to vary in a box uncertainty set and the eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ each to perturbate in a cone uncertainty with orthogonality preserved among them. Notice that the first standard basis vector

$$\mathbf{e}_1 = \mathbf{P}_i \mathbf{u}_i \tag{3.11}$$

where

$$\mathbf{P}_i = \prod_{j=1}^{n-1} \mathbf{G}_{ij}$$

with

$$\mathbf{G}_{i1} = \begin{bmatrix} c_{i1} & s_{i1} & \mathbf{0}_{1 \times (n-2)} \\ -s_{i1} & c_{i1} & \mathbf{0}_{1 \times (n-2)} \\ \mathbf{0}_{(n-2) \times 1} & \mathbf{0}_{(n-2) \times 1} & \mathbf{I}_{n-2} \end{bmatrix},$$

$$\mathbf{G}_{i,n-1} = \begin{bmatrix} \mathbf{I}_{n-2} & \mathbf{0}_{(n-2)\times 1} & \mathbf{0}_{(n-2)\times 1} \\ \mathbf{0}_{1\times(n-2)} & -s_{i,n-1} & c_{i,n-1} \\ \mathbf{0}_{1\times(n-2)} & c_{i,n-1} & s_{i,n-1,} \end{bmatrix}$$

and

$$\mathbf{G}_{ij} = \begin{bmatrix} \mathbf{I}_{j-1} & \mathbf{0}_{(j-1)\times 1} & \mathbf{0}_{(j-1)\times 1} & \mathbf{0}_{(j-1)\times(n-j-1)} \\ \mathbf{0}_{1\times(j-1)} & c_{ij} & s_{ij} & \mathbf{0}_{1\times(n-j-1)} \\ \mathbf{0}_{1\times(j-1)} & -s_{ij} & c_{ij} & \mathbf{0}_{1\times(n-j-1)} \\ \mathbf{0}_{(j-1)\times(j-1)} & \mathbf{0}_{(j-1)\times 1} & \mathbf{0}_{(j-1)\times 1} & \mathbf{I}_{n-j-1} \end{bmatrix},$$

such that $c_{ij} = h_{ij}\big/\sqrt{h_{ij}^2 + h_{i,j+1}^2}$, $s_{ij} = -h_{i,j+1}\big/\sqrt{h_{ij}^2 + h_{i,j+1}^2}$ and $\mathbf{h}_i = [h_{i1},\ldots,h_{i,n-1}]^{\mathrm{T}}$ is the vector which right multiplies $\mathbf{G}_{ij}$ in (3.11). Thus, the location-scale problem can be expressed as

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\sum_{i=1}^n \lambda_i(\mathbf{w}^{\mathrm{T}}\mathbf{P}_i^{-1}\mathbf{e}_1)^2}\right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w}\in\mathcal{W}_\ell}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\sum_{i=1}^n \lambda_i(\mathbf{e}_1^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2}\right\}.$$

The above equivalence is due to $\mathbf{P}_i$ being an orthogonal matrix, which essentially rotates $\mathbf{u}_i$ to $\mathbf{e}_1$ through a sequence of rotations $\mathbf{G}_{i,n-1},\ldots,\mathbf{G}_{i1}$, where $\mathbf{G}_{ij}$ rotates the vector it is right-multiplied with along the plane spanned by the jth and $(j+1)$th axes so that the resulting vector has a zero $(j+1)$th entry. The uncertainty set is

$$\mathcal{S} = \left\{\tilde{\Sigma} : \tilde{\Sigma} = \sum_{i=1}^n \tilde{\lambda}_i\mathbf{P}_i^{\mathrm{T}}\tilde{\mathbf{v}}\tilde{\mathbf{v}}^{\mathrm{T}}\mathbf{P}_i : \begin{array}{l} \tilde{\lambda}_i \in (\max\{0,\lambda_i - b_i\},\lambda_i + b_i] \;\forall i \\ \tilde{\mathbf{v}}^{\mathrm{T}}\mathbf{e}_1 \geq 1 - c, \|\tilde{\mathbf{v}}\|_2 = 1 \end{array}\right\}$$

where $\tilde{\mathbf{v}}$ is restricted to Euclidean length and varies within an acute cone of half-angle $\theta = \arccos(1 - c)$ such that $0 \leq c \leq 1$, while $\tilde{\lambda}_i$ perturbates within the interval $(\max\{0,\lambda_i - b_i\},\lambda_i + b_i]$ such that $b_i \geq 0$ for $i = 1,\ldots,n$. The (i,j)th entry of $\tilde{\Sigma}$ can be written as

$$\tilde{\sigma}_{ij} = \tilde{\mathbf{v}}^{\mathrm{T}}[\mathbf{P}_{1i},\ldots,\mathbf{P}_{ni}]\operatorname{diag}(\tilde{\boldsymbol{\lambda}})[\mathbf{P}_{1j},\ldots,\mathbf{P}_{nj}]^{\mathrm{T}}\tilde{\mathbf{v}}$$

where $\mathbf{P}_{ki}$ represents the ith column of $\mathbf{P}_k$. Thus, once an entry of $\tilde{\Sigma}$ is fixed through a choice of $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\mathbf{v}}$ in their respective uncertainty sets, the other entries are likewise determined. This is what makes the eigendecomposition uncertainty set much less conservative and sets it apart from the other covariance matrix uncertainty sets introduced previously in this chapter. The robust counterpart we would like to solve is

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{x}}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\sum_{i=1}^n \tilde{\lambda}_i((\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2)^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2} : \begin{array}{l} \tilde{\lambda}_i \in (\max\{0,\lambda_i - b_i\},\lambda_i + b_i] \;\forall i \\ (\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2)^{\mathrm{T}}\mathbf{e}_1 \geq 1 - c \end{array}\right\}$$

$$(3.12)$$

where $\tilde{\mathbf{v}}$ is replaced with $\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2$ so that the unit Euclidean length of the eigenvectors is ensured implicitly. The term within the squared root in (3.12) is always positive since assuming otherwise implies that $(\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2)^{\mathrm{T}}\mathbf{P}_i\mathbf{w} = 0$ and hence $\mathbf{w}$ is orthogonal to $\mathbf{P}_i^{\mathrm{T}}(\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2)$ for $i = 1,\ldots,n$, resulting in $n+1$ orthogonal vectors in $\mathbf{R}^n$, which is absurd.

Although (3.12) is non-convex in general, the next theorem shows that it can be converted into an SDP.

Theorem 3.5

Assume there exists $\mathbf{x}$ such that $(\mathbf{x}/\|\mathbf{x}\|_2)^{\mathrm{T}}\mathbf{e}_1 > 1 - c$, then the optimal $\mathbf{w}$ of (3.12) is equal to that of

$$\max_{(\mathbf{w},\tau,y)\in\mathcal{W}_\ell\times\mathbb{R}_+\times\mathbb{R}_{++}}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)y:\left[\begin{array}{c|c}(y+\tau(1-c)^2)\mathbf{I}_n-\tau\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} & \begin{array}{ccc}\mathbf{P}_1\mathbf{w} & \ldots & \mathbf{P}_n\mathbf{w}\end{array}\\\hline\begin{array}{c}\mathbf{w}^{\mathrm{T}}\mathbf{P}_1^{\mathrm{T}}\\\vdots\\\mathbf{w}^{\mathrm{T}}\mathbf{P}_n^{\mathrm{T}}\end{array} & \mathrm{diag}(\bar{\lambda}_1^{-1},\ldots,\bar{\lambda}_n^{-1})y\end{array}\right]\geq 0\right\}$$
(3.13)

where $\bar{\lambda}_i = \lambda_i + b_i$.

Proof: Notice that (3.12) can be written as

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{\tilde{\mathbf{x}}\in\mathbb{R}^n\backslash\{\mathbf{0}\}}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\sum_{i=1}^n\bar{\lambda}_i((\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2)^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2}:(\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2)^{\mathrm{T}}\mathbf{e}_1\geq 1-c\right\}$$

$$\Leftrightarrow\quad\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{\tilde{\mathbf{x}}\in\mathbb{R}^n\backslash\{\mathbf{0}\}}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\sum_{i=1}^n\bar{\lambda}_i((\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2)^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2}:(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{e}_1)^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}\geq(1-c)^2\right\}$$

$$\Leftrightarrow\quad\max_{\mathbf{w}\in\mathcal{W}_\ell}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\max_{\tilde{\mathbf{x}}\in\mathbb{R}^n\backslash\{\mathbf{0}\}}\left\{\sum_{i=1}^n(\bar{\lambda}_i(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}):(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{e}_1)^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}\geq(1-c)^2\right\}}\right\}$$ (3.14)

where we square both sides of the inequality constraint in the first equivalence since by doing so the set of objective values for the inner minimization problem remains unchanged although the set of $\tilde{\mathbf{x}}$'s for each $\mathbf{w}$ enlarges. We now set on proving that

$$\max_{\tilde{\mathbf{x}}\in\mathbb{R}^n\backslash\{\mathbf{0}\}}\left\{\sum_{i=1}^n(\bar{\lambda}_i(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}):(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{e}_1)^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}\geq(1-c)^2\right\}$$ (3.15)

is equivalent to

$$\min_{\tau\in\mathbb{R}_+,y\in\mathbb{R}}\left\{y^2:\mathbf{I}_n-\sum_{i=1}^n\frac{\bar{\lambda}_i\mathbf{P}_i\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{P}_i^{\mathrm{T}}}{y^2}-\frac{\tau}{y}\left(\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}}-(1-c)^2\mathbf{I}_n\right)\geq 0,y>0\right\}.$$ (3.16)

Let $\tau\geq 0$ and $y>0$ be arbitrary constants so that

$$\sum_{i=1}^n(\bar{\lambda}_i(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}})+y\tau((\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{e}_1)^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}-(1-c)^2)$$

is an upper bound of $\sum_{i=1}^{n}(\bar{\lambda}_i(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}})$ on the feasible set of (3.15) by construction. This means that

$$\max_{\tilde{\mathbf{x}}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \sum_{i=1}^{n}(\bar{\lambda}_i(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}) + y\tau((\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{e}_1)^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} - (1-c)^2)$$

is an upper bound to the optimal value of (3.15). Therefore, if

$$y^2 \geq \max_{\tilde{\mathbf{x}}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \sum_{i=1}^{n}(\bar{\lambda}_i(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}) + y\tau((\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{e}_1)^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} - (1-c)^2)$$

$$\Leftrightarrow \quad y^2 \geq \sum_{i=1}^{n}(\bar{\lambda}_i(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}) + y\tau((\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{e}_1)^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} - (1-c)^2) \quad \forall \tilde{\mathbf{x}} \in \mathbb{R}^n\setminus\{\mathbf{0}\}, \quad (3.17)$$

then y^2 is an upper bound to the optimal value of (3.15). Note that (3.17) can be written as

$$y^2\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} \geq \tilde{\mathbf{x}}^{\mathrm{T}}\left(\sum_{i=1}^{n}\bar{\lambda}_i\mathbf{P}_i\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{P}_i^{\mathrm{T}}\right)\tilde{\mathbf{x}} + y\tau(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}}\tilde{\mathbf{x}} - (1-c)^2\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}) \quad \forall \tilde{\mathbf{x}} \in \mathbb{R}^n\setminus\{\mathbf{0}\}$$

$$\Leftrightarrow \quad \tilde{\mathbf{x}}^{\mathrm{T}}\left(\mathbf{I}_n - \sum_{i=1}^{n}\frac{\bar{\lambda}_i\mathbf{P}_i\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{P}_i^{\mathrm{T}}}{y^2} - \frac{\tau}{y}\left(\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} - (1-c)^2\mathbf{I}_n\right)\right)\tilde{\mathbf{x}} \geq 0 \quad \forall \tilde{\mathbf{x}} \in \mathbb{R}^n\setminus\{\mathbf{0}\}$$

$$\Leftrightarrow \quad \mathbf{I}_n - \sum_{i=1}^{n}\frac{\bar{\lambda}_i\mathbf{P}_i\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{P}_i^{\mathrm{T}}}{y^2} - \frac{\tau}{y}\left(\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} - (1-c)^2\mathbf{I}_n\right) \geq 0,$$

so that the optimal value of (3.16) is greater than or equal to the optimal value of (3.15), which we denote as $V > 0$. We are left to prove that (3.16) has a feasible solution such that its corresponding objective function value is equal to V. In other words, we want to show that there exists $\tau_* \geq 0$ such that

$$\mathbf{I}_n - \sum_{i=1}^{n}\frac{\bar{\lambda}_i\mathbf{P}_i\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{P}_i^{\mathrm{T}}}{V} - \frac{\tau_*}{\sqrt{V}}\left(\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} - (1-c)^2\mathbf{I}_n\right) \geq 0. \quad (3.18)$$

Note that there exists an ϵ_0 and an $\mathbf{x}$ such that for every $\epsilon \in (0, \epsilon_0]$,

$$\mathbf{x}^{\mathrm{T}}\left(\frac{\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} - (1-c)^2\mathbf{I}_n}{\sqrt{V}}\right)\mathbf{x} \geq \epsilon\mathbf{x}^{\mathrm{T}}\mathbf{x} \quad (3.19)$$

due to the strict feasibility assumption. In addition, we have

$$\tilde{\mathbf{x}}^{\mathrm{T}}\left(\frac{\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} - (1-c)^2\mathbf{I}_n}{\sqrt{V}}\right)\tilde{\mathbf{x}} \geq \epsilon\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} \Rightarrow \tilde{\mathbf{x}}^{\mathrm{T}}\left(\mathbf{I}_n - \sum_{i=1}^{n}\frac{\bar{\lambda}_i\mathbf{P}_i\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{P}_i^{\mathrm{T}}}{V}\right)\tilde{\mathbf{x}} \geq 0 \quad \forall \epsilon \in (0, \epsilon_0]. \quad (3.20)$$

To see (3.20), assume that its "if" condition holds, then we have

$$\sum_{i=1}^{n}(\bar{\lambda}_i(\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2/\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}) \leq V$$

$$\Leftrightarrow \quad \tilde{\mathbf{x}}^{\mathrm{T}} \left(\sum_{i=1}^{n} \bar{\lambda}_i \mathbf{P}_i \mathbf{w} \mathbf{w}^{\mathrm{T}} \mathbf{P}_i^{\mathrm{T}} \right) \tilde{\mathbf{x}} \leq V \tilde{\mathbf{x}}^{\mathrm{T}} \tilde{\mathbf{x}}$$

$$\Leftrightarrow \quad \tilde{\mathbf{x}}^{\mathrm{T}} \tilde{\mathbf{x}} - \tilde{\mathbf{x}}^{\mathrm{T}} \left(\frac{\sum_{i=1}^{n} \bar{\lambda}_i \mathbf{P}_i \mathbf{w} \mathbf{w}^{\mathrm{T}} \mathbf{P}_i^{\mathrm{T}}}{V} \right) \tilde{\mathbf{x}} \geq 0$$

$$\Leftrightarrow \quad \tilde{\mathbf{x}}^{\mathrm{T}} \left(\mathbf{I}_n - \frac{\sum_{i=1}^{n} \bar{\lambda}_i \mathbf{P}_i \mathbf{w} \mathbf{w}^{\mathrm{T}} \mathbf{P}_i^{\mathrm{T}}}{V} \right) \tilde{\mathbf{x}} \geq 0.$$

Using the homogeneous $\mathcal{S}$-Lemma A.1 (ii), there exists $\tau_\epsilon \geq 0$ such that

$$\mathbf{I}_n - \sum_{i=1}^{n} \frac{\bar{\lambda}_i \mathbf{P}_i \mathbf{w} \mathbf{w}^{\mathrm{T}} \mathbf{P}_i^{\mathrm{T}}}{V} - \tau_\epsilon \left(\frac{\mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} - (1-c)^2 \mathbf{I}_n}{\sqrt{V}} - \epsilon \mathbf{I}_n \right) \geq 0$$

and in particular,

$$\mathbf{x}^{\mathrm{T}} \left(\mathbf{I}_n - \sum_{i=1}^{n} \frac{\bar{\lambda}_i \mathbf{P}_i \mathbf{w} \mathbf{w}^{\mathrm{T}} \mathbf{P}_i^{\mathrm{T}}}{V} \right) \mathbf{x} \geq \tau_\epsilon \mathbf{x}^{\mathrm{T}} \left(\frac{\mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} - (1-c)^2 \mathbf{I}_n}{\sqrt{V}} - \epsilon \mathbf{I}_n \right) \mathbf{x}$$

for each $\epsilon \in (0, \epsilon_0]$. The fact that $\mathbf{x}^{\mathrm{T}} \left(\frac{\mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} - (1-c)^2 \mathbf{I}_n}{\sqrt{V}} - \epsilon \mathbf{I}_n \right) \mathbf{x} > 0$ means that τ_ϵ stays bounded as $\epsilon \to 0$, which implies

$$\mathbf{I}_n - \sum_{i=1}^{n} \frac{\bar{\lambda}_i \mathbf{P}_i \mathbf{w} \mathbf{w}^{\mathrm{T}} \mathbf{P}_i^{\mathrm{T}}}{V} - \tau_{\epsilon_i} \left(\frac{\mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} - (1-c)^2 \mathbf{I}_n}{\sqrt{V}} - \epsilon_i \mathbf{I}_n \right) \geq 0 \qquad (3.21)$$

for a properly chosen sequence ϵ_i, $i = 1, 2, \ldots$ such that $\lim_{i \to \infty} \tau_{\epsilon_i} = \tau_*$ and $\lim_{i \to \infty} \epsilon_i = 0$. Taking limits on both sides of (3.21) as $i \to \infty$ obtains (3.18), which proves the equivalence between (3.15) and (3.16).

Since (3.16) can be reformulated as

$$\min_{\tau \in \mathbb{R}_+, y \in \mathbb{R}_{++}} \left\{ y^2 : \mathbf{I}_n - \frac{\tau}{y} \left(\mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} - (1-c)^2 \mathbf{I}_n \right) \right.$$

$$\left. - \left[\mathbf{P}_1(\mathbf{w}/y) \quad \ldots \quad \mathbf{P}_n(\mathbf{w}/y) \right] \operatorname{diag}(\bar{\lambda}_1, \ldots, \bar{\lambda}_n) \begin{bmatrix} (\mathbf{w}^{\mathrm{T}}/y) \mathbf{P}_1^{\mathrm{T}} \\ \vdots \\ (\mathbf{w}^{\mathrm{T}}/y) \mathbf{P}_n^{\mathrm{T}} \end{bmatrix} \geq 0 \right\}$$

$$\Leftrightarrow \min_{\tau \in \mathbb{R}_+, y \in \mathbb{R}_{++}} \left\{ y^2 : \left[\begin{array}{c|c} \mathbf{I}_n - \frac{\tau}{y} \left(\mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} - (1-c)^2 \mathbf{I}_n \right) & \mathbf{P}_1(\mathbf{w}/y) \quad \ldots \quad \mathbf{P}_n(\mathbf{w}/y) \\ \hline \begin{array}{c} (\mathbf{w}/y)^{\mathrm{T}} \mathbf{P}_1^{\mathrm{T}} \\ \vdots \\ (\mathbf{w}/y)^{\mathrm{T}} \mathbf{P}_n^{\mathrm{T}} \end{array} & \operatorname{diag}(\bar{\lambda}_1^{-1}, \ldots, \bar{\lambda}_n^{-1}) \end{array} \right] \geq 0 \right\}$$

$$\Leftrightarrow \min_{\tau \in \mathbb{R}_+, y \in \mathbb{R}_{++}} \left\{ y^2 : \left[\begin{array}{c|c} (y + \tau(1-c)^2) \mathbf{I}_n - \tau \mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} & \mathbf{P}_1 \mathbf{w} \quad \ldots \quad \mathbf{P}_n \mathbf{w} \\ \hline \begin{array}{c} \mathbf{w}^{\mathrm{T}} \mathbf{P}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{w}^{\mathrm{T}} \mathbf{P}_n^{\mathrm{T}} \end{array} & \operatorname{diag}(\bar{\lambda}_1^{-1} y, \ldots, \bar{\lambda}_n^{-1} y) \end{array} \right] \geq 0 \right\} \qquad (3.22)$$

where the first equivalence is by virtue of the Schur Complement Lemma A.2, substituting the inner maximization problem of (3.14) with (3.22) then yields (3.13), which completes the proof. □

From now on, we only consider the box uncertainty set for $\boldsymbol{\mu}$ and the eigendecomposition uncertainty set for Σ under elliptical distributions.

Chapter 4

Size of Uncertainty Sets and Numerical Experiments

If the uncertainty set is chosen too conservatively (too big), the asset allocation by robust optimization is typically uninteresting. If the uncertainty set is too small, the robustness we are looking for is lost. In this chapter, we tackle this conundrum by choosing the size of the uncertainty set based on the sensitivity of the robust optimal value to size changes. In particular, the median sensitivity is chosen for our numerical experiments. Nevertheless, let us first investigate the effects due to tail uncertainty in the next section.

4.1 Distributional Uncertainty

We focus on the problem

$$\max_{\mathbf{w} \in \mathcal{W}_\ell} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} - \kappa(\epsilon) \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} \right\} \tag{4.1}$$

with

$$\kappa(\epsilon) = F_{Z_2}^{-1}(1 - \epsilon)$$

which we label as type A, and with

$$\kappa(\epsilon) = \max_{\alpha \in \{1, 1.01, \dots, 2\}} \left\{ F_{Z_2}^{-1}(1 - \epsilon_*) F_{Z_\alpha}^{-1}(1 - \epsilon) / F_{Z_\alpha}^{-1}(1 - \epsilon_*) \right\} \tag{4.2}$$

for $\epsilon_* = 0.05$, 0.1 and 0.15 which we label as types B, C and D respectively, where we recall that $Z_\alpha \sim \text{Stable}(\alpha, 0, 1, 0)$. For type A, we find the optimal V@R$_\epsilon$ of the portfolio return $\mathbf{w}^{\mathrm{T}} \mathbf{R} \sim \text{Stable}\left(2, 0, \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}}, \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu}\right) = \mathcal{N}\left(\mathbf{w}^{\mathrm{T}} \boldsymbol{\mu}, 2\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}\right)$. For each of types B, C and D, we find the optimal V@R$_\epsilon$ of the portfolio return distributionally robust against all α-stable distributions where $\alpha \in \{1, 1.01, \dots, 2\}$, $\beta = 0$, $\gamma = F_{Z_2}^{-1}(1 - \epsilon_*)\sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}}/F_{Z_\alpha}^{-1}(1 - \epsilon_*)$ and $\delta = \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu}$, with γ scaled in such a way that the $100\epsilon_*$th percent quantile of the portfolio return is always equal to that for type A.

In Figure 4.1, each subfigure plots the objective values of (4.2) for type B ($\epsilon_* = 0.05$) against $\alpha \in \{1, 1.01, \dots, 2\}$ at a particular tolerance level ϵ, and the subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \dots, 0.49$. The maximum

in each subfigure is the optimal value of (4.2) for type B at its associated ϵ. The maximum occurs at $\alpha = 1$ if $\epsilon < \epsilon_*$ and $\alpha = 2$ if $\epsilon > \epsilon_*$, while it can be any value in $\{1, 1.01, \ldots, 2\}$ if $\epsilon = \epsilon_*$. Figures C.1 and C.2 in Appendix C are analogous plots for types C ($\epsilon_* = 0.10$) and D ($\epsilon_* = 0.15$) respectively, with similar observations.
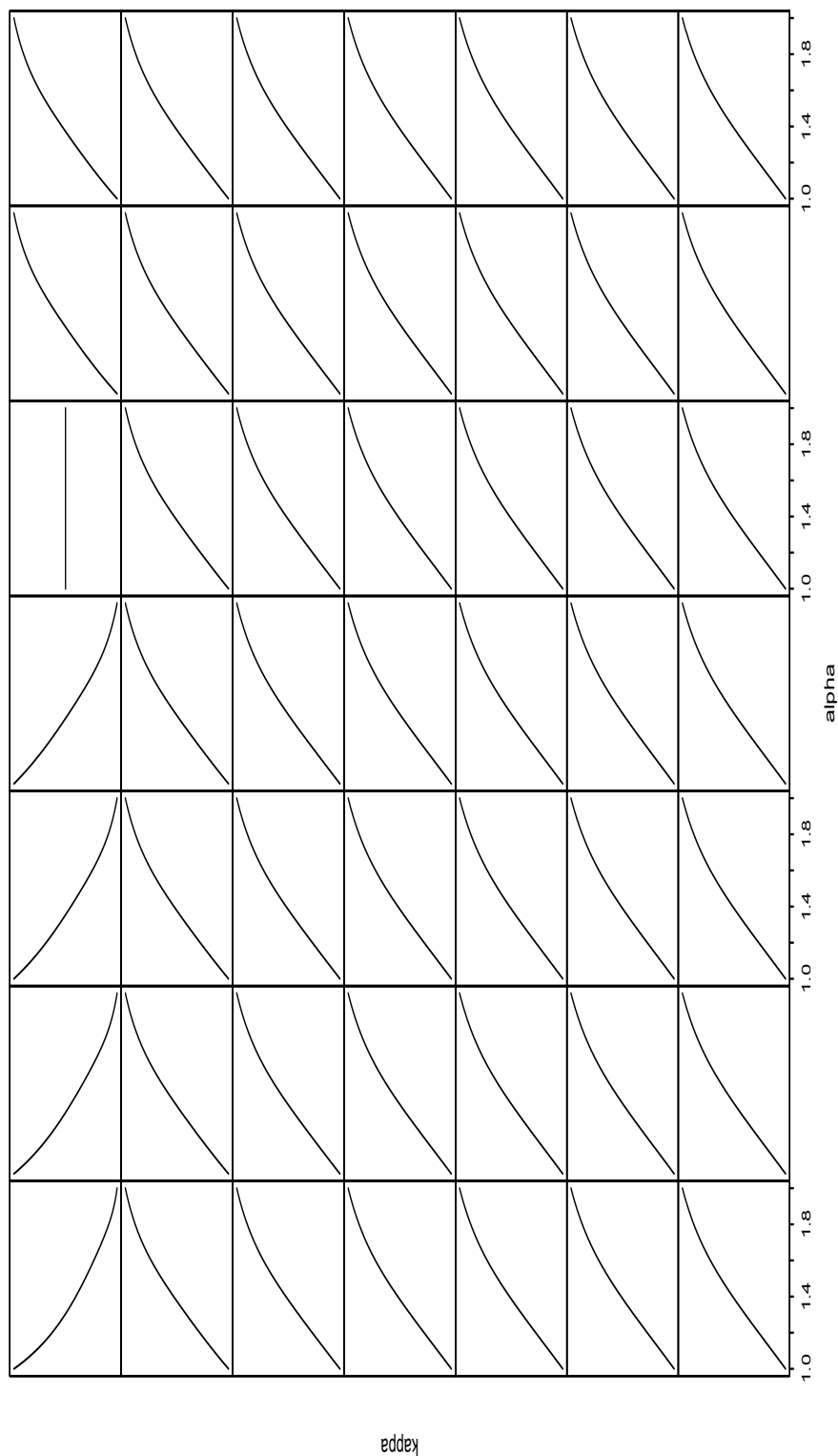
Figure 4.1: each subfigure plots values of the objective function $\kappa(\epsilon) = \max\limits_{\alpha \in \{1,1.01,...,2\}} \left\{ F_{Z_2}^{-1}(1-\epsilon_*) F_{Z_\alpha}^{-1}(1-\epsilon) / F_{Z_\alpha}^{-1}(1-\epsilon_*) \right\}$ for type B ($\epsilon_* = 0.05$) against $\alpha \in \{1, 1.01,...,2\}$ at a particular tolerance level ϵ, and subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02,...,0.49$.

4.1.1 Numerical Experiment

Daily stock returns of Netflix, Alphabet, Facebook, General Electric, Microsoft, Boeing, JP Morgan, Coca Cola, Pfizer, Delta Air Lines, Tesla, Abbot Laboratories, Cisco, Apple, Bank of America, Exxon Mobil, International Business Machines, Credit Suisse, Qualcomm and Hewlett Packard on either Nasdaq or NYSE for year 2015 are obtained from Google Finance, with the assets labeled 1 to 20 in the same order as they are written, where 1 indicates the highest and 20 the lowest location over scale ratio. The location vector and the scale matrix are estimated by multiplying a hundred to the sample mean and minimum covariance determinant (MCD) estimators of the daily returns, assumed to be independent and identically distributed, respectively. In Figure 4.2, each subfigure plots the optimal weights against the asset number for both the type A and B problems where short-selling is disallowed at a particular tolerance level ϵ, and has a red vertical line which separates assets with a positive location over scale ratio on the left from assets with a negative location over scale ratio on the right. The subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$. Figures C.3 and C.4 in Appendix C are analogs of Figure 4.2 for the type C and D problems respectively. Figure 4.3 (a) plots the optimal weights of the type B problem minus the optimal weights of the type A problem against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is disallowed. Figures 4.3 (b) and (c) are analogs of Figure 4.3 (a) but with the optimal weight of the type C and D problems respectively subtracting the optimal weight of type A problem instead. Expectedly, the optimal weights of the type A problem are the same as those of the type B, C and D problems when ϵ is greater than or equal to their corresponding ϵ_*'s, where the optimal value of (4.2) is exactly $F_{Z_2}^{-1}(1 - \epsilon)$. On the other hand, the optimal weights of the type A problem differ considerably to those of the type B, C and D problems when ϵ is less than their corresponding ϵ_*'s where $\alpha = 1$ is the optimal solution of (4.2). We thus see that interestingly, introducing heavy-tailedness into the distribution of returns causes weight to be moved from assets with higher location over scale ratio to those with lower. Figures 4.4 (a), (b) and (c) plot the optimal portfolio expected return of the type A problem and respectively the robust optimal portfolio expected return of the type B, C and D problems against the tolerance level, where short-selling is disallowed. Note that if ϵ is less than the corresponding ϵ_*'s, the optimal portfolio expected return for the type A problem is higher than the robust optimal portfolio expected return for the type B, C or D problem. Figure 4.5 is an analog of Figure 4.4 for the portfolio negated value-at-risk, and shows that if ϵ is less than the corresponding ϵ_*'s, the portfolio negated value-at-risk of the type A problem is higher than that of the type B,C or D problem, so that it is more likely to invest all wealth into the risky assets for the former than it is for the latter, which is not at all unintuitive since for the latter, the asset returns follow a heavy-tailed distribution so that the optimal allocation should be more conservative. Figures 4.6 - 4.9 are analogs of Figures 4.2 - 4.5 where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset, whereas Figures C.5 and C.6 in Appendix C are analogs of Figure 4.6 for the type C and D problems respectively. Similar conclusions can be drawn, but notice how in Figure 4.6, C.5 and C.6, if ϵ is less than the corresponding ϵ_*'s, short-selling is less encouraged for the type B, C or D problem than for the type A problem, which is again inutitive due to the same reason that for the former, the asset returns follow a heavy-tailed distribution so that the

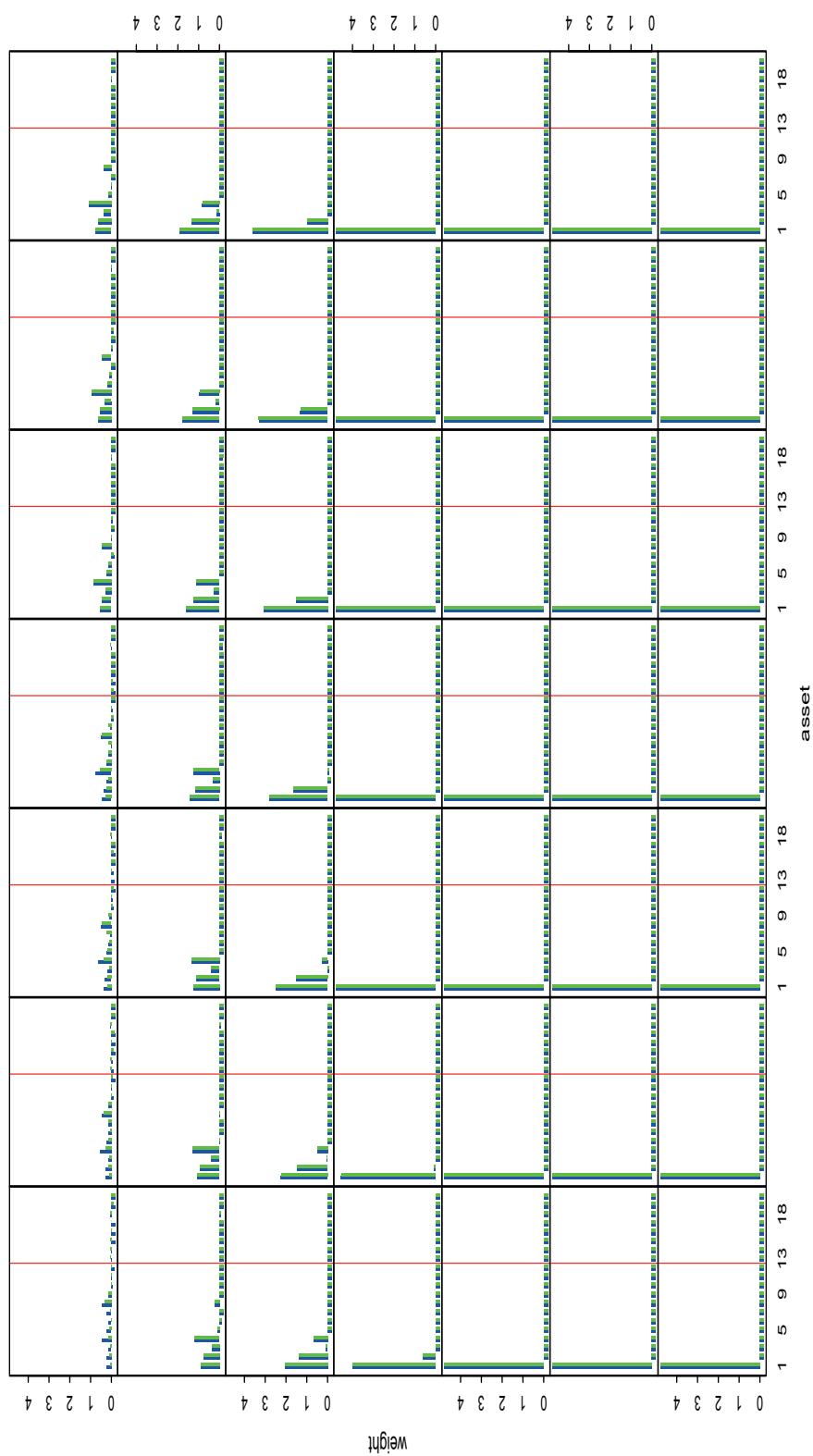optimal portfolio should be more conservative.
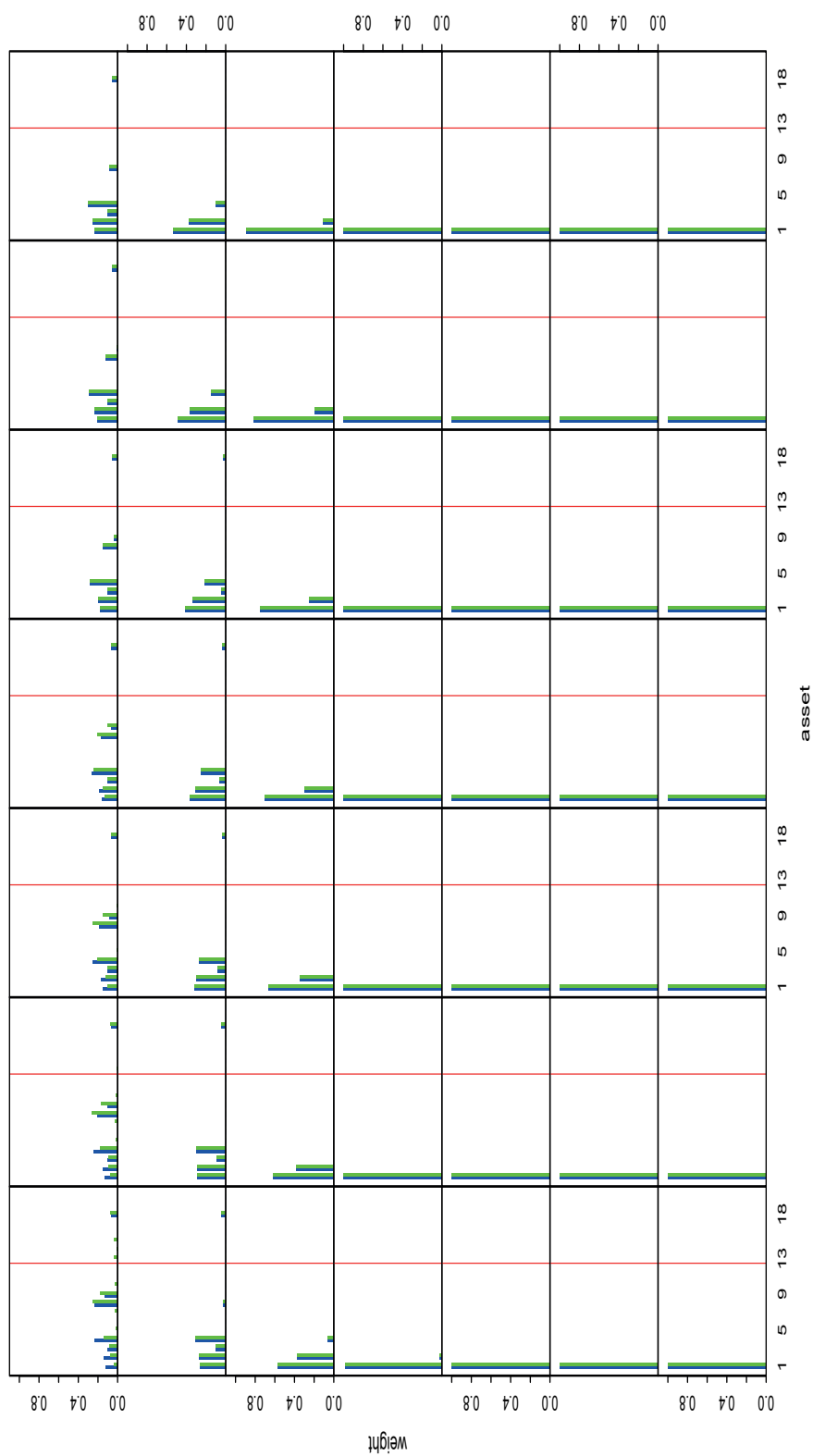
Figure 4.2: each subfigure plots the optimal weights against the asset number for both the type A (blue) and B (green) problems where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
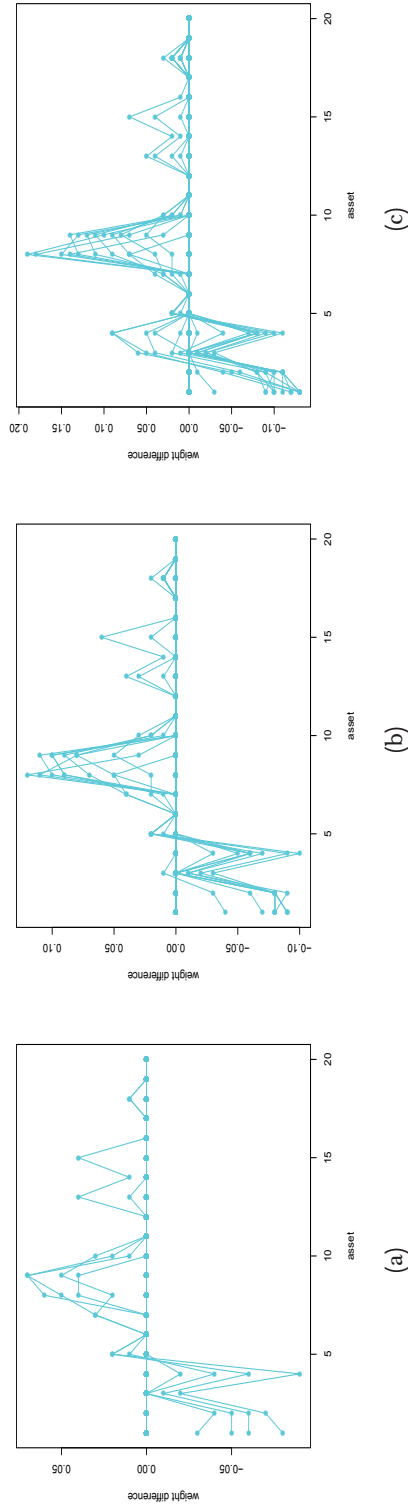
Figure 4.3: (a) plots the optimal weights of the type B problem minus the the optimal weights of the type A problem against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is disallowed; (b) and (c) are analogs of (a) but with the optimal weights of the type C and D problems respectively subtracting the optimal weights of the type A problem instead.
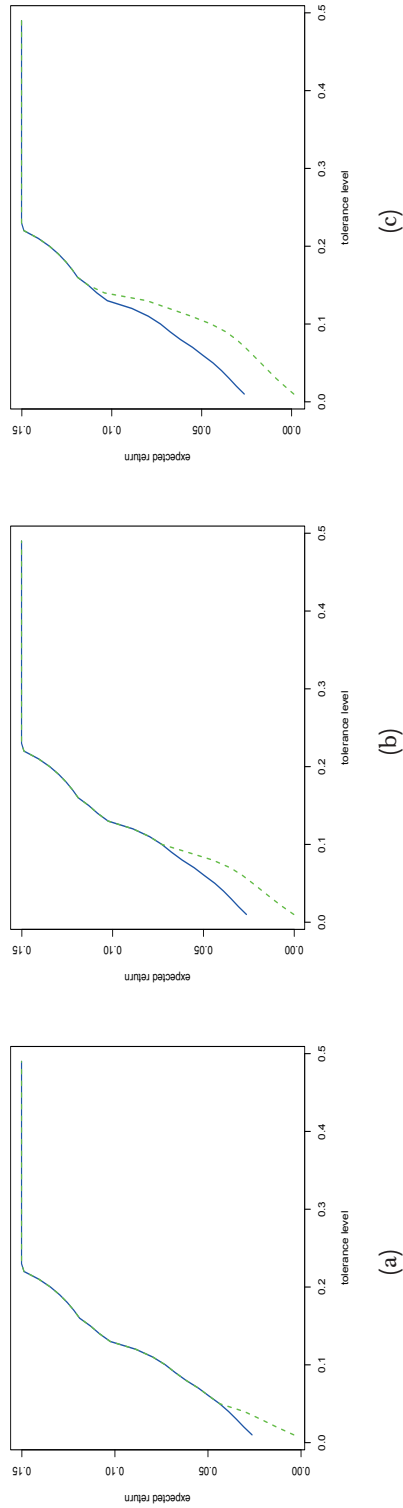
Figure 4.4: (a), (b) and (c) plot the optimal portfolio expected return of the type A problem (blue) and respectively the robust optimal portfolio expected return of the type B, C and D problems (green) against the tolerance level, where short-selling is disallowed.
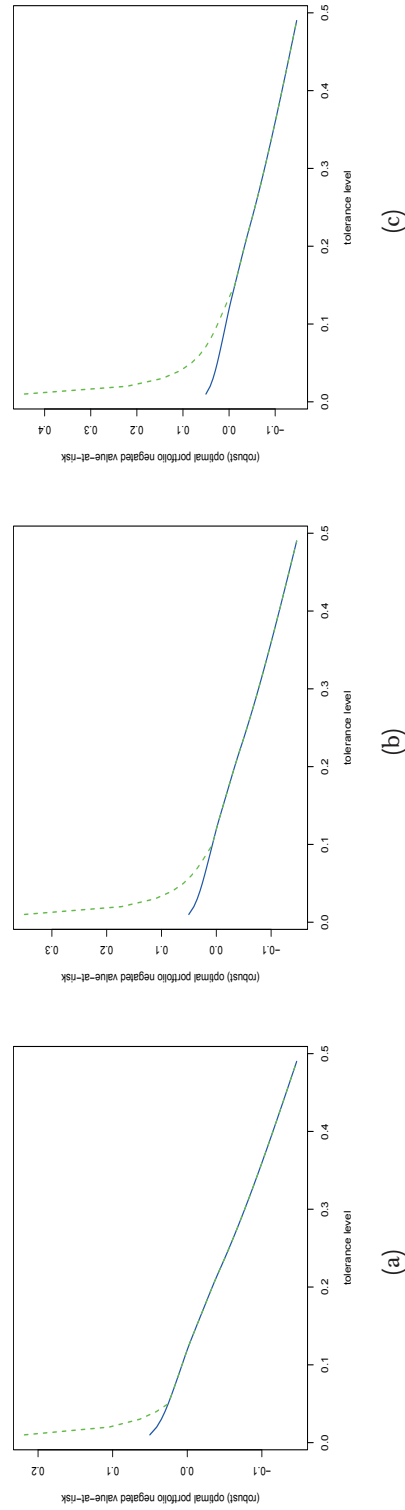
Figure 4.5: (a), (b) and (c) plot the optimal portfolio negated value-at-risk of the type A problem (blue) and respectively the robust optimal portfolio negated value-at-risk of the type B, C and D problems (green) against the tolerance level, where short-selling is disallowed.

Figure 4.6: each subfigure plots the optimal weights against the asset number for both the type A (blue) and B (green) problems where short-selling is allowed up to a maximum of one fifth the total wealth for each asset at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
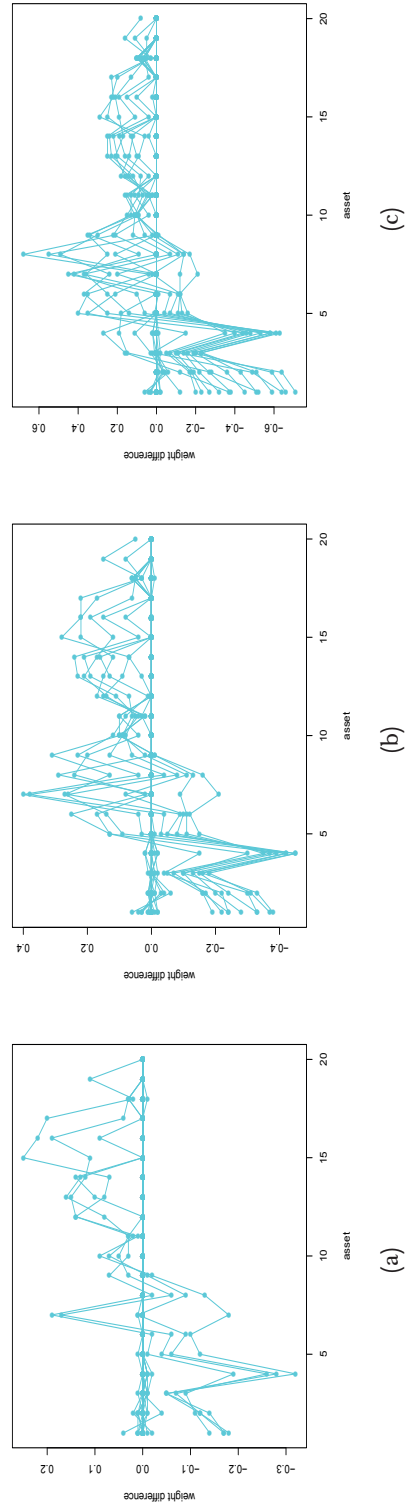
Figure 4.7: (a) plots the optimal weights of the type B problem minus the optimal weights of the type A problem against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is allowed up to a maximum of one fifth the total wealth for each asset; (b) and (c) are analogs of (a) but with the optimal weights of the type C and D problems respectively subtracting the optimal weights of the type A problem instead.
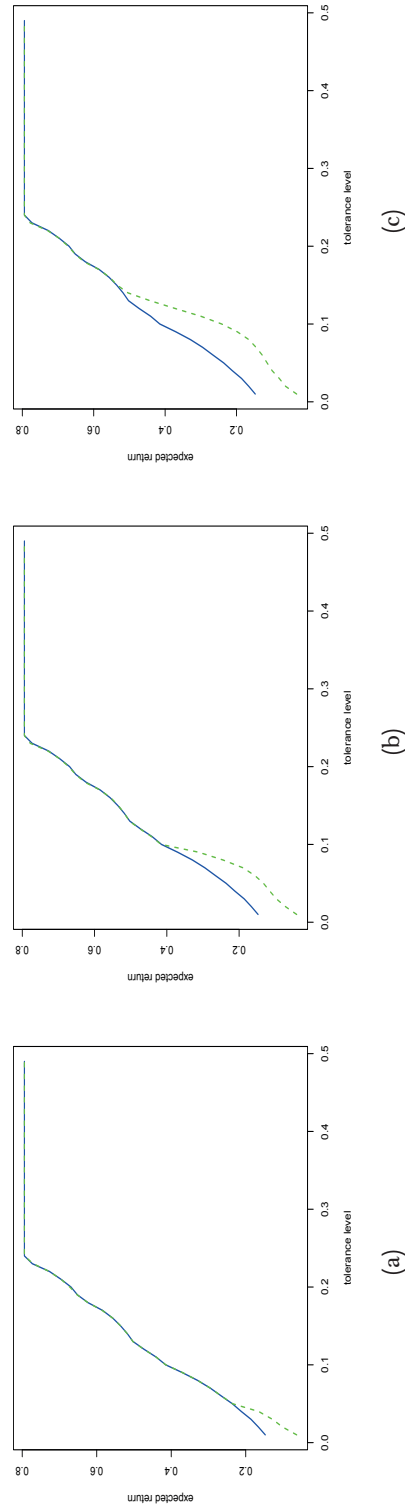
Figure 4.8: (a), (b) and (c) plot the optimal portfolio expected return of the type A problem (blue) and respectively the robust optimal portfolio expected return of the type B, C and D problems (green) against the tolerance level, where short-selling is allowed up to a maximum of one fifth the total wealth for each asset.
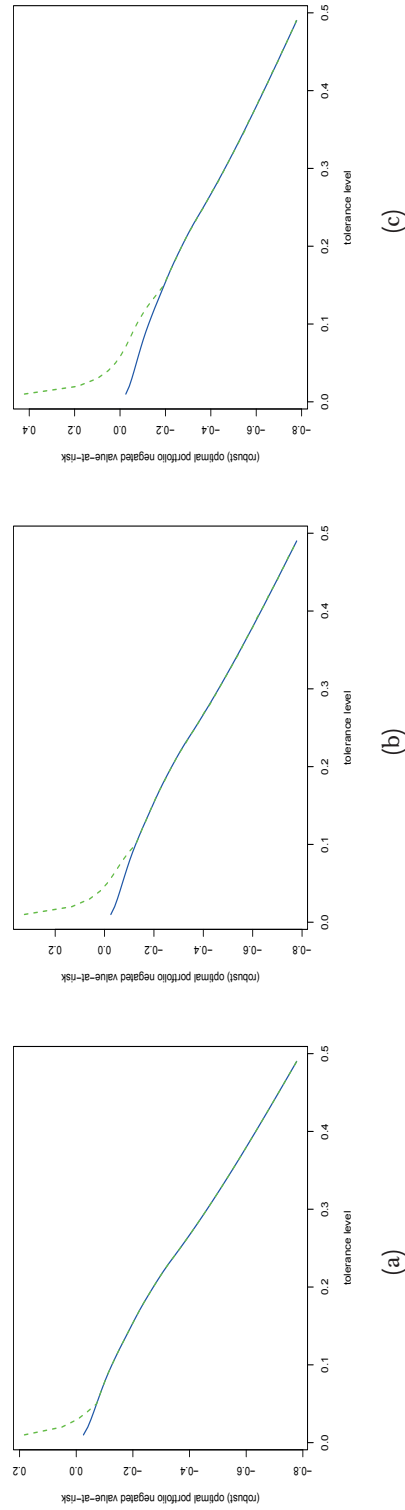
Figure 4.9: (a), (b) and (c) plot the optimal portfolio negated value-at-risk of the type A problem (blue) and respectively the robust optimal portfolio negated value-at-risk of the type B, C and D problems (green) against the tolerance level, where short-selling is allowed up to a maximum of one fifth the total wealth for each asset.

We now set on determining the 'right' size of the location box and the scale matrix eigendecomposition uncertainty sets. More precisely, we find the values of $\mathbf{a} = [a_1, \ldots, a_n]^\mathrm{T}$, $\mathbf{b} = [b_1, \ldots, b_n]^\mathrm{T}$ and c in the problem

$$\max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{\tilde{\mu},\tilde{\lambda},\tilde{\mathbf{x}}\in\mathbb{R}^n} \left\{ \mathbf{w}^\mathrm{T}\tilde{\mu} - \kappa(\epsilon)\sqrt{\sum_{i=1}^n \tilde{\lambda}_i((\mathbf{u}/\|\mathbf{u}\|_2)^\mathrm{T}\mathbf{P}_i\mathbf{w})^2} : \begin{array}{c} \tilde{\mu}_i \in [\mu_i - a_i, \mu_i + a_i] \ \forall i, \\ \tilde{\lambda}_i \in (\max\{0, \lambda_i - b_i\}, \lambda_i + b_i] \ \forall i, \\ (\tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_2)^\mathrm{T}\mathbf{e}_1 \geq 1 - c \end{array} \right\}$$

$$\Leftrightarrow \quad \max_{\mathbf{w}\in\mathcal{W}_\ell} \min_{\nu,\mathbf{u}\in\mathbb{R}^n} \left\{ \mathbf{w}^\mathrm{T}\nu - \kappa(\epsilon)\sqrt{\sum_{i=1}^n \bar{\lambda}_i((\mathbf{u}/\|\mathbf{u}\|_2)^\mathrm{T}\mathbf{P}_i\mathbf{w})^2} : \begin{array}{c} \nu_i \in [\mu_i - a_i, \mu_i + a_i] \ \forall i, \\ (\mathbf{u}/\|\mathbf{u}\|_2)^\mathrm{T}\mathbf{e}_1 \geq 1 - c \end{array} \right\} \quad (4.3)$$

where $\bar{\lambda}_i = \lambda_i + b_i$, based on some criteria.

4.2 Location Uncertainty With No Short-Selling

If we assume only location uncertainty with no short-selling, then the robust location-scale problem (4.3) can be written as

$$\max_{\mathbf{w}\in\mathcal{W}_0} \left\{ \mathbf{w}^\mathrm{T}(\mu - \mathbf{a}) - \kappa(\epsilon)\sqrt{\mathbf{w}^\mathrm{T}\Sigma\mathbf{w}} \right\}. \quad (4.4)$$

Theorem 4.6

(i) *Let $\mathbf{w}_*$ and V be the optimal solution and value of (4.4) respectively. Then,*

$$\frac{dV}{da_i} = -w_{i*}. \quad (4.5)$$

(ii) *We have*

$$\lim_{a_i\to\infty} \frac{dV}{da_i} = 0, \quad (4.6)$$

provided the limit exists.

(iii) *The second derivative of V with respect to each a_i is non-negative.*

Proof: First, note that since (4.4) is a convex optimization problem, there is zero duality gap so that the KKT conditions in Theorem A.23 hold. In particular, we have

$$\nabla_\mathbf{w} L(\mathbf{w}_*, \upsilon_*, \nu_*, \mathbf{a}) = \mathbf{0}, \quad (4.7)$$

where the Lagrangian L of (4.4) is a function of the optimal solution $\mathbf{w}_*$, the inequality and equality KKT multipliers υ_* and ν_* respectively obtained by solving the corresponding dual problem, and $\mathbf{a} = [a_1, \ldots, a_n]^\mathrm{T}$. The Jacobian of (4.7) with respect to $\mathbf{w}$ is exactly the Hessian of the concave objective function of (4.4), which is negative definite, so that it is non-singular. Therefore, the Implicit Function Theorem A.19 implies that w_{i*} is continuously differentiable with respect to a_j, for $i, j = 1, \ldots, n$. Moreover, the partial derivatives of the objective function of (4.4) with respect to w_i and a_i exist for $i = 1, \ldots, n$ at $\mathbf{w}_*$, so that we

can apply the Envelope Theorem A.24 to obtain (4.5). To prove (4.6), note that in effect we want to show $w_i^a = 0$ assuming w_{i*} converges to w_i^a as a_i tends to infinity. To that end, let

$$f_{a_i}(\mathbf{w}) = \frac{1}{a_i}\left(\sum_{j=1}^n w_j(\mu_j - a_j) - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathsf{T}}\Sigma\mathbf{w}}\right)$$

and denote its maximizer over $\mathcal{W}_0$ as $\mathbf{w}_*^{a_i}$. Note that $\mathbf{w}_*^{a_i} = \mathbf{w}_*$ so that by assumption, we have $w_{i*}^{a_i} \to w_i^a$ as $a_i \to \infty$. This implies that $\{w_{i*}^{a_i} : a_i \in \mathbb{R}_{++}\}$ is a bounded sequence in the positive real line, and there is a closed bounded interval $\mathcal{I} \subset \mathbb{R}$ containing $\{w_{i*}^{a_i} : a_i \in \mathbb{R}_{++}\}$, w_i^a and 0. Fix $w_j = w_{j*}^{a_i}$ for $j \neq i$, then $f_{a_i}(\mathbf{w})$ is a function of w_i which we denote as $g_{a_i}(w_i)$ and converges uniformly to $-w_i$ on $\mathcal{I}$ as $a_i \to \infty$. Therefore, for all $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that if $a_i \geq n$, then

$$|g_{a_i}(w_i) + w_i| \leq \epsilon \quad \forall w_i \in \mathcal{I},$$

and in particular,

$$|g_{a_i}(0)| \leq \epsilon \text{ and } |g_{a_i}(w_{i*}^{a_i}) + w_{i*}^{a_i}| \leq \epsilon$$
$$\Rightarrow \quad -\epsilon \leq g_{a_i}(0) \leq g_{a_i}(w_{i*}^{a_i}) \leq g_{a_i}(w_{i*}^{a_i}) + w_{i*}^{a_i} \leq \epsilon$$
$$\Rightarrow \quad |g_{a_i}(w_{i*}^{a_i})| \leq \epsilon. \tag{4.8}$$

Now we set on showing $w_i^a = 0$. Note that since $w_{i*}^{a_i} \to w_i^a$ as $a_i \to \infty$, we have that for all ϵ,

(i) there exists an $n_1 \in \mathbb{N}$ such that if $a_i \geq n_1$, then

$$\left|w_{i*}^{a_i} - w_i^a\right| \leq \frac{1}{3}\epsilon,$$

(ii) there exists an $n_2 \in \mathbb{N}$ such that if $a_i \geq n_2$, then

$$\left|-g_{a_i}(w_{i*}^{a_i})\right| \leq \frac{1}{3}\epsilon$$

by (4.8), and

(iii) there exists an $n_3 \in \mathbb{N}$ such that if $a_i \geq n_3$, then

$$\left|g_{a_i}(w_i) + w_i\right| \leq \frac{1}{3}\epsilon \quad \forall w_i \in \mathcal{I}$$

due to the uniform convergence of $g_{a_i}(w_i)$ to $-w_i$ on $\mathcal{I}$ as $a_i \to \infty$.

The above three statements imply that: For all $\epsilon > 0$, if $a_i \geq n_0 = \max\{n_1, n_2, n_3\}$, then

$$\left|w_{i*}^{a_i} - w_i^a\right| \leq \frac{1}{3}\epsilon, \ \left|-g_{a_i}(w_{i*}^{a_i})\right| \leq \frac{1}{3}\epsilon, \text{ and } \left|g_{a_i}(w_i) + w_i\right| \leq \frac{1}{3}\epsilon \quad \forall w_i \in \mathcal{I}$$
$$\Rightarrow \quad \left|-w_{i*}^{a_i} + w_i^a\right| + \left|-g_{a_i}(w_{i*}^{a_i})\right| + \left|g_{a_i}(w_{i*}^{a_i}) + w_{i*}^{a_i}\right| \leq \epsilon$$

$$\Rightarrow \quad |-w_i^a| \leq \epsilon, \hspace{4cm} \text{(by Triangle Inequality)}$$

which means that $w_i^a = 0$.

Finally, since w_{i*} is differentiable with respect to a_i, the second derivative of V with respect to a_i exists, for $i = 1, \ldots, n$, and is non-negative due to the Second-Order Envelope Theorem A.26. $\qquad \square$

Remark 4.4

The proof of Theorem 4.6 is based on the study of Kanniappan and Sastry (1983).

Part (iii) of Theorem 4.6 implies that the first derivative of V with respect to a_i is non-decreasing in a_i. Thus,

$$\min_{a_i \geq 0} \frac{dV}{da_i} = \left. \frac{dV}{da_i} \right|_{a_i=0} = -w_{i*}(\mathbf{a})|_{a_i=0}$$

and

$$\sup_{a_i \geq 0} \frac{dV}{da_i} = \lim_{a_i \to \infty} \frac{dV}{da_i} = 0,$$

so that we can choose the value of each a_i by solving

$$\min_{a_i \geq 0} \frac{dV}{da_i} + s_i \left(\sup_{a_i \geq 0} \frac{dV}{da_i} - \min_{a_i \geq 0} \frac{dV}{da_i} \right) = \frac{dV}{da_i}, \ i = 1, \ldots, n$$

$$\Leftrightarrow \quad (1 - s_i) w_{i*}(\mathbf{a})|_{a_i=0} = w_{i*}(\mathbf{a}), \ i = 1, \ldots, n \tag{4.9}$$

for $a_1, \ldots, a_n$ simultaneously, where $s_i \in (0, 1)$. The right-hand side of (4.9) can be interpreted as the sensitivity of V to changes in a_i while the left-hand side its targeted level. However, (4.9) is a set of n non-linear equations in n unknowns, which is not easily solvable. Therefore, we instead find the root a_{i*} of

$$(1 - s_i) w_{i*}(\mathbf{a})|_{\mathbf{a}=0} = w_{i*}(\mathbf{a})|_{a_j=0, j \neq i} \tag{4.10}$$

for $i = 1, \ldots, n$, separately. Of course, the solution $\mathbf{a}_*$ obtained by such an approach is not ideal since it does not satisfy (4.9), but since the right hand side of (4.10) is non-increasing in a_i, solving such an equation is not an issue numerically.

At first glance, it appears thatS scaling is a problem for this procedure. As we show next, this is not the case. Assume that each a_i is scaled by a parameter $k_i > 0$ in (4.4) so that it becomes

$$\max_{\mathbf{w} \in \mathcal{W}_0} \left\{ \sum_{i=1}^n w_i(\mu_i - k_i a_i) - \kappa(\epsilon)\sqrt{\mathbf{w}^T \Sigma \mathbf{w}} \right\}, \tag{4.11}$$

and denote the ith optimal weight and value of (4.11) as $\tilde{w}_{i*}$ and $\tilde{V}$ respectively. Then, we have that

$$\frac{d\tilde{V}}{da_i} = -k_i \tilde{w}_{i*},$$

with its limit as $a_i \to \infty$ being zero, and second derivative non-negative, as is the case without scaling. We solve

$$(1 - s_i) \tilde{w}_{i*}(\mathbf{a})|_{\mathbf{a}=0} = \tilde{w}_{i*}(\mathbf{a})|_{a_j=0, j \neq i} \tag{4.12}$$

$$\Leftrightarrow \quad (1-s_i)w_{i*}(\mathbf{a})|_{\mathbf{a}=\mathbf{0}} = w_{i*}(k_1 a_1, \ldots, k_n a_n)|_{a_j=0, j \neq i} \tag{4.13}$$

for $i = 1, \ldots, n$ separately to obtain $\mathbf{a}_{**} = [a_{1*}/k_1, \ldots, a_{n*}/k_n]^{\mathrm{T}}$ so that substituting it into (4.11) yields the exact same problem as if $\mathbf{a}_*$ is substituted into (4.4). That is to say, the optimal solution of (4.4) does not change with the scaling of each a_i. This scale invariance property is good news for us, because we can just assume each $k_i = 1$ without loss of generality.

4.2.1 Numerical Experiment Revisited

We go back to the numerical experiment in Section 4.1.1 but instead consider the robust location-scale problem (4.4) with

$$\kappa(\epsilon) = z_{1-\epsilon}$$

which we label as type I and

$$\kappa(\epsilon) = \max_{\alpha \in \{1, 1.01, \ldots, 2\}} \left\{ F_{Z_2}^{-1}(1 - \epsilon_*) F_{Z_\alpha}^{-1}(1 - \epsilon) / F_{Z_\alpha}^{-1}(1 - \epsilon_*) \right\}$$

for $\epsilon_* = 0.05$, 0.1 and 0.15 which we label as types II, III and IV respectively. Recall we always choose the median sensitivity as our targeted sensitivity level. In Figure 4.10, each subfigure plots the optimal weights of the type I problem both with added location uncertainty and without added uncertainty at a particular tolerance level ϵ against the asset number where short-selling is disallowed, and has a red vertical which separates assets with a positive location over scale ratio on the left from assets with a negative location over scale ratio on the right. The subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$. Figures C.9 - C.9 in Appendix C are analogs of Figure 4.10 for the type II, III and IV problems respectively. Figure 4.11 (a) plots the optimal weights of the type I problem with added location uncertainty minus the optimal weights of the same problem without added uncertainty against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is disallowed. Figures 4.11 (b), (c) and (d) are analogs of Figure 4.11 (a) but for the type II, III and IV problems respectively. From these figures just mentioned above, we see that added location uncertainty has a considerable influence on the optimal asset allocation and in particular, causes weight to be shifted from assets with higher loca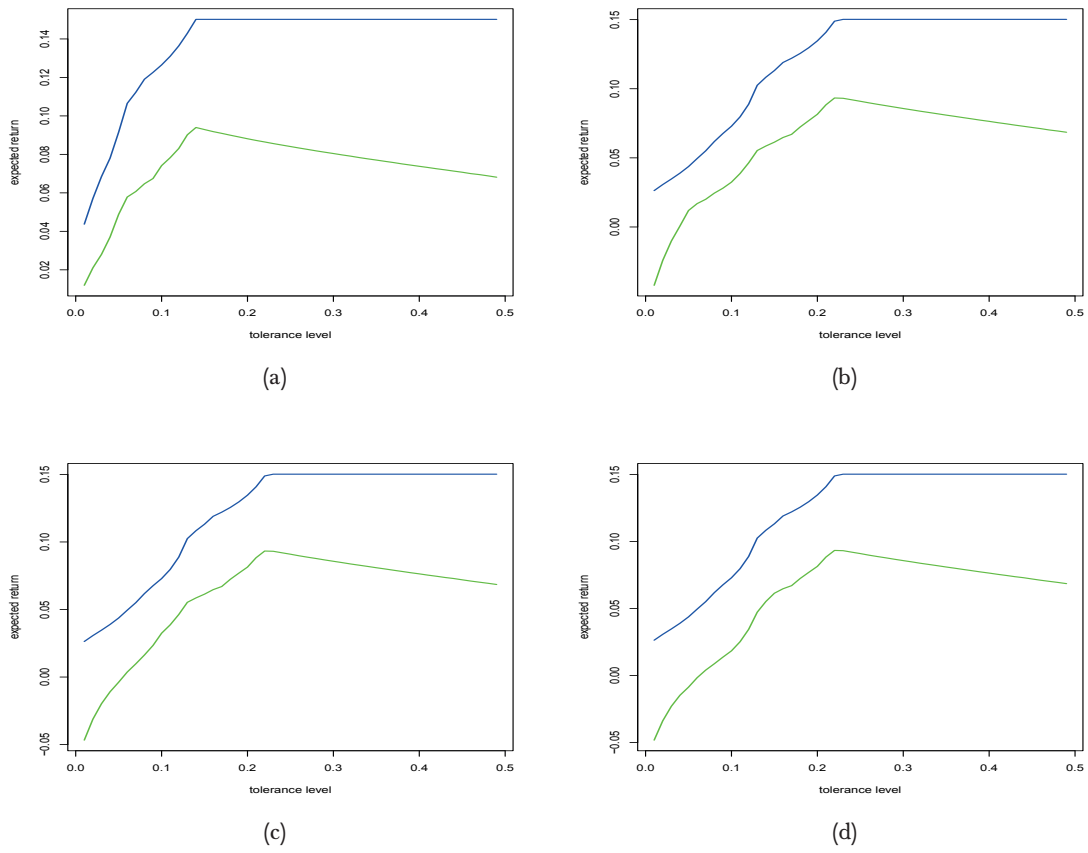tion over scale ratio to those with lower. In Figure 4.12, each subfigure plots the optimal portfolio expected return for a particular type of problem without uncertainty and the robust optimal portfolio expected return of the same problem with added location uncertainty against the tolerance level, where short-selling is disallowed. We observe that the optimal portfolio expected return of any type of problem without added uncertainty is never lower than the robust optimal portfolio expected return of the same problem with added mean uncertainty. Figure 4.13 is an analog of Figure 4.12 for the portfolio negated value-at-risk. Since the robust optimal portfolio negated value-at-risk of any type of problem with added location uncertainty is never lower than the optimal portfolio negated value-at-risk without added uncertainty by construction, it is less likely to invest all wealth in the risky assets with added location uncertainty as compared to when no unc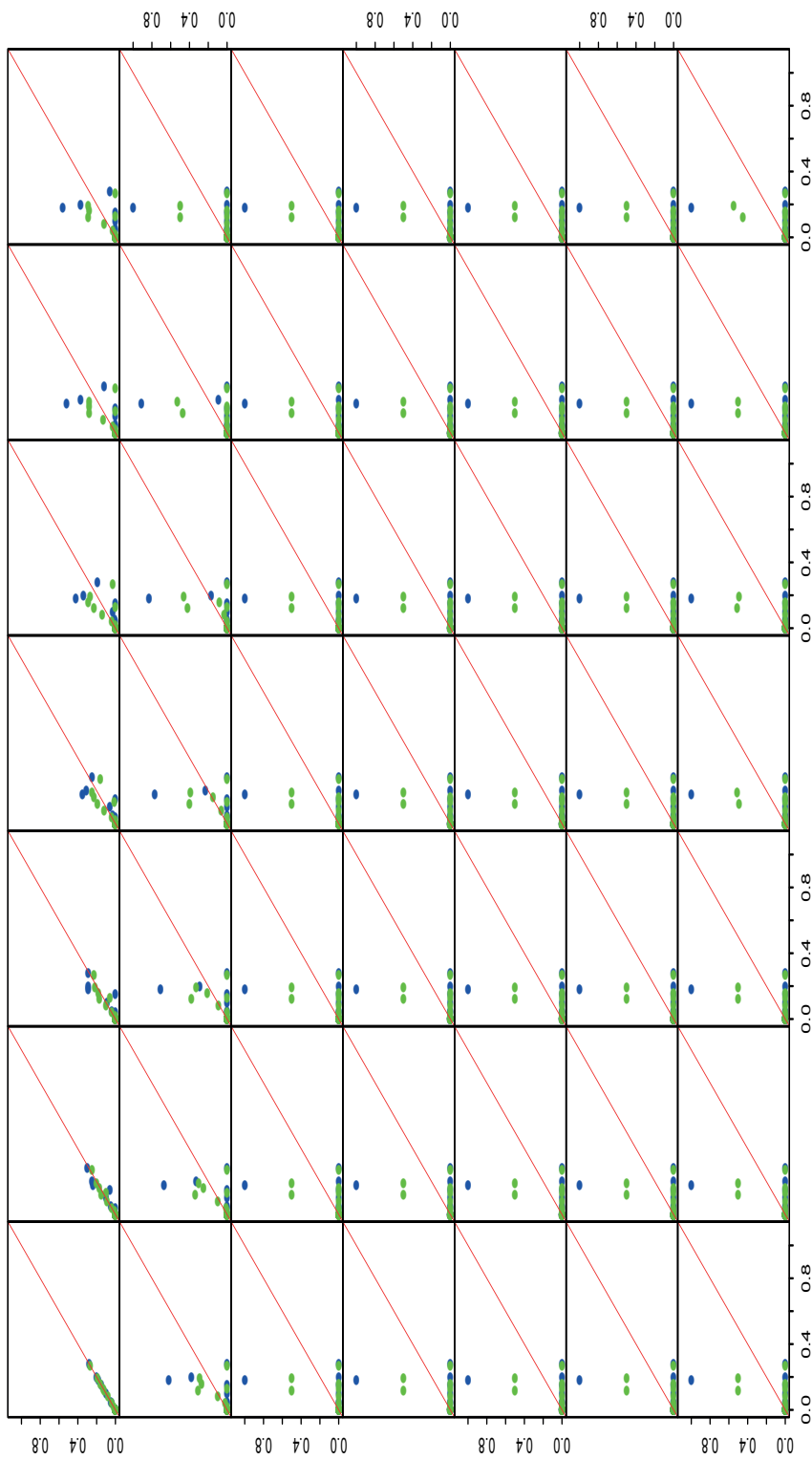ertainty is added, which certainly makes sense. In Figure 4.10, each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type I problem both with added location uncertainty and without added uncertainty, where short-selling is disallowed. The subfigures from

left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$. Figure 4.10 shows that the move towards a less diversified portfolio in the sense that it contains less assets is more gradual with added location uncertainty as compared to without added uncertainty, as can be seen by the fact that the green dots move away from the red diagonal line through the origin with unit gradient at a slower rate than the blue dots as ϵ increases. Figures C.7 - C.9 are analogs of Figure 4.10 for the type II, III and IV problems respectively, with similar observations.
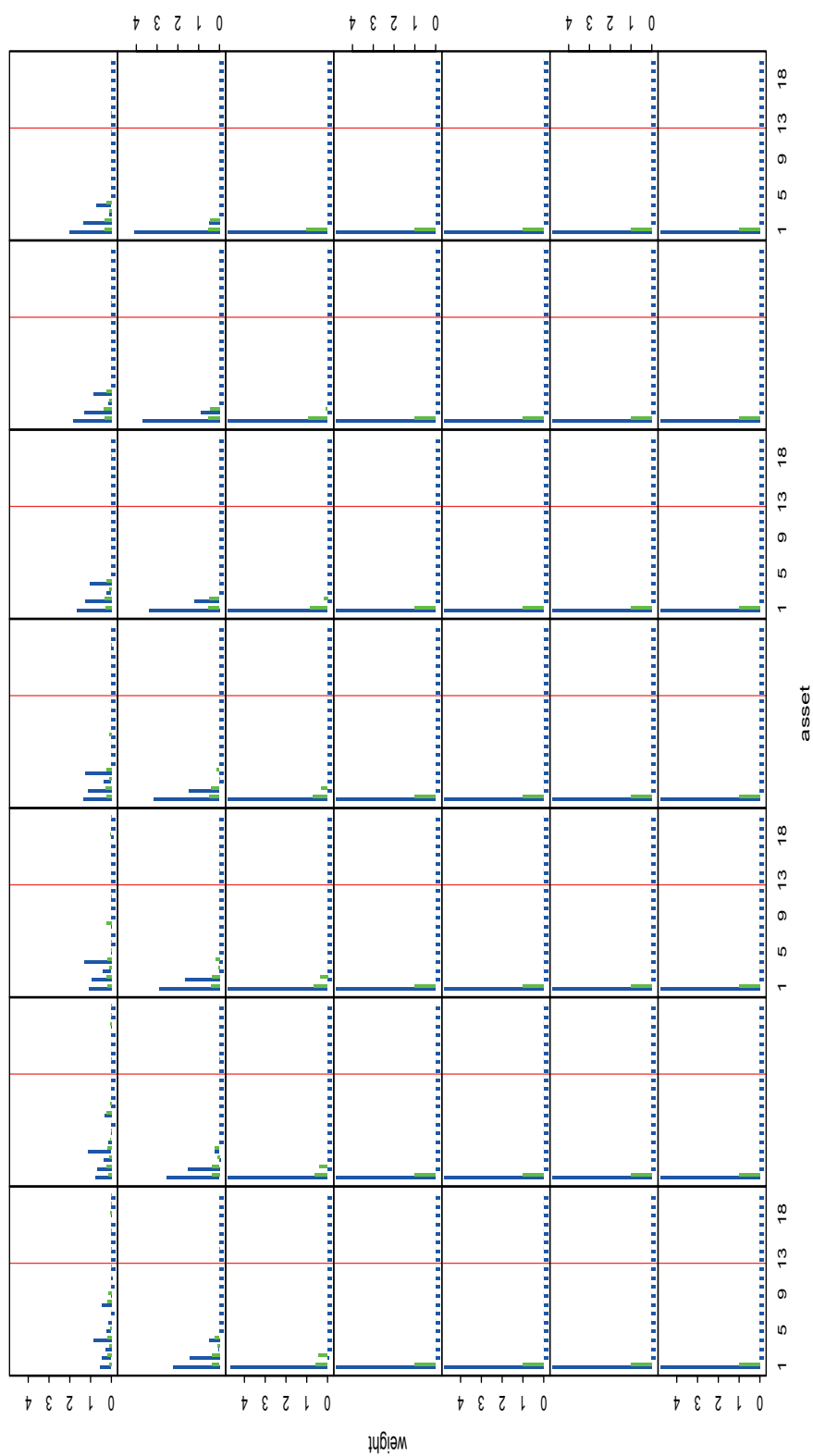
Figure 4.10: each subfigure plots the optimal weights of the type I problem both with added location uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
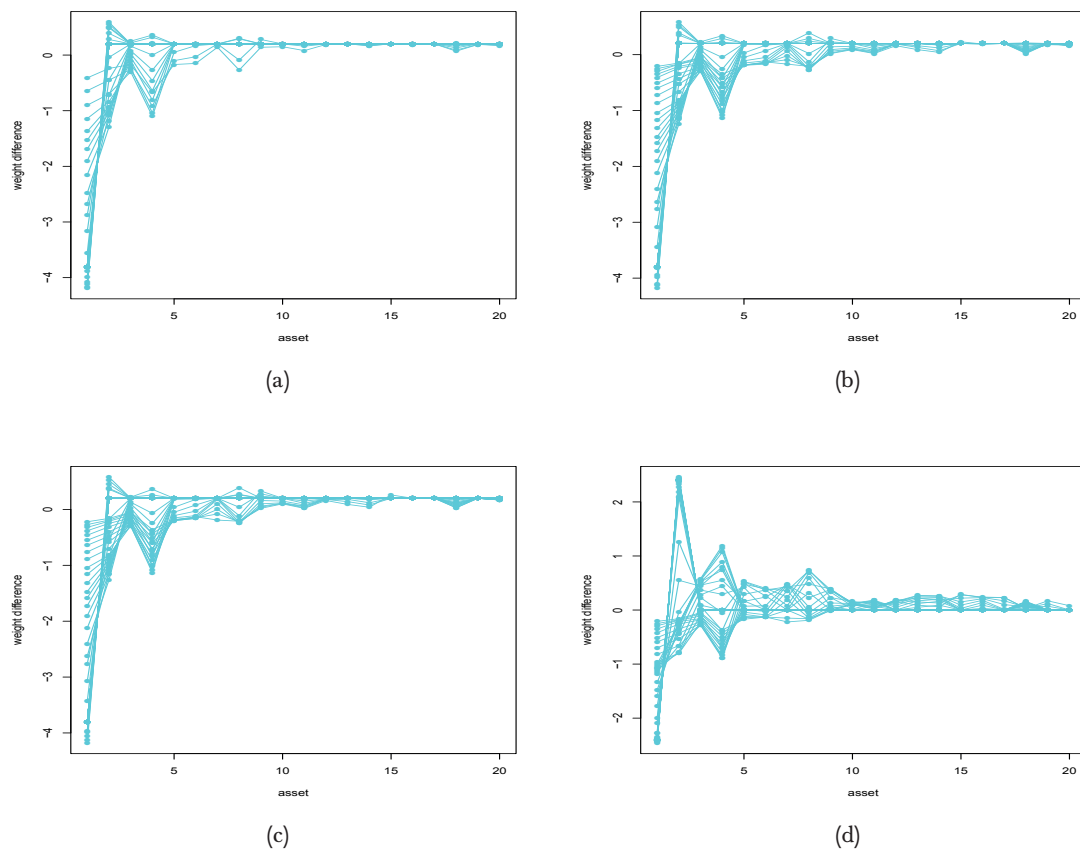
(a)

(b)

(c)

(d)
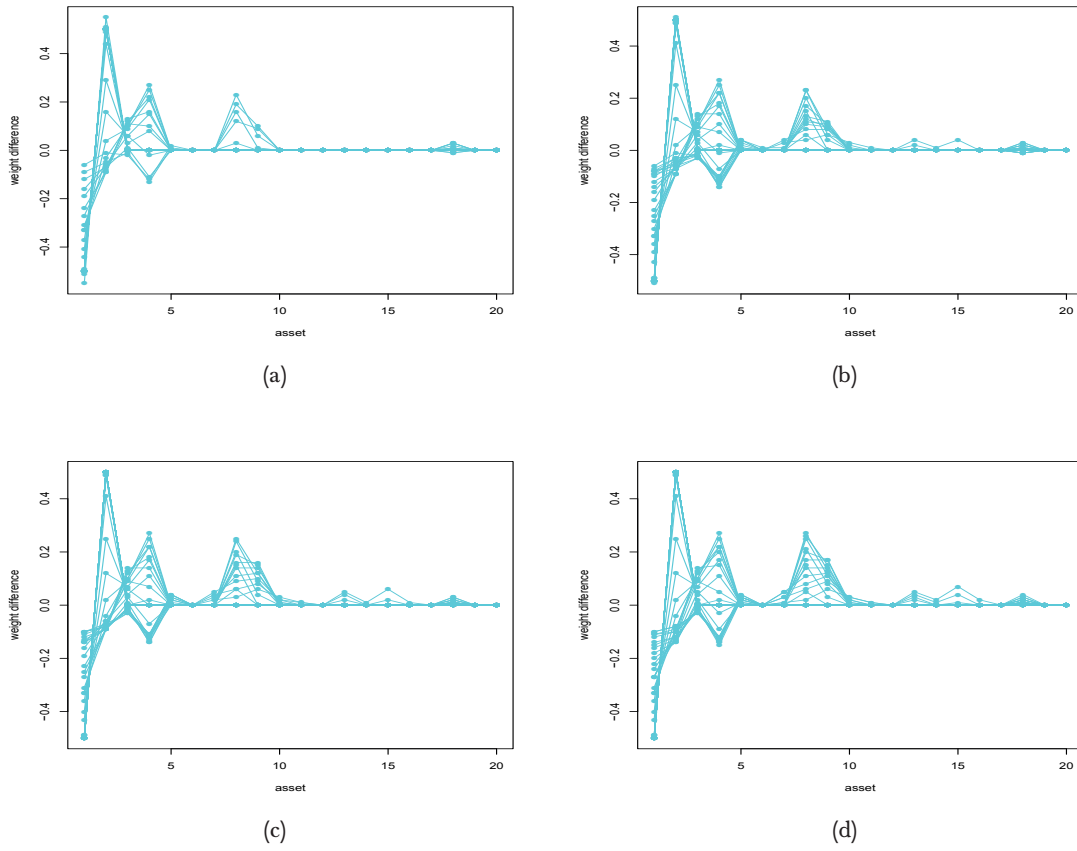
Figure 4.11: (a) plots the optimal weights of the type I problem with added location uncertainty minus the optimal wieghts of the same problem without added uncertainty against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.

Figure 4.12: (a) plots the optimal portfolio expected return for the type I problem without uncertainty (blue) and the robust optimal portfolio expected return of the same problem with added location uncertainty (green) against the tolerance level, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.

(a)

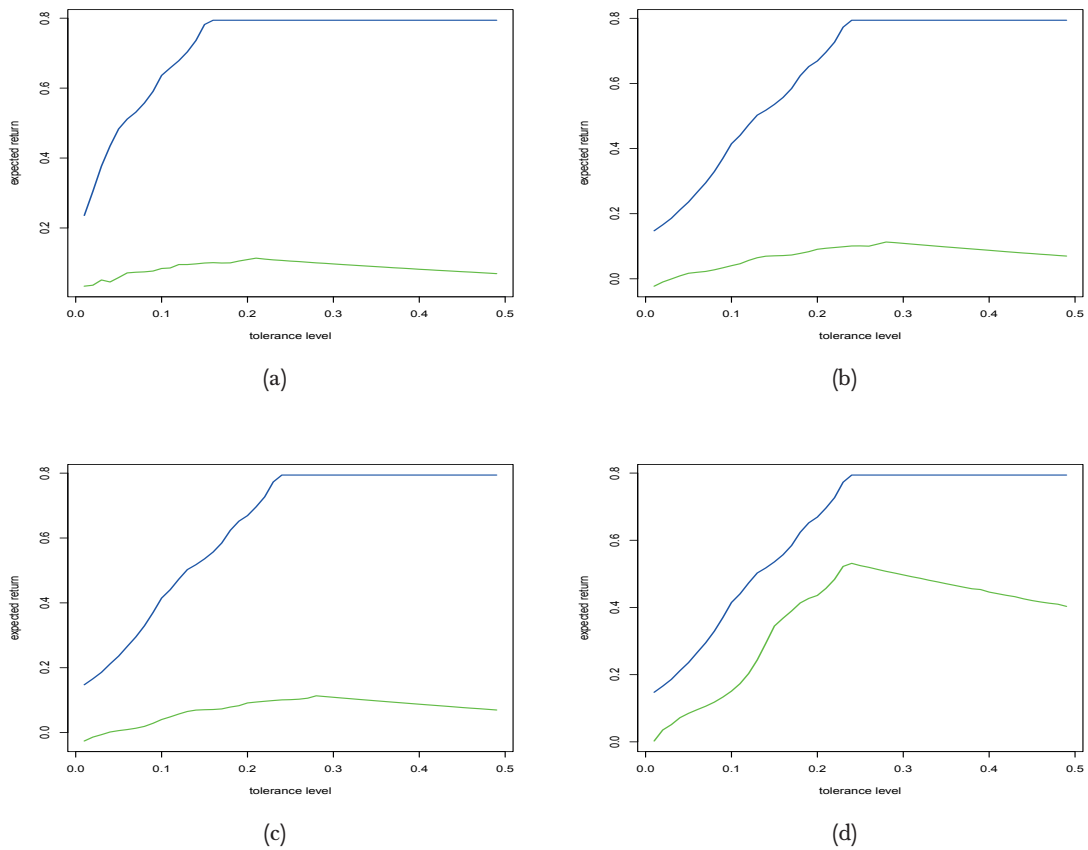(b)

(c)

(d)
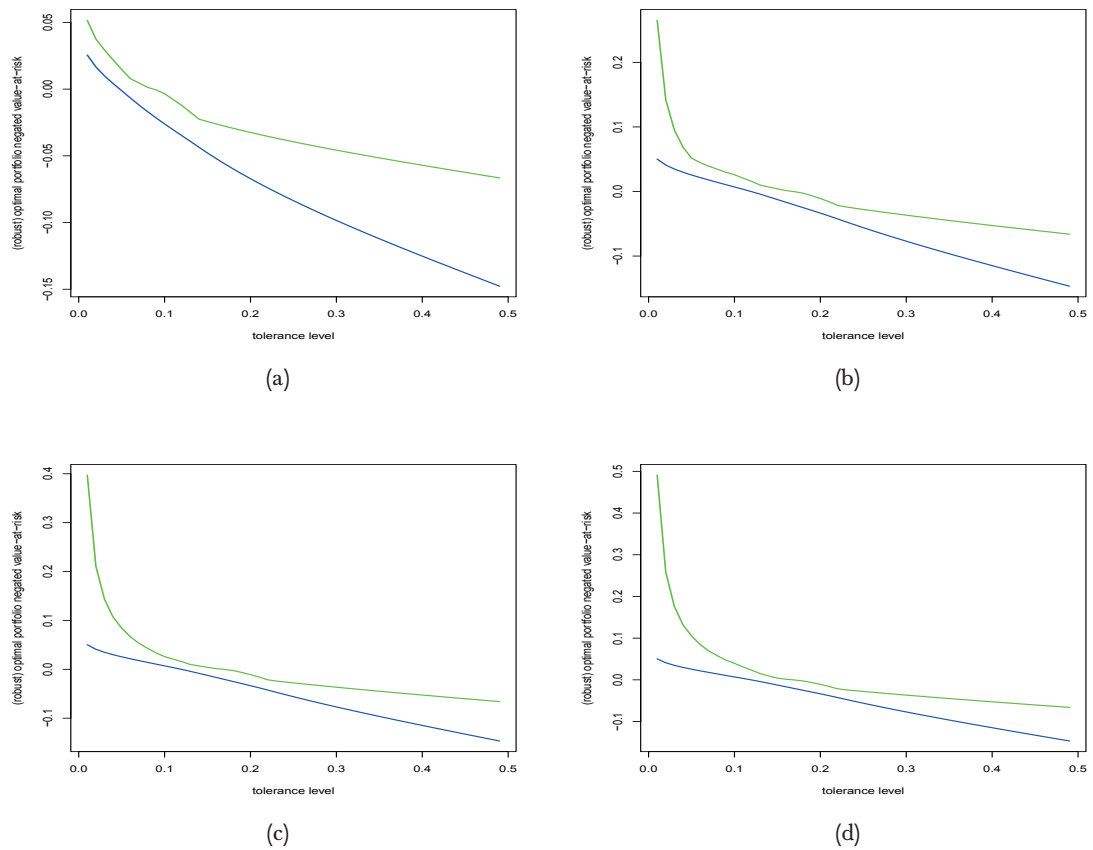
Figure 4.13: (a) plots the optimal portfolio negated value-at-risk for the type I problem without uncertainty (blue) and the robust optimal portfolio negated value-at-risk of the same problem with added location uncertainty (green) against the tolerance level, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.

Figure 4.14: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type I problem both with added location uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

4.3 Location Uncertainty With Short-Selling

If we only assume location uncertainty with short-selling, then the robust location-scale problem (4.3) can be written as

$$\max_{(\mathbf{w},\mathbf{x})\in\mathcal{W}_\ell\times\mathbb{R}^n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\mathbf{x}^{\mathrm{T}}\mathbf{a}-\kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}:-\mathbf{x}\le\mathbf{w}\le\mathbf{x}\right\}. \tag{4.14}$$

Theorem 4.7

(i) *Let $(\mathbf{w}_*,\mathbf{x}_*)$ and V be the optimal solution and value of (4.14) respectively. Then,*

$$\frac{dV}{da_i}=-x_{i*}. \tag{4.15}$$

(ii) *We have*

$$\lim_{a_i\to\infty}\frac{dV}{da_i}=0, \tag{4.16}$$

provided the limit exists.

(iii) *The second derivative of V with respect to each a_i is non-negative.*

Proof: The proof is completely analogous to that of Theorem 4.6, and thus omitted. □
  Due to part (iii) of Theorem 4.7,

$$\min_{a_i\ge0}\frac{dV}{da_i}=-x_{i*}(\mathbf{a})|_{a_i=0}$$

and

$$\sup_{a_i\ge0}\frac{dV}{da_i}=0.$$

We choose the value of each a_i with desired sensitivity level $s_i\in(0,1)$ by solving

$$(1-s_i)x_{i*}(\mathbf{a})|_{\mathbf{a}=\mathbf{0}}=x_{i*}(\mathbf{a})|_{a_j=0,j\ne i} \tag{4.17}$$

for $i=1,\dots,n$ separately. Again, note that solving (4.17) numerically is not a problem since its right-hand side is non-increasing.
  Scaling is also not a problem. Assume that each a_i is scaled by a parameter $k_i>0$ in (4.14) so that it becomes

$$\max_{(\mathbf{w},\mathbf{x})\in\mathcal{W}_\ell\times\mathbb{R}^n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\sum_{i=1}^n x_i k_i a_i-\kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}:-\mathbf{x}\le\mathbf{w}\le\mathbf{x}\right\}, \tag{4.18}$$

with the optimal solution of x_i and value denoted as $\tilde{x}_{i*}$ and $\tilde{V}$ respectively. Then, we have that

$$\frac{d\tilde{V}}{da_i}=-k_i\tilde{x}_{i*},$$

with its limit as $a_i \to \infty$ and second derivative zero and non-negative respectively. Finally, solving

$$(1 - s_i)\tilde{x}_{i*}(\mathbf{a})\big|_{\mathbf{a}=\mathbf{0}} = \tilde{x}_{i*}(\mathbf{a})\big|_{a_j=0, j \neq i} \tag{4.19}$$

$$\Leftrightarrow \quad (1 - s_i)x_{i*}(\mathbf{a})\big|_{\mathbf{a}=\mathbf{0}} = x_{i*}(k_1 a_1, \dots, k_n a_n)\big|_{a_j=0, j \neq i} \tag{4.20}$$

for $i = 1, \dots, n$ separately obtains $\mathbf{a}_{**} = [a_{1*}/k_1, \dots, a_{n*}/k_n]^{\mathrm{T}}$, which when being substituted into (4.18) yields the exact same problem as if $\mathbf{a}_*$ is substituted into (4.14). Therefore, we can just assume each $k_i = 1$ without loss of generality.

4.3.1 Numerical Experiment Revisited

Figures 4.15 - 4.19 correspond to the type I problem and are analogs of those in Section 4.2.1, with all else remaining constant except that short-selling is allowed up to a maximum of one-fifth the total wealth for each asset. Figures C.13 - C.15 (analogs of Figure 4.15) and Figures C.16 - C.18 (analogs of Figure 4.19) in Appendix C correspond to the types II, III and IV problems respectively. Note how added location uncertainty encourages less short-selling in Figures 4.15 and C.13 - C.15, but otherwise similar conclusions as in the case where short-selling is disallowed can be drawn.

Figure 4.15: each subfigure plots the optimal weights of the type I problem both with added location uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

(a)

(b)

(c)

(d)

Figure 4.16: (a) plots the optimal weights of the type I problem with added location uncertainty minus the optimal wieghts of the same problem without added uncertainty against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.

Figure 4.17: (a) plots the optimal portfolio expected return for the type I problem without uncertainty (blue) and the robust optimal portfolio expected return of the same problem with added location uncertainty (green) against the tolerance level, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
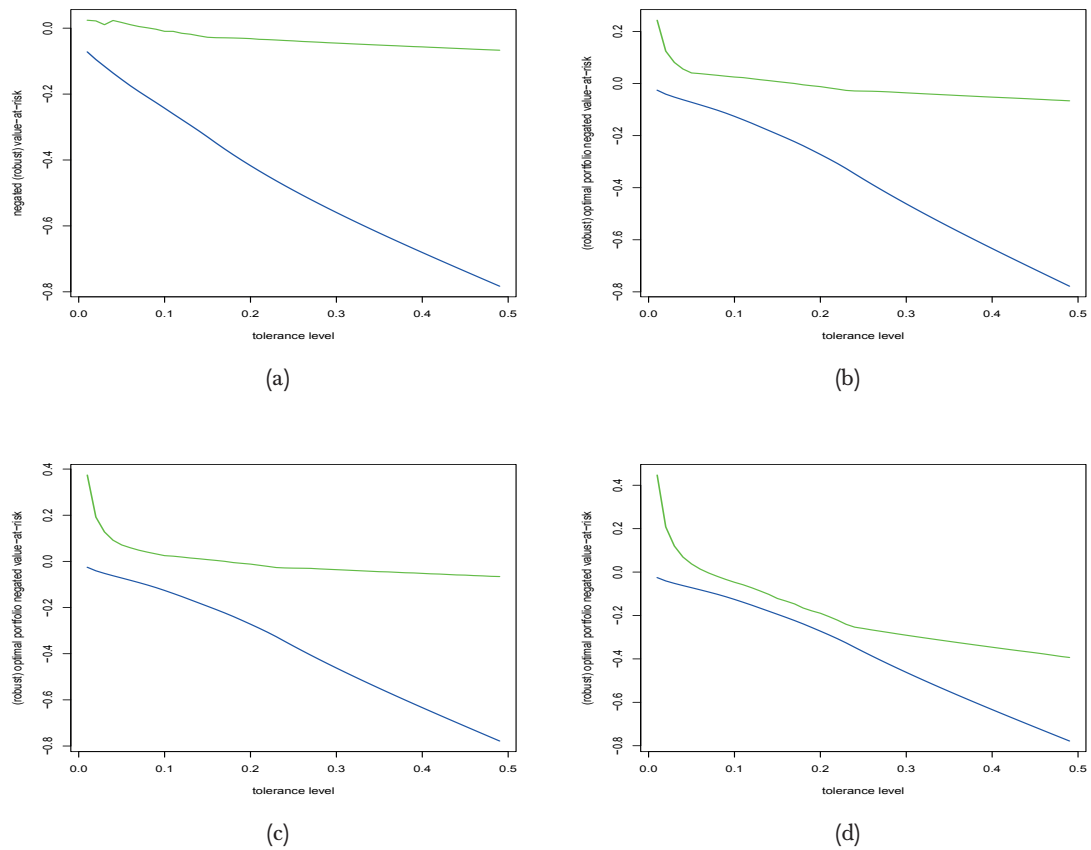
Figure 4.18: (a) plots the optimal portfolio negated value-at-risk for the type I problem without uncertainty (blue) and the robust optimal portfolio negated value-at-risk of the same problem with added location uncertainty (green) against the tolerance level, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
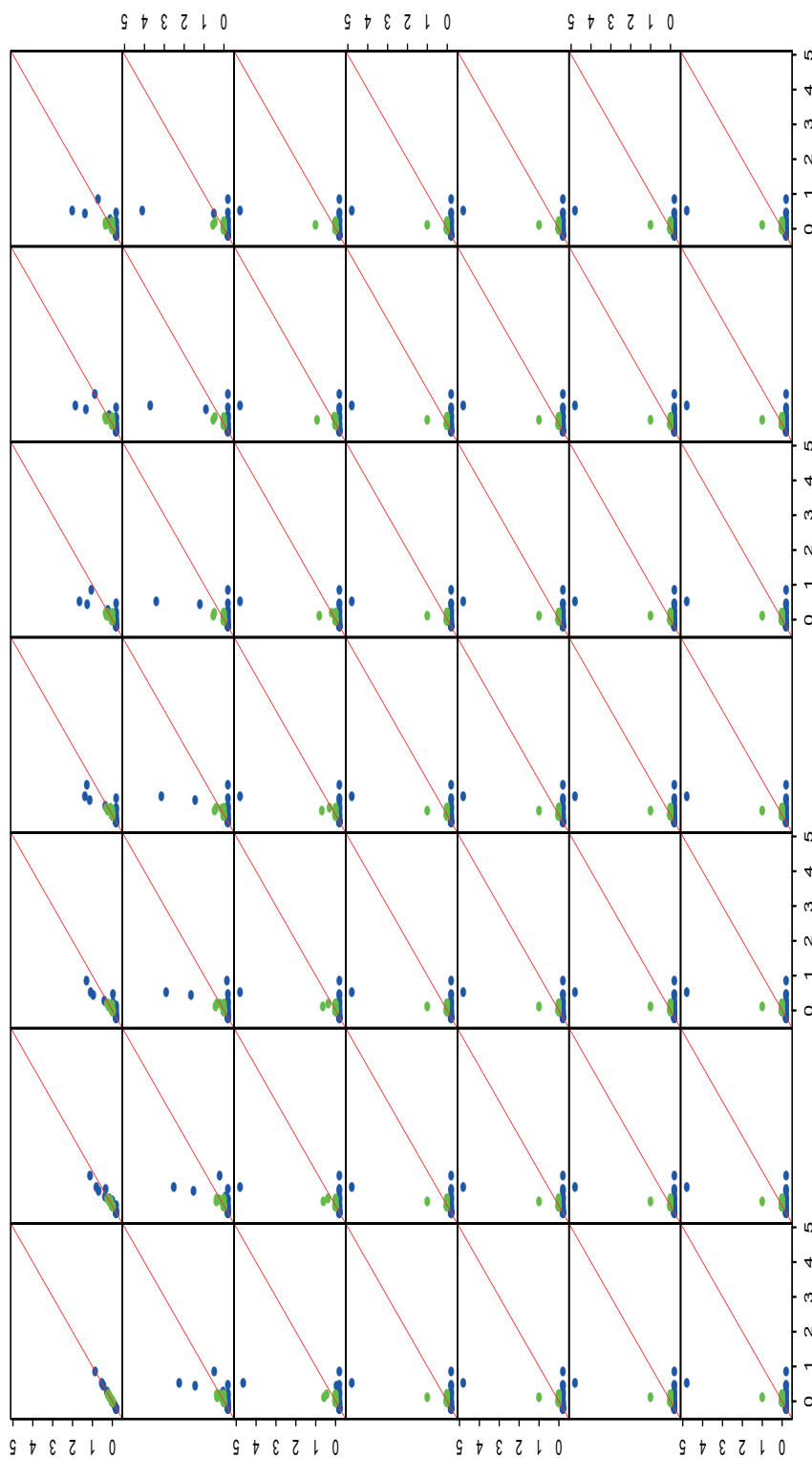
Figure 4.19: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type I problem both with added location uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

4.4 Eigenvalue Uncertainty

If we only assume eigenvalue uncertainty, then the robust location-scale problem (4.3) can be written as

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\left\{\mathbf{w}^\mathsf{T}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\sum_{i=1}^n(\lambda_i+b_i)(\mathbf{w}^\mathsf{T}\mathbf{u}_i)^2}\right\}. \tag{4.21}$$

Theorem 4.8

  (i) *Let $\mathbf{w}_*$ and V be the optimal solution and value of* (4.21) *respectively. Then,*

$$\frac{dV}{db_i}=-\frac{\kappa(\epsilon)(\mathbf{w}_*^\mathsf{T}\mathbf{u}_i)^2}{2\sqrt{\sum_{j=1}^n(\lambda_j+b_j)(\mathbf{w}_*^\mathsf{T}\mathbf{u}_j)^2}}. \tag{4.22}$$

 (ii) *We have*

$$\lim_{b_i\to\infty}\frac{dV}{db_i}=0, \tag{4.23}$$

  provided the limit exists.

(iii) *The second derivative of V with respect to each b_i is non-negative.*

Proof: The proof of (4.22) is analogous to that of (4.5) in Theorem 4.6.
  To show (4.23), assume that $\lim_{b_i\to\infty}\mathbf{w}_*=\mathbf{w}_{b_i}$ so that

$$\lim_{b_i\to\infty}\frac{dV}{db_i}=-\frac{\kappa(\epsilon)(\mathbf{w}_{b_i}^\mathsf{T}\mathbf{u}_i)^2}{2\sqrt{(\lambda_i+\lim_{b_i\to\infty}b_i)(\mathbf{w}_{b_i}^\mathsf{T}\mathbf{u}_i)^2+\sum_{j\neq i}(\lambda_j+b_j)(\mathbf{w}_{b_i}^\mathsf{T}\mathbf{u}_j)^2}}. \tag{4.24}$$

If $\mathbf{w}_{b_i}^\mathsf{T}\mathbf{u}_i=0$, then obviously (4.24) is zero, since its numerator is vanishing and its denominator is positive, otherwise there will be $n+1$ orthogonal vectors $\mathbf{w}_{b_i},\mathbf{u}_1,\dots,\mathbf{u}_n$ in $\mathbf{R}^n$, which is impossible. If $\mathbf{w}_{b_i}^\mathsf{T}\mathbf{u}_i\neq0$, then (4.24) is also zero since its numerator is finite and its denominator tends to infinity.

  It is easy to check that the second derivative of V with respect to each b_i exists just by observing its expression. Applying the Second-Order Envelope Theorem A.26 obtains

$$\frac{d^2V}{db_i^2}\geq\frac{\kappa(\epsilon)(\mathbf{w}_*^\mathsf{T}\mathbf{u}_i)^4}{4\left(\sum_{i=1}^n(\lambda_i+b_i)(\mathbf{w}_*^\mathsf{T}\mathbf{u}_i)^2\right)^{3/2}}\geq0,$$

which concludes our proof.   □

  Note that an infinitesimal change in b_i will not have any effect on V if $\mathbf{w}_*$ is orthogonal to $\mathbf{u}_i$. Each b_i is set to the root b_{i*} of

$$\min_{b_i\geq0}\left\{\frac{dV}{db_i}\Big|_{b_j=0.j\neq i}\right\}+s_i\left(\max_{b_i\geq0}\left\{\frac{dV}{db_i}\Big|_{b_j=0.j\neq i}\right\}-\min_{b_i\geq0}\left\{\frac{dV}{db_i}\Big|_{b_j=0.j\neq i}\right\}\right)=\frac{dV}{db_i}\Big|_{b_j=0,j\neq i}$$

$$\Leftrightarrow \quad (1-s_i)\left(\frac{dV}{db_i}\Big|_{b_1,\dots,b_n=0}\right) + s_i\left(\lim_{b_i\to\infty}\left\{\frac{dV}{db_i}\Big|_{b_j=0,j\neq i}\right\}\right) = \frac{dV}{db_i}\Big|_{b_j=0,j\neq i}$$

$$\Leftrightarrow \quad (1-s_i)\left(\frac{dV}{db_i}\Big|_{b_1,\dots,b_n=0}\right) = \frac{dV}{db_i}\Big|_{b_j=0,j\neq i},$$

where $s_i \in (0,1)$ is the chosen sensitivity level, the first equivalence is due to part (iii) of Theorem 4.8, and the second equivalence is by (4.23).

Like in the case of mean uncertanty, scaling does not pose a problem. Assume that each b_i is scaled by a parameter $k_i > 0$ in (4.21) so that it becomes

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\left\{\mathbf{w}^{\mathsf{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\sum_{i=1}^{n}(\lambda_i + k_ib_i)(\mathbf{w}^{\mathsf{T}}\mathbf{u}_i)^2}\right\}, \tag{4.25}$$

with the optimal solution and value denoted as $\tilde{\mathbf{w}}_*$ and $\tilde{V}$ respectively. Then, we have that

$$\frac{d\tilde{V}}{db_i} = -\frac{\kappa(\epsilon)(\tilde{\mathbf{w}}_*^{\mathsf{T}}\mathbf{u}_i)^2}{2\sqrt{\sum_{j=1}^{n}(\lambda_j + k_jb_j)(\tilde{\mathbf{w}}_*^{\mathsf{T}}\mathbf{u}_j)^2}},$$

with its limit as $b_i \to \infty$ and second derivative remaining zero and non-negative respectively.

Finally, we solve

$$(1-s_i)\left(\frac{d\tilde{V}(b_1,\dots,b_n)}{db_i}\Big|_{b_1,\dots,b_=0}\right) = \frac{d\tilde{V}(b_1,\dots,b_n)}{db_i}\Big|_{b_j=0,j\neq i}$$

$$\Leftrightarrow \quad (1-s_i)\left(\frac{dV(b_1,\dots,b_n)}{db_i}\Big|_{b_1,\dots,b_n=0}\right) = \frac{dV(k_1b_1,\dots,k_nb_n)}{db_i}\Big|_{b_j=0,j\neq i}$$

for $i = 1,\dots,n$ separately to obtain $\mathbf{b}_{**} = [b_{1*}/k_1,\dots,b_{n*}/k_n]^{\mathsf{T}}$, so that when being substituted into (4.25) yields the exact same problem as if $\mathbf{b}_* = [b_{1*},\dots,b_{n*}]^{\mathsf{T}}$ is substituted into (4.21). Therefore, we can just assume each $k_i = 1$ without loss of generality.

4.4.1 Numerical Experiment Revisited

Figures 4.20 - 4.24 are analogs of those in Section 4.2.1 while Figures 4.25 - 4.29 are analogs of those in Section 4.3.1, with all else remaining equal except that we now only consider eigenvalue uncertainty. Figures C.19 - C.21 (analogs of Figure 4.20), Figures C.22 - C.24 (analogs of Figure 4.24), Figures C.25 - C.27 (analogs of Figure 4.25) and Figures C.28 - C.30 (analogs of Figure 4.29) in Appendix C correspond to the types II, III and IV problems respectively. Similar conclusions as the case where only location uncertainty is considered can be drawn.
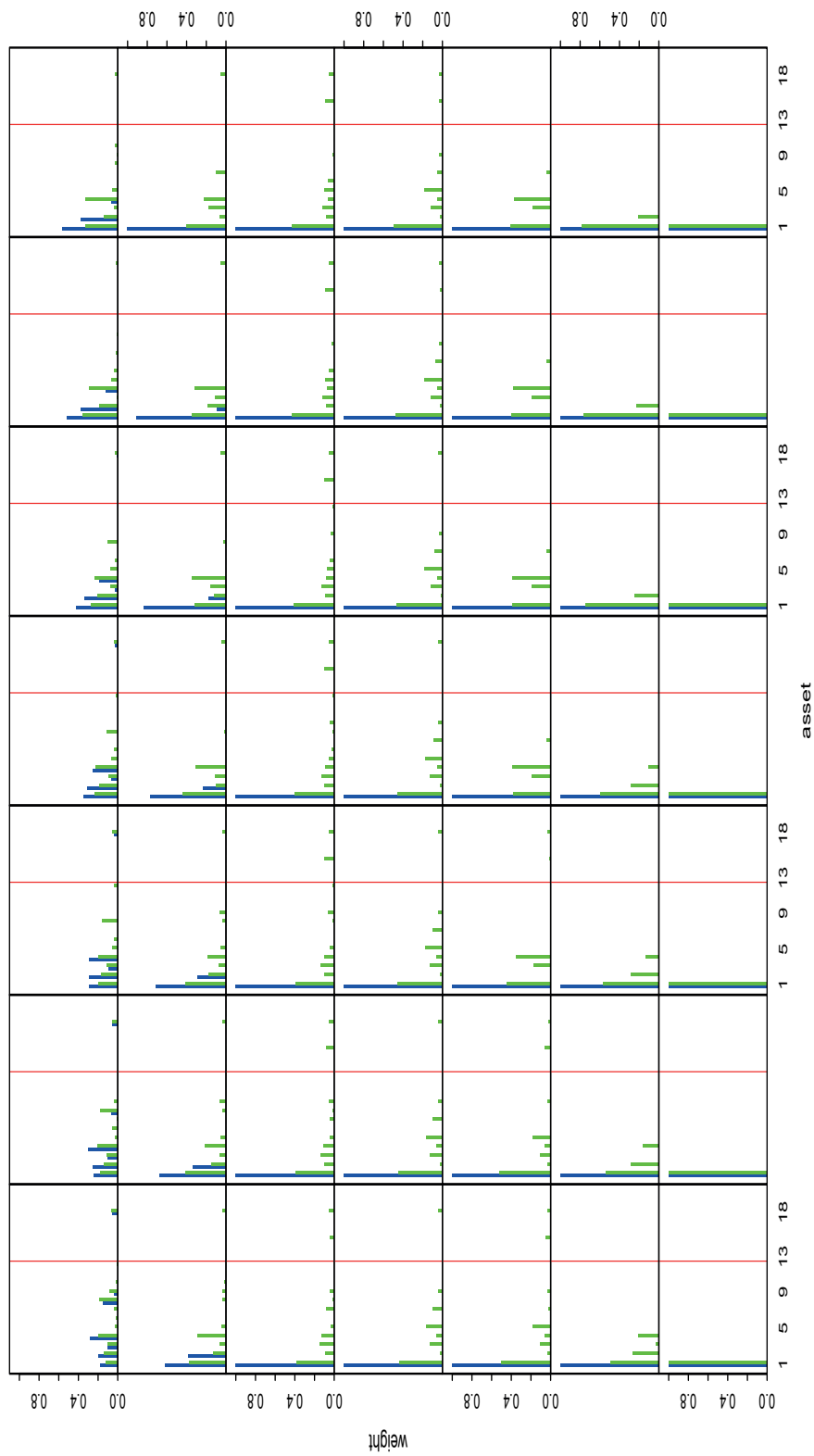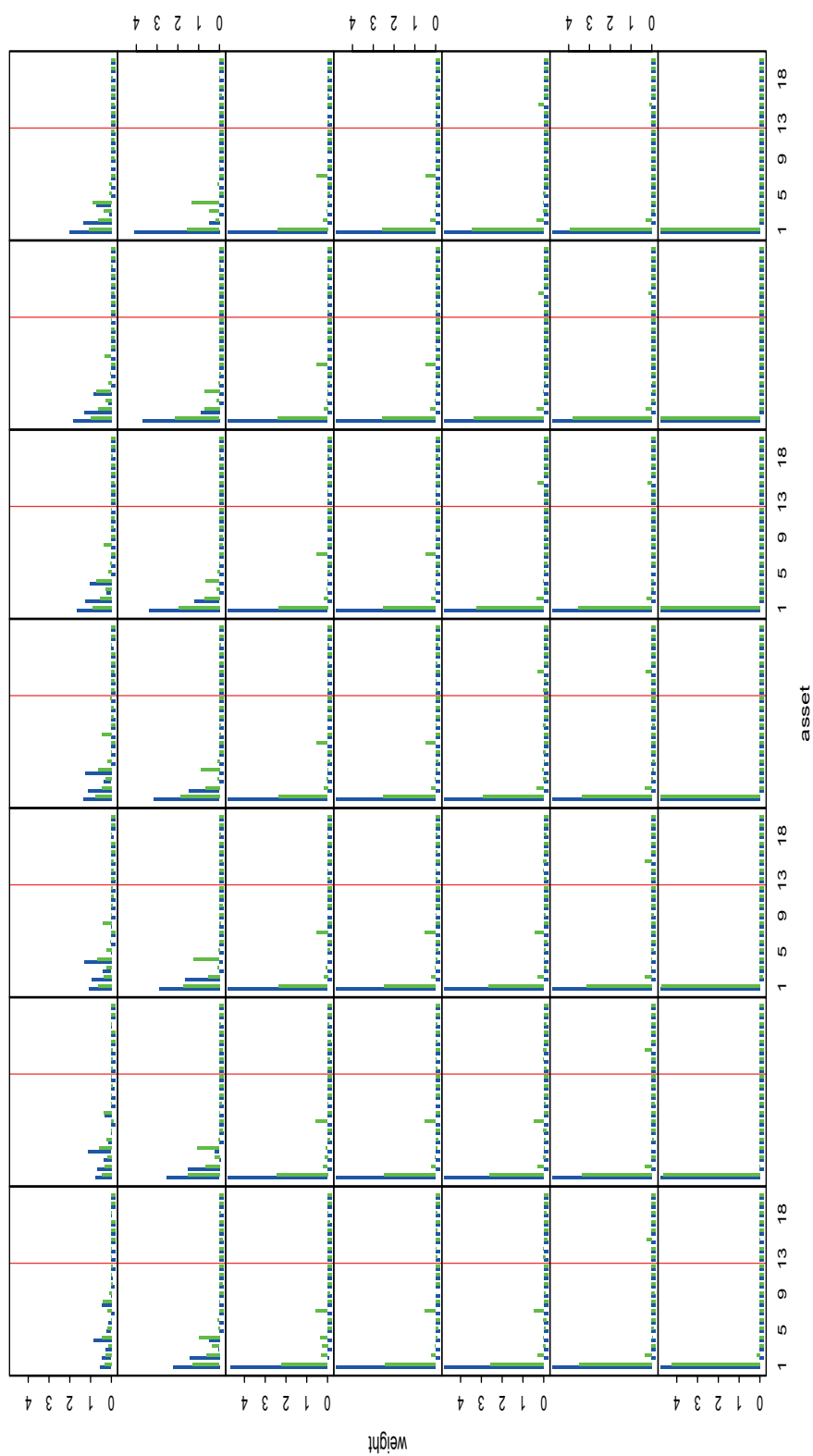
Figure 4.20: each subfigure plots the optimal weights of the type I problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
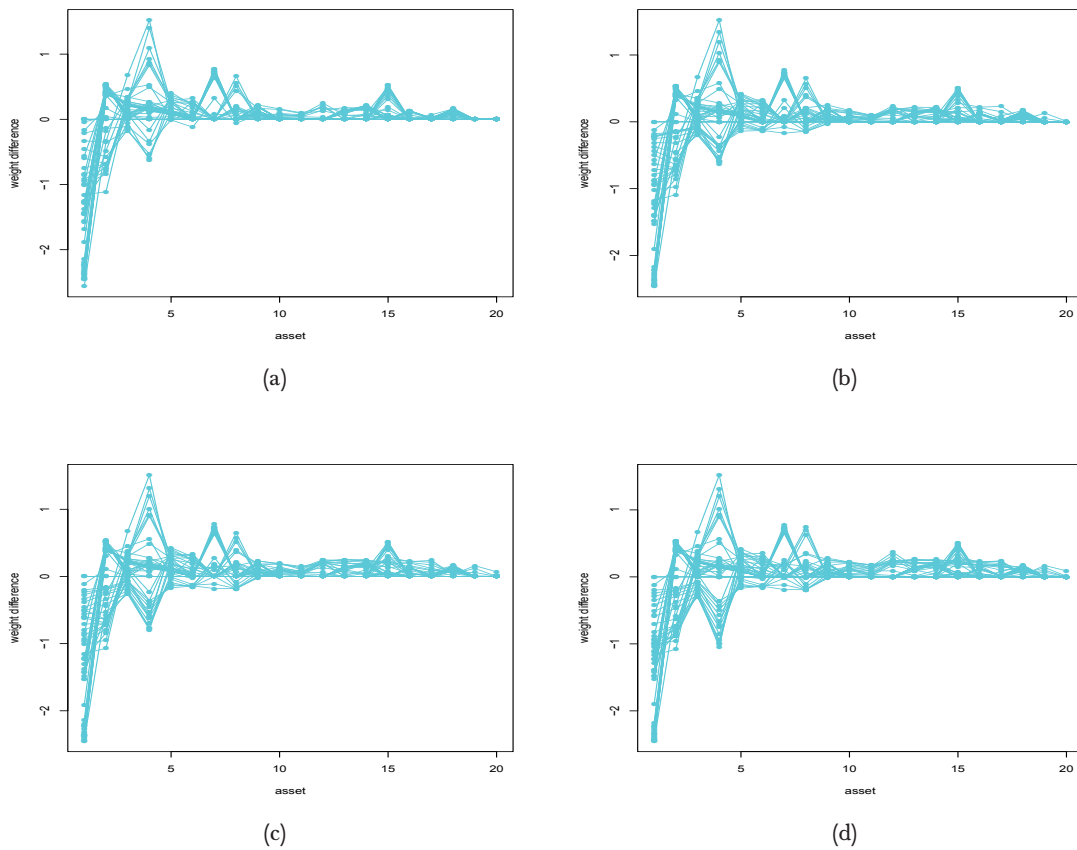
(a)

(b)

(c)

(d)
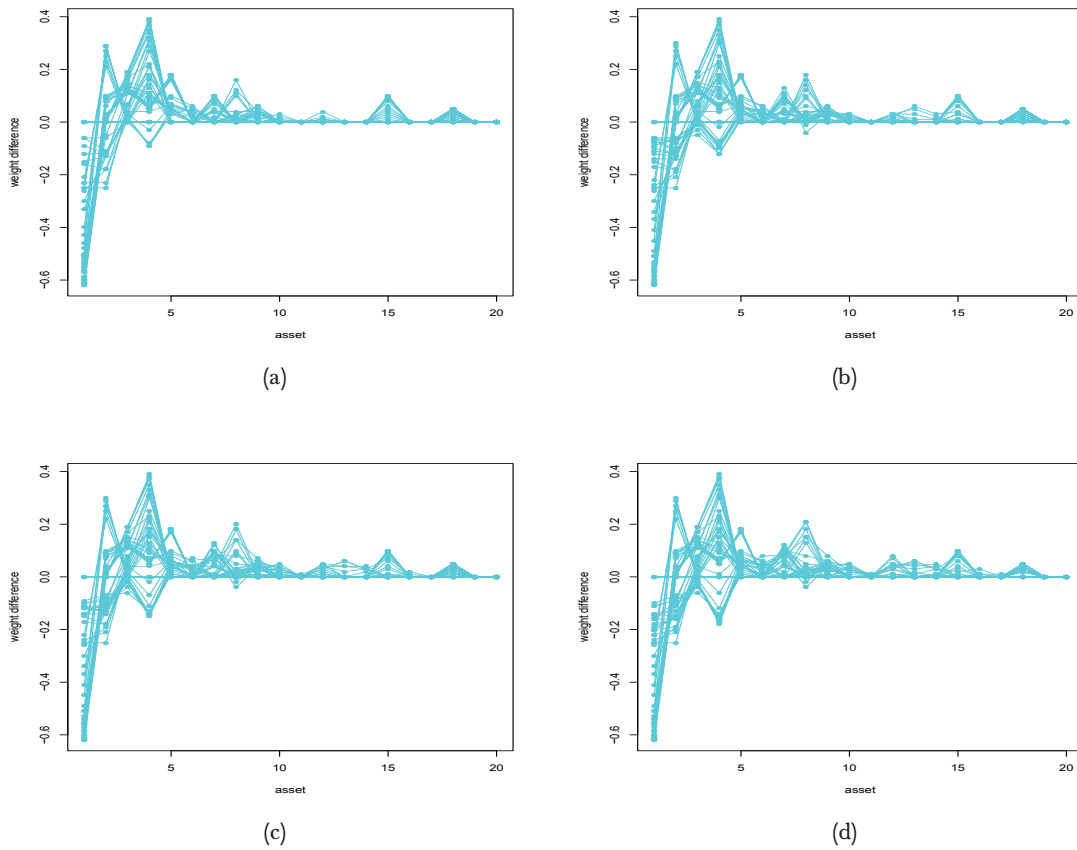
Figure 4.21: (a) plots the optimal weights of the type I problem with added eigenvalue uncertainty minus the optimal wieghts of the same problem without added uncertainty against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
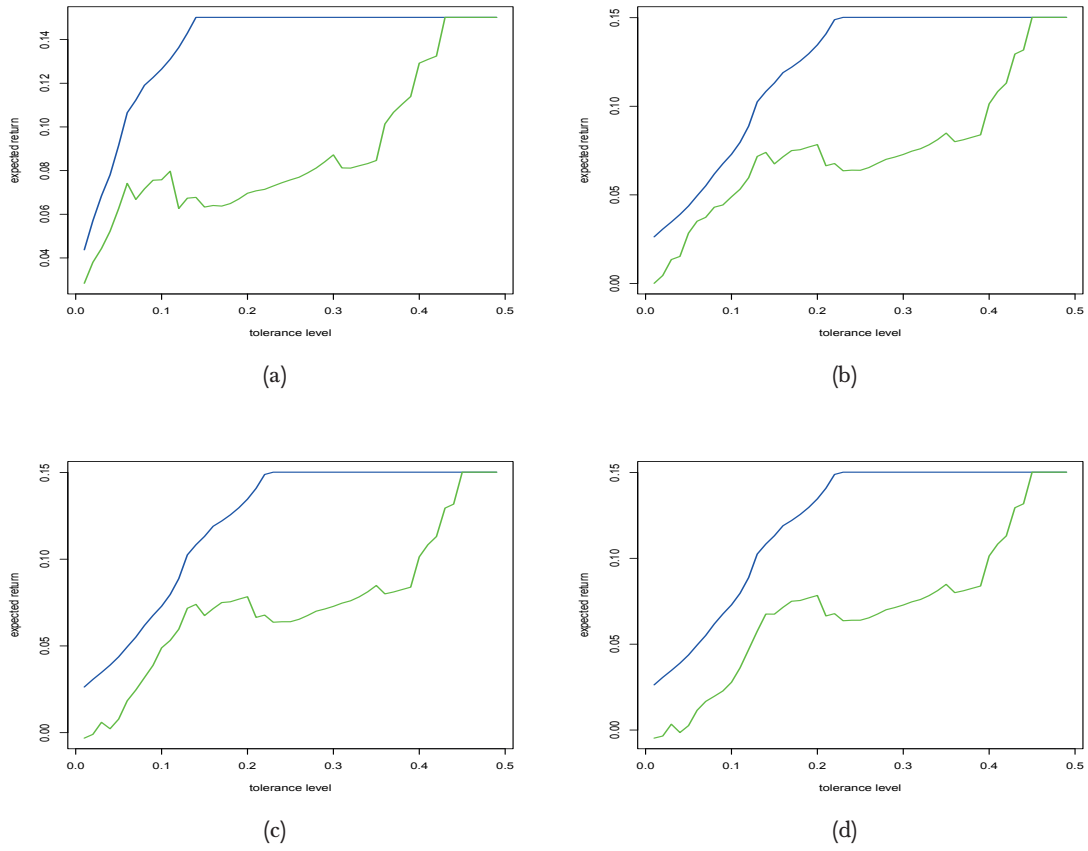
Figure 4.22: (a) plots the optimal portfolio expected return for the type I problem without uncertainty (blue) and the robust optimal portfolio expected return of the same problem with added eigenvalue uncertainty (green) against the tolerance level, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
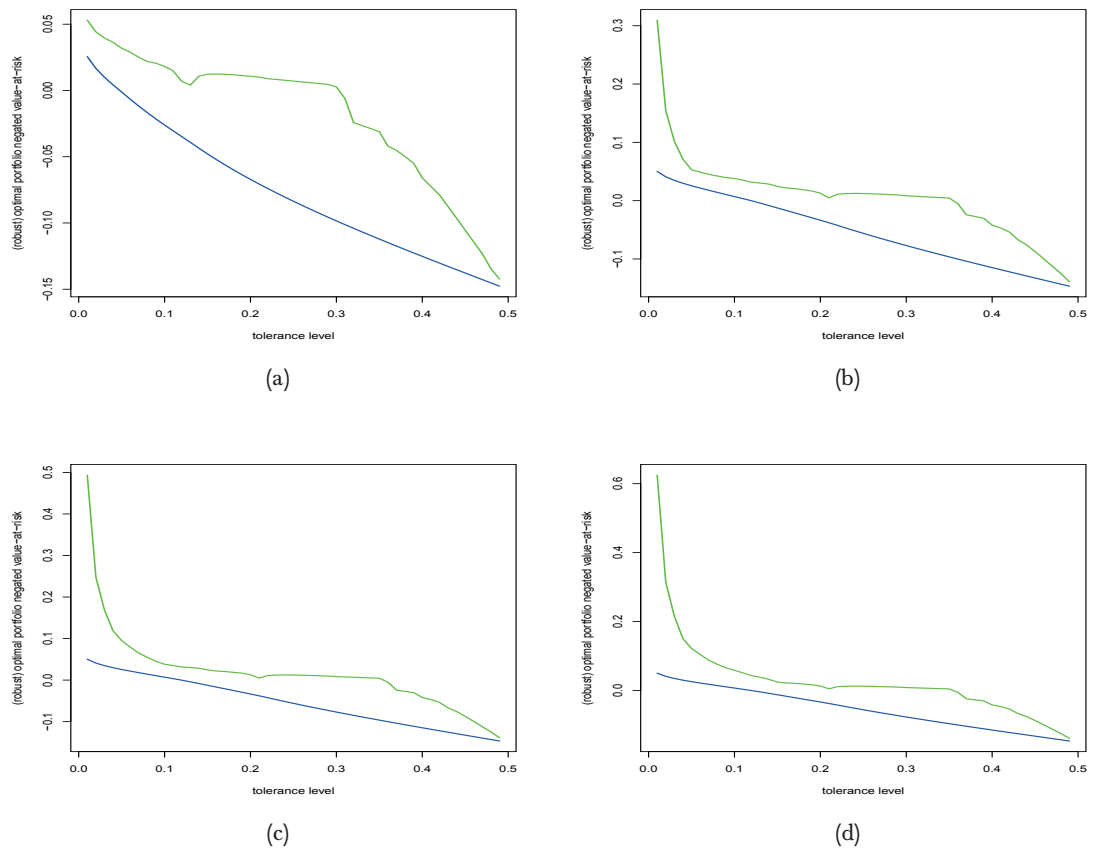
Figure 4.23: (a) plots the optimal portfolio negated value-at-risk for the type I problem without uncertainty (blue) and the robust optimal portfolio negated value-at-risk of the same problem with added eigenvalue uncertainty (green) against the tolerance level, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
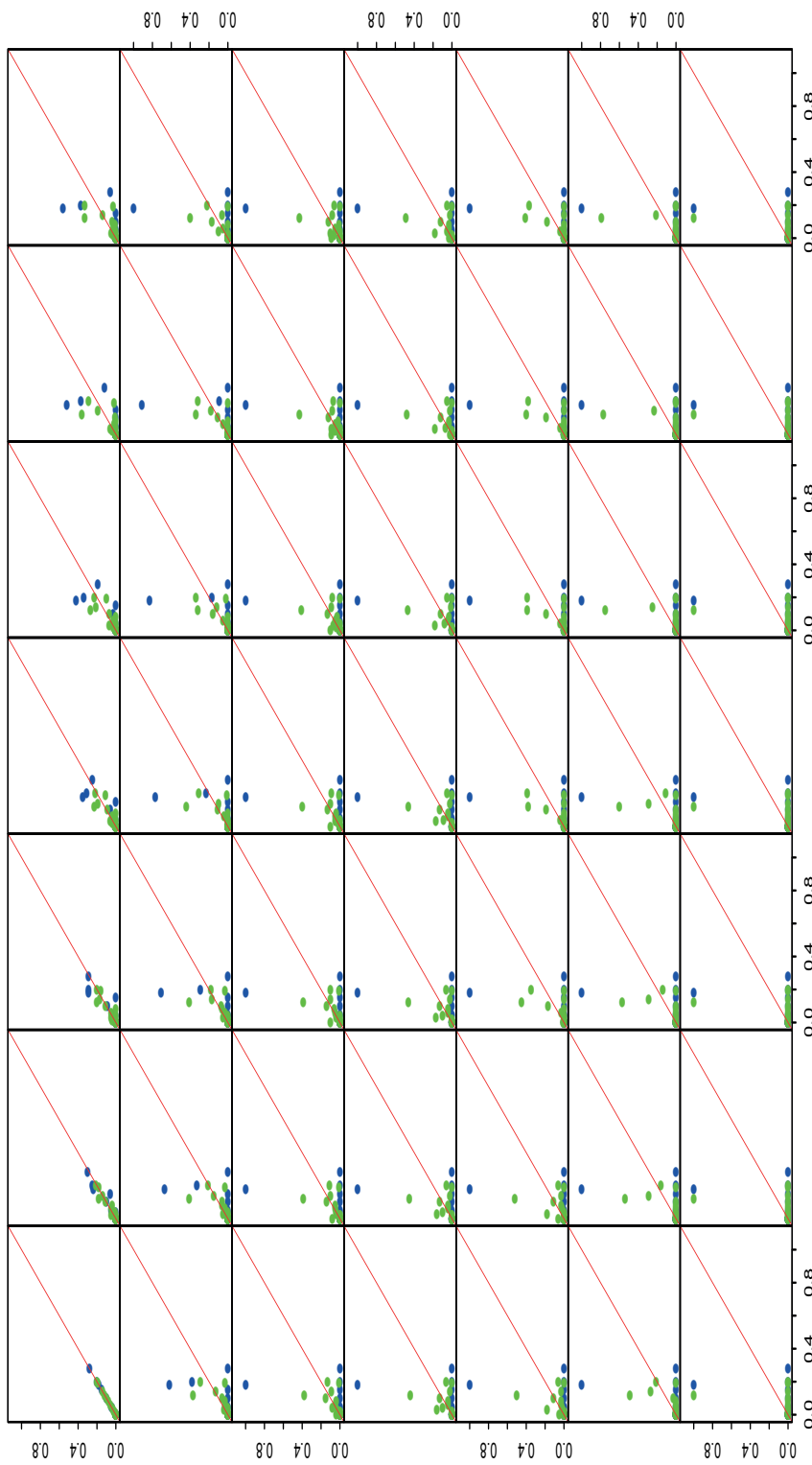
Figure 4.24: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type I problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure 4.25: each subfigure plots the optimal weights of the type I problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

(a)

(b)

(c)

(d)

Figure 4.26: (a) plots the optimal weights of the type I problem with added eigenvalue uncertainty minus the optimal wieghts of the same problem without added uncertainty against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
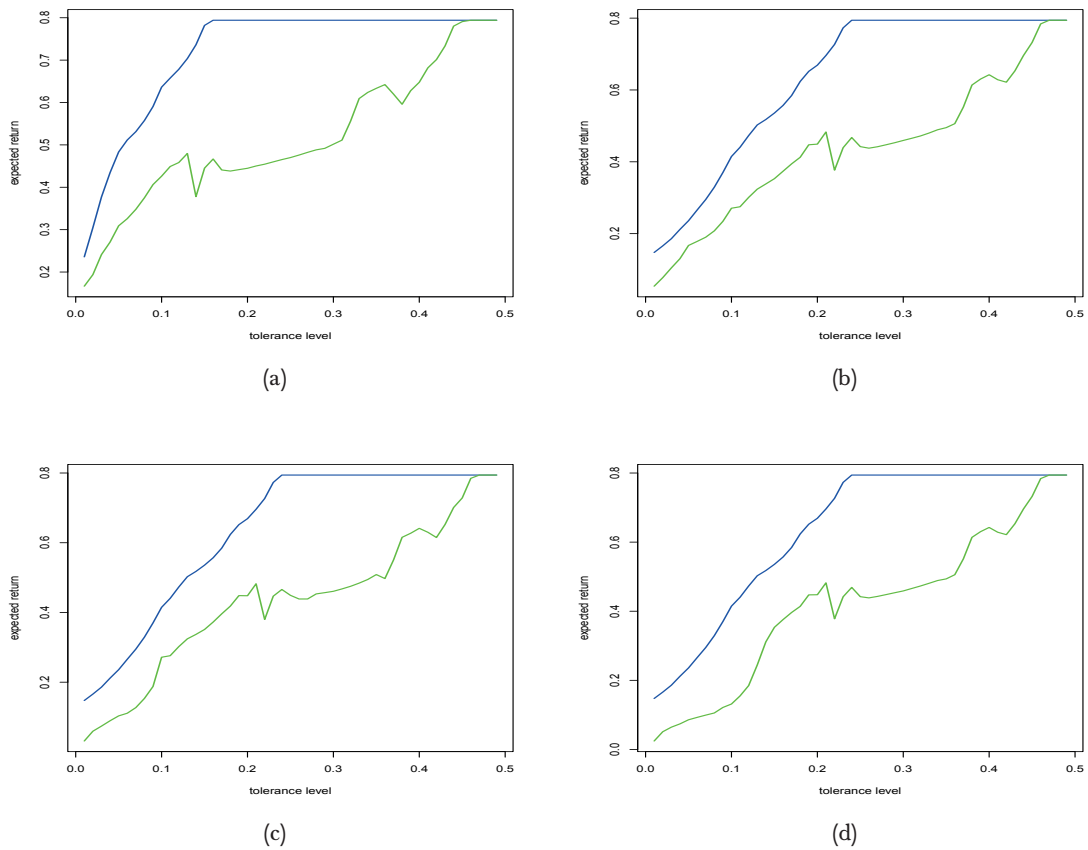
Figure 4.27: (a) plots the optimal portfolio expected return for the type I problem without uncertainty (blue) and the robust optimal portfolio expected return of the same problem with added eigenvalue uncertainty (green) against the tolerance level, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
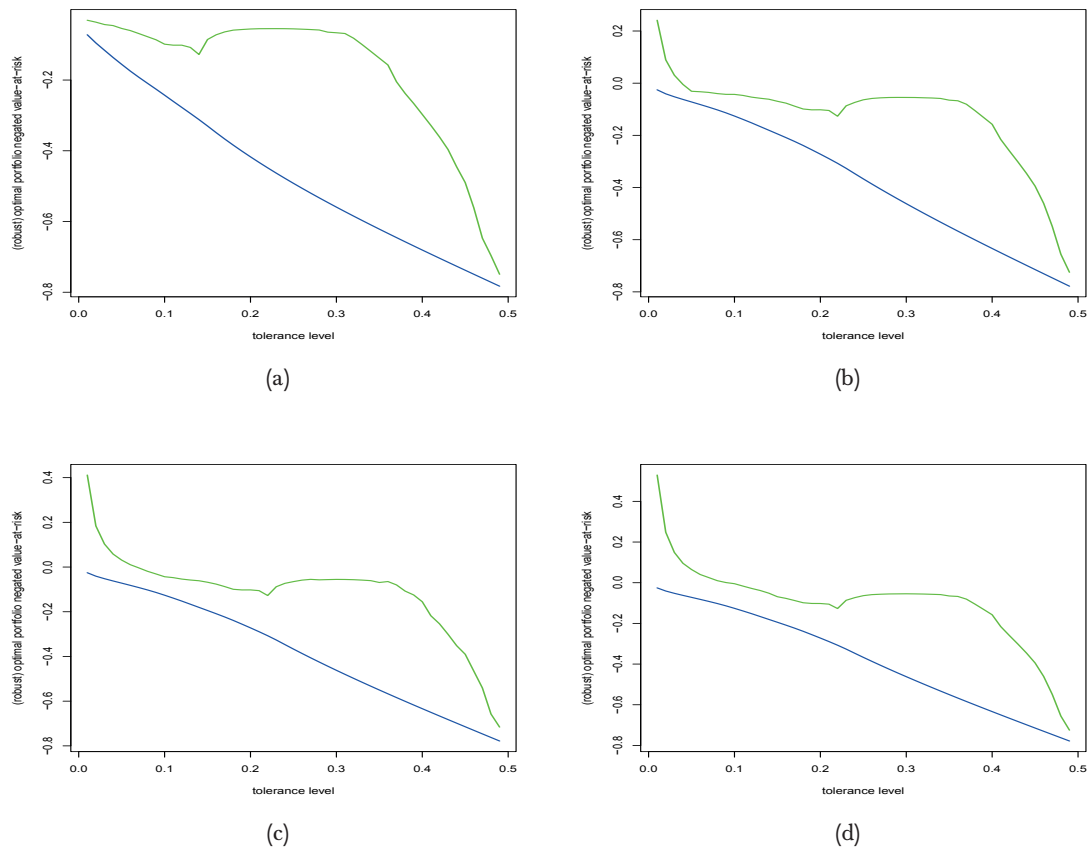
Figure 4.28: (a) plots the optimal portfolio negated value-at-risk for the type I problem without uncertainty (blue) and the robust optimal portfolio negated value-at-risk of the same problem with added eigenvalue uncertainty (green) against the tolerance level, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
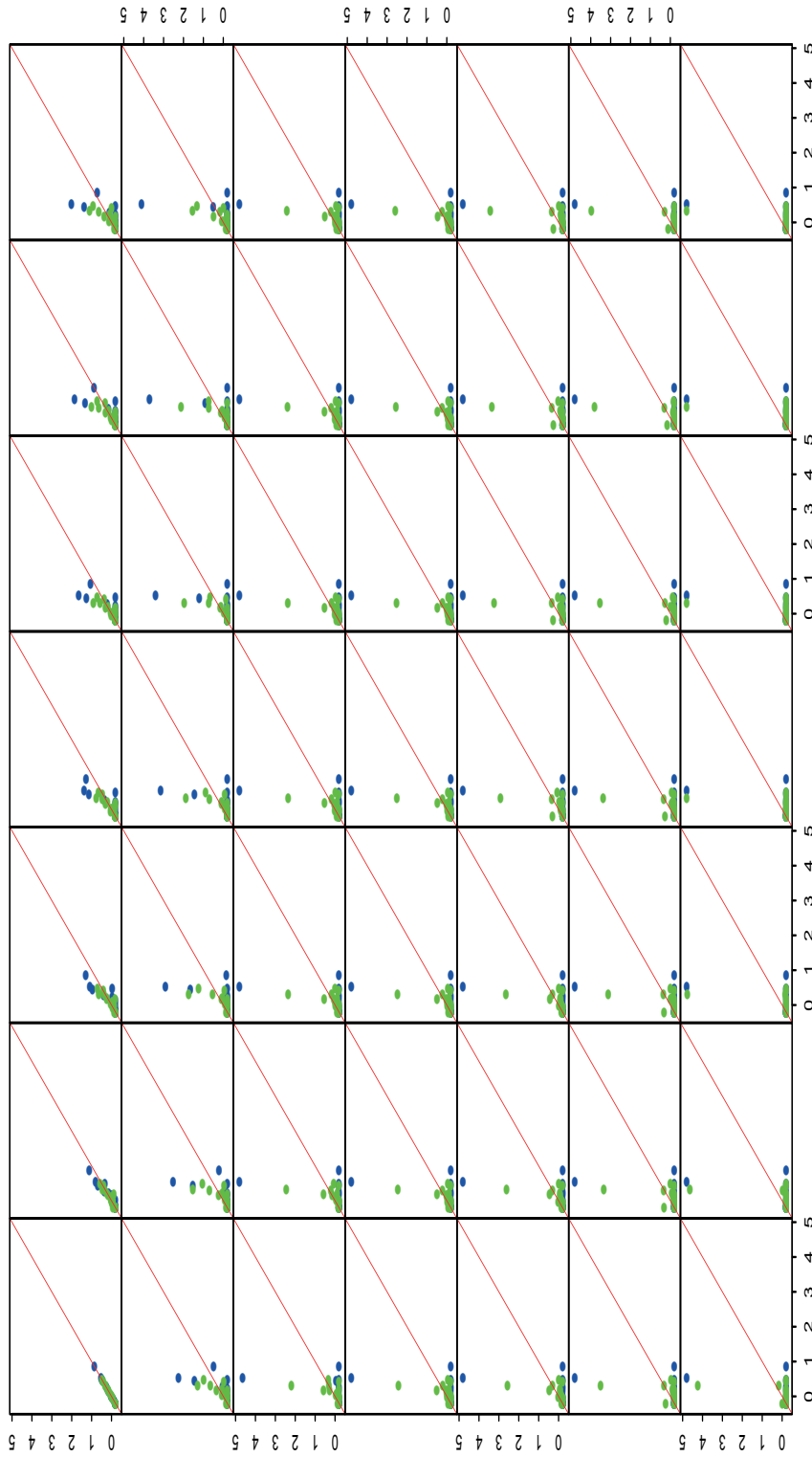
Figure 4.29: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type I problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

4.5 Eigenvector Uncertainty

If we only assume eigenvector uncertainty, then the optimal $\mathbf{w}$ of the robust location-scale problem (4.3) is approximately equal to that of

$$
\max_{(\mathbf{w},\tau,y)\in\mathcal{W}_\ell\times\mathbb{R}_+\times\mathbb{R}} \left\{ \mathbf{w}^\mathrm{T}\boldsymbol{\mu} - \kappa(\epsilon)y : \left[\begin{array}{c|c} (y+\tau(1-c)^2)\mathbf{I}_n - \tau\mathbf{e}_1\mathbf{e}_1^\mathrm{T} & \begin{array}{c} \mathbf{w}^\mathrm{T}\mathbf{P}_1^\mathrm{T} \\ \vdots \\ \mathbf{w}^\mathrm{T}\mathbf{P}_n^\mathrm{T} \end{array} \\ \hline \mathbf{P}_1\mathbf{w} \ \cdots \ \mathbf{P}_n\mathbf{w} & \mathrm{diag}(\lambda_1^{-1},\dots,\lambda_n^{-1})y \end{array} \right] \succeq 0, y\geq\delta \right\}
$$

$$(4.26)$$

by first using Theorem 3.5, then transposing the matrix in the semidefinite constraint (in which we lose nothing), and finally replacing the positivity constraint on y with $y\geq\delta$, where δ is a small positive number. Since the positive semi-definite space is a proper cone, a semi-definite constraint is also a generalized inequality. Therefore, by definition (A.20), the associated Lagrangian of (4.26) is

$$
L(\mathbf{w},\tau,y,\mathbf{Z},\nu.\upsilon) = \mathbf{w}^\mathrm{T}\boldsymbol{\mu} - \kappa(\epsilon)y
$$

$$
+ tr\left(\left[\begin{array}{c|c} (y+\tau(1-c)^2)\mathbf{I}_n - \tau\mathbf{e}_1\mathbf{e}_1^\mathrm{T} & \begin{array}{c} \mathbf{w}^\mathrm{T}\mathbf{P}_1^\mathrm{T} \\ \vdots \\ \mathbf{w}^\mathrm{T}\mathbf{P}_n^\mathrm{T} \end{array} \\ \hline \mathbf{P}_1\mathbf{w} \ \cdots \ \mathbf{P}_n\mathbf{w} & \mathrm{diag}(\lambda_1^{-1},\dots,\lambda_n^{-1})y \end{array} \right] \mathbf{Z} \right)
$$

$$
+ \nu(\mathbf{w}^\mathrm{T}\mathbb{1}-1) + \sum_{i=1}^n \upsilon_i(w_i-\ell_i) + \upsilon_{n+1}\tau + \upsilon_{n+2}(y-\delta)
$$

$$
= \sum_{i=1}^n w_i\left(\mu_i + \nu + \upsilon_i + tr\left(\left[\begin{array}{c|c} \mathbf{0}_{n\times n} & \begin{array}{c} \mathbf{p}_{1i}^\mathrm{T} \\ \vdots \\ \mathbf{p}_{ni}^\mathrm{T} \end{array} \\ \hline \mathbf{p}_{1i} \ \cdots \ \mathbf{p}_{ni} & \mathbf{0}_{n\times n} \end{array} \right] \mathbf{Z} \right) \right)
$$

$$
+ \tau\left(\upsilon_{n+1} + tr\left(\left[\begin{array}{c|c} (1-c)^2\mathbf{I}_n - \mathbf{e}_1\mathbf{e}_1^\mathrm{T} & \mathbf{0}_{n\times n} \\ \hline \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \end{array} \right] \mathbf{Z} \right) \right)
$$

$$
+ y\left(\upsilon_{n+2} - \kappa(\epsilon) + tr\left(\left[\begin{array}{c|c} \mathbf{I}_n & \mathbf{0}_{n\times n} \\ \hline \mathbf{0}_{n\times n} & \mathrm{diag}(\lambda_1,\dots,\lambda_n) \end{array} \right] \mathbf{Z} \right) \right)
$$

$$
- \nu - \sum_{i=1}^n \upsilon_i\ell_i - \delta\upsilon_{n+2},
$$

where $\mathbf{Z}\in\mathbb{S}^{2n}$ and $\mathbf{p}_{ji}$ represents the ith column of $\mathbf{P}_j$.

Denoting

$$
\mathbf{F}_i = \left[\begin{array}{c|c} \mathbf{0}_{n\times n} & \begin{array}{c} \mathbf{p}_{1i}^\mathrm{T} \\ \vdots \\ \mathbf{p}_{ni}^\mathrm{T} \end{array} \\ \hline \mathbf{p}_{1i} \ \cdots \ \mathbf{p}_{ni} & \mathbf{0}_{n\times n} \end{array} \right],
$$

$$\mathbf{G} = \left[\begin{array}{c|c} (1-c)^2\mathbf{I}_n - \mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} & \mathbf{0}_{n\times n} \\ \hline \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \end{array} \right],$$

$$\mathbf{H} = \left[\begin{array}{c|c} \mathbf{I}_n & \mathbf{0}_{n\times n} \\ \hline \mathbf{0}_{n\times n} & \mathrm{diag}(\lambda_1,\ldots,\lambda_n) \end{array} \right],$$

the Lagrange dual function is then

$$d(\mathbf{Z},\nu,\upsilon) = \sup_{(\mathbf{w},\tau,y)\in\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}} L(\mathbf{w},\tau,y,\mathbf{Z},\nu,\upsilon)$$

$$= \begin{cases} -\nu - \sum_{i=1}^n \upsilon_i\ell_i - \delta\upsilon_{n+2}, & \text{if } \begin{cases} \mu_i + \nu + \upsilon_i + tr(\mathbf{F}_i\mathbf{Z}) = 0, \ i = 1,\ldots,n, \\ \upsilon_{n+1} + tr(\mathbf{G}\mathbf{Z}) = 0, \\ \upsilon_{n+2} - \kappa(\epsilon) + tr(\mathbf{H}\mathbf{Z}) = 0, \end{cases} \\ \infty, & \text{otherwise,} \end{cases}$$

so that the Lagrange dual problem of (4.26) is

$$\min_{(\mathbf{Z},\nu,\upsilon)\in\mathbb{S}_+^{2n}\times\mathbb{R}\times\mathbb{R}_+^{n+2}} \left\{ -\nu - \sum_{i=1}^n \upsilon_i - \delta\upsilon_{n+2} : \begin{array}{c} \mu_i + \nu + \upsilon_i + tr(\mathbf{F}_i\mathbf{Z}) = 0, \ i = 1,\ldots,n \\ \upsilon_{n+1} + tr(\mathbf{G}\mathbf{Z}) = 0, \\ \upsilon_{n+2} - \kappa(\epsilon) + tr(\mathbf{H}\mathbf{Z}) = 0, \end{array} \right\} \qquad (4.27)$$

since the semi-positive definite cone is self-dual.

Theorem 4.9

Let $(\mathbf{w}_*,\tau_*,y_*)$ and $(\mathbf{Z}_*,\nu_*,\upsilon_*)$ be the optimal solutions of (4.26) and (4.27) respectively, and denote $\mathbf{z}_*$ the vectorized form of the upper half of $\mathbf{Z}_*$. Assume $[\mathbf{z}_*,\upsilon_{n+1*}]^{\mathrm{T}}$, $[\mathbf{z}_*,\upsilon_{n+2*}]^{\mathrm{T}}$ and each column of the matrix on the left-hand side of the semi-definite constraint in (4.26) evaluated at $(\mathbf{w}_*,\tau_*,y_*)$ are all non-zero vectors, and there exists a feasible solution of (4.26) such that the (matrix) inequality constraints hold strictly, then

(i) *denoting V as the optimal value of (4.26),*

$$\frac{dV}{dc} = -2(1-c)\sum_{i=1}^n Z_{ii*}\tau_*.$$

(ii) *the second derivative of V with respect to c is non-negative.*

Proof:

There is zero duality gap between (4.26) and (4.27) since by assumption, the Slater's condition is satisfied by for the former, whose equality and inequality constraints are linear, and the negative of

$$\left[\begin{array}{c|c} (y+\tau(1-c)^2)\mathbf{I}_n - \tau\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} & \begin{array}{c} \mathbf{w}^{\mathrm{T}}\mathbf{P}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{w}^{\mathrm{T}}\mathbf{P}_n^{\mathrm{T}} \end{array} \\ \hline \begin{array}{ccc} \mathbf{P}_1\mathbf{w} & \ldots & \mathbf{P}_n\mathbf{w} \end{array} & \mathrm{diag}(\lambda_1^{-1},\ldots,\lambda_n^{-1})y \end{array} \right]$$

is convex in $(\mathbf{w}, \tau, y)$ with respect to the semi-positive definite cone. Therefore, the KKT conditions in Theorem A.23 hold and in particular,

$$tr\left(\left[\begin{array}{ccc|c} & & & \mathbf{w}_*^{\mathsf{T}}\mathbf{P}_1^{\mathsf{T}} \\ (y_* + \tau_*(1-c)^2)\mathbf{I}_n - \tau_*\mathbf{e}_1\mathbf{e}_1^{\mathsf{T}} & & & \vdots \\ & & & \mathbf{w}_*^{\mathsf{T}}\mathbf{P}_n^{\mathsf{T}} \\ \hline \mathbf{P}_1\mathbf{w}_* & \dots & \mathbf{P}_n\mathbf{w}_* & \mathrm{diag}(\lambda_1^{-1},\dots,\lambda_n^{-1})y_* \end{array}\right]\mathbf{Z}_*\right) = 0, \tag{4.28}$$

$$v_{n+1*} + tr\left(\left[\begin{array}{c|c} (1-c)^2\mathbf{I}_n - \mathbf{e}_1\mathbf{e}_1^{\mathsf{T}} & \mathbf{0}_{n\times n} \\ \hline \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \end{array}\right]\mathbf{Z}_*\right) = 0, \tag{4.29}$$

$$\mathbf{w}_*^{\mathsf{T}}\mathbb{1} = 1, \tag{4.30}$$

$$v_{n+1*}\tau_* = 0, \tag{4.31}$$

$$v_{n+2*}(y_* - \delta) = 0. \tag{4.32}$$

Since

$$\left[\begin{array}{ccc|c} & & & \mathbf{w}_*^{\mathsf{T}}\mathbf{P}_1^{\mathsf{T}} \\ (y_* + \tau_*(1-c)^2)\mathbf{I}_n - \tau_*\mathbf{e}_1\mathbf{e}_1^{\mathsf{T}} & & & \vdots \\ & & & \mathbf{w}_*^{\mathsf{T}}\mathbf{P}_n^{\mathsf{T}} \\ \hline \mathbf{P}_1\mathbf{w}_* & \dots & \mathbf{P}_n\mathbf{w}_* & \mathrm{diag}(\lambda_1^{-1},\dots,\lambda_n^{-1})y_* \end{array}\right] \geq 0 \text{ and } \mathbf{Z}_* \geq 0,$$

we have

$$\left[\begin{array}{ccc|c} & & & \mathbf{w}_*^{\mathsf{T}}\mathbf{P}_1^{\mathsf{T}} \\ (y_* + \tau_*(1-c)^2)\mathbf{I}_n - \tau_*\mathbf{e}_1\mathbf{e}_1^{\mathsf{T}} & & & \vdots \\ & & & \mathbf{w}_*^{\mathsf{T}}\mathbf{P}_n^{\mathsf{T}} \\ \hline \mathbf{P}_1\mathbf{w}_* & \dots & \mathbf{P}_n\mathbf{w}_* & \mathrm{diag}(\lambda_1^{-1},\dots,\lambda_n^{-1})y_* \end{array}\right]\mathbf{Z}_* \tag{4.33}$$

is semi-positive definite[1] so that its diagonal entries are all non-negative. Together with (4.28), which in effect says that the sum of the diagonal entries of (4.33) is zero, the diagonal entries of (4.33) must all be zeros. This in turn forces all the off-diagonal entries of (4.33) to be zero, otherwise it is no longer semi-positive definite. In other words, (4.28) is equivalent to the matrix equality

$$\left[\begin{array}{ccc|c} & & & \mathbf{w}_*^{\mathsf{T}}\mathbf{P}_1^{\mathsf{T}} \\ (y_* + \tau_*(1-c)^2)\mathbf{I}_n - \tau_*\mathbf{e}_1\mathbf{e}_1^{\mathsf{T}} & & & \vdots \\ & & & \mathbf{w}_*^{\mathsf{T}}\mathbf{P}_n^{\mathsf{T}} \\ \hline \mathbf{P}_1\mathbf{w}_* & \dots & \mathbf{P}_n\mathbf{w}_* & \mathrm{diag}(\lambda_1^{-1},\dots,\lambda_n^{-1})y_* \end{array}\right]\mathbf{Z}_* = \mathbf{0}_{2n\times 2n}. \tag{4.34}$$

If $\mathbf{z}_* \neq \mathbf{0}$ and Z_{ij*} is any of its non-zero elements, then the partial derivative of the (i,j)th entry of the resulting matrix on the left-hand side of (4.34) with respect to τ_* obtains $Z_{ij*}((1-c)^2 - 1) > 0$. If $\mathbf{z}_* = \mathbf{0}$, then $v_{n+1*} > 0$ by assumption and the Jacobian of the left-hand side of (4.29) and (4.31) with respect to (v_{n+1*}, τ_*) is non-singular. In each case, the Implicit

---

[1]It is a well-known fact that if $\mathbf{A} \geq 0$ and $\mathbf{B} \geq 0$, then $\mathbf{AB} \geq 0$.

Function Theorem A.19 ensures that τ_* is continuously differentiable with respect to c. The same argument goes for y_* where (4.32) is used instead of (4.31). In similar fashion, the Jacobian of (4.29) and (4.30) with respect to (v_{n+1*}, w_{i*}) is also non-singular, so that by applying the Implicit Function Theorem w_{i*} is continuously differentiable with respect to c, for $i = 1,\dots,n$. We can now use Theorem A.25 to obtain the result of part (i). The partial derivative of the (i_0, j)th entry of the matrix on the left-hand side of (4.34) with respect to Z_{ij*} where (i_0, j) is the index of any non-zero element in the jth column of the matrix on the left-hand side of the semi-definite constraint in (4.26) evaluated at $(\mathbf{w}_*, \tau_*, y_*)$ yields the element itself, which is non-zero, for $i = 1,\dots,n$. This means that each entry of $\mathbf{Z}_*$ is continuously differentiable with respect to c, so that the result of part (ii) follows immediately by applying Theorem A.26. □

Assuming that the conditions of the above theorem are satisfied, the value of c is set to the root c_* of

$$\min_{0 \leq c \leq 1} \frac{dV}{dc} + s \left(\max_{0 \leq c \leq 1} \frac{dV}{dc} - \min_{0 \leq c \leq 1} \frac{dV}{dc} \right) = \frac{dV}{dc}$$

$$\Leftrightarrow \quad (1-s) \left. \frac{dV}{dc} \right|_{c=0} + s \left. \frac{dV}{dc} \right|_{c=1} = \frac{dV}{dc}$$

$$\Leftrightarrow \quad (1-s) \left. \frac{dV}{dc} \right|_{c=0} = \frac{dV}{dc}$$

where $s \in (0,1)$ is the chosen sensitivity level and the first equivalence is due to part (iii) of Theorem 4.9.

We shall show that scaling is not an issue as before. Assume that c is scaled by $k > 0$ in (4.26) so that it becomes

$$\max_{(\mathbf{w}, \tau, y) \in \mathcal{W}_\ell \times \mathbb{R}_+ \times \mathbb{R}_{++}} \left\{ \mathbf{w}^\mathsf{T} \boldsymbol{\mu} - \kappa(\epsilon)y : \left[\begin{array}{ccc|c} & & & \mathbf{w}^\mathsf{T} \mathbf{P}_1^\mathsf{T} \\ \multicolumn{3}{c|}{(y + \tau(1-kc)^2)\mathbf{I}_n - \tau \mathbf{e}_1 \mathbf{e}_1^\mathsf{T}} & \vdots \\ & & & \mathbf{w}^\mathsf{T} \mathbf{P}_n^\mathsf{T} \\ \hline \mathbf{P}_1 \mathbf{w} & \cdots & \mathbf{P}_n \mathbf{w} & \mathrm{diag}(\lambda_1^{-1}, \dots, \lambda_n^{-1})y \end{array} \right] \succeq 0 \right\}$$

(4.35)

with the optimal solution and value denoted as $(\mathbf{w}_{**}, \tau_{**}, y_{**})$ and $\tilde{V}$ respectively. The corresponding Lagrange dual problem of (4.35) is

$$\min_{(\mathbf{Z}, v, \mathbf{v}) \in \mathbb{S}_+^{2n} \times \mathbb{R} \times \mathbb{R}_+^2} \left\{ -v : \begin{array}{l} \mu_i + v + tr(\mathbf{F}_i \mathbf{Z}) = 0, \ i = 1,\dots,n \\ v_1 + tr(\mathbf{G}_k \mathbf{Z}) = 0, \\ v_2 - \kappa(\epsilon) + tr(\mathbf{H}\mathbf{Z}) = 0 \end{array} \right\}$$

(4.36)

where

$$\mathbf{G}_k = \left[\begin{array}{c|c} (1-kc)^2 \mathbf{I}_n - \mathbf{e}_1 \mathbf{e}_1^\mathsf{T} & \mathbf{0}_{n \times n} \\ \hline \mathbf{0}_{n \times n} & \mathbf{0}_{n \times n} \end{array} \right],$$

with optimal Lagrange multiplier matrix denoted as $\mathbf{Z}_{**}$. The derivative of $\tilde{V}$ with respect to c is

$$\frac{d\tilde{V}}{dc} = -2k(1-kc) \sum_{i=1}^{n} Z_{ii**} \tau_{**}.$$

and the second derivative of $\tilde{V}$ with respect to c is non-negative. The value c is now set to the root c_{**} of

$$\min_{0 \leq c \leq 1/k} \frac{d\tilde{V}(c)}{dc} + s\left(\max_{0 \leq c \leq 1/k} \frac{d\tilde{V}(c)}{dc} - \min_{0 \leq c \leq 1/k} \frac{d\tilde{V}(c)}{dc}\right) = \frac{d\tilde{V}(c)}{dc}$$

$$\Leftrightarrow \quad (1-s) \left.\frac{d\tilde{V}(c)}{dc}\right|_{c=0} + s \left.\frac{d\tilde{V}(c)}{dc}\right|_{c=1/k} = \frac{d\tilde{V}(c)}{dc}$$

$$\Leftrightarrow \quad (1-s) \left.\frac{d\tilde{V}(c)}{dc}\right|_{c=0} = \frac{d\tilde{V}(c)}{dc}$$

$$\Leftrightarrow \quad (1-s) \left.\frac{dV(c)}{dc}\right|_{c=0} = \frac{dV(kc)}{dc}.$$

Note that $c_{**} = c_*/k$, and when it is substituted into (4.35) obtains the same optimal solution as if c_* is substituted into (4.26). Thus, we may assume $k = 1$ without loss of generality.

4.5.1 Numerical Experiment Revisited

The targeted sensitivity level is always the median and we always assume that the regularity conditions of Theorem 4.9 hold. Figures 4.30 - 4.39 are analogs of those in Section 4.4.1 but with added eigenvector instead of eigenvalue uncertainty. Figures C.31 - C.33 (analogs of Figure 4.30), Figures C.34 - C.36 (analogs of Figure 4.34), Figures C.37 - C.39 (analogs of Figure 4.35) and Figures C.40 - C.42 (analogs of Figure 4.39) in Appendix C correspond to the types II, III and IV problems respectively. Similar conclusions as in the case where only location uncertainty is considered can be drawn. Numerical experiments for all the other different combinations of location, eigenvalue and eigenvector uncertainties are also done. Generally, for each combination conclusions are not much different from the previous cases, but the more types of uncertainty (location, eigenvalue and eigenvector) considered, the higher the robust optimal portfolio negated value-at-risk, so that it is less likely to invest all wealth into the risky assets. Each SDP is solved with SDPT3 by Toh (1999) using the MATLAB interface YALMIP by Löfberg (2004).
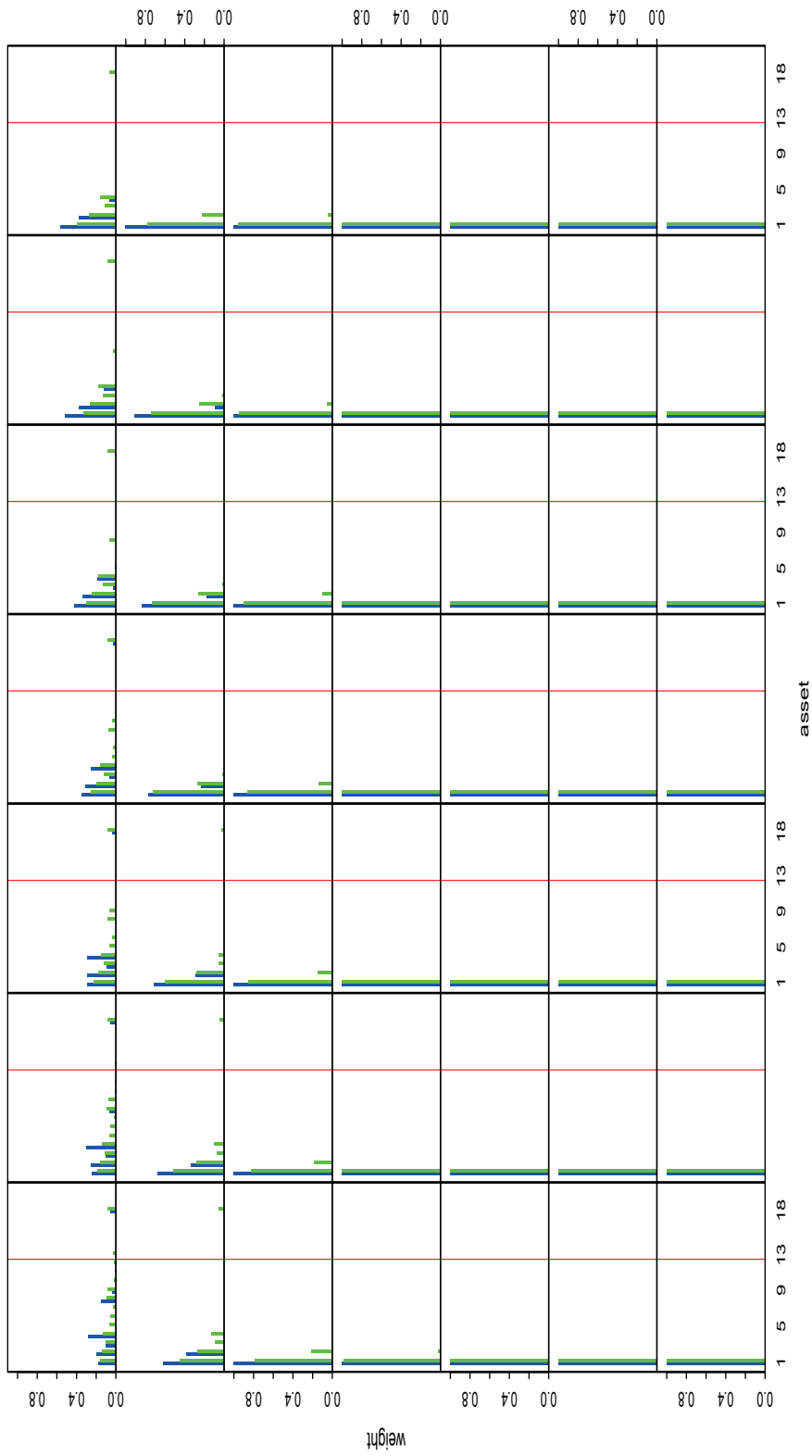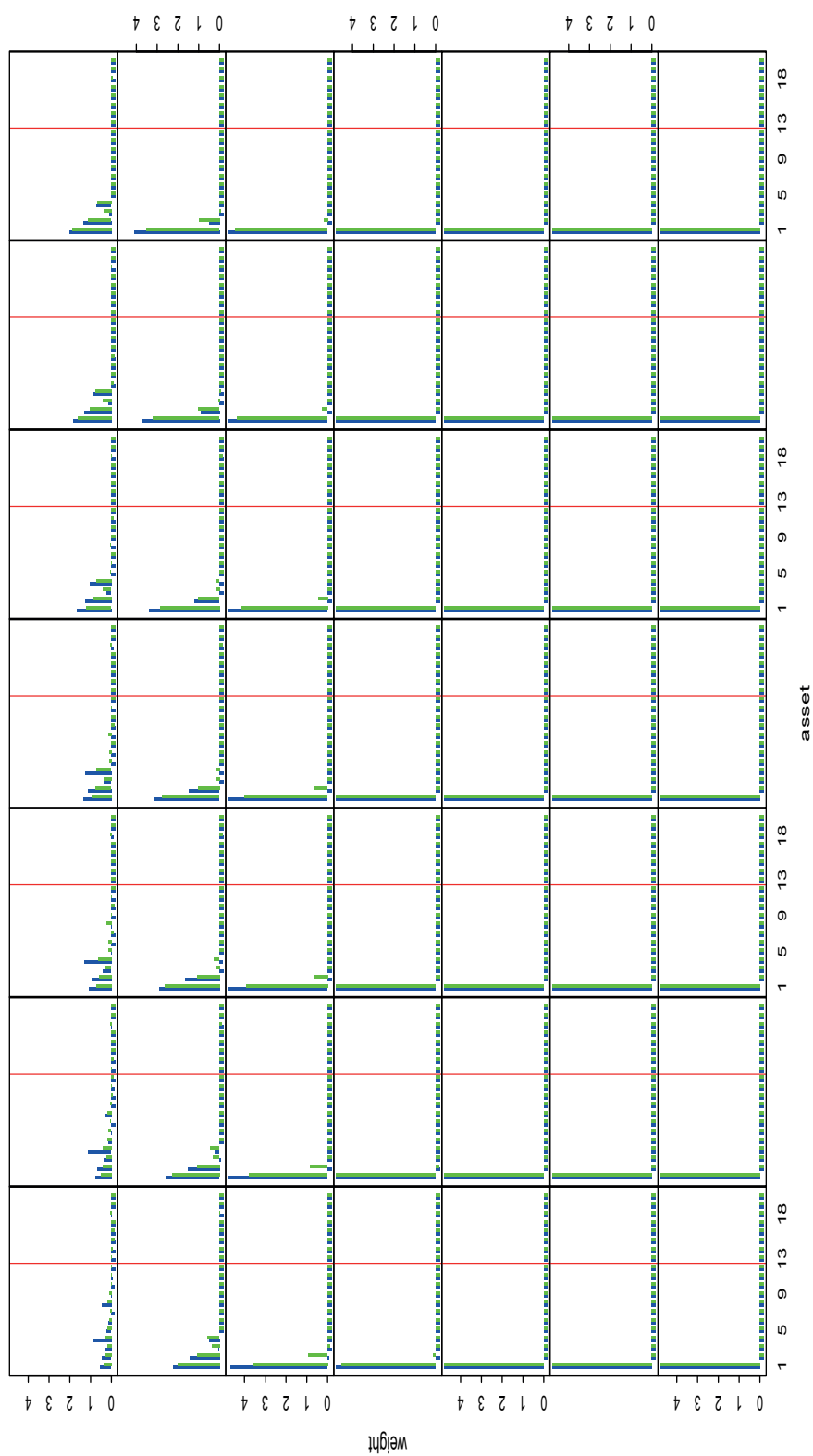
Figure 4.30: each subfigure plots the optimal weights of the type I problem both with added eigenvector uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
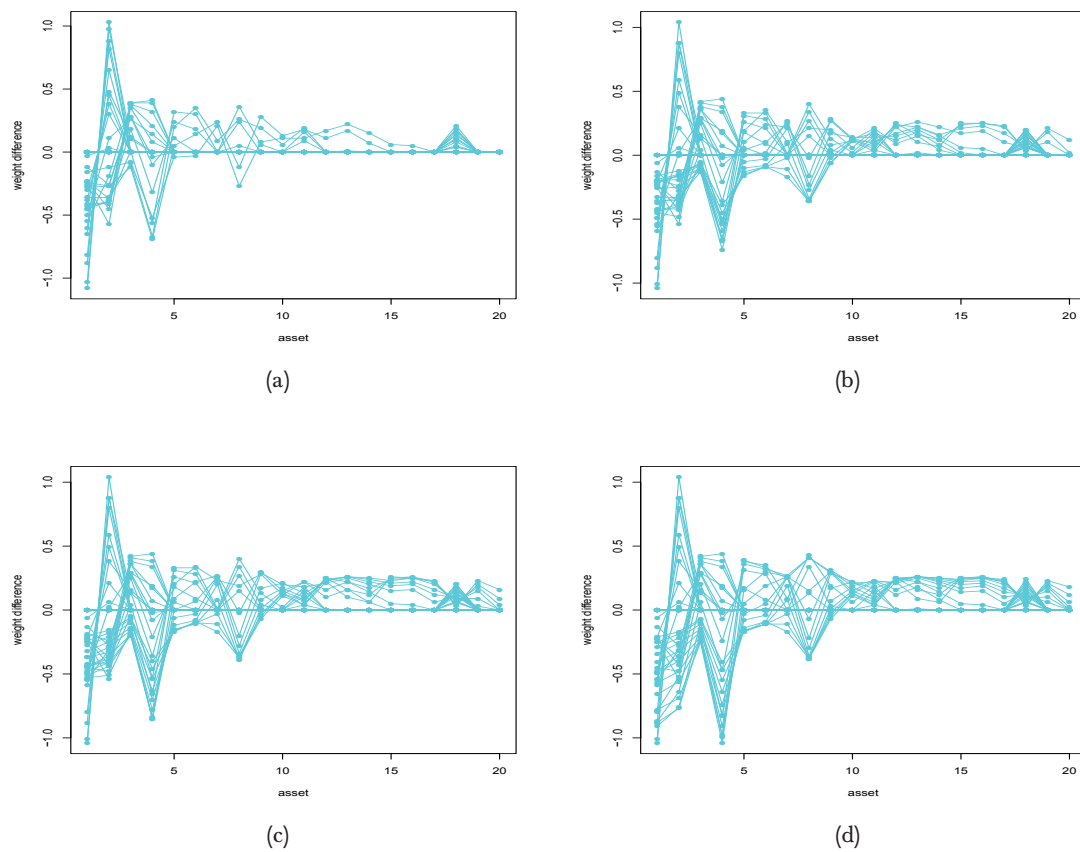
(a)

(b)

(c)

(d)
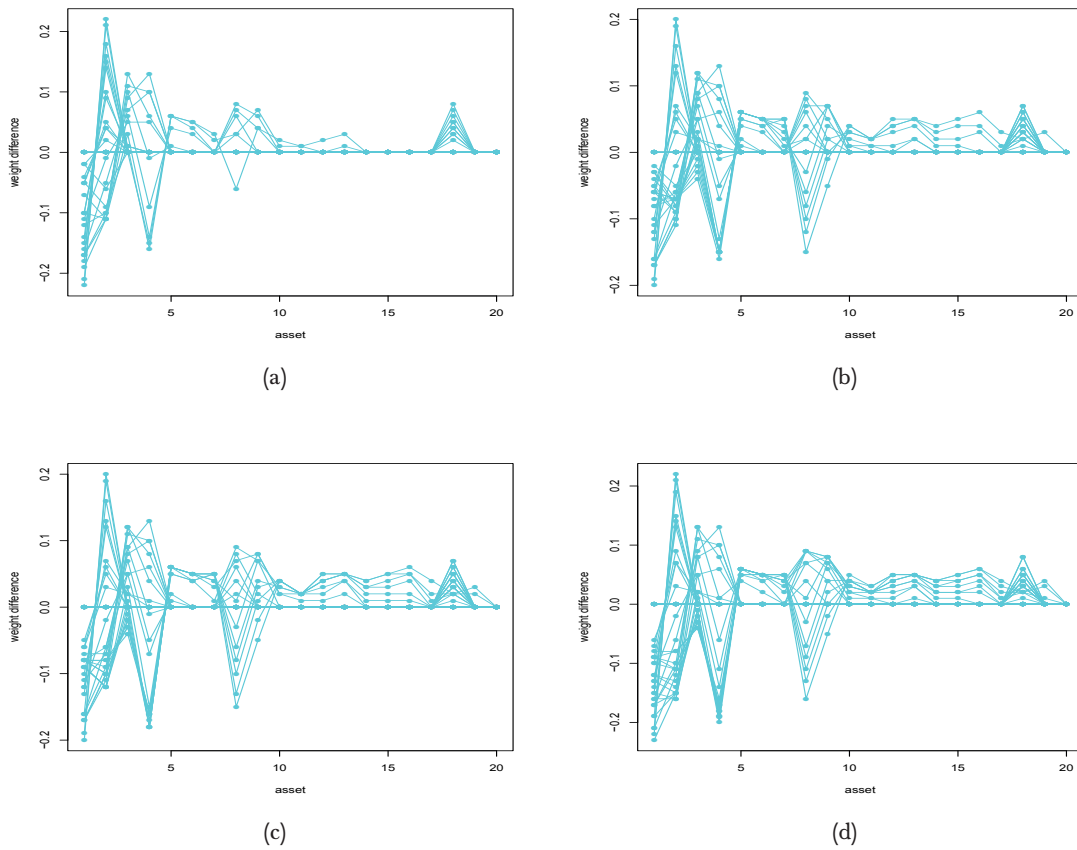
Figure 4.31: (a) plots the optimal weights of the type I problem with added eigenvector uncertainty minus the optimal wieghts of the same problem without added uncertainty against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
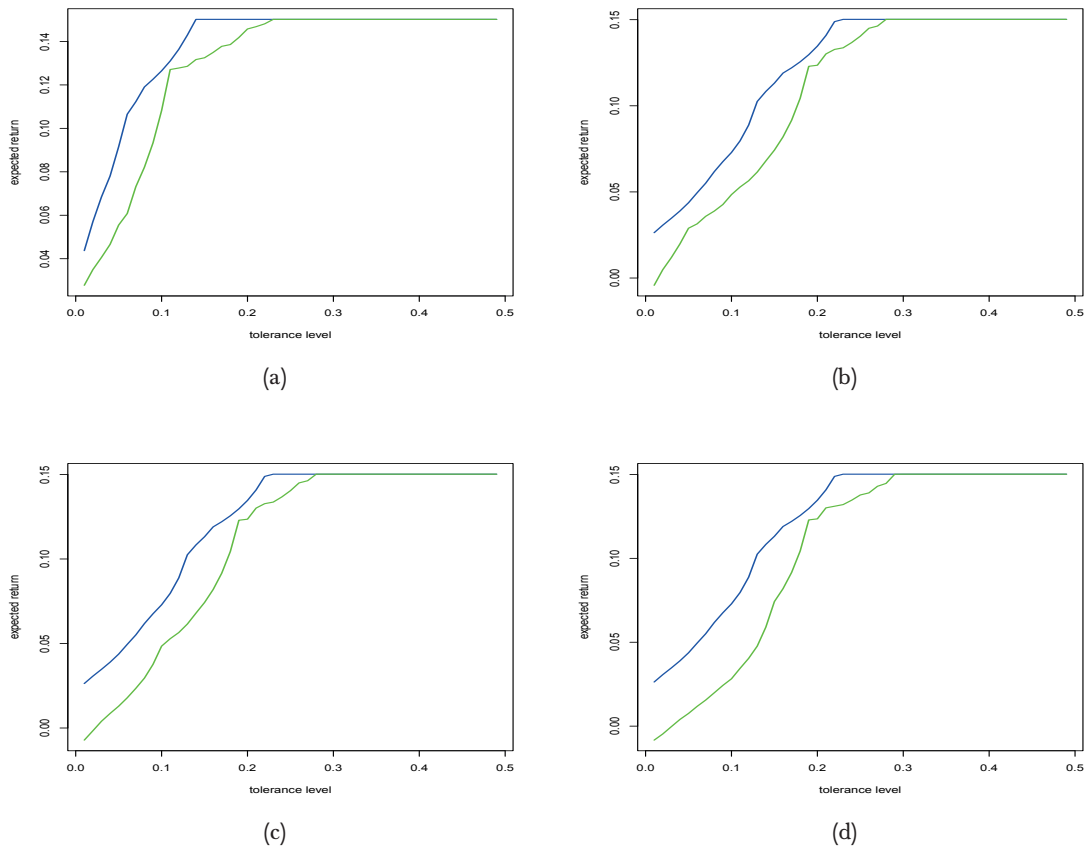
Figure 4.32: (a) plots the optimal portfolio expected return for the type I problem without uncertainty (blue) and the robust optimal portfolio expected return of the same problem with added eigenvector uncertainty (green) against the tolerance level, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
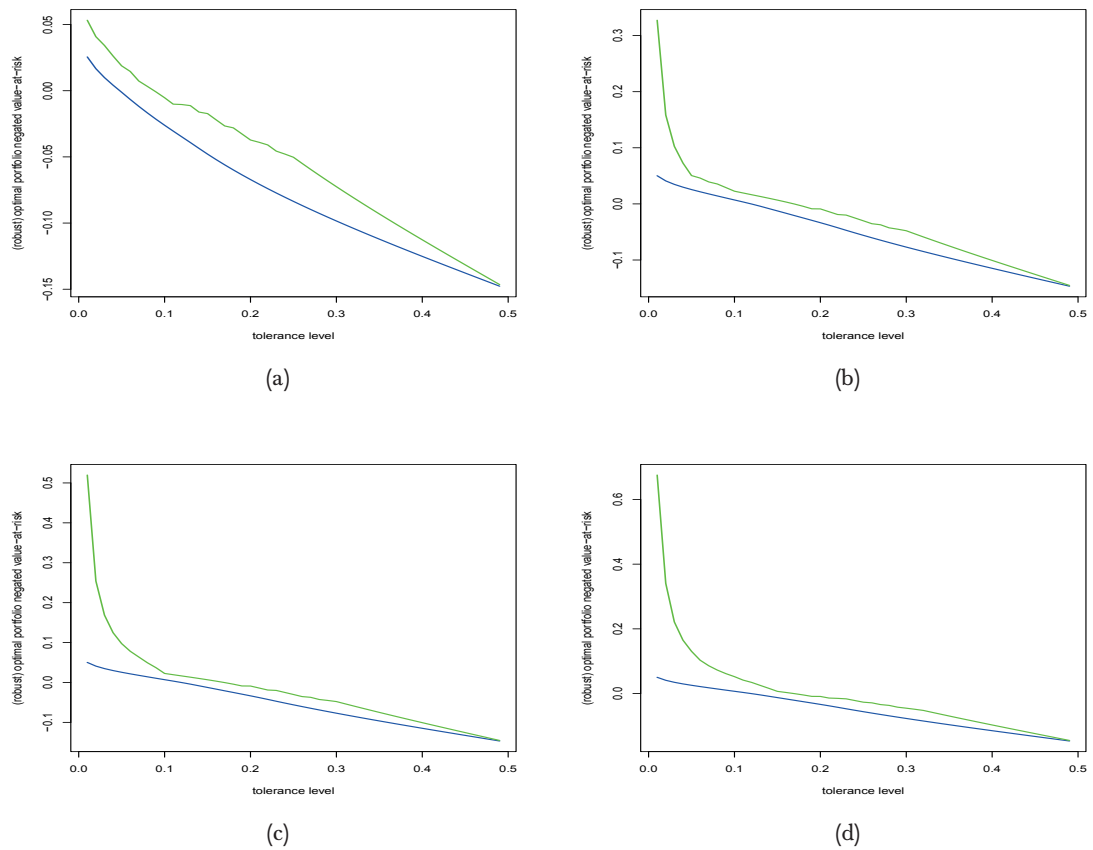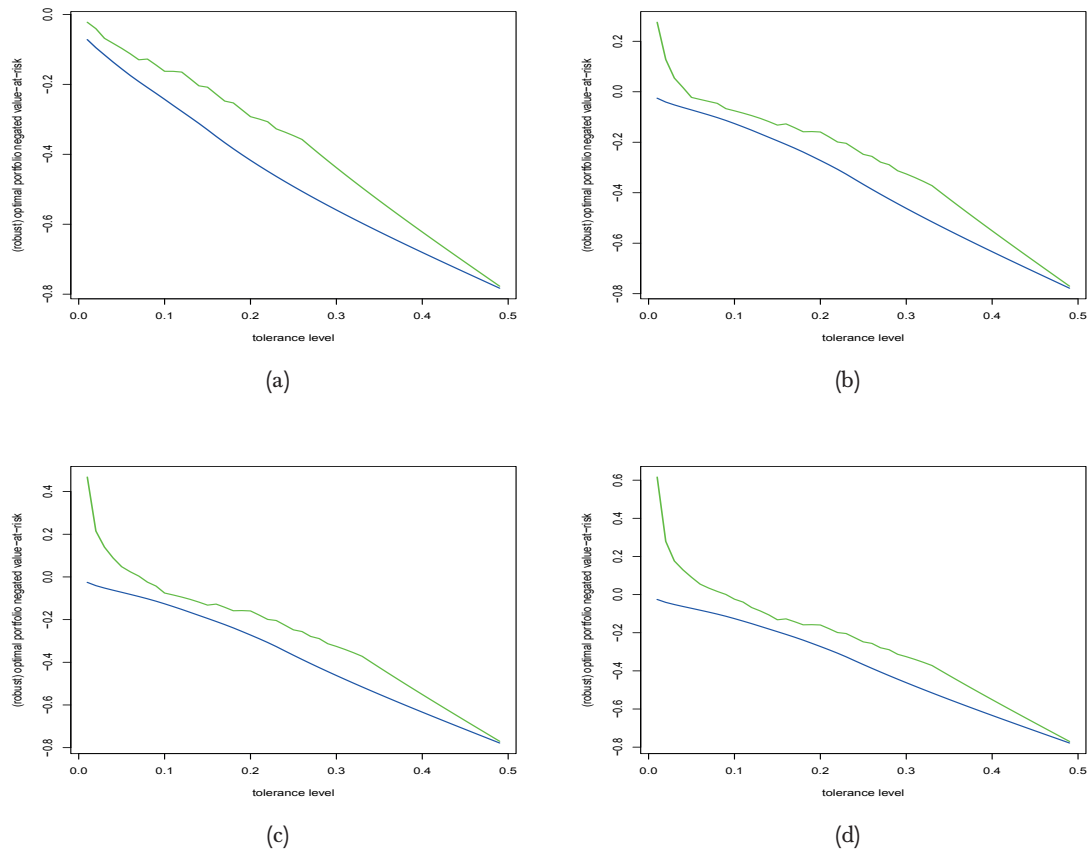
Figure 4.33: (a) plots the optimal portfolio negated value-at-risk for the type I problem without uncertainty (blue) and the robust optimal portfolio negated value-at-risk of the same problem with added eigenvector uncertainty (green) against the tolerance level, where short-selling is disallowed; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
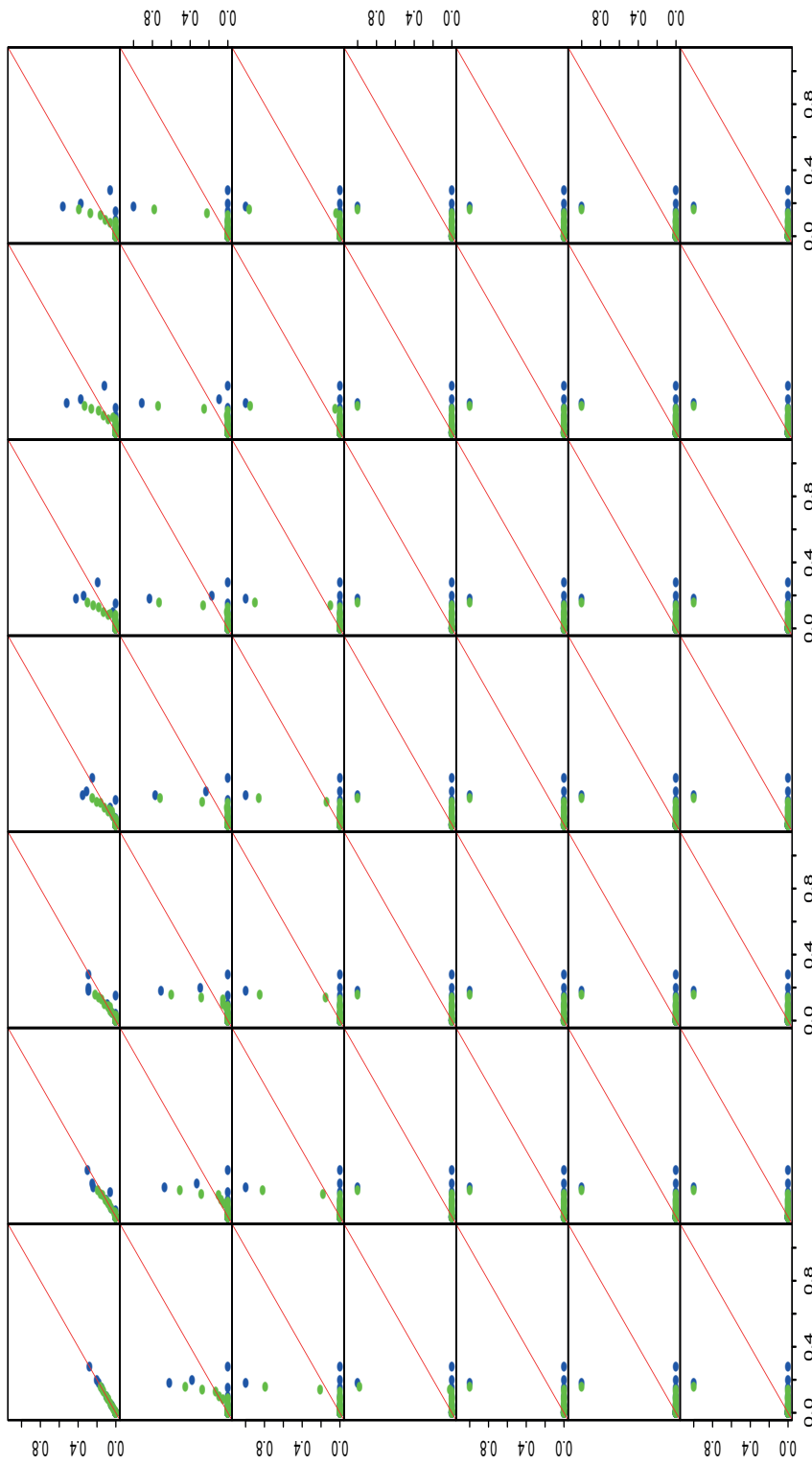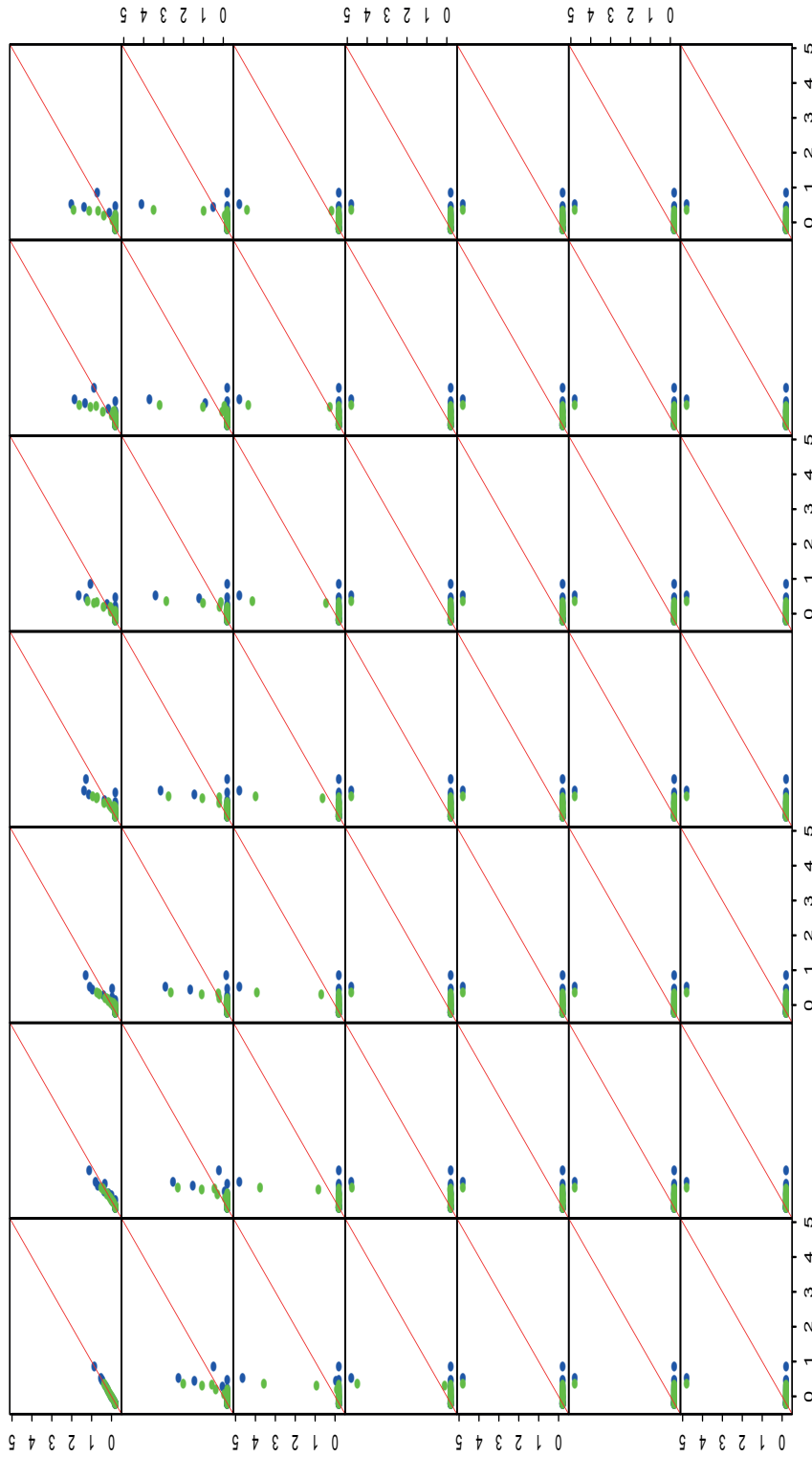
Figure 4.34: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type I problem both with added eigenvector uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure 4.35: each subfigure plots the optimal weights of the type I problem both with added eigenvector uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

(a)

(b)

(c)

(d)

Figure 4.36: (a) plots the optimal weights of the type I problem with added eigenvector uncertainty minus the optimal wieghts of the same problem without added uncertainty against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.

(a)

(b)

(c)

(d)
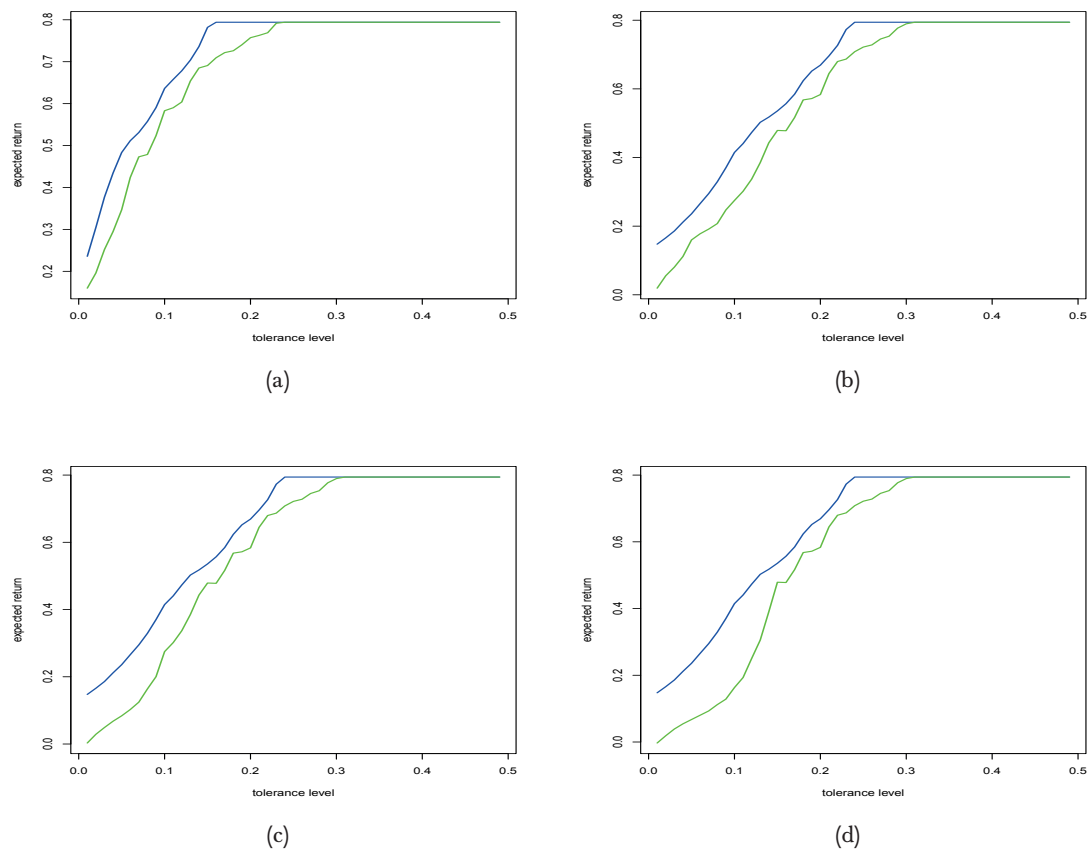
Figure 4.37: (a) plots the optimal portfolio expected return for the type I problem without uncertainty (blue) and the robust optimal portfolio expected return of the same problem with added eigenvector uncertainty (green) against the tolerance level, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.

Figure 4.38: (a) plots the optimal portfolio negated value-at-risk for the type I problem without uncertainty (blue) and the robust optimal portfolio negated value-at-risk of the same problem with added eigenvector uncertainty (green) against the tolerance level, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.

Figure 4.39: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type I problem both with added eigenvector uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Chapter 5

Trading Costs and Integer Constraints

In this chapter, we include trading costs in the robust location-scale problem considered in the last chapter. Since trading costs is a function of the asset weights, they render the problem non-convex. Previous work [126] to circumvent this issue is based on a heuristic algorithm, which is not exact. We convert the problem into a mixed-integer program. We also impose integer constraints on the trading quantities to make the asset allocation more realistic, and show that the resulting problem can likewise be expressed in the form of a mixed-integer program.

5.1 Trading Costs

Trading costs can be included in a portfolio optimization problem simply by adding the function

$$d(\mathbf{w}) = \sum_{i=1}^{n} d_i(w_i),$$

where $d_i(w_i)$ is the cost function of the ith asset, to the left-hand side of the budget constraint $\mathbf{w}^{\mathrm{T}}\mathbb{1} = 1$ to obtain

$$\mathbf{w}^{\mathrm{T}}\mathbb{1} + d(\mathbf{w}) = 1. \tag{5.1}$$

We only consider linear transaction with fixed costs[1], meaning that

$$d_i(w_i) = \begin{cases} 0, & w_i = 0, \\ \alpha_i^+ + \beta_i^+ w_i, & w_i > 0, \\ \alpha_i^- - \beta_i^- w_i, & w_i < 0, \end{cases}$$

where $\alpha_i^+ \geq 0$ and $\alpha_i^- \geq 0$ are the fixed costs, while $\beta_i^+ \geq 0$ and $\beta_i^- \geq 0$ are the cost rates of the long and short positions respectively. Observe that if we let $\mathbf{w}^+ = \mathbf{y}$ and $\mathbf{w}^- = \mathbf{z}$, then

---

[1]In practice, transaction costs may be more complicated, for which the constraint relaxation to be introduced subsequently can be easily adapted.
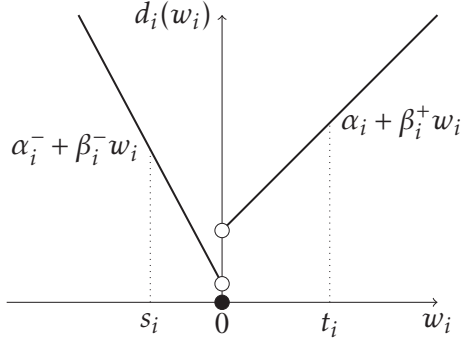
Figure 5.1: Linear transaction with fixed costs

(5.1) is equivalent to

$$(\mathbf{y} - \mathbf{z})^{\mathrm{T}} \mathbb{1} + \sum_{i=1}^{n} (\alpha_i^+ v_i + \beta_i^+ y_i + \alpha_i^- (1 - v_i) + \beta_i^- z_i) = 1,$$
$$y_i \le t_i v_i, \ z_i \le -s_i (1 - v_i),$$
$$v_i \in \{0, 1\}, \ i = 1, \dots, n,$$
$$\mathbf{y}, \mathbf{z} \ge \mathbf{0},$$

(5.2)

where $s_i = -\ell_i$ (the most the ith asset can be short-selled) and $t_i = \frac{1 - \alpha_i^+ - \sum_{j=1, j \ne i}^{n} \ell_j}{1 + \beta_i^+}$ are the largest values w_i^- and w_i^+ can take respectively. Note that t_i is obtained by solving $t_i + d_i(t_i) = 1 - \sum_{j=1, j \ne i}^{n} \ell_j$, and (5.2) allows only one of y_i and z_i to be non-zero, for $i = 1, \dots, n$. The portfolio return is

$$R_{\mathcal{P}} = \sum_{i=1}^{n} w_i (1 + R_i) - 1$$
$$= \mathbf{w}^{\mathrm{T}} \mathbf{R} + \mathbf{w}^{\mathrm{T}} \mathbb{1} - 1,$$

where the term $\mathbf{w}^{\mathrm{T}} \mathbb{1} - 1$ is not zero anymore but instead equal to the negative of $d(\mathbf{w})$. Therefore, if linear transaction with fixed costs are to be added to the robust location-scale problem involving any combination of the location, eigenvalue and eigenvector uncertainties from the last chapter, we need to solve a mixed integer program, obtained by first adding $\mathbf{w}^{\mathrm{T}} \mathbb{1}$ to the objective function, then replacing the budget constraint $\mathbf{w}^{\mathrm{T}} \mathbb{1} = 1$ with (5.2), before changing each $\mathbf{w}$ to $\mathbf{y} - \mathbf{z}$ in the problem.

5.1.1 Numerical Experiment Revisited

Figures 5.2 - 5.6 are analogs of Figures 4.35 - 4.39 in Section 4.5.1, with all else remaining constant except that we also include the eigenvalue and eigenvector uncertainties mentioned in the previous chapter as well as trading costs. Figures C.43 - C.45 (analogs of Figure 5.2) and Figures C.46 - C.48 (analogs of Figure 5.6) in Appendix C correspond to the type II, III and IV problems respectively. In particular, we fix $\alpha_i^+ = 0.0001$, $\alpha_i^- = 0.01$, $\beta_i^+ = 0.0002$ and $\beta_i^- = 0.02$ for $i = 1, \dots, 20$. Each mixed integer program is solved using the internal

branch and bound solver of the MATLAB interface YALMIP together with the lower bound solver SDPT3. Similar conclusions as the other cases can be drawn, but note that the optimal weights in each problem do not add up to the total wealth due to trading costs.
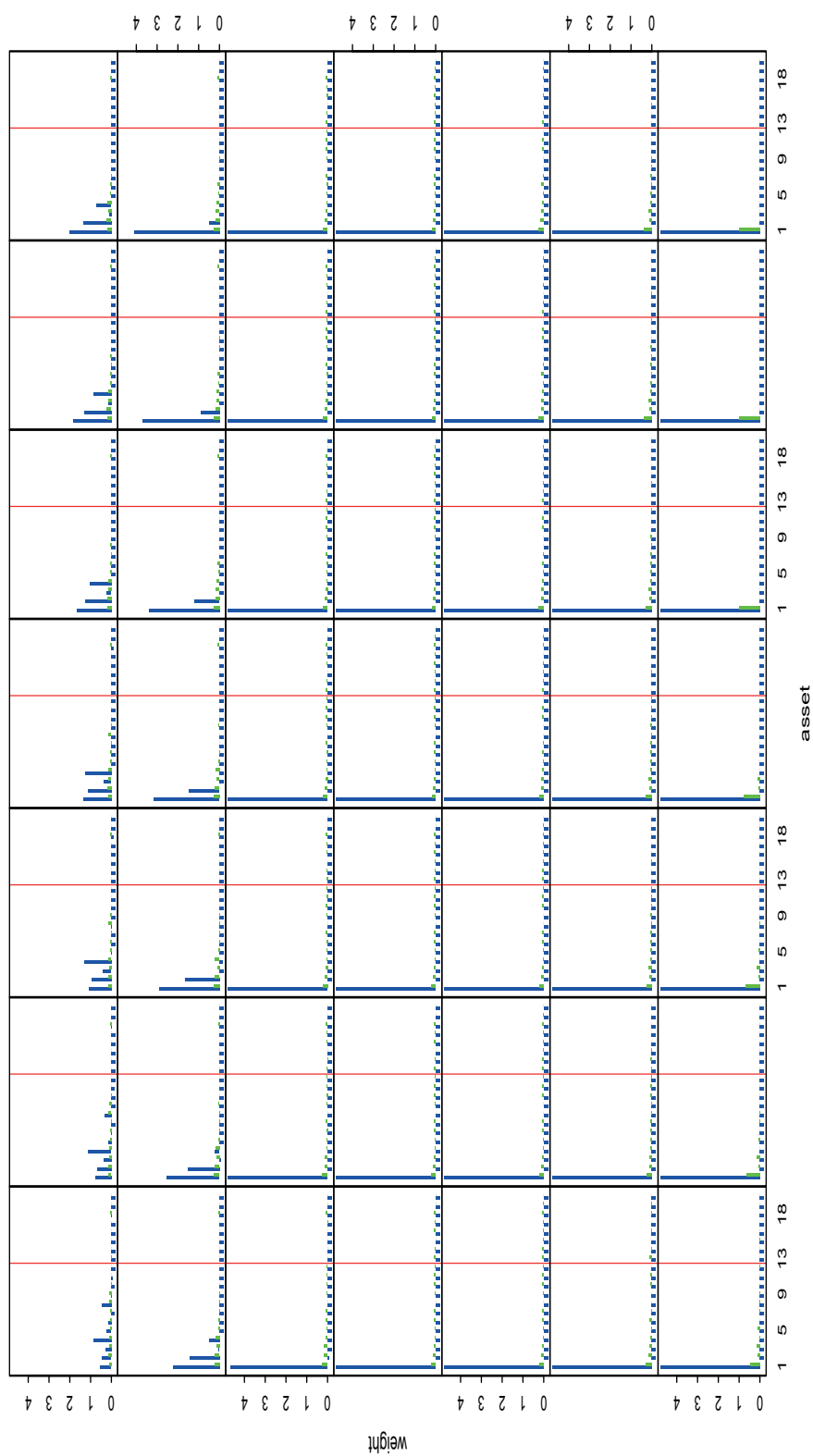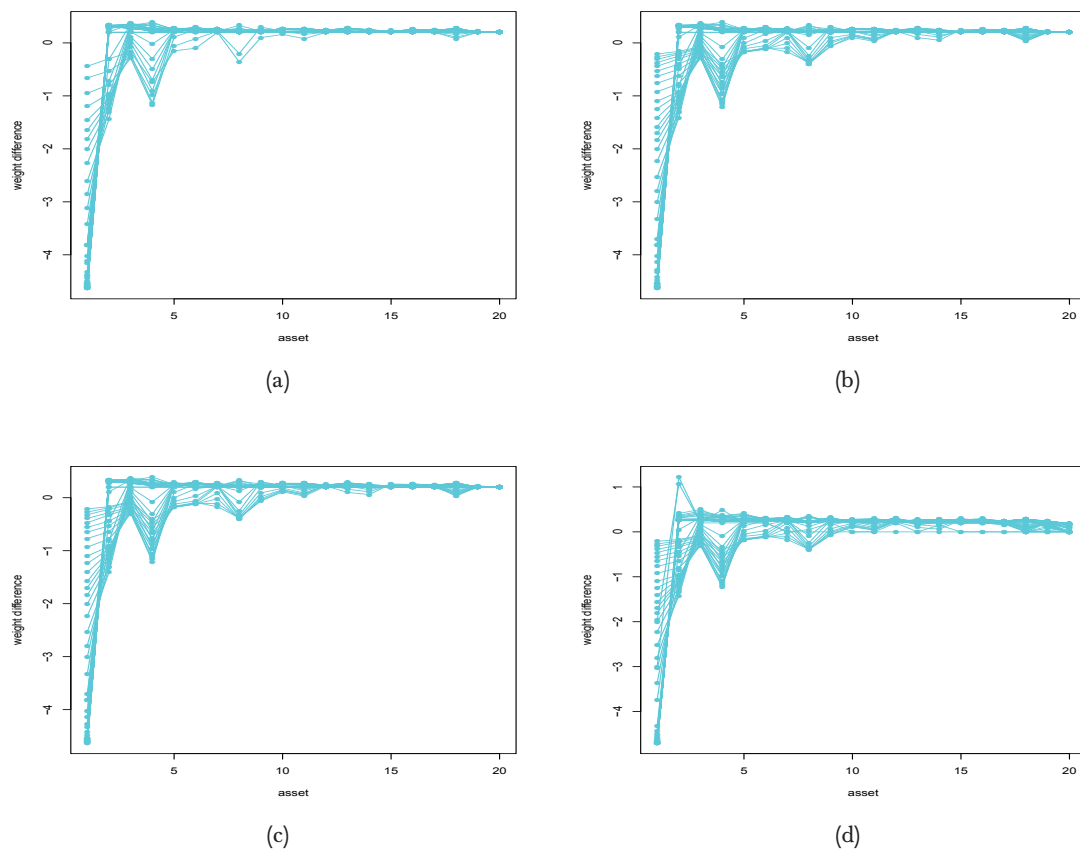
Figure 5.2: each subfigure plots the optimal weights of the type I problem both with added location, eigenvalue and eigenvector uncertainties plus trading costs (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure 5.3: (a) plots the optimal weights of the type I problem with added location, eigenvalue and eigenvector uncertainties plus trading costs minus the optimal wieghts of the same problem without added uncertainty against the asset number for $\epsilon = 0.01, 0.02, \ldots, 0.49$, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
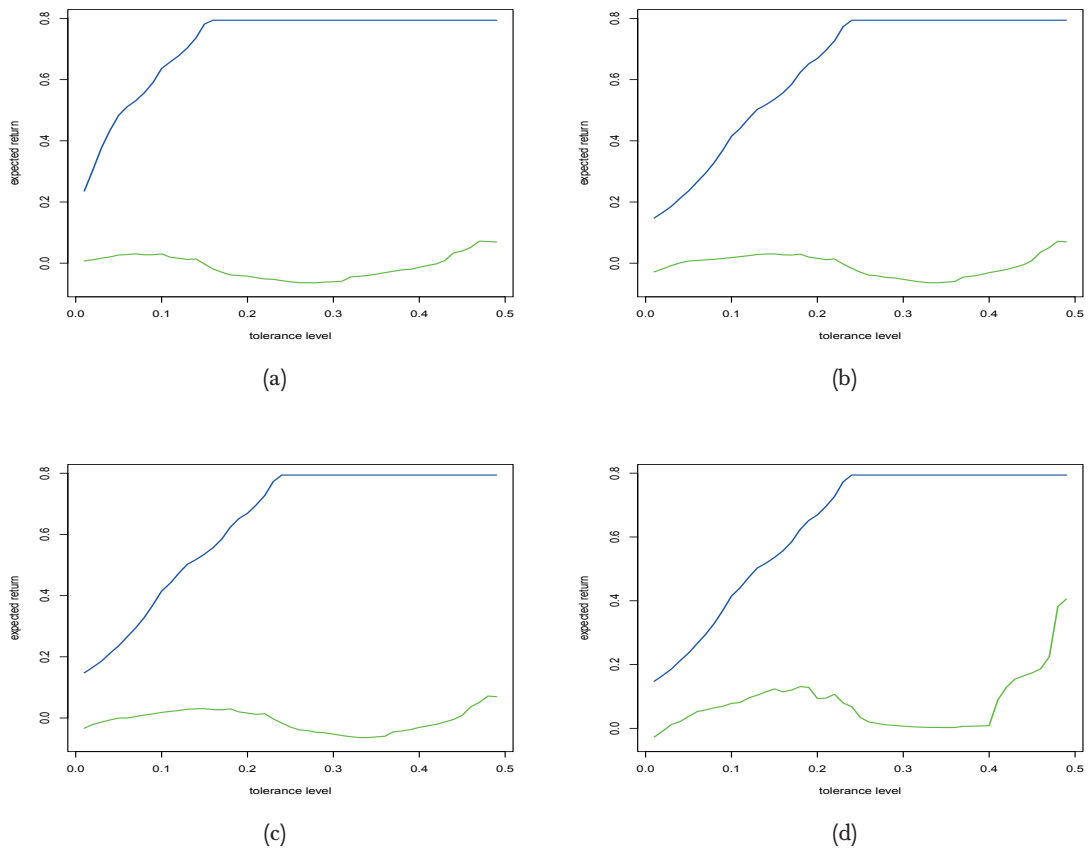
Figure 5.4: (a) plots the optimal portfolio expected return for the type I problem without uncertainty (blue) and the robust optimal portfolio expected return of the same problem with added location, eigenvalue and eigenvector uncertainties plus trading costs (green) against the tolerance level, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
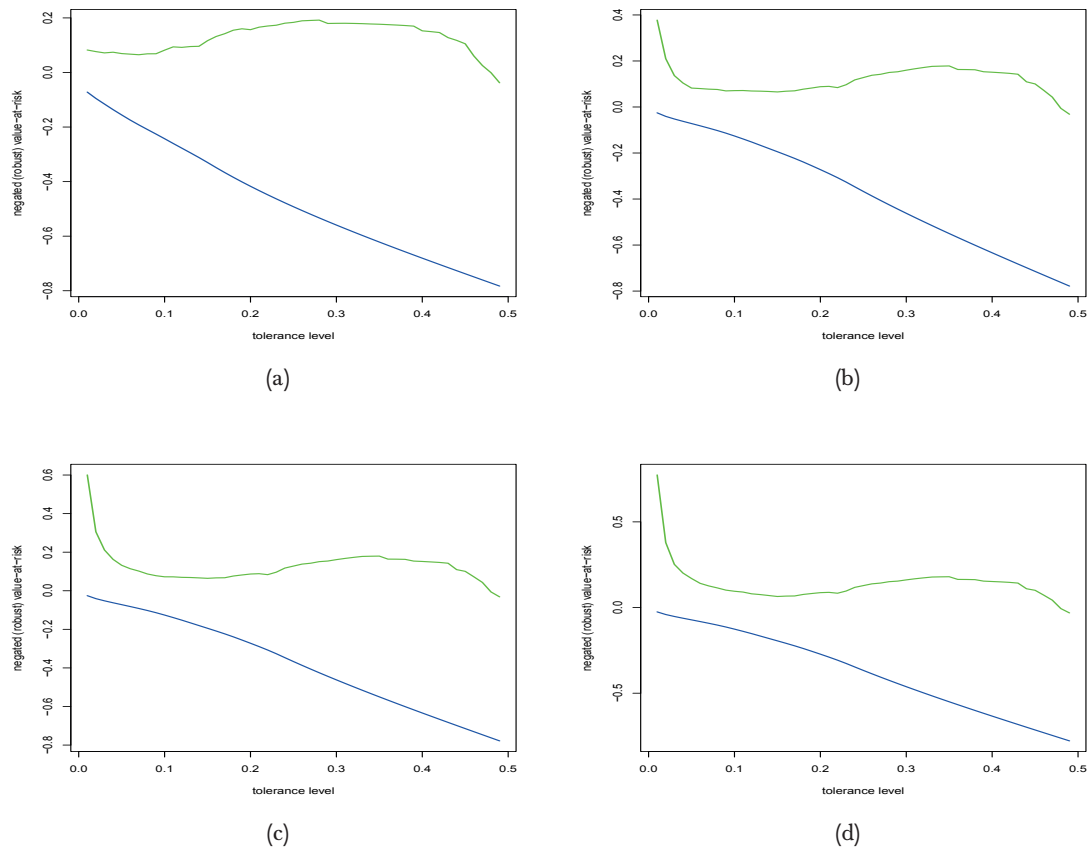
Figure 5.5: (a) plots the optimal portfolio negated value-at-risk for the type I problem without uncertainty (blue) and the robust optimal portfolio negated value-at-risk of the same problem with added location, eigenvalue and eigenvector uncertainties plus trading costs (green) against the tolerance level, where short-selling is allowed up to a maximum of one-fifth the total wealth; (b), (c) and (d) are analogs of (a) for the type II, III and IV problems respectively.
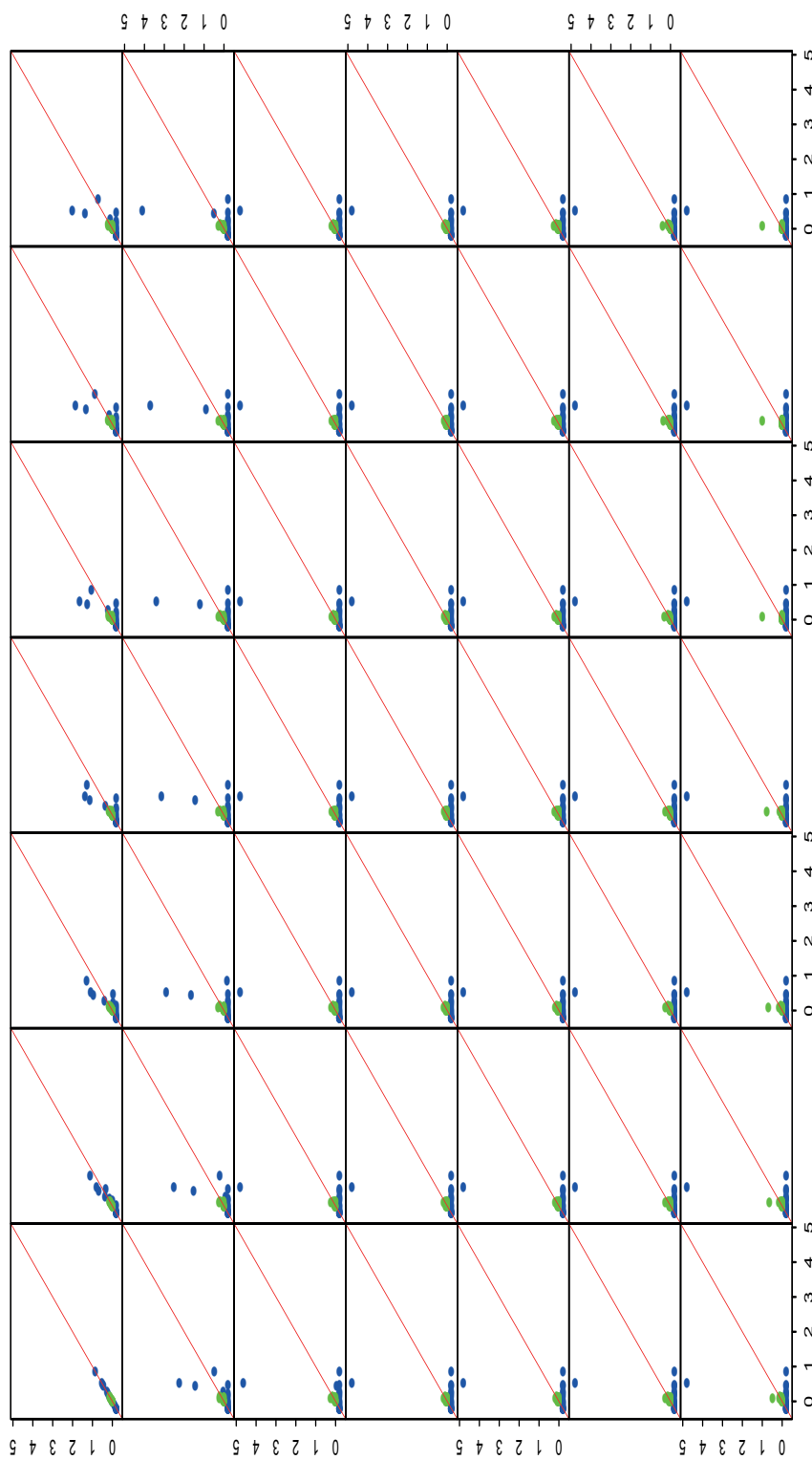
Figure 5.6: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type I problem both with added location, eigenvalue and eigenvector uncertainties plus trading costs (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

5.2 Integer Constraints

So far we assume that a fraction of an asset can be bought or sold, which is not true in reality. To take into account the indivisibility of shares, we impose integer constraints on the trading quantities. This means, in particular, that after investment in the risky assets there is usually remaining proportion of wealth $w_{n+1} \geq 0$ which can be put into a riskless asset with return μ_0 so that the portfolio return is

$$R_{\mathcal{P}} = \sum_{i=1}^{n} w_i(1 + R_i) + w_{n+1}(1 + \mu_0) - 1,$$

where

$$w_i = \frac{m_i S_i}{w_0}, \ i = 1, \ldots, n \tag{5.3}$$

such that $m_i \in \mathbb{Z}$ is the number of shares in the long or short position, S_i is the asset price at the beginning of the trading period, w_0 is the initial total wealth. The budget constraint then becomes

$$\mathbf{w}^{\mathrm{T}}\mathbb{1} + w_{n+1} - 1 = 0.$$

Therefore, if integer constraints are to be included in the robust location-scale problem involving any combination of the location, eigenvalue and eigenvector uncertainties from the last chapter, we need to solve a mixed integer program, obtained by first adding $w_{n+1}\mu_0$ to the objective function, then replacing $\mathbf{w}^{\mathrm{T}}\mathbb{1} = 1$ by (5.2), before letting $\mathbf{w}$ be defined as (5.3) in the problem. On the other hand, if linear transaction with fixed costs as well as integer constraints are to be included in the robust location-scale problem involving any combination of the location, eigenvalue and eigenvector uncertainties from the last chapter, we also need to solve a mixed integer program, obtained by first adding $w_{n+1}\mu_0 + \mathbf{w}^{\mathrm{T}}\mathbb{1}$ to the objective function, then replacing $\mathbf{w}^{\mathrm{T}}\mathbb{1} = 1$ with

$$(\mathbf{y} - \mathbf{z})^{\mathrm{T}}\mathbb{1} + \sum_{i=1}^{n}(\alpha_i^+ v_i + \beta_i^+ y_i + \alpha_i^-(1 - v_i) + \beta_i^- z_i) + w_{n+1} = 1,$$
$$y_i - z_i = \frac{m_i S_i}{w_0}, \ m_i \in \mathbb{Z},$$
$$y_i \leq t_i v_i, \ z_i \leq -s_i(1 - v_i),$$
$$v_i \in \{0, 1\}, \ i = 1, \ldots, n,$$
$$w_{n+1} \geq 0, \ \mathbf{y}, \mathbf{z} \geq \mathbf{0},$$

before letting $\mathbf{w}$ be defined as (5.3) in the problem. Results if we add integer constraints are very similiar to those of the previous numerical experiments and are thus omitted.

Chapter 6

Diversification

As we see in the last chapter, robust optimization tends to avoid allocating weight to assets with high location and low scale, which seems to indicate that robustness is linked to diversification. We will explore this topic in the following chapter.

6.1 Measure of Diversification

The Rao's Quadratic Entropy (RQE) [169] is extended to formally define the notion of portfolio diversification in this section.

Definition 6.3
[48, p. 5] Let Ω be a population of elements each characterized by a value X with probability measure P. The RQE of any two elements $\omega_1, \omega_2 \in \Omega$ corresponding to values X_1 and X_2 is defined as

$$H(P) = \int d(X_1, X_2) dP(X_1) dP(X_2) \tag{6.1}$$

where $d(\cdot, \cdot) \in \mathbb{R}_+$ is symmetric and represents the difference between ω_1 and ω_2.

Note that (6.1) can be interpreted as the average difference between X_1 and X_2. If X is discrete with n possible values $x_1, \ldots, x_n$, then

$$H(P) = \sum_{i,j=1}^{n} d_{ij} p_i p_j$$

where $p_i = \mathbb{P}(X = x_i)$, $d_{ij} = d_{ji}$ and $d_{ii} = 0$ for $i, j = 1, \ldots, n$.

6.1.1 Portfolio RQE

Definition 6.4
[48, p. 6] The portfolio RQE of a portfolio with n assets and weight vector $\mathbf{w} = [w_i, \ldots, w_n]^{\mathrm{T}}$ is defined as

$$\frac{1}{2} \mathbf{w}^{\mathrm{T}} \mathbf{D} \mathbf{w} =: H_{\mathbf{D}}(\mathbf{w}), \tag{6.2}$$

109

where the dissimilarity matrix $\mathbf{D}$ is symmetric such that its (i, j)th entry $d_{ij} \geq 0$ and its ith diagonal entry $d_{ii} = 0$.

Note that d_{ij} can be viewed as the amount of unshared information between assets i and j. Furthermore, the above definition implies that buying an asset and short-selling another dissimilar asset instead of buying both the assets should result in a more extreme portfolio return with less diversification and hence smaller portfolio RQE, which can also be seen as half the weighted difference between two randomly chosen assets with replacement. Since $\mathbf{D}$ is a Euclidean distance matrix, it is conditionally negative definite [176], meaning that $\mathbf{w}^{\mathrm{T}}\mathbf{D}\mathbf{w}$ is non-positive whenever $\mathbf{w}^{\mathrm{T}}\mathbb{1} = 0$, so that according to Rao and Nayak (1985),

$$H_{\mathbf{D}}(\mathbf{w}) = \frac{1}{2}\sum_{i,=1}^{n} w_i \mathbf{D}_{H_{\mathbf{D}}}(\mathbf{w}^i, \mathbf{w}) \tag{6.3}$$

where $w_i^i = 1$, $w_j^i = 0$ for $j \neq i$ and $\mathbf{D}_{H_{\mathbf{D}}}(\mathbf{w}^i, \mathbf{w}) = 2H_{\mathbf{D}}(\mathbf{w}^i, \mathbf{w}) - H_{\mathbf{D}}(\mathbf{w}^i) - H_{\mathbf{D}}(\mathbf{w})$ is the difference between the portfolios $\mathbf{w}$ and $\mathbf{w}^i$ such that $H_{\mathbf{D}}(\mathbf{w}^i, \mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{D}\mathbf{w}^i$. Without short-selling, (6.3) suggests that the more dissimilar the portfolio $\mathbf{w}$ is from the single asset portfolio $\mathbf{w}^i$, the higher the portfolio RQE. It can also be shown that the portfolio obtained by maximizing the portfolio RQE is equidissimlar to every asset if there is no restriction on short-selling. On the other hand, if each asset can be short-selled up to a certain bound, maximizing the portfolio RQE obtains a portfolio equidissimlar to each asset whose weight does not reach its bound. We refer the reader to Camarchael et al. (2015) for more properties of the portfolio RQE.

6.1.2 RQE as Unifying Diversification Measure

The portfolio RQE generalizes quite a number of diversification measures, and we mention a few of them.

6.1.2.1 Gini-Simpson Index

The Gini-Simpson Index (see for example [200] and [49]) is defined as

$$1 - \sum_{i=1}^{n} w_i^2 =: GS(\mathbf{w}),$$

and is a special case of the portfolio RQE where $d_{ij} = 1 - \delta_{ij}$ such that

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

To see this, note that

$$1 = \left(\sum_{i=1}^{n} w_i\right)^2$$

so that

$$1 - \sum_{i=1}^{n} w_i^2 = \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j.$$

6.1.2.2   Return Gap

When the portfolio consists of two equally weighted assets with scale matrix

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

and the (i, j)th entry of the dissimilarity matrix is

$$d_{ij} = \sigma^2 (1 - \rho)(1 - \delta_{ij}),$$

then

$$H_{\mathbf{D}}(\mathbf{w}) = \frac{1}{2} \sigma^2 (1 - \rho),$$

whose squared root is exactly the Return Gap (RG).

6.1.2.3   Diversification Return

Assume that $d_{ij} = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}$ where σ_i^2 and σ_{ij} are the (i, i)th and (i, j)th entries of the scale matrix Σ respectively. Then, the portfolio RQE is

$$
\begin{aligned}
H_{\mathbf{D}}(\mathbf{w}) &= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}) w_i w_j \\
&= \frac{1}{2} \left(\sum_{i=1}^{n} \sigma_i^2 w_i \sum_{j=1}^{n} w_j + \sum_{i=1}^{n} w_i \sum_{j=1}^{n} \sigma_j^2 w_j - 2 \sum_{i=1^n} \sum_{j=1}^{n} \sigma_{ij} w_i w_j \right) \\
&= \boldsymbol{\sigma}^2 \mathbf{w} - \mathbf{w}^{\mathsf{T}} \Sigma \mathbf{w}
\end{aligned}
\tag{6.4}
$$

where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \dots, \sigma_n^2]$. The expression (6.4), which we denote as $Dr(\mathbf{w})$, is also known as the diversifcation return, and first comes in the form of excess growth rate in Fernholz and Shay (1982) and appears as the difference between the portfolio compound return and weighted average asset compond return in Booth and Fama (1992).

6.1.3   Optimal Dissimilarity In Location-Scale Framework

[48, Section 5.2] We consider the optimal dissimilarity matrix $\mathbf{D}$ under different scenarios in the location-scale framework.

6.1.3.1   No Information

When no information is available, the location-scale optimal portfolio RQE is the GS index with $d_{ij} = d$ and $d_{ii} = 0$ for all $i = 1, \dots, n$ and $j \neq i$, since there is no way whatsoever to distinguish between any two assets.

6.1.3.2   Correlation Information

If only the correlation information is available, we may assume that each asset has unit variance without loss of generality. In this case, the (i, j)th entry of the optimal dissimilarity matrix is $d_{ij} = 2(1 - \rho_{ij})$, where ρ_{ij} is the correlation between assets i and j. The portfolio RQE then becomes

$$
\begin{aligned}
H_{\mathbf{D}}(\mathbf{w}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - \rho_{ij}) w_i w_j \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j - \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} w_i w_j \\
&= \left(\sum_{i=1}^{n} w_i \right)^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} w_i w_j \\
&= 1 - \mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w},
\end{aligned}
$$

so that minimizing the portfolio variance is equivalent to maximizing the portfolio RQE. Note that $H_{\mathbf{D}}(\mathbf{w})$ is a decreasing function of ρ_{ij}, which makes sense since a higher correlation among the assets will result in a lower portfolio RQE. In addition, if we let $\boldsymbol{\sigma}$ denote the squared root of $\boldsymbol{\sigma}^2$ element-wise, the diversification ratio is

$$
\begin{aligned}
DR(\mathbf{w}) &= \frac{\boldsymbol{\sigma} \mathbf{w}}{\sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}}} \\
&= \frac{1}{\sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}}}
\end{aligned}
$$

so that maximizing it is also equivalent to maximizing the portfolio RQE.

6.1.3.3   Scale Information

Maximizing the portfolio RQE results in a portfolio equidissmilar from each asset contained in the portfolio if no short-selling is allowed, or from each asset not fully short-selled if short-selling is allowed. Nevertheless, if the full scale matrix is available, it is better to place more weight on assets with a lower scale. Hence, maximizing the portfolio RQE is no longer location-scale optimal and should be combined with other measures to achieve better results. One way to do it is to minimize $\boldsymbol{\sigma}^2 \mathbf{w} - Dr(\mathbf{w})$, where the diversification return is coupled with the weighted average of the diagonal entries of the scale matrix. Another way is to maximize the $Dr(\mathbf{w})/\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}$ which is essentially the portfolio scale adjusted diversification return. We can also maximize the diversification ratio.

6.1.3.4   Location And Scale Information

If the location vector and scale matrix are fully known, then the portfolio RQE maximization is also location-scale sub-optimal, since naturally we would like to place more weight on

assets with higher location and lower scale instead of ignoring their effects. In this case, we may solve

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\}$$

$$\Leftrightarrow\quad \max_{\mathbf{w}\in\mathcal{W}_\ell}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\boldsymbol{\sigma}^2\mathbf{w}-(\boldsymbol{\sigma}^2\mathbf{w}-\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w})}\right\}$$

$$\Leftrightarrow\quad \max_{\mathbf{w}\in\mathcal{W}_\ell}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\boldsymbol{\sigma}^2\mathbf{w}-Dr(\mathbf{w})}\right\},$$

where the diversification return is combined with the portfolio location and weighted average of the diagonal entries of the scale matrix to achieve a better outcome. If there is imperfect information of the location vector and scale matrix, then we may solve

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{(\boldsymbol{\mu},\Sigma)\in\mathcal{M}\times\mathcal{S}}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\},$$

which is exactly the robust location-scale problem we are interested in.

6.2 Sensitivity of Diversification

Having shown that the natural portfolio RQE measure of the portfolio obtained by solving the location-scale problem is the diversification return, let us look at how it behaves as uncertainty is introduced into the model. As we will see later, with the introduction of uncertainty into the location-scale problem, no simple pattern of how the portfolio diversification return changes may be found.

6.2.1 Location Uncertainty

We first consider the effects of location uncertainty on the portfolio diversification return. To obtain analytical results later, we do not let w_i to be either 0 or ℓ_i, so that the robust location-scale problem (4.3) can be rewritten as

$$\max_{w_i\in\mathbb{R}\setminus\{0,\ell_i\},i=1,\dots,n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\sum_{i=1}^{n}|w_i|a_i-\kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}:\mathbf{w}^{\mathrm{T}}\mathbb{1}=1,\mathbf{w}\geq\boldsymbol{\ell}\right\}. \tag{6.5}$$

Note that the impact of removing 0 and ℓ_i from the feasible domain of w_i is negligible due to continuity of the original problem and the optimal solution of (6.5) can be very well approximated in practice by solving

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\sum_{i=1}^{n}|w_i|a_i-\kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}:\begin{array}{c}\mathbf{w}^{\mathrm{T}}\mathbb{1}=1,\\ \boldsymbol{\ell}+\delta\mathbb{1}\leq\mathbf{w}\leq-\delta\mathbb{1},\\ \mathbf{w}\geq\delta\mathbb{1}\end{array}\right\}$$

where δ is a small positive number. Before moving on further, we state a theorem providing us with the optimal solution of (6.5) in analytical form.

Theorem 6.10

Let V be the optimal value of (6.5), then its optimal solution is

$$\mathbf{w}_* = \frac{\Sigma^{-1}\begin{bmatrix}(\mu_1 - \mathrm{sign}(w_{1*})a_1)\\ \vdots \\ (\mu_n - \mathrm{sign}(w_{n*})a_n)\end{bmatrix} - V\Sigma^{-1}\mathbb{1}}{\mathbb{1}^T\Sigma^{-1}\begin{bmatrix}(\mu_1 - \mathrm{sign}(w_{1*})a_1)\\ \vdots \\ (\mu_n - \mathrm{sign}(w_{n*})a_n)\end{bmatrix} - V\mathbb{1}^T\Sigma^{-1}\mathbb{1}}.$$

Proof: Note that solving (6.5) is equivalent to solving

$$\max_{x_i\in\mathbb{R}\setminus\{0,\pm\ell_i\},i=1,\dots,n}\left\{\begin{array}{l}[|x_1|+\ell_1,\dots,|x_n|+\ell_n]\boldsymbol{\mu} - \sum_{i=1}^n\||x_i|+\ell_i|a_i\\ -\kappa(\epsilon)\sqrt{[|x_1|+\ell_1,\dots,|x_n|+\ell_n]\Sigma\begin{bmatrix}|x_1|+\ell_1\\ \vdots \\ |x_n|+\ell_n\end{bmatrix}}\\ :[|x_1|+\ell_1,\dots,|x_n|+\ell_n]^T\mathbb{1}=1\end{array}\right\}\quad(6.6)$$

for $\begin{bmatrix}|x_1|+\ell_1\\ \vdots \\ |x_n|+\ell_n\end{bmatrix}$. Let $\mathbf{x}_*$ be the optimal solution of (6.6). Since x_i cannot be 0 or $\pm\ell_i$, the objective and constraint functions are continuously differentiable so that there exists a Lagrange multiplier ν_* which satisfies

$$\begin{bmatrix}\mathrm{sign}(x_{1*})(\mu_1 - \mathrm{sign}(|x_{1*}|+\ell_1)a_1)\\ \vdots \\ \mathrm{sign}(x_{n*})(\mu_n - \mathrm{sign}(|x_{n*}|+\ell_n)a_n)\end{bmatrix} - \kappa(\epsilon)\frac{\mathrm{diag}(\mathrm{sign}(x_{1*}),\dots,\mathrm{sign}(x_{n*}))\Sigma\begin{bmatrix}|x_{1*}|+\ell_1\\ \vdots \\ |x_{n*}|+\ell_n\end{bmatrix}}{\sqrt{[|x_{1*}|+\ell_1,\dots,|x_{n*}|+\ell_n]\Sigma\begin{bmatrix}|x_{1*}|+\ell_1\\ \vdots \\ |x_{n*}|+\ell_n\end{bmatrix}}}$$

$$-\nu_*\begin{bmatrix}\mathrm{sign}(x_{1*})\\ \vdots \\ \mathrm{sign}(x_{n*})\end{bmatrix} = \mathbf{0}_{n\times1}.$$

(6.7)

and

$$[|x_{1*}|+\ell_1,\dots,|x_{n*}|+\ell_n]^T\mathbb{1}=1\qquad(6.8)$$

Multiplying (6.7) on the left by $[x_{1*}+\mathrm{sign}(x_{1*})\ell_1,\dots,x_{n*}+\mathrm{sign}(x_{n*})\ell_n]$, we obtain

$$[|x_{1*}|+\ell_1,\dots,|x_{n*}|+\ell_n]\boldsymbol{\mu} - \sum_{i=1}^n\||x_{i*}|+\ell_i|a_i$$

$$
-\kappa(\epsilon)\sqrt{[|x_{1*}|+\ell_1,\ldots,|x_{n*}|+\ell_n]\Sigma\begin{bmatrix}|x_{1*}|+\ell_1\\ \vdots\\ |x_{n*}|+\ell_n\end{bmatrix}}
$$

$$
-\nu_*[|x_{1*}|+\ell_1,\ldots,|x_{n*}|+\ell_n]^{\mathsf{T}}\mathbb{1}=0
$$

which implies

$$
\nu_* = V
$$

by (6.8) and the fact that both (6.5) and (6.6) have the same optimal value.

Multiplying (6.7) on the left by $\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathrm{diag}(\mathrm{sign}(x_{1*}),\ldots,\mathrm{sign}(x_{n*}))$ obtains

$$
\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\begin{bmatrix}(\mu_1-\mathrm{sign}(|x_{1*}|+\ell_1)a_1)\\ \vdots\\ (\mu_n-\mathrm{sign}(|x_{n*}|+\ell_n)a_n)\end{bmatrix}-\kappa(\epsilon)\frac{\mathbb{1}^{\mathsf{T}}\begin{bmatrix}|x_{1*}|+\ell_1\\ \vdots\\ |x_{n*}|+\ell_n\end{bmatrix}}{\sqrt{[|x_{1*}|+\ell_1,\ldots,|x_{n*}|+\ell_n]\Sigma\begin{bmatrix}|x_{1*}|+\ell_1\\ \vdots\\ |x_{n*}|+\ell_n\end{bmatrix}}}
$$

$$
-\nu_*\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}=0
$$

$$
\Rightarrow \quad \kappa(\epsilon)=\left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\begin{bmatrix}(\mu_1-\mathrm{sign}(|x_{1*}|+\ell_1)a_1)\\ \vdots\\ (\mu_n-\mathrm{sign}(|x_{n*}|+\ell_n)a_n)\end{bmatrix}-\nu_*\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)
$$

$$
\sqrt{[|x_{1*}|+\ell_1,\ldots,|x_{n*}|+\ell_n]\Sigma\begin{bmatrix}|x_{1*}|+\ell_1\\ \vdots\\ |x_{n*}|+\ell_n\end{bmatrix}}
$$

by using (6.8) again. Substituting the above into (6.7) before multiplying the resulting equation on the left by $\Sigma^{-1}\mathrm{diag}(\mathrm{sign}(x_{1*}),\ldots,\mathrm{sign}(x_{n*}))$ result in

$$
\Sigma^{-1}\begin{bmatrix}(\mu_1-\mathrm{sign}(|x_{1*}|+\ell_1)a_1)\\ \vdots\\ (\mu_n-\mathrm{sign}(|x_{n*}|+\ell_n)a_n)\end{bmatrix}-\nu_*\Sigma^{-1}\mathbb{1}
$$

$$
-\begin{bmatrix}|x_{1*}|+\ell_1\\ \vdots\\ |x_{n*}|+\ell_n\end{bmatrix}\left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\begin{bmatrix}(\mu_1-\mathrm{sign}(|x_{1*}|+\ell_1)a_1)\\ \vdots\\ (\mu_n-\mathrm{sign}(|x_{n*}|+\ell_n)a_n)\end{bmatrix}-\nu_*\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)=0
$$

so that rearranging we get

$$
\begin{bmatrix}|x_{1*}|+\ell_1\\ \vdots\\ |x_{n*}|+\ell_n\end{bmatrix}=\frac{\Sigma^{-1}\begin{bmatrix}(\mu_1-\mathrm{sign}(|x_{1*}|+\ell_1)a_1)\\ \vdots\\ (\mu_n-\mathrm{sign}(|x_{n*}|+\ell_n)a_n)\end{bmatrix}-\nu_*\Sigma^{-1}\mathbb{1}}{\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\begin{bmatrix}(\mu_1-\mathrm{sign}(|x_{1*}|+\ell_1)a_1)\\ \vdots\\ (\mu_n-\mathrm{sign}(|x_{n*}|+\ell_n)a_n)\end{bmatrix}-\nu_*\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}}.
$$

It remains to substitute $\mathbf{w}_* = \begin{bmatrix} |x_{1*}| + \ell_1 \\ \vdots \\ |x_{n*}| + \ell_n \end{bmatrix}$ and $\nu_* = V$ into the above equation to obtain the desired result. $\qquad\square$

6.2.1.1 Individual Location Uncertainty

We are now ready to provide a necessary and sufficient condition (6.9) in the following theorem for the diversification return of the portfolio obtained by solving

$$\max_{w_i \in \mathbb{R} \setminus \{0, \ell_i\}, i = 1, \dots, n} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} - \kappa(\epsilon) \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell} \right\}$$

to increase when the jth asset location changes slightly.

Theorem 6.11

Let $\mathbf{w}_*$ be the optimal solution of

$$\max_{w_i \in \mathbb{R} \setminus \{0, \ell_i\}, i = 1, \dots, n} \left\{ \mathbf{w}^{T} \boldsymbol{\mu} - \sum_{i=1}^{n} |w_i| a_i - \kappa(\epsilon) \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell} \right\}$$

where $a_i = p_i \sigma_i$ for $i = 1, \dots, n$. Then, a necessary and sufficient condition for $\left. \frac{dDr(\mathbf{w}_*)}{dp_j} \right|_{\mathbf{p}=\mathbf{0}}$ to be positive is

$$(AB - D(F - \boldsymbol{\mu}^{\mathrm{T}} \bar{\mathbf{w}})) \bar{w}_j - (E_j(B-2) - DC_j)\bar{V} - 2\bar{w}_j \bar{V} - AC_j + E_j(F - 2\boldsymbol{\mu}^{\mathrm{T}} \bar{\mathbf{w}}) > 0, \qquad (6.9)$$

where $\bar{\mathbf{w}}$ and $\bar{V}$ are the optimal solution and value of

$$\max_{w_i \in \mathbb{R} \setminus \{0, \ell_i\}, i = 1, \dots, n} \left\{ \mathbf{w}^{T} \boldsymbol{\mu} - \kappa(\epsilon) \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell} \right\}$$

respectively, $A = \mathbb{1}^{\mathrm{T}} \Sigma^{-1} \boldsymbol{\mu}$, $B = \boldsymbol{\sigma}^2 \Sigma^{-1} \mathbb{1}$, $C_j = \boldsymbol{\sigma}^2 \Sigma^{-1} \mathbf{e}_j$, $D = \mathbb{1}^{\mathrm{T}} \Sigma^{-1} \mathbb{1}$, $E_j = \mathbb{1}^{\mathrm{T}} \Sigma^{-1} \mathbf{e}_j$ and $F = \boldsymbol{\sigma}^2 \Sigma^{-1} \boldsymbol{\mu}$.

Proof: First, note that w_{i*} is differentiable with respect to p_j for all $i, j = 1, \dots, n$ by using arguments similar to those found in the proof of Theorem 4.6. Therefore, we have that

$$\frac{dDr(\mathbf{w}_*)}{dp_j} = \left(\boldsymbol{\sigma}^2 - 2\mathbf{w}_*^{\mathrm{T}} \Sigma \right) \frac{d\mathbf{w}_*}{dp_j}$$

where

$$\frac{d\mathbf{w}_*}{dp_j} = \left(\left(\mathbb{1}^{\mathrm{T}} \Sigma^{-1} \begin{bmatrix} (\mu_1 - \mathrm{sign}(w_{1*}) p_1 \sigma_1) \\ \vdots \\ (\mu_n - \mathrm{sign}(w_{n*}) p_n \sigma_n) \end{bmatrix} - V \mathbb{1}^{\mathrm{T}} \Sigma^{-1} \mathbb{1} \right) \left(-\Sigma^{-1} \mathrm{sign}(w_{j*}) \sigma_j \mathbf{e}_j + |w_{j*}| \sigma_j \Sigma^{-1} \mathbb{1} \right) \right.$$
$$\left. - \left(\Sigma^{-1} \begin{bmatrix} (\mu_1 - \mathrm{sign}(w_{1*}) p_1 \sigma_1) \\ \vdots \\ (\mu_n - \mathrm{sign}(w_{n*}) p_n \sigma_n) \end{bmatrix} - V \Sigma^{-1} \mathbb{1} \right) \left(-\mathbb{1}^{\mathrm{T}} \Sigma^{-1} \mathrm{sign}(w_{j*}) \sigma_j \mathbf{e}_j + |w_{j*}| \sigma_j \mathbb{1}^{\mathrm{T}} \Sigma^{-1} \mathbb{1} \right) \right)$$

$$\left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\begin{bmatrix} (\mu_1 - \text{sign}(w_{1*})p_1\sigma_1) \\ \vdots \\ (\mu_n - \text{sign}(w_{n*})p_n\sigma_n) \end{bmatrix} - V\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)^2$$

by using Theorem 6.10 and the Envelope Theorem A.24, so that

$$\begin{aligned}
\left.\frac{d\mathbf{w}_*}{dp_j}\right|_{\mathbf{p}=0} &= \left(\left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)\left(-\Sigma^{-1}\text{sign}(\bar{w}_j)\sigma_j\mathbf{e}_j + |\bar{w}_j|\sigma_j\Sigma^{-1}\mathbb{1}\right)\right. \\
&\quad \left. - \left(\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\Sigma^{-1}\mathbb{1}\right)\left(-\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\text{sign}(\bar{w}_j)\sigma_j\mathbf{e}_j + |\bar{w}_j|\sigma_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)\right) \\
&\quad \left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)^2 \\
&= \left(|\bar{w}_j|\sigma_j\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \Sigma^{-1}\text{sign}(\bar{w}_j)\sigma_j\mathbf{e}_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} + \bar{V}\Sigma^{-1}\text{sign}(\bar{w}_j)\sigma_j\mathbf{e}_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right. \\
&\quad \left. + \Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\text{sign}(\bar{w}_j)\sigma_j\mathbf{e}_j - |\bar{w}_j|\sigma_j\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1} - \bar{V}\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\text{sign}(\bar{w}_j)\sigma_j\mathbf{e}_j\right) \\
&\quad \left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)^2
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{dDr(\mathbf{w}_*)}{dp_j}\right|_{\mathbf{p}=0} &= \left(|\bar{w}_j|\sigma_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}\left(\sigma^2\Sigma^{-1}\mathbb{1} - 2\right) - \text{sign}(\bar{w}_j)\sigma_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}\left(\sigma^2\Sigma^{-1}\mathbf{e}_j - 2\bar{w}_j\right)\right. \\
&\quad + \bar{V}\text{sign}(\bar{w}_j)\sigma_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\left(\sigma^2\Sigma^{-1}\mathbf{e}_j - 2\bar{w}_j\right) + \text{sign}(\bar{w}_j)\sigma_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbf{e}_j\left(\sigma^2\Sigma^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^{\mathsf{T}}\bar{\mathbf{w}}\right) \\
&\quad \left. - |\bar{w}_j|\sigma_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\left(\sigma^2\Sigma^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^{\mathsf{T}}\bar{\mathbf{w}}\right) - \bar{V}\text{sign}(\bar{w}_j)\sigma_j\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbf{e}_j\left(\sigma^2\Sigma^{-1}\mathbb{1} - 2\right)\right) \\
&\quad \left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)^{-2} \\
&= \left(|\bar{w}_j|\sigma_j A(B-2) - \text{sign}(\bar{w}_j)\sigma_j A\left(C_j - 2\bar{w}_j\right)\right. \\
&\quad + \bar{V}\text{sign}(\bar{w}_j)\sigma_j D\left(C_j - 2\bar{w}_j\right) + \text{sign}(\bar{w}_j)\sigma_j E_j\left(F - 2\boldsymbol{\mu}^{\mathsf{T}}\bar{\mathbf{w}}\right) \\
&\quad \left. - |\bar{w}_j|\sigma_j D\left(F - 2\boldsymbol{\mu}^{\mathsf{T}}\bar{\mathbf{w}}\right) - \bar{V}\text{sign}(\bar{w}_j)\sigma_j E_j(B-2)\right) \\
&\quad \left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}.
\end{aligned}$$

The theorem follows immediately by noting that a necessary and sufficient condition for the above expression to be positive is for its numerator to be as well. $\square$

6.2.1.2 Simultaneous Location Uncertainty

Next, we provide a necessary and sufficient condition (6.10) in the following theorem for the diversification return of the portfolio obtained by solving

$$\max_{w_i \in \mathbb{R}\setminus\{0,\ell_i\}, i=1,\dots,n}\left\{\mathbf{w}^{\mathsf{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathsf{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\mathsf{T}}\mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell}\right\}$$

to increase when the each asset location changes slightly.

Theorem 6.12

Let $\mathbf{w}_*$ be the optimal solution of

$$\max_{w_i \in \mathbb{R} \setminus \{0, \ell_i\}, i=1,\ldots,n} \left\{ \mathbf{w}^T \boldsymbol{\mu} - \sum_{i=1}^{n} |w_i| a_i - \kappa(\epsilon) \sqrt{\mathbf{w}^T \Sigma \mathbf{w}} : \mathbf{w}^T \mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell} \right\}$$

where $a_i = p\sigma_i$ for $i = 1,\ldots,n$. Then, a necessary and sufficient condition for $\frac{dDr(\mathbf{w}_*)}{dp}\Big|_{\mathbf{p}=0}$ to be positive is

$$AB - C + \left(E - 2\sum_{i=1}^{n} |\bar{w}_i| \sigma_i \right) \left(F - 2\boldsymbol{\mu}^T \bar{\mathbf{w}} \right) + \bar{V} \left(D \left(C - 2\sum_{i=1}^{n} |\bar{w}_i| \sigma_i \right) - E(B-2) \right) > 0, \quad (6.10)$$

where $\bar{\mathbf{w}}$ and $\bar{V}$ are the optimal solution and value of

$$\max_{w_i \in \mathbb{R} \setminus \{0, \ell_i\}, i=1,\ldots,n} \left\{ \mathbf{w}^T \boldsymbol{\mu} - \kappa(\epsilon) \sqrt{\mathbf{w}^T \Sigma \mathbf{w}} : \mathbf{w}^T \mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell} \right\}$$

respectively, $A = \mathbb{1}^T \Sigma^{-1} \boldsymbol{\mu}$, $B = \boldsymbol{\sigma}^2 \Sigma^{-1} \mathbb{1}$, $C_i = \boldsymbol{\sigma}^2 \Sigma^{-1} \mathbf{e}_i$, $D = \mathbb{1}^T \Sigma^{-1} \mathbb{1}$, $E_i = \mathbb{1}^T \Sigma^{-1} \mathbf{e}_i$ and $F = \boldsymbol{\sigma}^2 \Sigma^{-1} \boldsymbol{\mu}$.

Proof: First, note that w_{i*} is differentiable with respect to p for $i = 1,\ldots,n$ by using arguments similar to those found in the proof of Theorem 4.6. Therefore, we have that

$$\frac{dDr(\mathbf{w}_*)}{dp} = \left(\boldsymbol{\sigma}^2 - 2\mathbf{w}_*^T \Sigma \right) \frac{d\mathbf{w}_*}{dp}$$

where

$$\frac{d\mathbf{w}_*}{dp} = \left(\left(\mathbb{1}^T \Sigma^{-1} \begin{bmatrix} (\mu_1 - \text{sign}(w_{1*})p\sigma_1) \\ \vdots \\ (\mu_n - \text{sign}(w_{n*})p\sigma_n) \end{bmatrix} - V \mathbb{1}^T \Sigma^{-1} \mathbb{1} \right) \right.$$

$$\left(-\Sigma^{-1} \begin{bmatrix} \text{sign}(w_{1*})\sigma_1 \\ \vdots \\ \text{sign}(w_{n*})\sigma_n \end{bmatrix} + \left(\sum_{i=1}^{n} |w_{i*}| \sigma_i \right) \Sigma^{-1} \mathbb{1} \right)$$

$$- \left(\Sigma^{-1} \begin{bmatrix} (\mu_1 - \text{sign}(w_{1*})p\sigma_1) \\ \vdots \\ (\mu_n - \text{sign}(w_{n*})p\sigma_n) \end{bmatrix} - V\Sigma^{-1} \mathbb{1} \right)$$

$$\left(-\mathbb{1}^T \Sigma^{-1} \begin{bmatrix} \text{sign}(w_{1*})\sigma_1 \\ \vdots \\ \text{sign}(w_{n*})\sigma_n \end{bmatrix} + \left(\sum_{i=1}^{n} |w_{i*}| \sigma_i \right) \mathbb{1}^T \Sigma^{-1} \mathbb{1} \right) \right)$$

$$\left(\mathbb{1}^T \Sigma^{-1} \begin{bmatrix} (\mu_1 - \text{sign}(w_{1*})p_1\sigma_1) \\ \vdots \\ (\mu_n - \text{sign}(w_{n*})p_n\sigma_n) \end{bmatrix} - V \mathbb{1}^T \Sigma^{-1} \mathbb{1} \right)^{-2}$$

by using Theorem 6.10 and the Envelope Theorem A.24, so that

$$
\frac{d\mathbf{w}_*}{dp}\bigg|_{p=0} = \left(\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)\left(-\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix} + \left(\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)\Sigma^{-1}\mathbb{1}\right)\right.
$$
$$
\left. - \left(\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\Sigma^{-1}\mathbb{1}\right)\left(-\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix} + \left(\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)\right)
$$
$$
\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)^2
$$
$$
= \left(\left(\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}\right.
$$
$$
+ \bar{V}\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1} + \Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix}
$$
$$
\left. - \left(\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1} - \bar{V}\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix}\right)
$$
$$
\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}
$$

and

$$
\frac{dDr(\mathbf{w}_*)}{dp}\bigg|_{p=0} = \left(\left(\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}\left(\sigma^2\Sigma^{-1}\mathbb{1} - 2\right)\right.
$$
$$
- \mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}\left(\sigma^2\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix} - 2\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)
$$
$$
+ \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\left(\sigma^2\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix} - 2\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)
$$
$$
+ \mathbb{1}^{\mathrm{T}}\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix}\left(\sigma^2\Sigma^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^{\mathrm{T}}\bar{\mathbf{w}}\right)
$$
$$
- \left(\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\left(\sigma^2\Sigma^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^{\mathrm{T}}\bar{\mathbf{w}}\right)
$$
$$
\left. - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\begin{bmatrix} \mathrm{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \mathrm{sign}(\bar{w}_n)\sigma_n \end{bmatrix}\left(\sigma^2\Sigma^{-1}\mathbb{1} - 2\right)\right)
$$

119

$$\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}$$

$$= \left(\left(\left(\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)A\left(B-2\right) - \left(C - 2\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)\right.\right.$$

$$+ \bar{V}D\left(C - 2\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right) + E\left(F - 2\boldsymbol{\mu}^{\mathrm{T}}\bar{\mathbf{w}}\right)$$

$$\left.- \left(\sum_{i=1}^{n}|\bar{w}_i|\sigma_i\right)D\left(F - 2\boldsymbol{\mu}^{\mathrm{T}}\bar{\mathbf{w}}\right) - \bar{V}E\left(B-2\right)\right)$$

$$\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}.$$

Since a necessary and sufficient condition for a fraction to be positive is that its numerator is greater than or equal to zero, (6.10) follows immediately. $\qquad\square$

6.2.2 Eigenvalue Uncertainty

We consider the effects of eigenvalue uncertainty on the portfolio diversification return. To obtain analytical results later, we do not let w_i to be ℓ_i, so that the robust location-scale problem (4.3) can be rewritten as

$$\max_{w_i\in\mathbb{R}\backslash\{\ell_i\},i=1,\ldots,n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\left(\sum_{i=1}^{n}(\lambda_i + b_i)\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbf{w}} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell}\right\}. \tag{6.11}$$

Note that the impact of removing ℓ_i from the feasible domain of w_i is negligible due to continuity of the original problem and the optimal solution of (6.11) can be very well approximated in practice by solving

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\left(\sum_{i=1}^{n}(\lambda_i + b_i)\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbf{w}} : \begin{array}{l}\mathbf{w}^{\mathrm{T}}\mathbb{1} = 1,\\ \mathbf{w} \geq \boldsymbol{\ell} + \delta\mathbb{1}\end{array}\right\}$$

where δ is a small positive number. Like in the case of location uncertainty, we state a theorem providing us with the optimal solution of (6.11) in analytical form.

Theorem 6.13

Let V be the optimal value of (6.11), then its optimal solution is

$$\mathbf{w}_* = \frac{\left(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu} - V\left(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}}{\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu} - V\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}}.$$

Proof: Note that solving (6.11) is equivalent to solving

$$\max_{x_i \in \mathbb{R} \setminus \{0\}, i=1,\ldots,n} \left\{ \begin{array}{l} [|x_1| + \ell_1, \ldots, |x_n| + \ell_n]\,\boldsymbol{\mu} \\[6pt] -\kappa(\epsilon)\sqrt{[|x_1| + \ell_1, \ldots, |x_n| + \ell_n]\Big(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\Big)\begin{bmatrix} |x_1| + \ell_1 \\ \vdots \\ |x_n| + \ell_n \end{bmatrix}} \\[6pt] : [|x_1| + \ell_1, \ldots, |x_n| + \ell_n]^{\mathrm{T}}\,\mathbb{1} = 1 \end{array} \right\} \quad (6.12)$$

for $\begin{bmatrix} |x_1| + \ell_1 \\ \vdots \\ |x_n| + \ell_n \end{bmatrix}$. Let $\mathbf{x}_*$ be the optimal solution of (6.12). Since x_i cannot be 0, the objective and constraint functions are continuously differentiable so that there exists a Lagrange multiplier ν_* which satisfies

$$\begin{bmatrix} \text{sign}(x_{1*})\mu_1 \\ \vdots \\ \text{sign}(x_{n*})\mu_n \end{bmatrix} - \kappa(\epsilon)\frac{\text{diag}(\text{sign}(x_{1*}),\ldots,\text{sign}(x_{n*}))\Big(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\Big)\begin{bmatrix} |x_{1*}| + \ell_1 \\ \vdots \\ |x_{n*}| + \ell_n \end{bmatrix}}{\sqrt{[|x_{1*}| + \ell_1, \ldots, |x_{n*}| + \ell_n]\Big(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\Big)\begin{bmatrix} |x_{1*}| + \ell_1 \\ \vdots \\ |x_{n*}| + \ell_n \end{bmatrix}}}$$

$$- \nu_*\begin{bmatrix} \text{sign}(x_{1*}) \\ \vdots \\ \text{sign}(x_{n*}) \end{bmatrix} = \mathbf{0}_{n\times 1}.$$

$$(6.13)$$

and

$$[|x_{1*}| + \ell_1, \ldots, |x_{n*}| + \ell_n]^{\mathrm{T}}\,\mathbb{1} = 1 \qquad (6.14)$$

Multiplying (6.13) on the left by $[x_{1*} + \text{sign}(x_{1*})\ell_1, \ldots, x_{n*} + \text{sign}(x_{n*})\ell_n]$, we obtain

$$[|x_{1*}| + \ell_1, \ldots, |x_{n*}| + \ell_n]\,\boldsymbol{\mu}$$

$$- \kappa(\epsilon)\sqrt{[|x_{1*}| + \ell_1, \ldots, |x_{n*}| + \ell_n]\left(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\begin{bmatrix} |x_{1*}| + \ell_1 \\ \vdots \\ |x_{n*}| + \ell_n \end{bmatrix}}$$

$$- \nu_*[|x_{1*}| + \ell_1, \ldots, |x_{n*}| + \ell_n]^{\mathrm{T}}\,\mathbb{1} = 0$$

which implies

$$\nu_* = V$$

by (6.14) and the fact that both (6.11) and (6.12) have the same optimal value. Multiplying (6.13) on the left by $\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\text{diag}(\text{sign}(x_{1*}),\ldots,\text{sign}(x_{n*}))$ obtains

$$\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^{n}(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu}$$

$$
-\kappa(\epsilon)\frac{\mathbb{1}^{\mathrm{T}}\begin{bmatrix}|x_{1*}|+\ell_1\\\vdots\\|x_{n*}|+\ell_n\end{bmatrix}}{\sqrt{[|x_{1*}|+\ell_1,\ldots,|x_{n*}|+\ell_n]\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\begin{bmatrix}|x_{1*}|+\ell_1\\\vdots\\|x_{n*}|+\ell_n\end{bmatrix}}}
$$

$$
-\nu_*\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}=0
$$

$$
\Rightarrow\quad \kappa(\epsilon)=\frac{\left(\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu}-\nu_*\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}\right)}{\sqrt{[|x_{1*}|+\ell_1,\ldots,|x_{n*}|+\ell_n]\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\begin{bmatrix}|x_{1*}|+\ell_1\\\vdots\\|x_{n*}|+\ell_n\end{bmatrix}}}
$$

by using (6.14) again. Substituting the above into (6.13) before multiplying the resulting equation on the left by $\Sigma^{-1}\mathrm{diag}(\mathrm{sign}(x_{1*}),\ldots,\mathrm{sign}(x_{n*}))$ result in

$$
\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu}-\nu_*\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}
$$

$$
-\begin{bmatrix}|x_{1*}|+\ell_1\\\vdots\\|x_{n*}|+\ell_n\end{bmatrix}\left(\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu}-\nu_*\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}\right)=0
$$

so that rearranging we get

$$
\begin{bmatrix}|x_{1*}|+\ell_1\\\vdots\\|x_{n*}|+\ell_n\end{bmatrix}=\frac{\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu}-\nu_*\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}}{\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu}-\nu_*\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n(\lambda_i+b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}}.
$$

It remains to substitute $\mathbf{w}_*=\begin{bmatrix}|x_{1*}|+\ell_1\\\vdots\\|x_{n*}|+\ell_n\end{bmatrix}$ and $\nu_*=V$ into the above equation to obtain the desired result. $\qquad\square$

6.2.2.1 Individual Eigenvalue Uncertainty

We are now ready to provide a necessary and sufficient condition (6.15) in the following theorem for the diversification return of the portfolio obtained by solving

$$
\max_{w_i\in\mathbb{R}\setminus\{\ell_i\},i=1,\ldots,n}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}-\kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}:\mathbf{w}^{\mathrm{T}}\mathbb{1}=1,\mathbf{w}\geq\boldsymbol{\ell}\right\}
$$

to increase when the ith eigenvalue of the scale matrix changes slightly.

Theorem 6.14

Let $\mathbf{w}_*$ be the optimal solution of

$$\mathbf{w}_* = \frac{\left(\sum_{i=1}^n (\lambda_i + b_i)^{-1} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu} - V\left(\sum_{i=1}^n (\lambda_i + b_i)^{-1} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}}{\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n (\lambda_i + b_i)^{-1} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}\right)\boldsymbol{\mu} - V\mathbb{1}^{\mathrm{T}}\left(\sum_{i=1}^n (\lambda_i + b_i)^{-1} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}\right)\mathbb{1}}.$$

Then, a necessary and sufficient condition for $\left.\frac{dDr(\mathbf{w}_*)}{db_j}\right|_{\mathbf{b}=0}$ to be positive is

$$A\bar{V}^2 + B\bar{V} + C > 0 \tag{6.15}$$

where

$$A = \left(\sigma^2 - 2\bar{\mathbf{w}}^{\mathrm{T}}\Sigma\right)\left(\frac{\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1} - \mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}}{\lambda_j^2}\right),$$

$$B = \left(\sigma^2 - 2\bar{\mathbf{w}}^{\mathrm{T}}\Sigma\right)\left(\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} + \mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1} - \Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1} - \Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_j^2}\right),$$

$$C = \left(\sigma^2 - 2\bar{\mathbf{w}}^{\mathrm{T}}\Sigma\right)\left(\frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu} - \mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_j^2}\right.$$

$$\left. + \left.\frac{dV}{db_j}\right|_{\mathbf{b}=0}\left(\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1} - \Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}\right)\right).$$

Proof: First, note that w_{i*} is differentiable with respect to b_j for all $i, j = 1, \ldots, n$ by using arguments similar to those found in the proof of Theorem 4.6. Therefore, the derivative of the diversification return of the portfolio obtained by solving (6.11) with respect to b_j is

$$\frac{dDr(\mathbf{w}_*)}{db_j} = \left(\left[\sum_{i=1}^n (\lambda_i + b_i)u_{i1}^2, \ldots, \sum_{i=1}^n (\lambda_i + b_i)u_{in}^2\right] - 2\mathbf{w}_*^{\mathrm{T}}\sum_{i=1}^n (\lambda_i + b_i)\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\frac{d\mathbf{w}_*}{db_j}$$

where

$$\frac{d\mathbf{w}_*}{db_j} = \left(\left(\sum_{i=1}^n \frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_i + b_i} - V\sum_{i=1}^n \frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i + b_i}\right)\left(-\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}}{(\lambda_j + b_j)^2} + V\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}}{(\lambda_j + b_j)^2} - \frac{dV}{db_j}\sum_{i=1}^n \frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i + b_i}\right)\right.$$

$$\left. - \left(\sum_{i=1}^n \frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_i + b_i} - V\sum_{i=1}^n \frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i + b_i}\right)\left(-\frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}}{(\lambda_j + b_j)^2} + V\frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}}{(\lambda_j + b_j)^2} - \frac{dV}{db_j}\sum_{i=1}^n \frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i + b_i}\right)\right)$$

$$\left(\sum_{i=1}^n \frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_i + b_i} - V\sum_{i=1}^n \frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i + b_i}\right)^{-2}.$$

such that

$$\frac{dV}{db_j} = -\frac{\kappa(\epsilon)(\mathbf{u}_j^{\mathrm{T}}\mathbf{w}_*)^2}{\sqrt{\sum_{i=1}^n (\lambda_i + b_i)(\mathbf{u}_i^{\mathrm{T}}\mathbf{w}_*)^2}},$$

by using the Envelope Theorem A.24. Thus, we have

$$
\begin{aligned}
\frac{d\mathbf{w}_*}{db_j}\bigg|_{\mathbf{b}=0} =& \Bigg(\Big(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\Big)\Big(-\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_j^2} + \bar{V}\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}}{\lambda_j^2} - \frac{dV}{db_j}\bigg|_{\mathbf{b}=0}\Sigma^{-1}\mathbb{1}\Big) \\
& -\Big(\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\Sigma^{-1}\mathbb{1}\Big)\Big(-\frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_j^2} + \bar{V}\frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}}{\lambda_j^2} - \frac{dV}{db_j}\bigg|_{\mathbf{b}=0}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\Big)\Bigg) \\
& \Big(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\Big)^{-2} \\
=& \Bigg(-\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_j^2} + \bar{V}\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_j^2} - \frac{dV}{db_j}\bigg|_{\mathbf{b}=0}\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} \\
& +\bar{V}\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}}{\lambda_j^2} - \bar{V}^2\frac{\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}}{\lambda_j^2} + \frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_j^2} \\
& -\bar{V}\frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}}{\lambda_j^2} + \frac{dV}{db_j}\bigg|_{\mathbf{b}=0}\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1} - \bar{V}\frac{\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_j^2} \\
& +\bar{V}^2\frac{\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbb{1}}{\lambda_j^2}\Bigg)\Big(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\Big)^{-2}
\end{aligned}
$$

so that

$$
\begin{aligned}
\frac{dDr(\mathbf{w}_*)}{db_j}\bigg|_{\mathbf{b}=0} &= \Big(\sigma^2 - 2\bar{\mathbf{w}}^{\mathrm{T}}\Sigma\Big)\Big(\frac{d\mathbf{w}_*}{db_j}\bigg|_{\mathbf{b}=0}\Big) \\
&= \Big(A\bar{V}^2 + B\bar{V} + C\Big)\Big(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\Big)^{-2}.
\end{aligned}
$$

The theorem follows by noting that a necessary and sufficient condition for the above expression to be positive is for its numerator to be as well. □

6.2.2.2 Simultaneous Eigenvalue Uncertainty

We now provide a necessary and sufficient condition (6.16) in the following theorem for the diversification return of the portfolio obtained by solving

$$
\max_{w_i \in \mathbb{R}\setminus\{\ell_i\}, i=1,\dots,n}\Big\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}} : \mathbf{w}^{\mathrm{T}}\mathbb{1} = 1, \mathbf{w} \geq \boldsymbol{\ell}\Big\}
$$

to increase when the each eigenvalue of the scale matrix changes slightly.

Theorem 6.15

Let $\mathbf{w}_*$ be the optimal solution of

$$
\mathbf{w}_* = \frac{\Big(\sum_{i=1}^n(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\Big)\boldsymbol{\mu} - V\Big(\sum_{i=1}^n(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\Big)\mathbb{1}}{\mathbb{1}^{\mathrm{T}}\Big(\sum_{i=1}^n(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\Big)\boldsymbol{\mu} - V\mathbb{1}^{\mathrm{T}}\Big(\sum_{i=1}^n(\lambda_i + b_i)^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\Big)\mathbb{1}}
$$

where $b_i = b$ for $i = 1,\ldots,n$. Then, a necessary and sufficient condition for $\left.\frac{dDr(\mathbf{w}_*)}{db}\right|_{b=0}$ to be positive is

$$A\bar{V}^2 + B\bar{V} + C > 0 \tag{6.16}$$

where

$$A = \left(\sigma^2 - \bar{\mathbf{w}}^{\mathsf{T}}\Sigma\right)\left(\sum_{i=1}^{n} \frac{\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1} - \mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}}{\lambda_i^2}\right),$$

$$B = \left(\sigma^2 - 2\bar{\mathbf{w}}^{\mathsf{T}}\Sigma\right)\left(\sum_{i=1}^{n} \frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} + \mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1} - \Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1} - \Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{\lambda_i^2}\right),$$

$$C = \left(\sigma^2 - 2\bar{\mathbf{w}}^{\mathsf{T}}\Sigma\right)\left(\sum_{i=1}^{n} \frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu} - \mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_i^2}\right.$$
$$\left. + \left.\frac{dV}{db}\right|_{b=0}\left(\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1} - \Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}\right)\right).$$

Proof: First, note that w_{i*} is differentiable with respect to b for $i = 1,\ldots,n$ by using arguments similar to those found in the proof of Theorem 4.6. Therefore, we have that

$$\frac{dDr(\mathbf{w}_*)}{db} = \left(\left[\sum_{i=1}^{n}(\lambda_i + b)u_{i1}^2,\ldots,\sum_{i=1}^{n}(\lambda_i + b)u_{in}^2\right] - 2\mathbf{w}_*^{\mathsf{T}}\sum_{i=1}^{n}(\lambda_i + b)\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\right)\frac{d\mathbf{w}_*}{db}$$

where

$$\frac{d\mathbf{w}_*}{db} = \left(\left(\sum_{i=1}^{n} \frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{\lambda_i + b} - V\sum_{i=1}^{n} \frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i + b}\right)\left(\sum_{i=1}^{n} -\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{(\lambda_i + b)^2} + V\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{(\lambda_i + b)^2} - \frac{dV}{db}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i + b}\right)\right.$$
$$\left. - \left(\sum_{i=1}^{n} \frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{\lambda_i + b} - V\sum_{i=1}^{n} \frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i + b}\right)\left(\sum_{i=1}^{n} -\frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{(\lambda_i + b)^2} + V\frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{(\lambda_i + b)^2} - \frac{dV}{db}\frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i + b_i}\right)\right)$$
$$\left(\sum_{i=1}^{n} \frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{\lambda_i + b} - V\sum_{i=1}^{n} \frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i + b}\right)^{-2}$$

such that

$$\frac{dV}{db} = -\frac{\kappa(\epsilon)\sum_{i=1}^{n}(\mathbf{u}_i^{\mathsf{T}}\mathbf{w}_*)^2}{\sqrt{\sum_{i=1}^{n}(\lambda_i + b)(\mathbf{u}_i^{\mathsf{T}}\mathbf{w}_*)^2}},$$

by using the Envelope Theorem A.24. Thus, we have

$$\left.\frac{d\mathbf{w}_*}{db}\right|_{b=0} = \left(\left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)\left(\sum_{i=1}^{n}\left(-\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{\lambda_i^2} + \bar{V}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i^2}\right) - \left.\frac{dV}{db}\right|_{b=0}\Sigma^{-1}\mathbb{1}\right)\right.$$
$$\left. - \left(\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\Sigma^{-1}\mathbb{1}\right)\left(\sum_{i=1}^{n}\left(-\frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{\lambda_i^2} + \bar{V}\frac{\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i^2}\right) - \left.\frac{dV}{db}\right|_{b=0}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)\right)$$

$$\left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}$$

$$= \Bigg(-\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_i^2} + \bar{V}\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_i^2} - \frac{dV}{db}\bigg|_{b=0}\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}$$

$$+ \bar{V}\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}}{\lambda_i^2} - \bar{V}^2\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}}{\lambda_i^2} + \sum_{i=1}^{n}\frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{\lambda_i^2}$$

$$- \bar{V}\sum_{i=1}^{n}\frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i^2} + \frac{dV}{db}\bigg|_{b=0}\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1} - \bar{V}\sum_{i=1}^{n}\frac{\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\boldsymbol{\mu}}{\lambda_i^2}$$

$$+ \bar{V}^2\sum_{i=1}^{n}\frac{\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbb{1}}{\lambda_i^2}\Bigg)\left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}$$

so that

$$\frac{dDr(\mathbf{w}_*)}{db}\bigg|_{b=0} = \left(\boldsymbol{\sigma}^2 - 2\bar{\mathbf{w}}^{\mathsf{T}}\Sigma\right)\left(\frac{d\mathbf{w}_*}{db}\bigg|_{b=0}\right)$$

$$= \left(A\bar{V}^2 + B\bar{V} + C\right)\left(\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathsf{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}.$$

The theorem follows immediately by noting that a necessary and sufficient condition for the above expression to be positive is for its numerator to be as well. $\qquad\square$

6.2.3 Location and Eigenvalue Uncertainties

We consider the effects of both location and eigenvalue uncertainties on the portfolio diversification return. To obtain analytical results later, we do not let w_i to be either 0 or ℓ_i, so that the robust location-scale problem (4.3) can be rewritten as

$$\max_{w_i\in\mathbb{R}\backslash\{0,\ell_i\},i=1,\dots,n}\left\{\mathbf{w}^{\mathsf{T}}\boldsymbol{\mu} - \sum_{i=1}^{n}|w_i|a_i - \kappa(\epsilon)\sqrt{\sum_{i=1}^{n}(\lambda_i + b_i)(\mathbf{w}^{\mathsf{T}}\mathbf{u}_i)^2} : \mathbf{w}^{\mathsf{T}}\mathbb{1} = 1, \mathbf{w}\geq\boldsymbol{\ell}\right\}. \quad (6.17)$$

Note that the impact of removing 0 and ℓ_i from the feasible domain of w_i is negligible due to the continuity of the original problem. The optimal solution of (6.17) can be very well approximated in practice by solving

$$\max_{\mathbf{w}\in\mathbb{R}^n}\left\{\mathbf{w}^{\mathsf{T}}\boldsymbol{\mu} - \sum_{i=1}^{n}|w_i|a_i - \kappa(\epsilon)\sqrt{\sum_{i=1}^{n}(\lambda_i + b_i)(\mathbf{w}^{\mathsf{T}}\mathbf{u}_i)^2} : \begin{array}{c}\mathbf{w}^{\mathsf{T}}\mathbb{1} = 1, \\ \boldsymbol{\ell} + \delta\mathbb{1}\leq\mathbf{w}\leq-\delta\mathbb{1}, \\ \mathbf{w}\geq\delta\mathbb{1}\end{array}\right\}$$

where δ is a small positive number. We state a theorem providing us with the optimal solution of (6.11) in analytical form.

Theorem 6.16

The optimal solution of (6.17) *is*

$$
\mathbf{w}_* = \frac{\sum_{i=1}^{n} \frac{\mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}}{\lambda_i + b_i} \left(\begin{bmatrix} (\mu_1 - \mathrm{sign}(w_{1*})a_1) \\ \vdots \\ (\mu_n - \mathrm{sign}(w_{n*})a_n) \end{bmatrix} - V\mathbb{1} \right)}{\sum_{i=1}^{n} \frac{\mathbb{1}^{\mathrm{T}} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}}{\lambda_i + b_i} \left(\begin{bmatrix} (\mu_1 - \mathrm{sign}(w_{1*})a_1) \\ \vdots \\ (\mu_n - \mathrm{sign}(w_{n*})a_n) \end{bmatrix} - V\mathbb{1} \right)}
$$

where V is the associated optimal value.

Proof: The proof is completely analogous with that of Theorem 6.10 with the covariance matrix replaced by its eigendecomposition. □

Subsequently, we provide a necessary and sufficient condition (6.18) in the following theorem for the diversification return of the portfolio obtained by solving

$$
\max_{w_i \in \mathbb{R} \setminus \{0, \ell_i\}, i=1,\ldots,n} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} - \kappa(\epsilon) \sqrt{\mathbf{w}^{\mathrm{T}} \Sigma \mathbf{w}} : \mathbf{w}^{\mathrm{T}} \mathbb{1} = 1, \mathbf{w} \geq \ell \right\}
$$

to increase when each asset location and each eigenvalue of the scale matrix changes slightly.

Theorem 6.17

Let $\mathbf{w}_*$ be the optimal solution of

$$
\mathbf{w}_* = \frac{\sum_{i=1}^{n} \frac{\mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}}{\lambda_i + b_i} \left(\begin{bmatrix} (\mu_1 - \mathrm{sign}(w_{1*})a_1) \\ \vdots \\ (\mu_n - \mathrm{sign}(w_{n*})a_n) \end{bmatrix} - V\mathbb{1} \right)}{\sum_{i=1}^{n} \frac{\mathbb{1}^{\mathrm{T}} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}}{\lambda_i + b_i} \left(\begin{bmatrix} (\mu_1 - \mathrm{sign}(w_{1*})a_1) \\ \vdots \\ (\mu_n - \mathrm{sign}(w_{n*})a_n) \end{bmatrix} - V\mathbb{1} \right)}
$$

$a_i = p\sigma_i$ *and* $b_i = p$ *for* $i = 1, \ldots, n$. *Then, a necessary and sufficient condition for* $\left. \frac{d Dr(\mathbf{w}_*)}{dp} \right|_{p=0}$ *to be positive is*

$$
A\bar{V}^2 + B\bar{V} + C > 0, \tag{6.18}
$$

where

$$
A = \left(\sigma^2 - \bar{\mathbf{w}}^{\mathrm{T}} \Sigma \right) \left(\sum_{i=1}^{n} \frac{\Sigma^{-1} \mathbb{1} \mathbb{1}^{\mathrm{T}} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \mathbb{1} - \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \mathbb{1} \mathbb{1}^{\mathrm{T}} \Sigma^{-1} \mathbb{1}}{\lambda_i^2} \right),
$$

$$
B = \left(\sigma^2 - 2\bar{\mathbf{w}}^{\mathrm{T}} \Sigma \right) \left(\sum_{i=1}^{n} \frac{\mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \mathbb{1} \mathbb{1}^{\mathrm{T}} \Sigma^{-1} \boldsymbol{\mu} + \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \boldsymbol{\mu} \mathbb{1}^{\mathrm{T}} \Sigma^{-1} \mathbb{1} - \Sigma^{-1} \boldsymbol{\mu} \mathbb{1}^{\mathrm{T}} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \mathbb{1} - \Sigma^{-1} \mathbb{1} \mathbb{1}^{\mathrm{T}} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \boldsymbol{\mu}}{\lambda_i^2} \right.
$$

$$+\Sigma^{-1}\begin{bmatrix} \text{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \text{sign}(\bar{w}_n)\sigma_n \end{bmatrix}\mathbb{1}^{\text{T}}\Sigma^{-1}\mathbb{1} - \Sigma^{-1}\mathbb{1}\mathbb{1}^{\text{T}}\Sigma^{-1}\begin{bmatrix} \text{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \text{sign}(\bar{w}_n)\sigma_n \end{bmatrix}\Bigg),$$

$$C = \left(\sigma^2 - 2\bar{\mathbf{w}}^{\text{T}}\Sigma\right)\Bigg(\sum_{i=1}^{n}\frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\text{T}}\mathbf{u}_i\mathbf{u}_i^{\text{T}}\boldsymbol{\mu} - \mathbf{u}_i\mathbf{u}_i^{\text{T}}\boldsymbol{\mu}\mathbb{1}^{\text{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_i^2}$$

$$+\frac{dV}{dp}\bigg|_{p=0}\left(\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\text{T}}\Sigma^{-1}\mathbb{1} - \Sigma^{-1}\mathbb{1}\mathbb{1}^{\text{T}}\Sigma^{-1}\boldsymbol{\mu}\right) + \Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\text{T}}\Sigma^{-1}\begin{bmatrix} \text{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \text{sign}(\bar{w}_n)\sigma_n \end{bmatrix}$$

$$-\Sigma^{-1}\begin{bmatrix} \text{sign}(\bar{w}_1)\sigma_1 \\ \vdots \\ \text{sign}(\bar{w}_n)\sigma_n \end{bmatrix}\mathbb{1}^{\text{T}}\Sigma^{-1}\boldsymbol{\mu}\Bigg).$$

Proof: First, note that w_{i*} is differentiable with respect to p for $i = 1,\ldots,n$ by using arguments similar to those found in the proof of Theorem 4.6. Therefore, we have that

$$\frac{dDr(\mathbf{w}_*)}{dp} = \left(\left[\sum_{i=1}^{n}(\lambda_i + p)u_{i1}^2,\ldots,\sum_{i=1}^{n}(\lambda_i + p)u_{in}^2\right] - 2\mathbf{w}_*^{\text{T}}\sum_{i=1}^{n}(\lambda_i + p)\mathbf{u}_i\mathbf{u}_i^{\text{T}}\right)\frac{d\mathbf{w}_*}{dp}$$

where

$$\frac{d\mathbf{w}_*}{dp} = \Bigg(\Bigg(\sum_{i=1}^{n}\frac{\mathbb{1}^{\text{T}}\mathbf{u}_i\mathbf{u}_i^{\text{T}}}{\lambda_i + p}\left(\begin{bmatrix} \mu_1 - \text{sign}(w_{1*})p\sigma_1 \\ \vdots \\ \mu_n - \text{sign}(w_{n*})p\sigma_n \end{bmatrix} - V\mathbb{1}\right)\Bigg)$$

$$\left(\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\text{T}}}{\lambda_i + p}\left(\begin{bmatrix} -\text{sign}(w_{1*})\sigma_1 \\ \vdots \\ -\text{sign}(w_{n*})\sigma_n \end{bmatrix} - \frac{dV}{dp}\mathbb{1}\right)\right)$$

$$-\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\text{T}}}{(\lambda_i + p)^2}\left(\begin{bmatrix} \mu_1 - \text{sign}(w_{1*})p\sigma_1 \\ \vdots \\ \mu_n - \text{sign}(w_{n*})p\sigma_n \end{bmatrix} - V\mathbb{1}\right)\Bigg)$$

$$-\left(\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\text{T}}}{\lambda_i + p}\left(\begin{bmatrix} \mu_1 - \text{sign}(w_{1*})p\sigma_1 \\ \vdots \\ \mu_n - \text{sign}(w_{n*})p\sigma_n \end{bmatrix} - V\mathbb{1}\right)\right)$$

$$\left(\sum_{i=1}^{n}\frac{\mathbb{1}^{\text{T}}\mathbf{u}_i\mathbf{u}_i^{\text{T}}}{\lambda_i + p}\left(\begin{bmatrix} -\text{sign}(w_{1*})\sigma_1 \\ \vdots \\ -\text{sign}(w_{n*})\sigma_n \end{bmatrix} - \frac{dV}{dp}\mathbb{1}\right)\right)$$

$$-\sum_{i=1}^{n}\frac{\mathbb{1}^{\text{T}}\mathbf{u}_i\mathbf{u}_i^{\text{T}}}{(\lambda_i + p)^2}\left(\begin{bmatrix} \mu_1 - \text{sign}(w_{1*})p\sigma_1 \\ \vdots \\ \mu_n - \text{sign}(w_{n*})p\sigma_n \end{bmatrix} - V\mathbb{1}\right)\Bigg)\Bigg)$$

$$\left(\sum_{i=1}^{n}\frac{\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}}{\lambda_i+b_i}\left(\left[\begin{array}{c}(\mu_1-\operatorname{sign}(w_{1*})a_1)\\ \vdots\\ (\mu_n-\operatorname{sign}(w_{n*})a_n)\end{array}\right]-V\mathbb{1}\right)\right)^{-2}$$

such that

$$\frac{dV}{dp}=-\sum_{i=1}^{n}|w_{i*}|\sigma_i-\frac{\kappa(\epsilon)\sum_{i=1}^{n}(\mathbf{u}_i^{\mathrm{T}}\mathbf{w}_*)^2}{\sqrt{\sum_{i=1}^{n}(\lambda_i+p)(\mathbf{u}_i^{\mathrm{T}}\mathbf{w}_*)^2}},$$

by using the Envelope Theorem A.24. Thus, we have

$$\begin{aligned}
\frac{d\mathbf{w}_*}{dp}\bigg|_{p=0}=&\Bigg(\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}-\bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)\left(-\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_i^2}+\bar{V}\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i^2}\right.\\
&-\Sigma^{-1}\left[\begin{array}{c}\operatorname{sign}(\bar{w}_1)\sigma_1\\ \cdots\\ \operatorname{sign}(\bar{w}_n)\sigma_n\end{array}\right]-\frac{dV}{dp}\bigg|_{p=0}\Sigma^{-1}\mathbb{1}\Bigg)\\
&-\left(\Sigma^{-1}\boldsymbol{\mu}-\bar{V}\Sigma^{-1}\mathbb{1}\right)\left(-\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_i^2}+\bar{V}\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i^2}\right.\\
&-\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\left[\begin{array}{c}\operatorname{sign}(\bar{w}_1)\sigma_1\\ \cdots\\ \operatorname{sign}(\bar{w}_n)\sigma_n\end{array}\right]-\frac{dV}{dp}\bigg|_{p=0}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\Bigg)\Bigg)\\
&\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}-\bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}\\
=&\Bigg(-\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_i^2}+\bar{V}\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}}{\lambda_i^2}-\frac{dV}{dp}\bigg|_{p=0}\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}\\
&+\bar{V}\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}}{\lambda_i^2}-\bar{V}^2\sum_{i=1}^{n}\frac{\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}}{\lambda_i^2}+\sum_{i=1}^{n}\frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_i^2}\\
&-\bar{V}\sum_{i=1}^{n}\frac{\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i^2}+\frac{dV}{dp}\bigg|_{p=0}\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}-\bar{V}\sum_{i=1}^{n}\frac{\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\mu}}{\lambda_i^2}\\
&+\bar{V}^2\sum_{i=1}^{n}\frac{\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbb{1}}{\lambda_i^2}-\Sigma^{-1}\left[\begin{array}{c}\operatorname{sign}(\bar{w}_1)\sigma_1\\ \vdots\\ \operatorname{sign}(\bar{w}_n)\sigma_n\end{array}\right]\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}\\
&+\bar{V}\Sigma^{-1}\left[\begin{array}{c}\operatorname{sign}(\bar{w}_1)\sigma_1\\ \vdots\\ \operatorname{sign}(\bar{w}_n)\sigma_n\end{array}\right]\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}+\Sigma^{-1}\boldsymbol{\mu}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\left[\begin{array}{c}\operatorname{sign}(\bar{w}_1)\sigma_1\\ \vdots\\ \operatorname{sign}(\bar{w}_n)\sigma_n\end{array}\right]\\
&-\bar{V}\Sigma^{-1}\mathbb{1}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\left[\begin{array}{c}\operatorname{sign}(\bar{w}_1)\sigma_1\\ \vdots\\ \operatorname{sign}(\bar{w}_n)\sigma_n\end{array}\right]\Bigg)\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}-\bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}
\end{aligned}$$

so that

$$\frac{dDr(\mathbf{w}_*)}{dp}\bigg|_{p=0} = \left(\boldsymbol{\sigma}^2 - 2\bar{\mathbf{w}}^{\mathrm{T}}\Sigma\right)\left(\frac{d\mathbf{w}_*}{dp}\bigg|_{p=0}\right)$$
$$= \left(A\bar{V}^2 + B\bar{V} + C\right)\left(\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \bar{V}\mathbb{1}^{\mathrm{T}}\Sigma^{-1}\mathbb{1}\right)^{-2}.$$

The theorem follows immediately by noting that a necessary and sufficient condition for the above expression to be positive is for its numerator to be as well. □

6.2.4 Eigenvector Uncertainty

Finding conditions for the diversification return of the portfolio obtained by solving the location-scale problem to increase if the eigenvectors of the scale matrix Σ change direction slightly with their orthogonality preserved is non-trivial, although there is an expression (6.19) to describe the behavior of this change in the following theorem, which can be easily modified to include location uncertainty and/or eigenvalue uncertainty.

Theorem 6.18
Denote $\bar{\mathbf{w}}$ the optimal solution of

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)\sqrt{\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}}\right\}.$$

Let $(\mathbf{w}_*, \tau_*, y_*)$ and $(\mathbf{Z}_*\nu_*, \upsilon_*)$ be the optimal solution of

$$\max_{(\mathbf{w},\tau,y)\in\mathcal{W}_\ell\times\mathbb{R}_+\times\mathbb{R}}\left\{\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} - \kappa(\epsilon)y : \left[\begin{array}{ccc|c} & & & \mathbf{w}^{\mathrm{T}}\mathbf{P}_1^{\mathrm{T}} \\ & (y+\tau(1-c)^2)\mathbf{I}_n - \tau\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} & & \vdots \\ & & & \mathbf{w}^{\mathrm{T}}\mathbf{P}_n^{\mathrm{T}} \\ \hline \mathbf{P}_1\mathbf{w} & \cdots & \mathbf{P}_n\mathbf{w} & \mathrm{diag}(\lambda_1^{-1},\ldots,\lambda_n^{-1})y \end{array}\right] \succeq 0, y \geq \delta\right\}$$

and its Lagrange dual problem

$$\min_{(\mathbf{Z},\nu,\upsilon)\in\mathbb{S}_+^{2n}\times\mathbb{R}\times\mathbb{R}_+^{n+2}}\left\{-\nu - \sum_{i=1}^{n}\upsilon_i - \delta\upsilon_{n+2} : \begin{array}{l}\mu_i + \nu + \upsilon_i + tr(\mathbf{F}_i\mathbf{Z}) = 0, \ i = 1,\ldots,n \\ \upsilon_{n+1} + tr(\mathbf{G}\mathbf{Z}) = 0, \\ \upsilon_{n+2} - \kappa(\epsilon) + tr(\mathbf{H}\mathbf{Z}) = 0,\end{array}\right\}$$

respectively, where $\mathbf{w}_*$ is also the optimal solution of the robust location-scale problem

$$\max_{\mathbf{w}\in\mathcal{W}_\ell}\min_{\mathbf{u}\in\mathbb{R}^n}\left\{\mathbf{w}^T\boldsymbol{\nu} - \kappa(\epsilon)\sqrt{\sum_{i=1}^{n}\lambda_i((\mathbf{u}/\|\mathbf{u}\|_2)^{\mathrm{T}}\mathbf{P}_i\mathbf{w})^2} : (\mathbf{u}/\|\mathbf{u}\|_2)^{\mathrm{T}}\mathbf{e}_1 \geq 1 - c\right\}$$

using Theorem 3.5.

    Denote $\mathbf{z}_*$ the vectorized form of the upper half of $\mathbf{Z}_*$. Assume $[\mathbf{z}_*, \upsilon_{n+1*}]^{\mathrm{T}}$, $[\mathbf{z}_*, \upsilon_{n+2*}]^{\mathrm{T}}$ and each column of the matrix on the left-hand side of the semi-definite constraint in (4.26) evaluated at $(\mathbf{w}_*, \tau_*, y_*)$ are all non-zero vectors. If there exists a feasible solution of (4.26) such that the

(matrix) inequality constraints hold strictly, then the derivative of the diversification return of the portfolio obtained by solving (4.26) *evaluated at $c = 0$ is*

$$\frac{dDr(\mathbf{w}_*)}{dc}\bigg|_{c=0} = (\boldsymbol{\sigma}^2 - 2\bar{\mathbf{w}}^{\mathrm{T}}\Sigma)\frac{d\mathbf{w}_*}{dc}\bigg|_{c=0} \tag{6.19}$$

where $\frac{d\mathbf{w}_*}{dc}$ is the first n entries of

$$-\begin{bmatrix} \frac{\partial f_1}{\partial w_{1*}} \cdots \frac{\partial f_1}{\partial w_{n*}} \frac{\partial f_1}{\partial \tau_*} \frac{\partial f_1}{\partial y_*} \frac{\partial f_1}{\partial Z_{11*}} \cdots \frac{\partial f_1}{\partial Z_{1,2n*}} \cdots \cdots \frac{\partial f_1}{\partial Z_{2n,1*}} \cdots \frac{\partial f_1}{\partial Z_{2n,2n*}} \frac{\partial f_1}{\partial v_*} \frac{\partial f_1}{\partial v_{1*}} \cdots \frac{\partial f_1}{\partial v_{n+2*}} \\ \vdots \\ \frac{\partial f_N}{\partial w_{1*}} \cdots \frac{\partial f_N}{\partial w_{n*}} \frac{\partial f_N}{\partial \tau_*} \frac{\partial f_N}{\partial y_*} \frac{\partial f_N}{\partial Z_{11*}} \cdots \frac{\partial f_N}{\partial Z_{1,2n*}} \cdots \cdots \frac{\partial f_N}{\partial Z_{2n,1*}} \cdots \frac{\partial f_N}{\partial Z_{2n,2n*}} \frac{\partial f_N}{\partial v_*} \frac{\partial f_N}{\partial v_{1*}} \cdots \frac{\partial f_N}{\partial v_{n+2*}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1}{\partial c} \\ \vdots \\ \frac{\partial f_N}{\partial c} \end{bmatrix} \tag{6.20}$$

provided the inverse exists, such that each of $f_1,\ldots,f_{N=4n^2+2n+5}$ represents the left-hand side of one of the equations

$$\left(\begin{bmatrix} (y_* + \tau_*(1-c)^2)\mathbf{I}_n - \tau_*\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} & \begin{array}{c} \mathbf{w}_*^{\mathrm{T}}\mathbf{P}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{w}_*^{\mathrm{T}}\mathbf{P}_n^{\mathrm{T}} \end{array} \\ \hline \mathbf{P}_1\mathbf{w}_* \quad \cdots \quad \mathbf{P}_n\mathbf{w}_* & \mathrm{diag}(\bar{\lambda}_1^{-1},\ldots,\bar{\lambda}_n^{-1})y_* \end{bmatrix}\mathbf{Z}_*\right)_{ij} = 0, \ i,j = 1,\ldots,2n, \tag{6.21}$$

$$\mu_i + \nu_* + \upsilon_{i*} + tr\left(\begin{bmatrix} \mathbf{0}_{n\times n} & \begin{array}{c} \mathbf{p}_{1i}^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_{ni}^{\mathrm{T}} \end{array} \\ \hline \mathbf{p}_{1i} \quad \cdots \quad \mathbf{p}_{ni} & \mathbf{0}_{n\times n} \end{bmatrix}\mathbf{Z}_*\right) = 0, \ i = 1,\ldots,n, \tag{6.22}$$

$$\upsilon_{n+1*} + tr\left(\begin{bmatrix} (1-c)^2\mathbf{I}_n - \mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} & \mathbf{0}_{n\times n} \\ \hline \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \end{bmatrix}\mathbf{Z}_*\right) = 0, \tag{6.23}$$

$$\upsilon_{n+2*} - \kappa(\epsilon) + tr\left(\begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n\times n} \\ \hline \mathbf{0}_{n\times n} & \mathrm{diag}(\bar{\lambda}_1,\ldots,\bar{\lambda}_n) \end{bmatrix}\mathbf{Z}_*\right) = 0, \tag{6.24}$$

$$\mathbf{w}_*^{\mathrm{T}}\mathbb{1} - 1 = 0, \tag{6.25}$$

$$\upsilon_{i*}w_{i*} = 0, \ i = 1,\ldots,n, \tag{6.26}$$

$$\upsilon_{n+1*}\tau_* = 0, \tag{6.27}$$

$$\upsilon_{n+2*}y_* = 0. \tag{6.28}$$

Proof: Due to the same reasons as in the proof of Theorem 4.9, the KKT conditions (6.21)-(6.28) hold, and $(\mathbf{w}_*, \tau_*, y_*)$ and $\mathbf{Z}_*$ are continuously differentiable with respect to c. Since the partial derivative of (6.23) with respect to υ_{n+1*} is one, and the Jacobian of the left-hand side of (6.22)-(6.23) with respect to (υ_{n+1*}, ν_*), the Jacobian of the left-hand side of (6.22)-(6.23) with respect to $(\upsilon_{n+1*}, \upsilon_{i*})$ for $i = 1,\ldots,n$, and the Jacobian of the left-hand side of (6.23)-(6.24) with respect to $(\upsilon_{n+1*}, \upsilon_{n+2*})$ are all non-singular, we have that ν_* and

v_* are continuously differentiable with respect to c as well by the Implicit Function Theorem A.19. Thus, differentiating each equation in (6.21)-(6.28) with respect to c obtains

$$
\begin{bmatrix}
\frac{\partial f_1}{\partial w_{1*}} \cdots \frac{\partial f_1}{\partial w_{n*}} \frac{\partial f_1}{\partial \tau_*} \frac{\partial f_1}{\partial y_*} \frac{\partial f_1}{\partial Z_{11*}} \cdots \frac{\partial f_1}{\partial Z_{1,2n*}} \cdots\cdots \frac{\partial f_1}{\partial Z_{2n,1*}} \cdots \frac{\partial f_1}{\partial Z_{2n,2n*}} \frac{\partial f_1}{\partial v_*} \frac{\partial f_1}{\partial v_{1*}} \cdots \frac{\partial f_1}{\partial v_{n+2*}} \\
\vdots \\
\frac{\partial f_N}{\partial w_{1*}} \cdots \frac{\partial f_N}{\partial w_{n*}} \frac{\partial f_N}{\partial \tau_*} \frac{\partial f_N}{\partial y_*} \frac{\partial f_N}{\partial Z_{11*}} \cdots \frac{\partial f_N}{\partial Z_{1,2n*}} \cdots\cdots \frac{\partial f_N}{\partial Z_{2n,1*}} \cdots \frac{\partial f_N}{\partial Z_{2n,2n*}} \frac{\partial f_N}{\partial v_*} \frac{\partial f_N}{\partial v_{1*}} \cdots \frac{\partial f_N}{\partial v_{n+2*}}
\end{bmatrix}
$$

$$
\begin{bmatrix} \frac{dw_{1*}}{dc} \cdots \frac{dw_{n*}}{dc} \frac{d\tau_*}{dc} \frac{dy_*}{dc} \frac{dZ_{11*}}{dc} \cdots \frac{dZ_{1,2n*}}{dc} \cdots\cdots \frac{dZ_{2n,1*}}{dc} \cdots \frac{dZ_{2n,2n*}}{dc} \frac{dv_*}{dc} \frac{dv_{1*}}{dc} \cdots \frac{dv_{n+2*}}{dc} \end{bmatrix}^{\mathrm{T}}
$$

$$
+ \begin{bmatrix} \frac{\partial f_1}{\partial c} \\ \vdots \\ \frac{\partial f_N}{\partial c} \end{bmatrix} = \mathbf{0}_{N \times 1},
$$

from which it follows that $\frac{d\mathbf{w}_*}{dc}$ is indeed the first n entries of (6.20) provided the inverse exists. The proof concludes by noting (6.19) as an obvious fact. $\qquad\square$

Chapter 7

Conclusions

In Chapter 1, we give the motivation of our thesis and introduce the Markowitz model before showing that it is equivalent to the portfolio return value-at-risk optimization problem if we assume the risky asset returns follow a multivariate normal distribution, in which case a solution of the latter results in a portfolio expected return and standard deviation that lies exactly on a point of the Markowitz efficient frontier. If a riskless asset is added and the amounts to be invested in the risky assets and riskless asset are predetermined exogeneously, then the optimal portfolio with riskless asset has an expected return and standard deviation that falls on the line joining the point represented by the expected return and standard deviation of the optimal portfolio without riskless asset and $(0, \mu_0)$, where μ_0 is the risk-free asset return. If the amount to be invested in the risky assets and riskless asset are to be determined endogeneously, then three things can happen. First, if the optimal portfolio value-at-risk is lower than the risk-free asset return, then all wealth is kept in the riskless asset. Second, if the optimal portfolio value-at-risk is higher than the risk-free asset return, then all wealth is placed on the risky assets. Third, if the optimal portfolio value-at-risk is equal to the risk-free asset return, then the allocation of wealth has to be decided exogeneously. We also mention the failure of the Markowitz model to include model uncertainty and higher moments, but since it is equivalent to the portfolio return value-at-risk optimization problem under multivariate normality of risky asset returns, these features can be added into the former implicitly by incorporating them into the latter. However, doing so generally causes the resulting problem to be non-convex and existing methods in the literature to overcome this issue are too conservative and/or intractable.

In Chapter 2, we introduce a spline approximation method where the minimal sample value-at-risk or any other risk measure is smoothed via a quadratic B-spline, which can then be maximized. Gaivoronski and Pflug (2005) introduce a similar method, but theirs does not take into account model uncertainty. The cardinality of the discretization set of the feasible domain and the number of basis parameters to be estimated only increases polynomially and linearly respectively with the number of assets. Simulation results are reasonably accurate in the two-dimensional setting.

In Chapter 3, we show that the robust portfolio return value-at-risk optimization problem under elliptical distributions possesses a location-scale form. We introduce the box and ellipsoidal uncertainty sets for the location vector, where in particular under the former

uncertainty set the location-scale problem is equivalent to an SOCP. We introduce the box, ellipsoidal, correlation coefficient, specific portfolio scale and one-factor model uncertainty sets for the scale matrix, under any of which uncertainty sets the location-scale problem either is very conservative, requires a positive definite constraint on the scale matrix (since the uncertainty set does not live in the positive definite space), has a high computational complexity or is in any combination of the three situations just mentioned. A novel eigen-decomposition uncertainty set for the scale matrix is then introduced where the eigenvalues vary in a box uncertainty set and the eigenvectors each varies in a cone uncertainty set with orthogonality preserved among them, such that the scale matrix lives in the positive definite space and all its other entries are determined by fixing any one of them, thus greatly reducing conservativeness. Although the robust location-scale problem with the eigende-composition uncertainty set for the scale matrix is non-convex in general, we can convert it into an SDP which is solvable in polynomial time.

In Chapter 4, we introduce a scale invariant method to determine the size of the box uncertainty set for the location vector, the box uncertainty set for the eigenvalues and the cone uncertainty set for each eigenvector, based on a chosen level of sensitivity. We perform some numerical experiments using data obtained from Nasdaq and NYSE under different combinations of these uncertainty sets, where the median sensitivity level is always chosen and the portfolio return value-at-risk is maximized assuming that the returns follow a multivariate normal distribution and several other comparable elliptical distributions. In particular, we see that with uncertainty, the move towards a less diversified portfolio (in the sense that fewer assets are included) is more gradual as the value-at-risk level increases.

In Chapter 5, we include trading costs and integer constraints so that the robust location-scale optiimization problem involving any combination of the location, eigenvalue and eigen-vector uncertainty sets from the previous chapter can be reformulated as a mixed integer program. Numerical experiments analogous to those conducted previously are performed where similar conclusions can be drawn.

In Chapter 6, we introduce the portfolio RQE as a unifying diversification measure, and interpret the robust location-scale problem as maximizing the diversification return (a special case of the portfolio RQE) combined with other measures so that the asset allo-cation performance is improved, under imperfect information of the model. We also find expressions for the sensitivity of the diversification return of the optimal portfolio obtained by solving the location-scale problem to various types of uncertainty, and in some cases provide conditions for them to be positive.

In conclusion, we provide a probabilistic method and a deterministic model within the location-scale framework to robustly optimize the portfolio return value-at-risk, with the flexibility of extending to other risk measures. In the future, we hope to move outside the location-scale framework without compromising on computational complexity and conser-vativeness, as well as consider the multi-period setting. On a final note, what we present in this thesis is only our humble opinion of robust portfolio optimization, which for sure is a burgeoning field of research with many more exciting years to come.

Appendix A

A.1 Mathematical Background

Definition A.5
*A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a **norm** if*

   (i) *f is non-negative: $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$,*

   (ii) *f is definite: $f(\mathbf{x}) = 0$ only if $\mathbf{x} = \mathbf{0}$,*

   (iii) *f is homogeneous: $f(t\mathbf{x}) = |t| f(\mathbf{x}) \; \forall t \in \mathbb{R}, \, \mathbf{x}, \in \mathbb{R}^n$, and*

   (iv) *f satisfies the triangle inequality: $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.*

Remark A.5
*We use the notation $\|\cdot\|$ to represent a norm hereafter. The **distance** between two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as the norm of its difference $\|\mathbf{x} - \mathbf{y}\|$. Furthermore, if $\|\cdot\|_a$ and $\|\cdot\|_b$ represent two different norms defined on $\mathbf{R}^n$, then there exists α and β such that for all $\mathbf{x} \in \mathbb{R}^n$,*

$$\alpha \|\mathbf{x}\|_a \leq \|\mathbf{x}\|_b \leq \beta \|\mathbf{x}\|_a.$$

*Therefore, we say that all the norms of any finite-dimensional Euclidean space are **equivalent**.*

Definition A.6
*A set $\mathcal{C} \subseteq \mathbb{R}^n$ is **convex** if for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and $\theta \in [0, 1]$, then $\theta \mathbf{x} + (1 - \theta)\mathbf{y} \in \mathcal{C}$.*

Definition A.7
*A **cone** is a set $\mathcal{C} \subseteq \mathbb{R}^n$ such that if $\mathbf{x} \in \mathcal{C}$ and $\theta \geq 0$, then $\theta \mathbf{x} \in \mathcal{C}$.*

Definition A.8
*An element $\mathbf{x} \in \mathcal{C} \subseteq \mathbb{R}^n$ is an **interior point** if there exists an $\epsilon > 0$ such that all points whose Euclidean distance is less than or equal to ϵ from $\mathbf{x}$ lies completely in $\mathcal{C}$, that is,*

$$\{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\|_2 \leq \epsilon\} \subseteq \mathcal{C}.$$

*The **interior** of $\mathcal{C}$, which we denote as $\mathrm{int}(\mathcal{C})$, is the set of the interior points of $\mathcal{C}$.*

Remark A.6

Due to the equivalence of norms in the Euclidean space, all norms generate the same set of interior points. Therefore, we choose the Euclidean norm $\|\cdot\|_2$ without loss of generality.

Definition A.9

*An set $\mathcal{C} \subseteq \mathbb{R}^n$ is **open** if every of its points is in its interior, that is, $\mathrm{int}(\mathcal{C}) = \mathcal{C}$.*

Definition A.10

*A set $\mathcal{C} \subseteq \mathbb{R}^n$ is **closed** if its complement $\mathbf{R}^n \backslash \mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \notin \mathcal{C}\}$ is open.*

Remark A.7

Note that we can also define closed sets in terms of convergent sequence and limit points. A set $\mathcal{C} \subseteq \mathbb{R}^n$ is closed if and only if $\mathbf{x} \in \mathbb{R}^n$ for any sequence of points $\mathbf{x}_1, \mathbf{x}_2, \ldots$ which converges to $\mathbf{x}$.

Definition A.11

*The **affine hull** of a set $\mathcal{C} \subseteq \mathbb{R}^n$ is denoted $\mathbf{aff}\ \mathcal{C}$ and defined as*

$$\mathbf{aff}\ \mathcal{C} = \left\{ \sum_{i=1}^{k} \theta_i \mathbf{x}_i : \sum_{i=1}^{k} \theta_i = 1 \right\}.$$

Definition A.12

*The **relative interior** of a set $\mathcal{C}$ is denoted $\mathbf{relint}\ \mathcal{C}$ and defined as*

$$\{\mathbf{x} : \mathcal{B}(\mathbf{x}, r) \cap \mathbf{aff}\ \mathcal{C} \subseteq \mathcal{C} \text{ for some } r \geq 0\}.$$

where $\mathcal{B}(\mathbf{x}, r) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\|_2 \leq r\}$.

Remark A.8

Analogous to the definition of the interior of a set, any norm on $\mathbf{R}^n$ defines the same relative interior, so that we choose the Euclidean norm $\|\cdot\|_2$ without loss of generality.

Theorem A.19 (Implicit Function Theorem)

Let $f_1, \ldots, f_m : \mathbb{R}^{m+n} \to \mathbb{R}$ be continuously differentiable functions. Consider the system of equations

$$f_1(\mathbf{y}, \mathbf{x}) = c_1$$
$$\vdots \tag{A.1}$$
$$f_m(\mathbf{y}, \mathbf{x}) = c_m$$

where $\mathbf{y} = [y_1, \ldots, y_m]^\mathrm{T}$ a $\mathbf{x} = [x_1, \ldots, x_n]^\mathrm{T}$. Suppose $\mathbf{y}_* = [y_{1*}, \ldots, y_{m*}]^\mathrm{T}$ and $\mathbf{x}_* = [x_{1*}, \ldots, x_{n*}]^\mathrm{T}$ form a solution of (A.1), and the determinant of

$$\frac{\partial(f_1, \ldots, f_m)}{\partial(y_1, \ldots, y_m)} := \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{bmatrix}$$

evaluated at $(\mathbf{y}_*, \mathbf{x}_*)$ *is non-vanishing, then there exist continuously differentiable functions* $g_1(\mathbf{x}), \ldots, g_m(\mathbf{x})$ *such that*

$$f_1(g_1(\mathbf{x}), \ldots, g_m(\mathbf{x}), \mathbf{x}) = c_1$$

$$\vdots$$

$$f_m(g_1(\mathbf{x}), \ldots, g_m(\mathbf{x}), \mathbf{x}) = c_m$$

for all $\mathbf{x}$ *in a ball centered around* $\mathbf{x}_*$, *and*

$$y_{1*} = g_1(\mathbf{x}_*)$$

$$\vdots$$

$$y_{m*} = g_m(\mathbf{x}_*).$$

In addition, $\frac{\partial g_k}{\partial x_h}(\mathbf{x}_*)$ *can be computed by letting* $dx_h = 1$ *and* $dx_j = 0, j \neq h$ *in*

$$\frac{\partial f_1}{\partial y_1} dy_1 + \ldots + \frac{\partial f_1}{\partial y_m} dy_m + \frac{\partial f_1}{\partial x_1} dx_1 + \ldots + \frac{\partial f_1}{\partial x_n} dx_n = 0$$

$$\vdots$$

$$\frac{\partial f_m}{\partial y_1} dy_1 + \ldots + \frac{\partial f_m}{\partial y_m} dy_m + \frac{\partial f_m}{\partial x_1} dx_1 + \ldots + \frac{\partial f_m}{\partial x_n} dx_n = 0$$

and solving for dy_k, *where each partial derivative is evaluated at* $(g_1(\mathbf{x}_*), \ldots, g_m(\mathbf{x}_*), \mathbf{x}_*)$.

Remark A.9
Refer to [183, Chapter 15] for a discussion on the Implicit Function Theorem.

Lemma A.1 ($\mathcal{S}$-Lemma)

(i) *(homogeneous version) Let* $\mathbf{A}$, $\mathbf{B}$ *be symmetric matrices of the same size such that* $\mathbf{x}^\mathsf{T} \mathbf{A} \mathbf{x} > 0$ *for some* $\mathbf{x}$. *Then*

$$\mathbf{x}^\mathsf{T} \mathbf{A} \mathbf{x} \geq 0 \Rightarrow \mathbf{x}^\mathsf{T} \mathbf{B} \mathbf{x} \geq 0$$

holds if and only if

$$\exists \lambda \geq 0 : \mathbf{B} \succeq \lambda \mathbf{A}.$$

(ii) *(inhomogeneous version) Let* $\mathbf{A}$, $\mathbf{B}$ *be symmetric matrices of the same size and the quadratic form*

$$\mathbf{x}^\mathsf{T} \mathbf{A} \mathbf{x} + 2\mathbf{a}^\mathsf{T} \mathbf{x} + \alpha \Rightarrow \mathbf{x}^\mathsf{T} \mathbf{B} \mathbf{x} + 2\mathbf{b}^\mathsf{T} \mathbf{x} + \beta \geq 0$$

holds if and only if

$$\exists \lambda \geq 0 : \left[\begin{array}{c|c} \mathbf{B} - \lambda \mathbf{A} & \mathbf{b} - \lambda \mathbf{a} \\ \hline \mathbf{b}^\mathsf{T} - \lambda \mathbf{a}^\mathsf{T} & \beta - \lambda \alpha \end{array} \right] \succeq 0.$$

Proof: Refer to [30, section 4.3.5].   □

Lemma A.2 (Schur Complement Lemma)

A symmetric block matrix

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{P} & \mathbf{Q}^T \\ \hline \mathbf{Q} & \mathbf{R} \end{array} \right]$$

with $\mathbf{R}$ *being positive (semi)-definite if and only if*

$$\mathbf{P} - \mathbf{Q}^T \mathbf{R}^{-1} \mathbf{Q}$$

is positive (semi)-definite.

Proof: $\mathbf{A}$ is positive semi-definite if and only if

$$\forall \mathbf{u}, \mathbf{v} : \mathbf{u}^\mathsf{T} \mathbf{P} \mathbf{u} + 2 \mathbf{u}^\mathsf{T} \mathbf{Q}^\mathsf{T} \mathbf{v} + \mathbf{v}^\mathsf{T} \mathbf{R} \mathbf{v} \geq 0$$
$$\Leftrightarrow \quad \forall \mathbf{u} : \min_{\mathbf{v}} \left\{ \mathbf{u}^\mathsf{T} \mathbf{P} \mathbf{u} + 2 \mathbf{u}^\mathsf{T} \mathbf{Q}^\mathsf{T} \mathbf{v} + \mathbf{v}^\mathsf{T} \mathbf{R} \mathbf{v} \right\} \geq 0. \tag{A.2}$$

Since $\mathbf{R}$ is positive semi-definite, the optimal value of the above optimization problem with an objective function of quadratic form occurs exactly when the first-order condition is satisfied, that is, when $\mathbf{v} = \mathbf{R}^{-1} \mathbf{Q} \mathbf{u}$ which, upon substituting into (A.2) obtains

$$\forall \mathbf{u} : \mathbf{u}^\mathsf{T} (\mathbf{P} - \mathbf{Q}^\mathsf{T} \mathbf{R}^{-1} \mathbf{Q}) \mathbf{u} \geq 0,$$

equivalent to the positive semi-definiteness of $\mathbf{P} - \mathbf{Q}^\mathsf{T} \mathbf{R}^{-1} \mathbf{Q}$. The same argument applies for the positive definite case.   □

Theorem A.20

Let $\mathbf{x} \in \mathbb{R}^n$ *and* $y, z \in \mathbb{R}_+$. *Then*

$$\mathbf{x}^\mathsf{T} \mathbf{x} \leq yz$$

if and only if

$$\left\| \begin{bmatrix} 2\mathbf{x} \\ y - z \end{bmatrix} \right\|_2 \leq y + z.$$

where $\| \cdot \|_2$ *represents the Euclidean norm.*

Proof: If y and z are non-negative, then

$$\left\| \begin{bmatrix} 2\mathbf{x} \\ y - z \end{bmatrix} \right\|_2 \leq y + z$$
$$\Leftrightarrow \quad \left\| \begin{bmatrix} 2\mathbf{x} \\ y - z \end{bmatrix} \right\|_2^2 \leq (y + z)^2$$
$$\Leftrightarrow \quad 4\mathbf{x}^\mathsf{T} \mathbf{x} + (y - z)^2 \leq y^2 + z^2 + 2yz$$
$$\Leftrightarrow \quad \mathbf{x}^\mathsf{T} \mathbf{x} \leq yz.$$

□

A.2 Generalized Inequality Constrained Optimization

We first state two definitions which we will use in this section.

Definition A.13
*A **non-strict partial ordering** over a set $\mathcal{C} \subseteq \mathbb{R}^n$, denoted as $\preceq_{\mathcal{C}}$, is a binary relation such that $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{C}$ satisfy the following properties:*

(i) $\mathbf{x} \preceq_{\mathcal{C}} \mathbf{x}$ *(reflexivity)*,

(ii) *if $\mathbf{x} \preceq_{\mathcal{C}} \mathbf{y}$ and $\mathbf{y} \preceq_{\mathcal{C}} \mathbf{x}$, then $\mathbf{x} = \mathbf{y}$ (antisymmetry)*,

(iii) *if $\mathbf{x} \preceq_{\mathcal{C}} \mathbf{y}$ and $\mathbf{y} \preceq_{\mathcal{C}} \mathbf{z}$, then $\mathbf{x} \preceq_{\mathcal{C}} \mathbf{z}$ (transitivity)*.

Definition A.14
*A **strict partial ordering** over a set $\mathcal{C} \subseteq \mathbb{R}^n$, denoted as $\prec_{\mathcal{C}}$, is a binary relation such that $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{K}$ satisfy the following properties:*

(i) $\mathbf{x} \prec_{\mathcal{C}} \mathbf{x}$ *(irreflexivity)*,

(ii) *if $\mathbf{x} \prec_{\mathcal{C}} \mathbf{y}$, then $\mathbf{y} \nprec_{\mathcal{C}} \mathbf{x}$, (asymmetry)*,

(iii) *if $\mathbf{x} \prec_{\mathcal{C}} \mathbf{y}$ and $\mathbf{y} \prec_{\mathcal{C}} \mathbf{z}$, then $\mathbf{x} \prec_{\mathcal{C}} \mathbf{z}$ (transitivity)*.

A.2.1 Generalized Inequality Constraints

Definition A.15
*A cone $\mathcal{K} \subseteq \mathbb{R}^n$ is **proper** if it is convex, closed, solid (has nonempty interior), and pointed (contains no line, meaning that if $\mathbf{x}, -\mathbf{x} \in \mathcal{K}$, then $\mathbf{x} = \mathbf{0}$).*

Definition A.16
*A **generalized non-strict inequality** associated with a proper cone $\mathcal{K} \in \mathbb{R}^n$ is the non-strict partial ordering defined by*

$$\mathbf{x} \preceq_{\mathcal{K}} \mathbf{y} \Leftrightarrow \mathbf{y} - \mathbf{x} \in \mathcal{K}.$$

Definition A.17
*A **generalized strict inequality** associated with a proper cone $\mathcal{K} \in \mathbb{R}^n$ is the strict partial ordering defined by*

$$\mathbf{x} \prec_{\mathcal{K}} \mathbf{y} \Leftrightarrow \mathbf{y} - \mathbf{x} \in \text{int}(\mathcal{K}).$$

Remark A.10
We also write $\mathbf{y} \succeq_{\mathcal{K}} \mathbf{x}$ and $\mathbf{y} \succ_{\mathcal{K}} \mathbf{x}$ for $\mathbf{x} \preceq_{\mathcal{K}} \mathbf{y}$ and $\mathbf{x} \prec_{\mathcal{K}} \mathbf{y}$ respectively.

A.2.2 Duality

Definition A.18

The set

$$\mathcal{K}_* = \{\mathbf{y} : \mathbf{x}^\mathsf{T}\mathbf{y} \geq 0 \ \forall \mathbf{x} \in \mathcal{K}\}$$

*where $\mathcal{K}$ is a cone is called the **dual cone** of $\mathcal{K}$.*

Remark A.11

Note that $\mathcal{K}_*$ is always convex, even though it may not be so for $\mathcal{K}$.

Definition A.19

*A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{K}$-**convex** if*

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \preceq_\mathcal{K} \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y}).$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\theta \in [0,1]$, where $\mathcal{K} \subseteq \mathbb{R}^m$ is a proper cone associated with the generalized inequality $\preceq_\mathcal{K}$.

Definition A.20

Consider the constrained optimization problem with generalized inequality constraints

$$\max_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}) : g_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{0}, h_j(\mathbf{x}) = 0, i = 1, \ldots, m, j = 1, \ldots, p \right\} \tag{A.3}$$

with $\mathcal{K}_i \subseteq \mathbb{R}^{c_i}$ being a proper cone and a non-empty domain $\mathcal{D} = \mathbf{dom}\, f \cap_{i=1}^m \mathbf{dom}\, g_i \cap_{j=1}^p \mathbf{dom}\, h_j$. The associated Lagrangian is defined as

$$L(\mathbf{x}, \upsilon_1, \ldots, \upsilon_m, \boldsymbol{\nu}) = f(\mathbf{x}) - \sum_{i=1}^m \upsilon_i^\mathsf{T} g_i(\mathbf{x}) - \sum_{j=1}^p \nu_j h_j(\mathbf{x}),$$

*such that $\upsilon_i \in \mathbb{R}^{c_i}$ and $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_p]$. The **Lagrange dual function** is defined as*

$$d(\upsilon_1, \ldots, \upsilon_m, \boldsymbol{\nu}) = \sup_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \upsilon_1, \ldots, \upsilon_m, \boldsymbol{\nu}).$$

Remark A.12

Since $d(\upsilon_1, \ldots, \upsilon_m, \boldsymbol{\nu})$ is a pointwise supremum of the Lagrangian which is affine in $(\upsilon_1, \ldots, \upsilon_m, \boldsymbol{\nu})$, it is always convex.[1]

Definition A.21

*The **Lagrange dual problem** of (A.3) is*

$$\min_{\upsilon_1, \ldots, \upsilon_m \in \prod_{i=1}^m \mathbb{R}^{c_i}} \left\{ d(\upsilon_1, \ldots, \upsilon_m, \boldsymbol{\nu}) : \upsilon_i \succeq_{\mathcal{K}_{i*}} \mathbf{0}, i = 1, \ldots, m \right\}. \tag{A.4}$$

---

[1]It is a well-known fact that if $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for each $\mathbf{y}$ in an arbitrary set $\mathcal{A}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{A}} f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$.

Definition A.22
Let p_* and d_* be the optimal values of (A.3) *and* (A.4) *respectively. If $p_* \leq d_*$, we say that **weak duality** holds. If $p_* = d_*$, we say that **strong duality** holds, or there is **zero duality gap** between the primal and dual problems* (A.3) *and* (A.4) *respectively.*

Theorem A.21
Weak duality always holds for (A.3).

Proof: Since $v_i \succeq_{\mathcal{K}_{i*}} \mathbf{0}$ and $g_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{0}$ for any primal and dual feasible $\mathbf{x}$ and v_i respectively, we have $v_i^{\mathsf{T}} g_i(\mathbf{x}) \leq 0$, so that

$$f(\mathbf{x}) \leq f(\mathbf{x}) - \sum_{i=1}^{m} v_i^{\mathsf{T}} g_i(\mathbf{x}) - \sum_{j=1}^{p} v_j h_i(\mathbf{x})$$

due to the third term on the right being zero. Taking the supremum on both sides over $\mathbf{x}$ yields

$$p_* \leq d(v_1, \ldots, v_m, \nu),$$

from which the result follows immediately. $\square$

Definition A.23
If there exists an $\mathbf{x} \in \mathbf{relint}\, \mathcal{D}$ of (A.3) *such that $g_i(\mathbf{x}) \preceq_{\mathcal{K}_i}$ and $\mathbf{0}, h_j(\mathbf{x}) = 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, p$, then we say that **Slater's condition** holds.*

Theorem A.22
If f is convex, g_i is $\mathcal{K}_i$-convex and Slater's condition is satisfied for (A.3), *then **strong duality** holds.*

Proof: Refer to [170, p. 47]. $\square$

Theorem A.23 (KKT Optimality Conditions)
Let $\mathbf{x}_*$ and $(v_{1*}, \ldots, v_{m*}, \nu_*)$ be the primal and dual optimal solutions of (A.3) *and* (A.4) *respectively with zero duality gap. In addition, assume that g_i and h_j are differentiable for $i = 1, \ldots, m$ and $j = 1, \ldots, p$. Then the following conditions, called the Karush-Kuhn-Tucker (KKT) conditions, are satisfied:*

(i) $g_i(\mathbf{x}_*) \preceq_{\mathcal{K}_i} \mathbf{0}$, $i = 1, \ldots, m$,

(ii) $h_j(\mathbf{x}_*) = 0$, $j = 1, \ldots, p$,

(iii) $v_{i*} \succeq_{\mathcal{K}_{i*}} \mathbf{0}$, $i = 1, \ldots, m$,

(iv) $v_{i*}^{\mathsf{T}} g_i(\mathbf{x}_*) = 0$, $i = 1, \ldots, m$, and

(v) $\nabla f(\mathbf{x}_*) + \sum_{i=1}^{m} Dg_i^{\mathsf{T}}(\mathbf{x}_*) v_i^* + \sum_{j=1}^{p} \nu_{j*} \nabla h_j(\mathbf{x}_*) = \mathbf{0}$.

Proof: The first three conditions follow straightaway from the definitions of the primal and dual problems. For the fourth and fifth conditions, note that

$$f(\mathbf{x}_*) = d(\boldsymbol{\upsilon}_{1*}, \ldots, \boldsymbol{\upsilon}_{m*}, \boldsymbol{\nu}_*)$$

$$\geq f(\mathbf{x}_*) - \sum_{i=1}^{m} \boldsymbol{\upsilon}_{i*}^{\mathrm{T}} g_i(\mathbf{x}_*) - \sum_{j=1}^{p} \nu_{i*} h_i(\mathbf{x}_*) \tag{A.5}$$

$$\geq f(\mathbf{x}_*),$$

where the equality is due to the zero duality gap, the first inequality is by definition of the Lagrange dual problem, and the second inequality is due to the fact that $h_i(\mathbf{x}_*) = 0$ and $\boldsymbol{\upsilon}_{i*}^{\mathrm{T}} g_i(\mathbf{x}_*) \leq 0$, with the latter being the case because $\boldsymbol{\upsilon}_* \succeq_{\mathcal{K}_{i*}} \mathbf{0}$ and $g_i(\mathbf{x}_*) \preceq_{\mathcal{K}_i}$ so that $\boldsymbol{\upsilon}_{i*}^{\mathrm{T}} g_i(\mathbf{x}_*) \leq 0$ by definition of the dual cone. This implies that

$$f(\mathbf{x}_*) - \sum_{i=1}^{m} \boldsymbol{\upsilon}_{i*}^{\mathrm{T}} g_i(\mathbf{x}_*) - \sum_{j=1}^{p} \nu_{i*} h_i(\mathbf{x}_*) = f(\mathbf{x}_*),$$

and that we must have $\sum_{i=1}^{m} \boldsymbol{\upsilon}_{i*}^{\mathrm{T}} g_i(\mathbf{x}_*) = 0$, where each summand is non-positive, from which we conclude the fourth condition. We can also draw from (A.5) that $\mathbf{x}_*$ minimizes $L(\mathbf{x}, \boldsymbol{\upsilon}_{1*}, \ldots, \boldsymbol{\upsilon}_{m*}, \boldsymbol{\nu}_*)$, from which the last condition follows. $\qquad\square$

A.3 Envelope Theorems

Theorem A.24
[142] Consider the optimization problem

$$\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$$

where $\mathcal{X}$ be the set of feasible solutions and $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a parameterized objective function. Let $\mathbf{x}_* = [x_{1*}, \ldots, x_{n*}]^{\mathrm{T}}$ be an optimal solution and assume that the optimal value $V(\theta) = f(\mathbf{x}_*, \theta)$ is differentiable at θ, that is,

$$\frac{dV}{d\theta} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_*, \theta) \frac{dx_{i*}}{d\theta} + \frac{\partial f}{\partial \theta}(\mathbf{x}_*, \theta)$$

exists, then

$$\frac{dV}{d\theta} = \frac{\partial f}{\partial \theta}(\mathbf{x}_*, \theta),$$

where $\frac{\partial f}{\partial x_i}(\mathbf{x}_*, \theta)$ and $\frac{\partial f}{\partial x_i}(\mathbf{x}_*, \theta)$ are partial derivatives with respect to x_i and θ evaluated at $\mathbf{x}_*$ respectively.

Proof: Since V is differentiable, we have that

$$\frac{dV}{d\theta} = \frac{dV}{d\theta_-} = \frac{dV}{d\theta_+}. \tag{A.6}$$

Now by the definitions of $V(\theta)$ and $\mathcal{X}_*(\theta)$, it is obvious that

$$f(\mathbf{x}_*, \theta') - f(\mathbf{x}_*, \theta) \le V(\theta') - V(\theta) \tag{A.7}$$

for any $\theta' > 0$. Assuming $\theta' - \theta > 0$, dividing both sides of (A.7) by it and taking their limits as $\theta' \to \theta_+$, we obtain

$$\lim_{\theta' \to \theta_+} \frac{f(\mathbf{x}_*, \theta') - f(\mathbf{x}_*, \theta)}{\theta' - \theta} \le \lim_{\theta' \to \theta_+} \frac{V(\theta') - V(\theta)}{\theta' - \theta}$$

$$\frac{\partial f}{\partial \theta}(x_*, \theta) \le \frac{dV}{d\theta_+}.$$

Analogously, assuming $\theta' - \theta < 0$, dividing both sides of (A.7) by it and taking their limits as $\theta' \to \theta_-$ yields $\frac{\partial f}{\partial \theta}(\mathbf{x}_*, \theta) \ge \frac{dV}{d\theta_-}$. This implies

$$\frac{dV}{d\theta_-} \le \frac{\partial f}{\partial \theta}(\mathbf{x}_*, \theta) \le \frac{dV}{d\theta_+}$$

$$\Rightarrow \quad \frac{dV}{d\theta} \le \frac{\partial f}{\partial \theta}(\mathbf{x}_*, \theta) \le \frac{dV}{d\theta} \tag{by (A.6)}$$

$$\Rightarrow \quad \frac{dV}{d\theta} = \frac{\partial f}{\partial \theta}(\mathbf{x}_*, \theta).$$

$\square$

Remark A.13

The above is the Envelope Theorem for an optimzation problem with a parameterized objective function.

Theorem A.25

Let $\mathbf{x}_*$ be an optimal solution of the primal problem

$$\max_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}) : g_i(\mathbf{x}, a_i) \preceq_{\mathcal{K}_i} \mathbf{0}, h_j(\mathbf{x}, b_j) = 0, i = 1, \dots, m, j = 1, \dots, p \right\}$$

with a non-empty domain $\mathcal{D} = \mathbf{dom}\, f \cap_{i=1}^{m} \mathbf{dom}\, g_i \cap_{j=1}^{p} \mathbf{dom}\, h_j$, where f, g_i and h_j are all differentiable and $\mathcal{K}_i \subseteq \mathbb{R}^{c_i}$ are proper cones. Furthermore, let $(v_{1*}, \dots, v_{m*}, \nu_*)$ be an optimal solution of the dual problem

$$\min_{v_1, \dots, v_m \in \prod_{i=1}^{m} \mathbb{R}^{c_i}} \left\{ \sup_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x}) - \sum_{i=1}^{m} v_i^{\mathrm{T}} g_i(\mathbf{x}, a_i) - \sum_{j=1}^{p} v_j h_j(\mathbf{x}, b_j) : v_i \succeq_{\mathcal{K}_{i*}} \mathbf{0}, i = 1, \dots, m \right\}.$$

with zero duality gap. If $\mathbf{x}_*$ is continuously differentiable at $(\mathbf{a}, \mathbf{b})$ where $\mathbf{a} = [a_1, \dots, a_m]^{\mathrm{T}}$ and $\mathbf{b} = [b_1, \dots, b_p]^{\mathrm{T}}$, then

$$\frac{df}{da_i}(\mathbf{x}_*) = -\frac{\partial}{\partial a_i} \sum_{k=1}^{c_i} v_{ik} g_{ik}(\mathbf{x}_*, a_i), \quad i = 1, \dots, m,$$

$$\frac{df}{db_i}(\mathbf{x}_*) = -v_{j*} \frac{\partial h_i}{\partial b_i}(\mathbf{x}_*, b_i), \quad i = 1, \dots, p.$$

Proof: First, note that all the conditions of Theorem A.23 are satisfied. Using the last condition of Theorem A.23,

$$\frac{\partial}{\partial x_\ell} f(\mathbf{x}_*) - \sum_{j=1}^{m} \sum_{k=1}^{c_j} v_{jk*} \frac{\partial}{\partial x_\ell} g_{jk}(\mathbf{x}_*, a_j) - \sum_{j=1}^{p} v_{j*} \frac{\partial}{\partial x_\ell} h_j(\mathbf{x}_*, b_j) = 0, \quad \ell = 1, \dots, n$$

which implies

$$\frac{\partial}{\partial x_\ell} f(\mathbf{x}_*) = \sum_{j=1}^{m} \sum_{k=1}^{c_j} v_{jk*} \frac{\partial}{\partial x_\ell} g_{jk}(\mathbf{x}_*, a_j) + \sum_{j=1}^{p} v_{j*} \frac{\partial}{\partial x_\ell} h_j(\mathbf{x}_*, b_j). \tag{A.8}$$

Now, by the Chain Rule, we have for $i = 1, \dots, p$

$$\frac{df}{db_i}(\mathbf{x}_*) = \sum_{\ell=1}^{n} \frac{\partial}{\partial x_\ell} f(\mathbf{x}_*) \frac{dx_{\ell*}}{db_i}$$

$$= \sum_{\ell=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{c_j} v_{jk*} \frac{\partial}{\partial x_\ell} g_{jk}(\mathbf{x}_*, a_j) \frac{dx_{\ell*}}{db_i} + \sum_{\ell=1}^{n} \sum_{j=1}^{p} v_{j*} \frac{\partial}{\partial x_\ell} h_j(\mathbf{x}_*, b_j) \frac{dx_{\ell*}}{db_i} \quad \text{(by (A.8))}$$

$$= v_{i*} \left(\sum_{\ell=1}^{n} \frac{\partial}{\partial x_\ell} h_i(\mathbf{x}_*, b_i) \frac{dx_{\ell*}}{db_i} \right) + \sum_{j=1, j\neq i}^{p} v_{j*} \left(\sum_{\ell=1}^{n} \frac{d}{dx_\ell} h_j(\mathbf{x}_*, b_j) \frac{dx_{\ell*}}{db_i} \right)$$

$$\sum_{j=1}^{m} \sum_{\ell=1}^{n} \sum_{k=1}^{c_j} v_{jk*} \frac{\partial}{\partial x_\ell} g_{jk}(\mathbf{x}_*, a_j) \frac{dx_{\ell*}}{db_i}$$

$$= -v_{i*} \frac{\partial}{\partial b_i} h_i(\mathbf{x}_*, b_i),$$

where the last equality holds because performing implicit differentiation with respect to b_i on both sides of each equation in condition (ii) of Theorem A.23 obtains

$$\sum_{\ell=1}^{n} \frac{\partial}{\partial x_\ell} h_i(\mathbf{x}_*, b_i) \frac{dx_{\ell*}}{db_i} + \frac{\partial}{\partial b_i} h_i(\mathbf{x}_*, b_i) = 0,$$

$$\sum_{\ell=1}^{n} \frac{\partial h_j}{\partial x_\ell}(\mathbf{x}_*, b_j) \frac{dx_{\ell*}}{db_i} = 0, \quad \forall j \neq i$$

and doing the same for condition (iv) of Theorem A.23 yields

$$\sum_{\ell=1}^{n} \sum_{k=1}^{c_j} v_{jk*} \frac{\partial}{\partial x_\ell} g_{jk}(\mathbf{x}_*, a_j) \frac{dx_{\ell*}}{db_i} = 0, \quad j = 1, \dots, m.$$

An analogous argument yields

$$\frac{df}{da_i}(\mathbf{x}_*) = -\frac{\partial}{\partial a_i} \sum_{k=1}^{c_i} v_{ik} g_{ik}(\mathbf{x}_*, a_i), \quad i = 1, \dots, m.$$

$\square$

Remark A.14

Theorem A.25 is the Envelope Theorem for an optimization problem with parameterized constraint functions and a generalization of Theorem 19.5 in [183].

Theorem A.26

Consider the optimization problem

$$\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$$

where $\mathcal{X}$ is the set of feasible solutions and $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a parameterized objective function. Let $\mathbf{x}_*$ be an optimal solution and assume that the optimal value $V(\theta) = f(\mathbf{x}_*, \theta)$ is twice differentiable at θ, then

$$\frac{d^2 V}{d\theta^2} \geq \frac{\partial^2 f}{\partial \theta^2}(\mathbf{x}_*, \theta),$$

where $\frac{\partial^2 f}{\partial \theta^2}(\mathbf{x}_*, \theta)$ is the second-order partial derivative with respect to θ evaluated at $\mathbf{x}_*$.

Proof: Note that

$$V(\tilde{\theta}) \geq f(\mathbf{x}_*, \tilde{\theta}) \tag{A.9}$$

for $\tilde{\theta} \neq \theta$, since $\mathbf{x}_*$ is computed at θ and suboptimal at $\tilde{\theta}$. The Taylor's expansion of $V(\tilde{\theta})$ about θ is

$$V(\tilde{\theta}) = V(\theta) + \frac{dV(\theta)}{d\tilde{\theta}}(\tilde{\theta} - \theta) + \frac{1}{2}\frac{d^2 V(\theta)}{d\tilde{\theta}^2}(\tilde{\theta} - \theta)^2 + \dots$$

while that of $f(\mathbf{x}_*, \tilde{\theta})$ with respect to $\tilde{\theta}$ about θ is

$$f(\mathbf{x}_*, \tilde{\theta}) = f(\mathbf{x}_*, \theta) + \frac{\partial f}{\partial \tilde{\theta}}(\mathbf{x}_*, \theta)(\tilde{\theta} - \theta) + \frac{1}{2}\frac{\partial^2 f}{\partial \tilde{\theta}^2}(\mathbf{x}_*, \theta)(\tilde{\theta} - \theta)^2 + \dots$$

$$= V(\theta) + \frac{dV(\theta)}{d\tilde{\theta}}(\tilde{\theta} - \theta) + \frac{1}{2}\frac{\partial^2 f}{\partial \tilde{\theta}^2}(\mathbf{x}_*, \theta)(\tilde{\theta} - \theta)^2 + \dots,$$

where $\frac{\partial f}{\partial \tilde{\theta}}(\mathbf{x}_*, \theta) = \frac{dV(\theta)}{d\tilde{\theta}}$ by the Envelope Theorem A.24. Therefore, inequality (A.9) becomes

$$V(\theta) + \frac{dV(\theta)}{d\tilde{\theta}}(\tilde{\theta} - \theta) + \frac{1}{2}\frac{d^2 V(\theta)}{d\tilde{\theta}^2}(\tilde{\theta} - \theta)^2 + \dots \geq V(\theta) + \frac{dV(\theta)}{d\tilde{\theta}}(\tilde{\theta} - \theta) + \frac{1}{2}\frac{\partial^2 f}{\partial \tilde{\theta}^2}(\mathbf{x}_*, \theta)(\tilde{\theta} - \theta)^2 + \dots, \tag{A.10}$$

which implies

$$\frac{d^2 V}{d\theta^2} \geq \frac{\partial^2 f}{\partial \theta^2}(\mathbf{x}_*, \theta)$$

if $\tilde{\theta}$ is close enough to θ so that third and higher order terms in the Taylor's expansions on both sides of (A.10) can be ignored. $\qquad\square$

Remark A.15

Theorem A.26 is called the Second-Order Envelope Theorem.

Appendix B

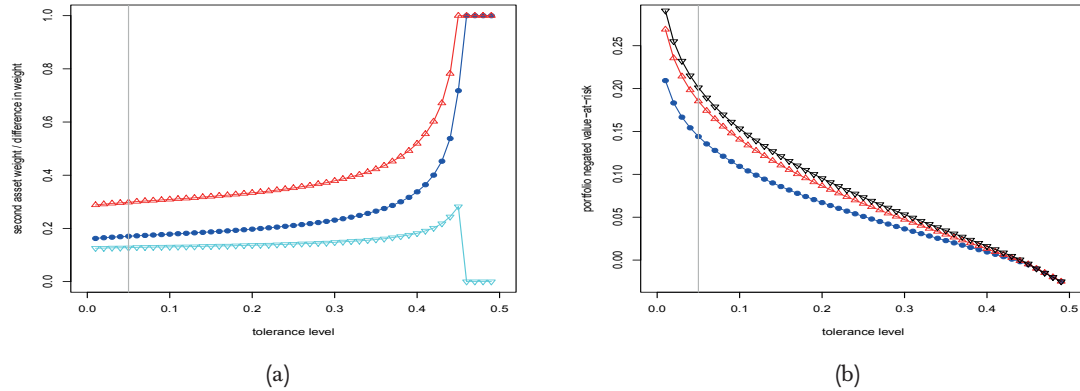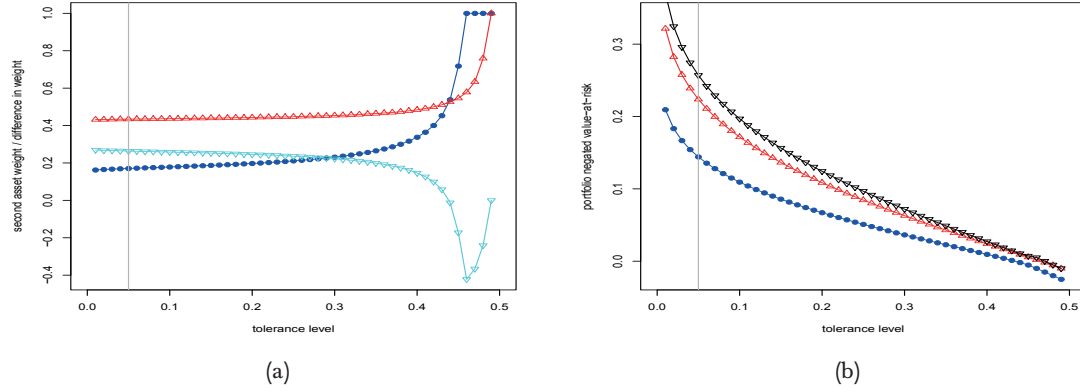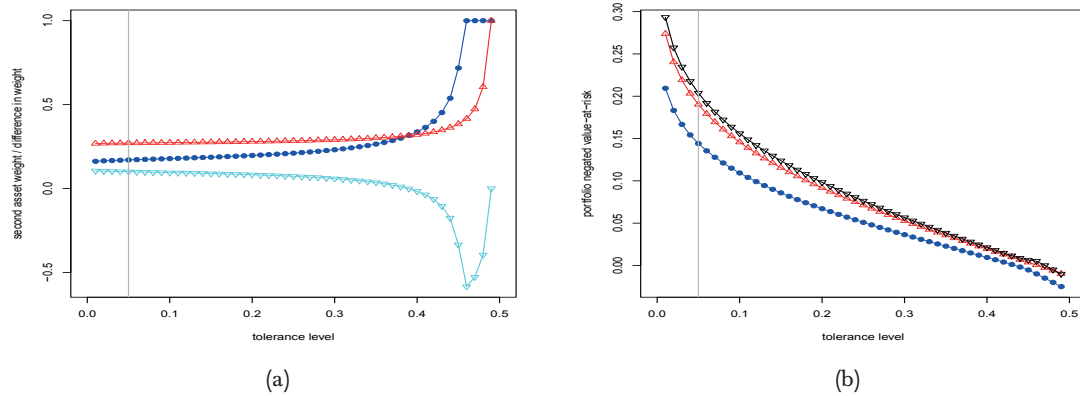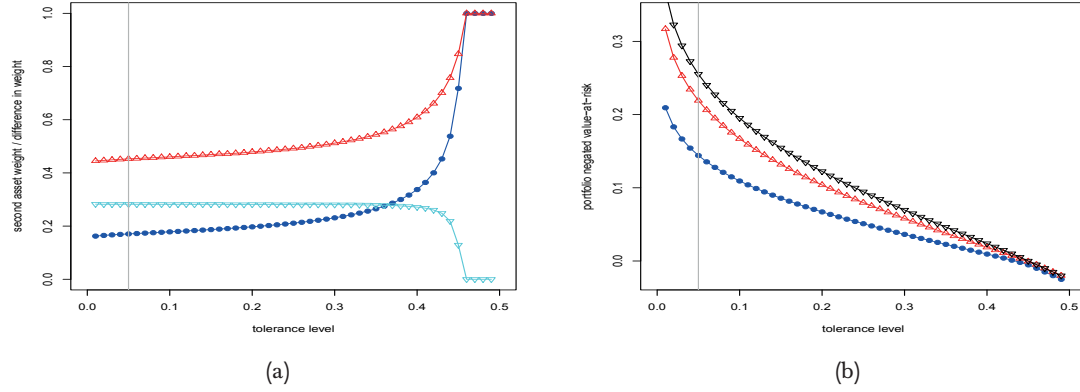We place all the figures referred to in Section [2.3](#) here.
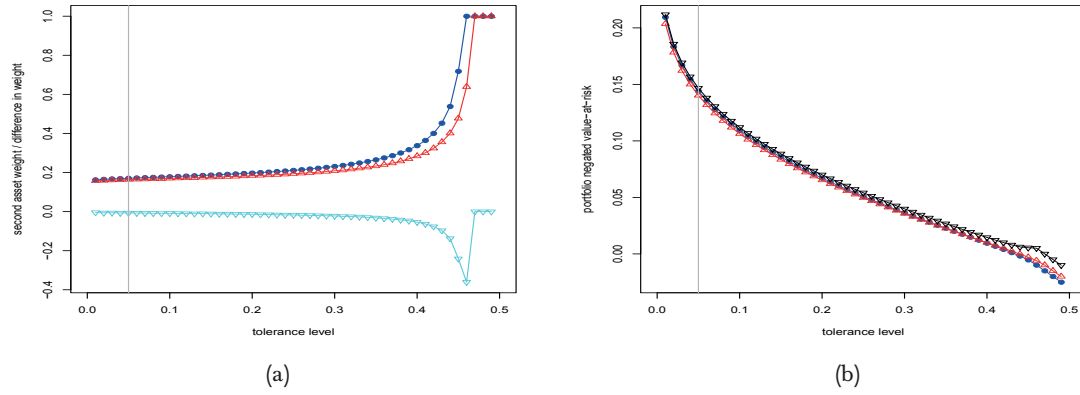
Figure B.1: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2] \in \{\mathbf{0.01}, \mathbf{0.015}\} \times \{\mathbf{0.025}, \mathbf{0.03}\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
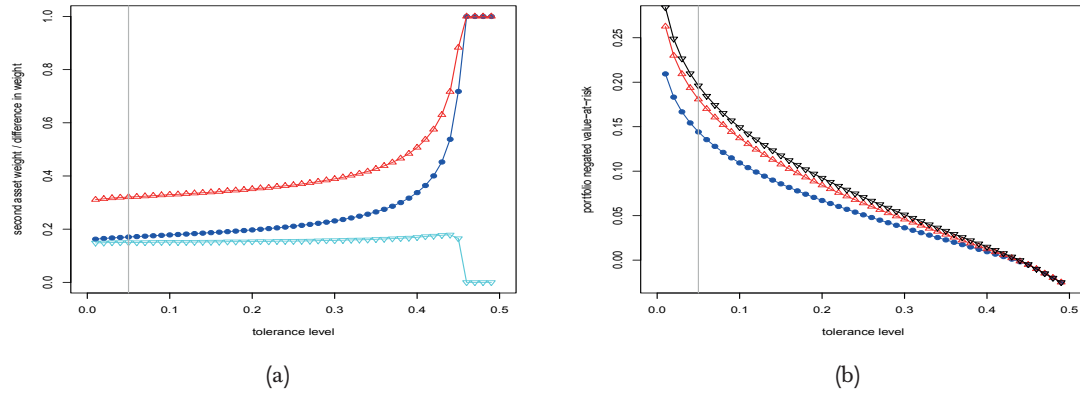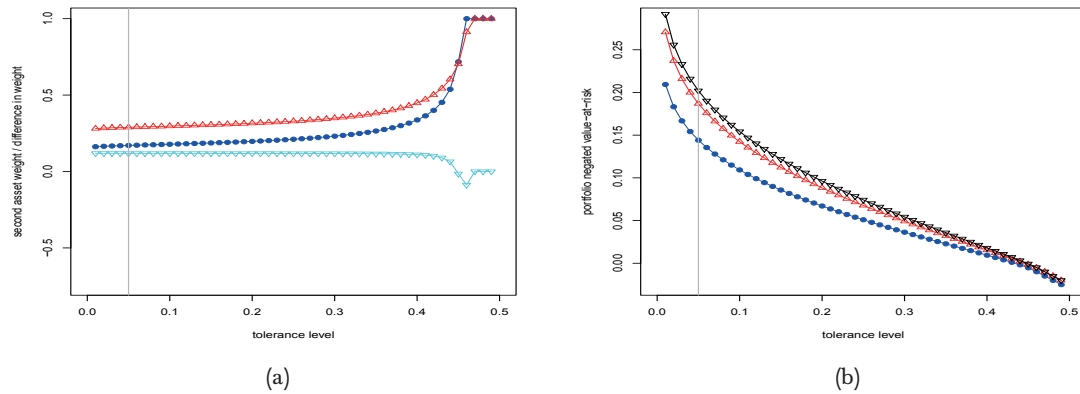

Figure B.2: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\sigma_1, \sigma_2] \in \{\mathbf{0.1}, \mathbf{0.12}, \mathbf{0.14}\} \times \{\mathbf{0.14}, \mathbf{0.16}, \mathbf{0.2}\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
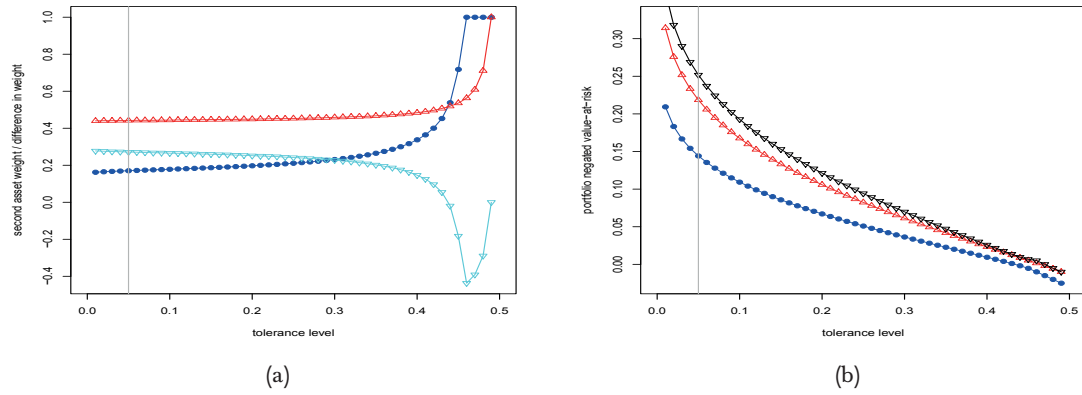
Figure B.3: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2] \in \{0.01, 0.015, 0.02, 0.025\} \times \{0.015, 0.02, 0.025, 0.03\} \times \{0.1, 0.12, \ldots, 0.18\} \times \{0.12, 0.14, \ldots, 0.2\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
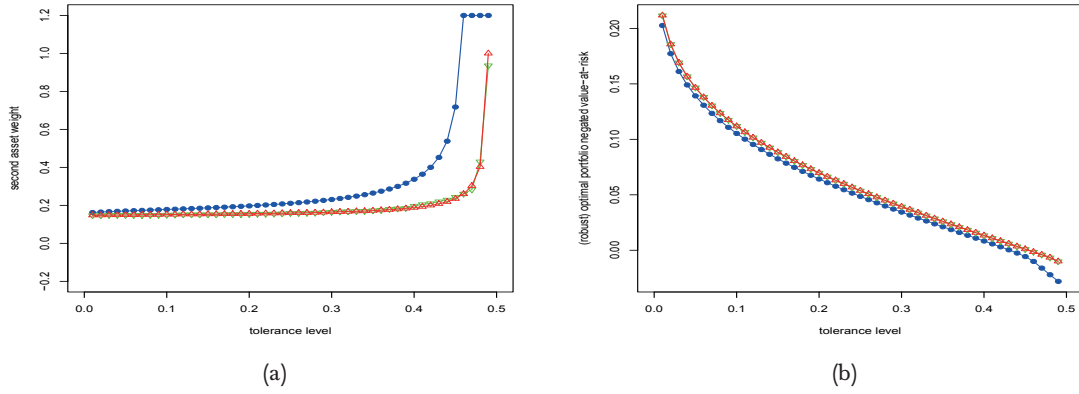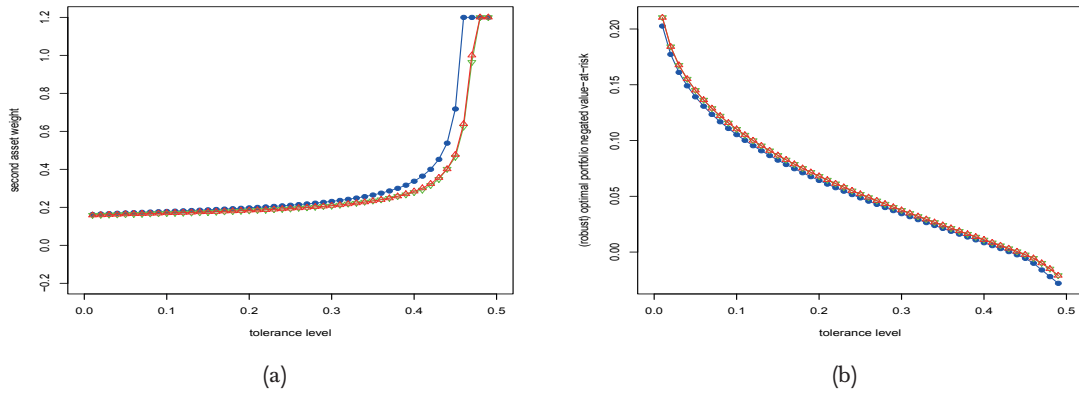

Figure B.4: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2] \in \{0.01, 0.015, 0.02, 0.025\} \times \{0.015, 0.02, 0.025, 0.03\} \times \{0.1, 0.12, 0.14\} \times \{0.16, 0.18, 0.2\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
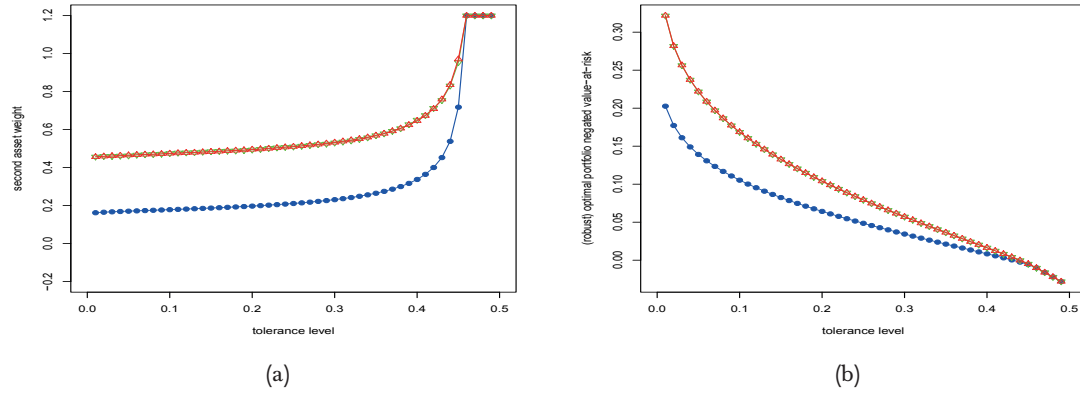
Figure B.5: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.12, \ldots, 0.18\} \times \{0.12, 0.14, \ldots, 0.2\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
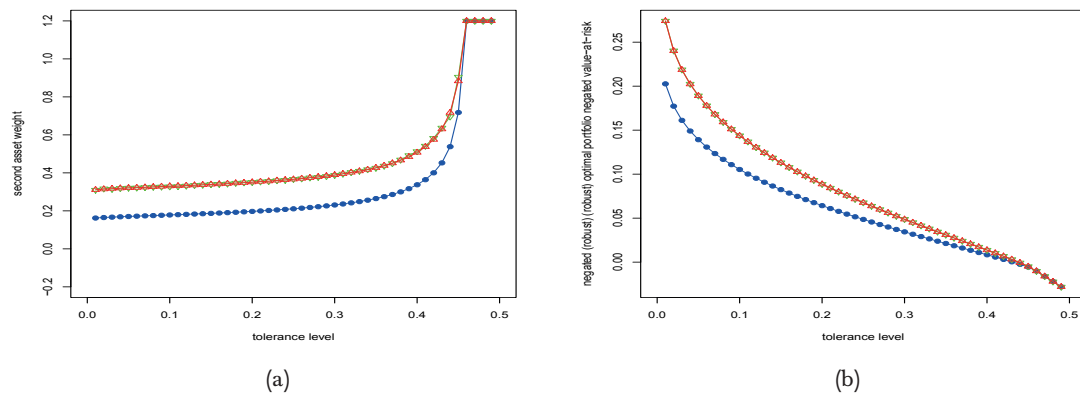

Figure B.6: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.12, 0.14\} \times \{0.16, 0.18, 0.2\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
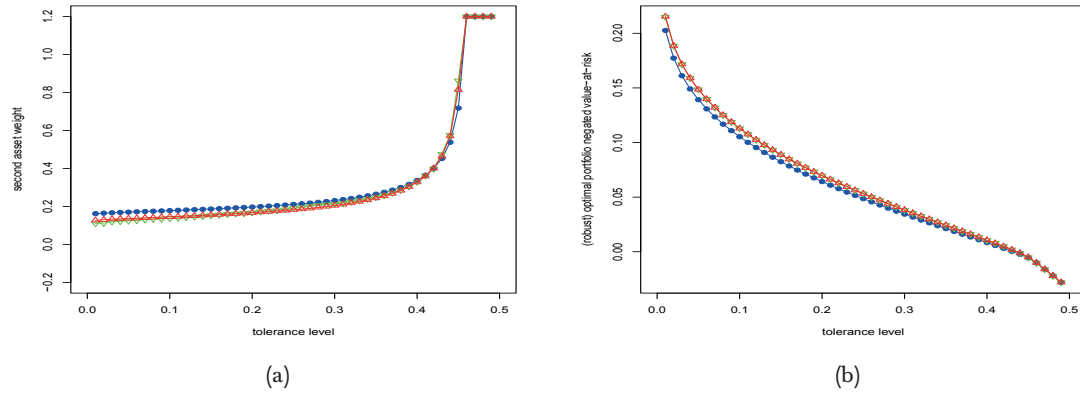
(a)

(b)
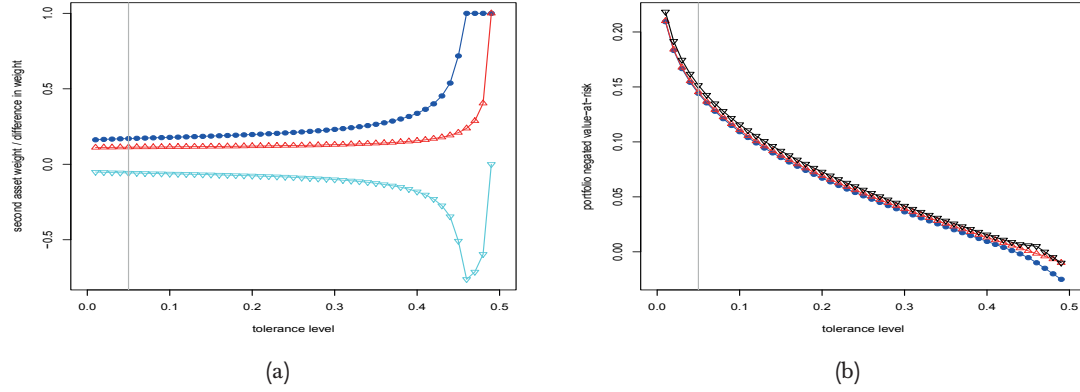
Figure B.7: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \rho] \in \{0.01, 0.015, 0.02, 0.025\} \times \{0.015, 0.02, 0.025, 0.03\} \times \{0.1, 0.2, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
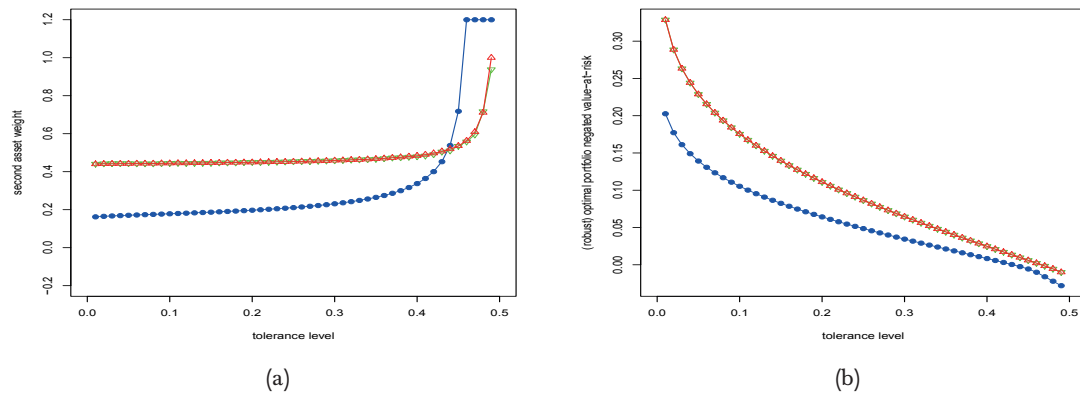

(a)

(b)
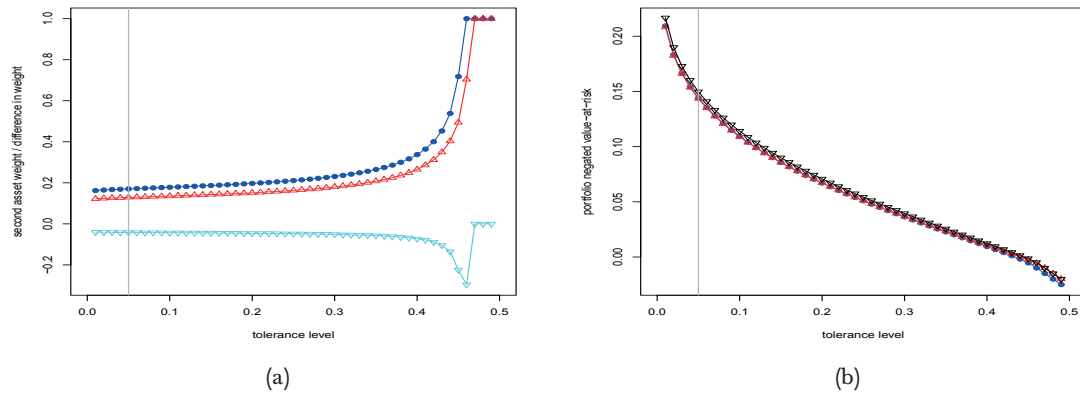
Figure B.8: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \rho] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.2, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
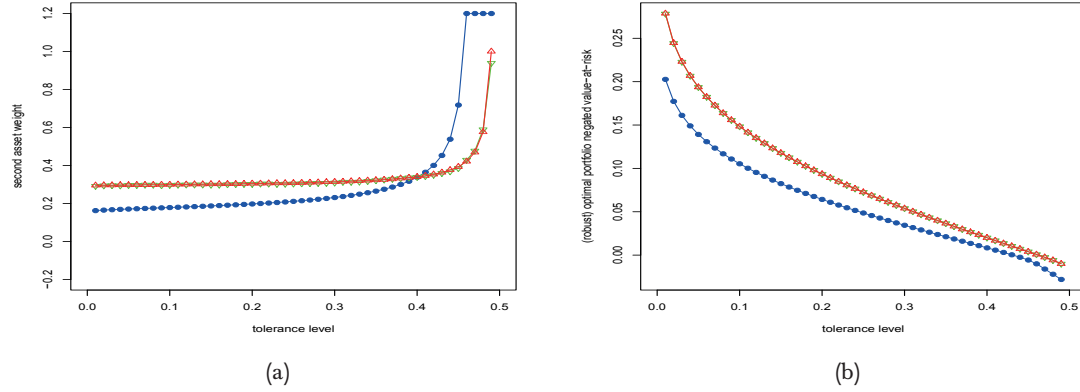
(a)

(b)
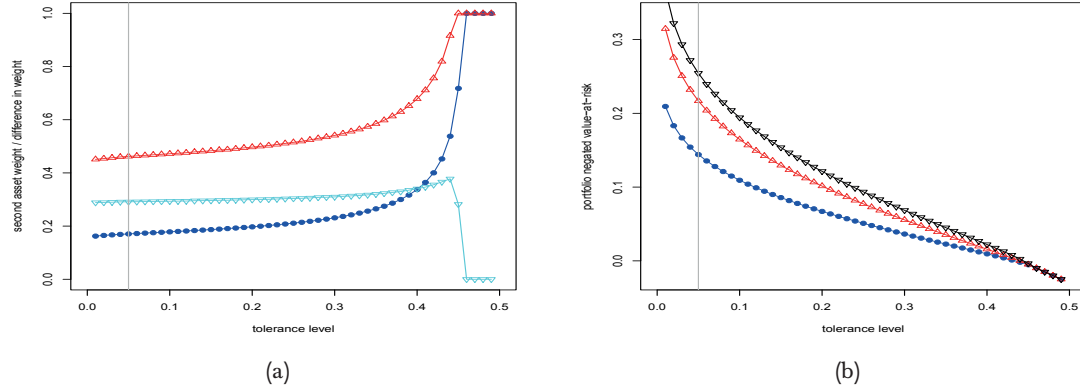
Figure B.9: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\sigma_1, \sigma_2, \rho] \in \{0.1, 0.12, \ldots, 0.18\} \times \{0.12, 0.14, \ldots, 0.2\} \times \{0.1, 0.2, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
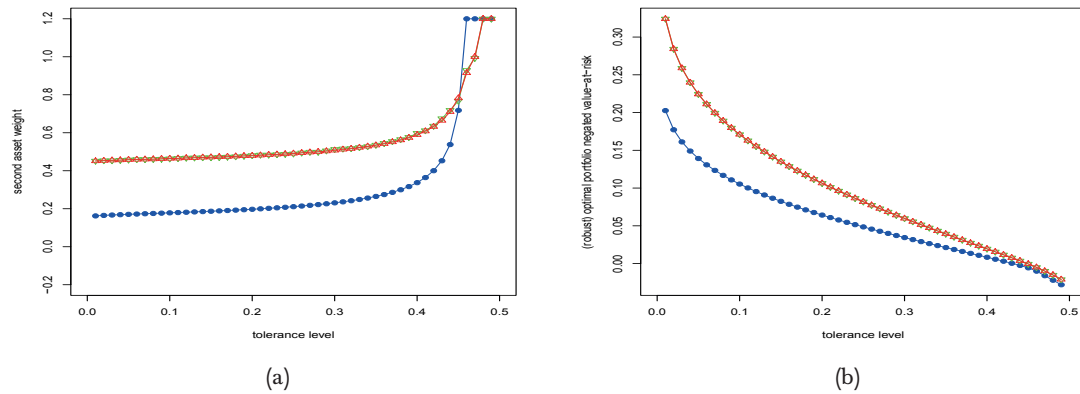

(a)

(b)

Figure B.10: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\sigma_1, \sigma_2, \rho] \in \{0.1, 0.12, 0.14\} \times \{0.14, 0.16, 0.2\} \times \{0.1, 0.2, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
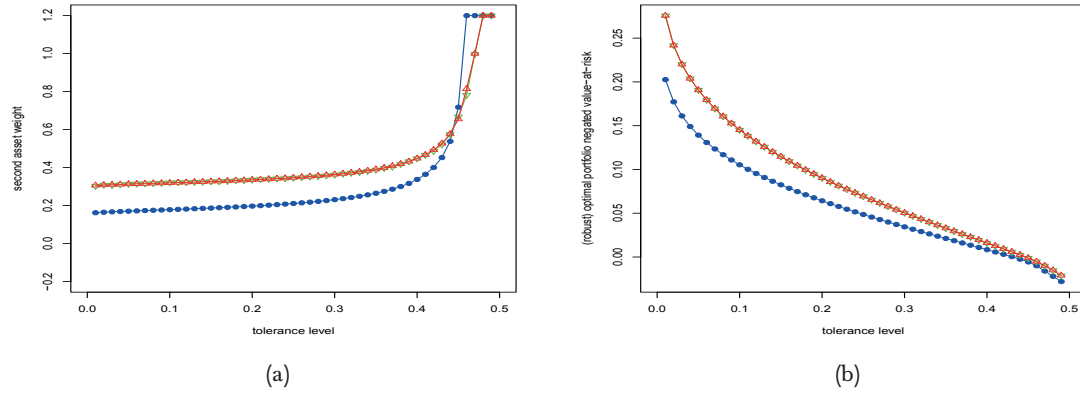
(a) (b)

Figure B.11: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho] \in \{0.01, 0.025\} \times \{0.015, 0.03\} \times \{0.1, 0.18\} \times \{0.12, 0.2\} \times \{0.1, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
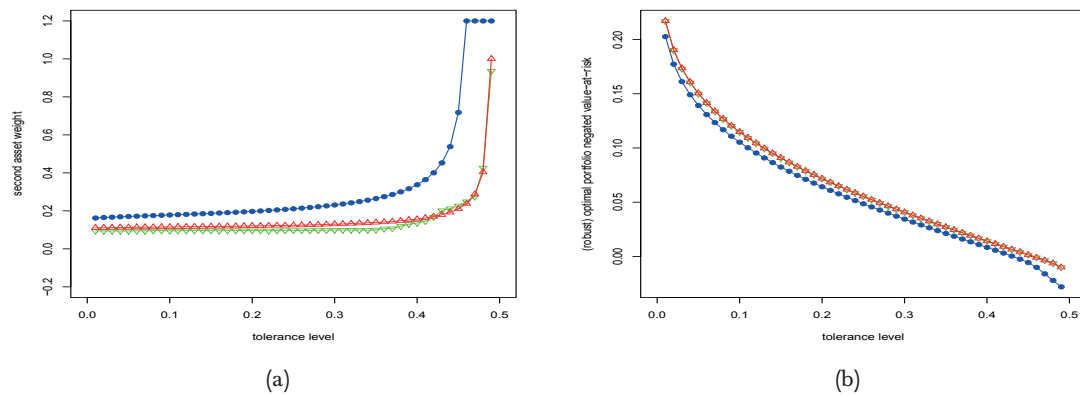

(a) (b)

Figure B.12: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho] \in \{0.01, 0.025\} \times \{0.015, 0.03\} \times \{0.1, 0.14\} \times \{0.16, 0.2\} \times \{0.1, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
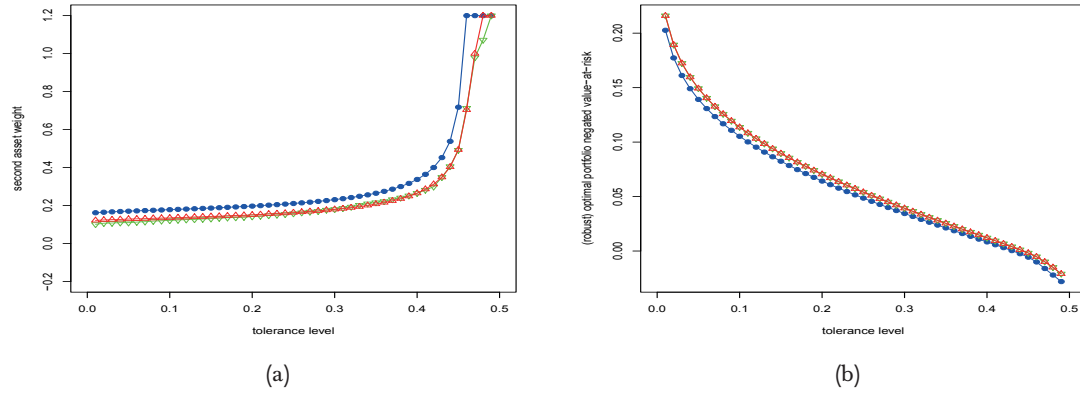
(a)

(b)

Figure B.13: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.18\} \times \{0.12, 0.2\} \times \{0.1, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
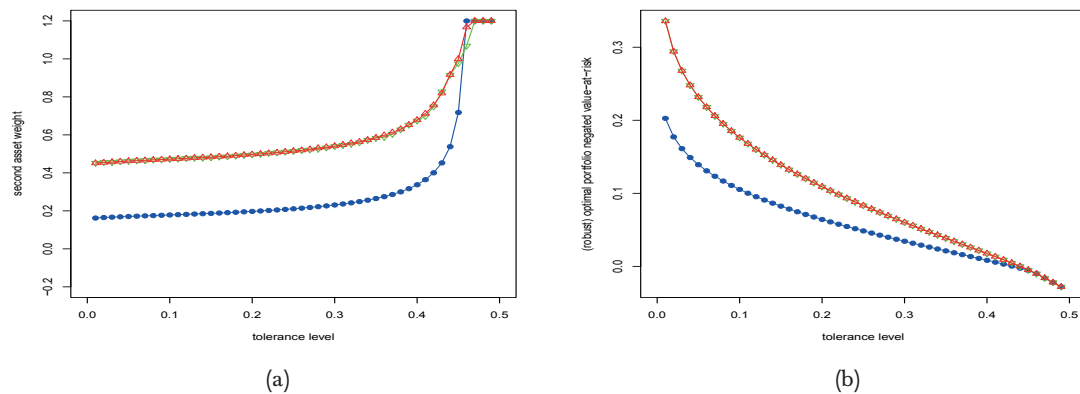

(a)

(b)

Figure B.14: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is disallowed; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.14\} \times \{0.16, 0.2\} \times \{0.1, 0.3\}$; cyan in (a) - difference between red and blue lines; black in (b) - based on solution without uncertainty and parameters that give the highest possible portfolio negated value-at-risk there.
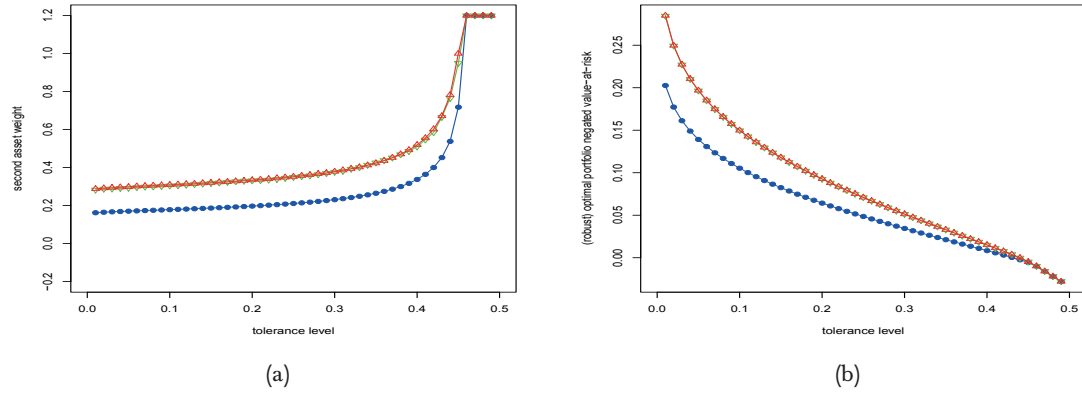
Figure B.15: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2] \in \{0.01, 0.015, 0.02, 0.025\} \times \{0.015, 0.02, 0.025, 0.03\}$.
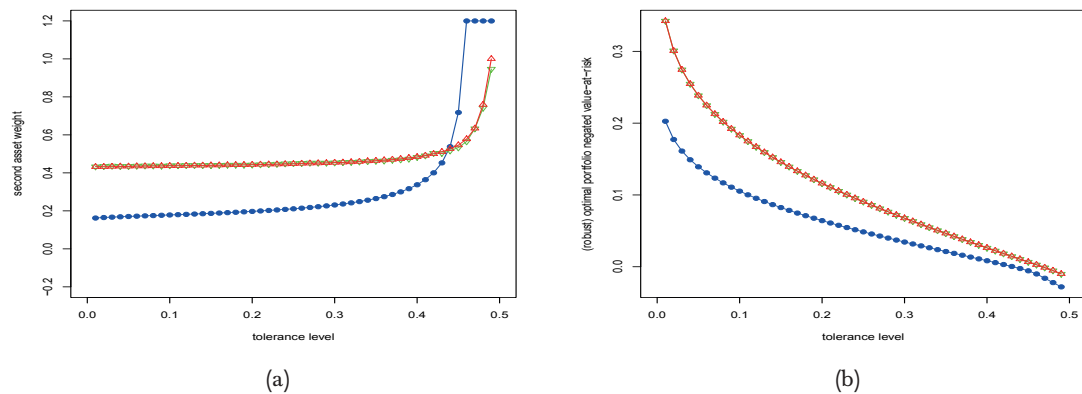

Figure B.16: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2] \in \{0.01, 0.015\} \times \{0.025, 0.03\}$.

(a)

(b)

Figure B.17: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\sigma_1, \sigma_2] \in \{0.1, 0.12, \ldots, 0.18\} \times \{0.12, 0.14, \ldots, 0.2\}$.
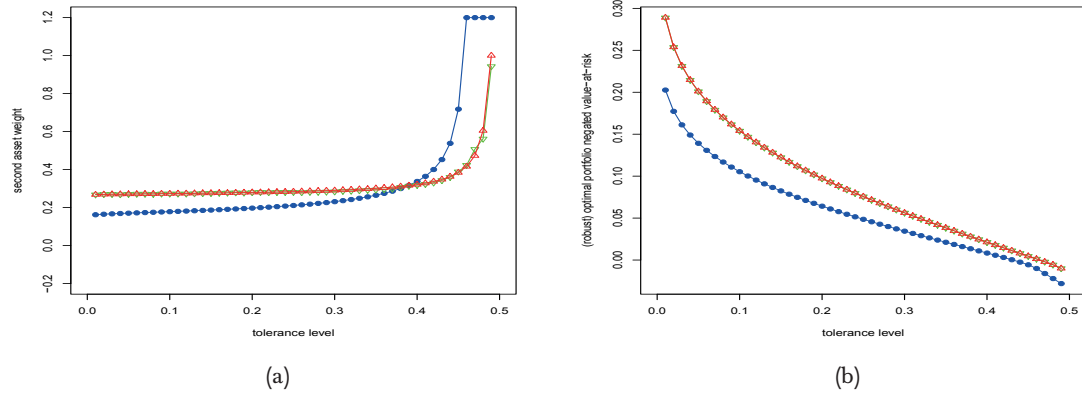

(a)

(b)

Figure B.18: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\sigma_1, \sigma_2] \in \{0.1, 0.12, 0.14\} \times \{0.14, 0.16, 0.2\}$.
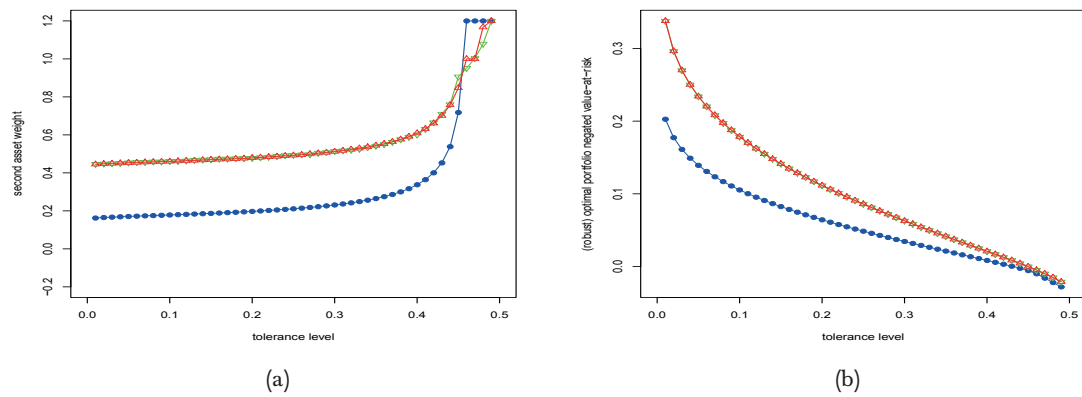
(a)

(b)

Figure B.19: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $\rho \in \{0.1, 0.2, 0.3\}$.

(a)

(b)

Figure B.20: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2] \in \{0.01, 0.015, 0.02, 0.025\} \times \{0.015, 0.02, 0.025, 0.03\} \times \{0.1, 0.12, \ldots, 0.18\} \times \{0.12, 0.14, \ldots, 0.2\}$.
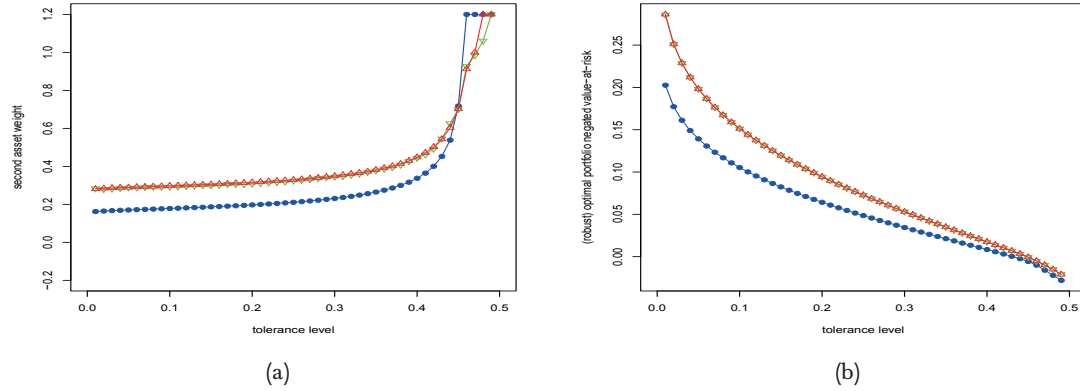
(a)

(b)

Figure B.21: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2] \in \{0.01, 0.015, 0.02, 0.025\} \times \{0.015, 0.02, 0.025, 0.03\} \times \{0.1, 0.12, 0.14\} \times \{0.16, 0.18, 0.2\}$.

(a)

(b)

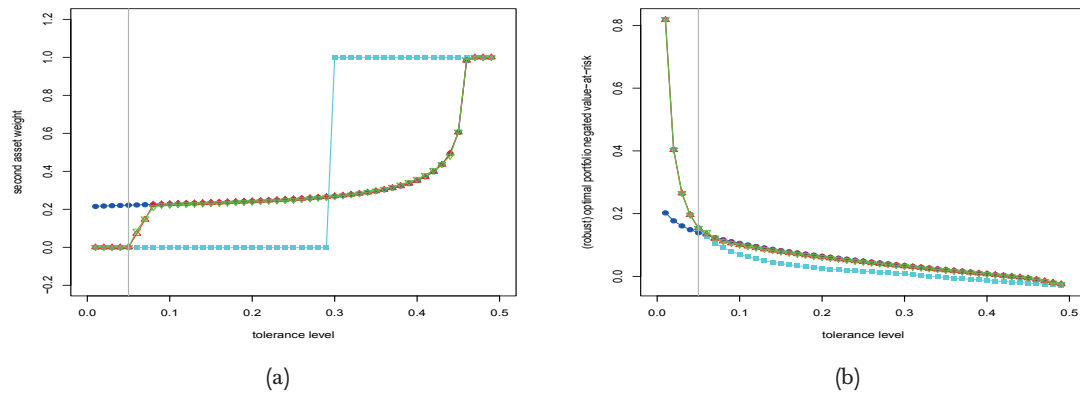Figure B.22: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.12, \ldots, 0.18\} \times \{0.12, 0.14, \ldots, 0.2\}$.
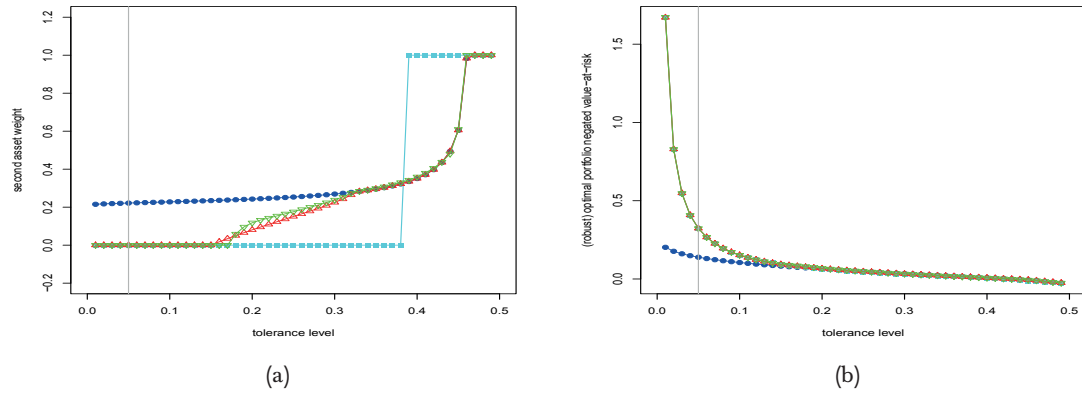
(a)

(b)

Figure B.23: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.12, 0.14\} \times \{0.16, 0.18, 0.2\}$.

(a)

(b)

Figure B.24: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \rho] \in \{0.01, 0.015, 0.02, 0.025\} \times \{0.015, 0.02, 0.025, 0.03\} \times \{0.1, 0.2, 0.3\}$.

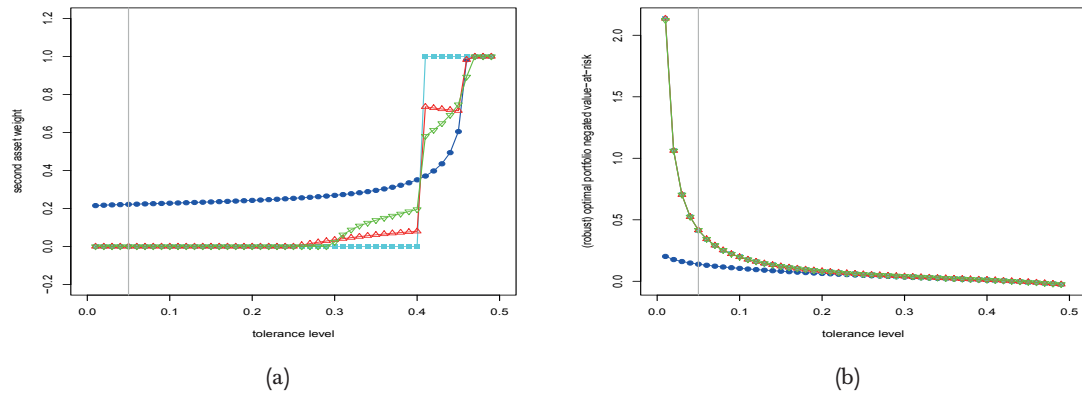Figure B.25: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \rho] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.2, 0.3\}$.

Figure B.26: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\sigma_1, \sigma_2, \rho] \in \{0.1, 0.12, \ldots, 0.18\} \times \{0.12, 0.14, \ldots, 0.2\} \times \{0.1, 0.2, 0.3\}$.

(a)

(b)

Figure B.27: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\sigma_1, \sigma_2, \rho] \in \{0.1, 0.12, 0.14\} \times \{0.14, 0.16, 0.2\} \times \{0.1, 0.2, 0.3\}$.

(a)

(b)

Figure B.28: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho] \in \{0.01, 0.025\} \times \{0.015, 0.03\} \times \{0.1, 0.18\} \times \{0.12, 0.2\} \times \{0.1, 0.3\}$.

Figure B.29: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho] \in \{0.01, 0.025\} \times \{0.015, 0.03\} \times \{0.1, 0.14\} \times \{0.16, 0.2\} \times \{0.1, 0.3\}$.

Figure B.30: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.18\} \times \{0.12, 0.2\} \times \{0.1, 0.3\}$.

Figure B.31: (a) optimal weight on second asset/difference in optimal weight between assets and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow a bivariate normal with $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$, $\sigma_2 = 0.2$ and $\rho = 0.2$; red - robust counterpart where $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho] \in \{0.01, 0.015\} \times \{0.025, 0.03\} \times \{0.1, 0.14\} \times \{0.16, 0.2\} \times \{0.1, 0.3\}$.

Figure B.32: (a) optimal weight on second asset and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow an independent bivariate normal distribution where $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$; cyan - returns follow an independent bivariate Cauchy distribution where $m_1 = 0.01$, $m_2 = 0.03$, $\gamma_1 = 0.026$ and $\gamma_2 = 0.052$ such that the fifth percent quantile of each of its marginals coincides with that of the corresponding marginal of the aforementioned bivariate normal distribution; red - robust against both bivariate distributions; green - spline approximation of red case.

(a)

(b)

Figure B.33: (a) optimal weight on second asset and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow an independent bivariate normal distribution where $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$; cyan - returns follow an independent bivariate Cauchy distribution where $m_1 = 0.01$, $m_2 = 0.03$, $\gamma_1 = \mathbf{0.053}$ and $\gamma_2 = \mathbf{0.106}$ such that the fifteenth percent quantile of each of its marginals coincides with that of the corresponding marginal of the aforementioned bivariate normal distribution; red - robust against both bivariate distributions; green - spline approximation of red case.

(a)

(b)

Figure B.34: (a) optimal weight on second asset and (b) portfolio negated value-at-risk against tolerance level where short-selling is allowed up to a maximum of one-fifth the total wealth for each asset; blue - returns follow an independent bivariate normal distribution where $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$; cyan - returns follow an independent bivariate Cauchy distribution where $m_1 = 0.01$, $m_2 = 0.03$, $\gamma_1 = \mathbf{0.067}$ and $\gamma_2 = \mathbf{0.135}$ such that the twenty fifth percent quantile of each of its marginals coincides with that of the corresponding marginal of the aforementioned bivariate normal distribution; red - robust against both bivariate distributions; green - spline approximation of red case.

Appendix C

We place all the figures referred to in Chapter 4 here. Figures C.1 - C.6 correspond to Section 4.1.1, Figures C.7 - C.12 to Section 4.2.1, Figures C.13 - C.18 to Section 4.3.1, Figures C.19 - C.30 to Section 4.4.1, Figures C.31 - C.42 to Section 4.5.1, and Figures C.43 - C.48 to Section 5.1.1.
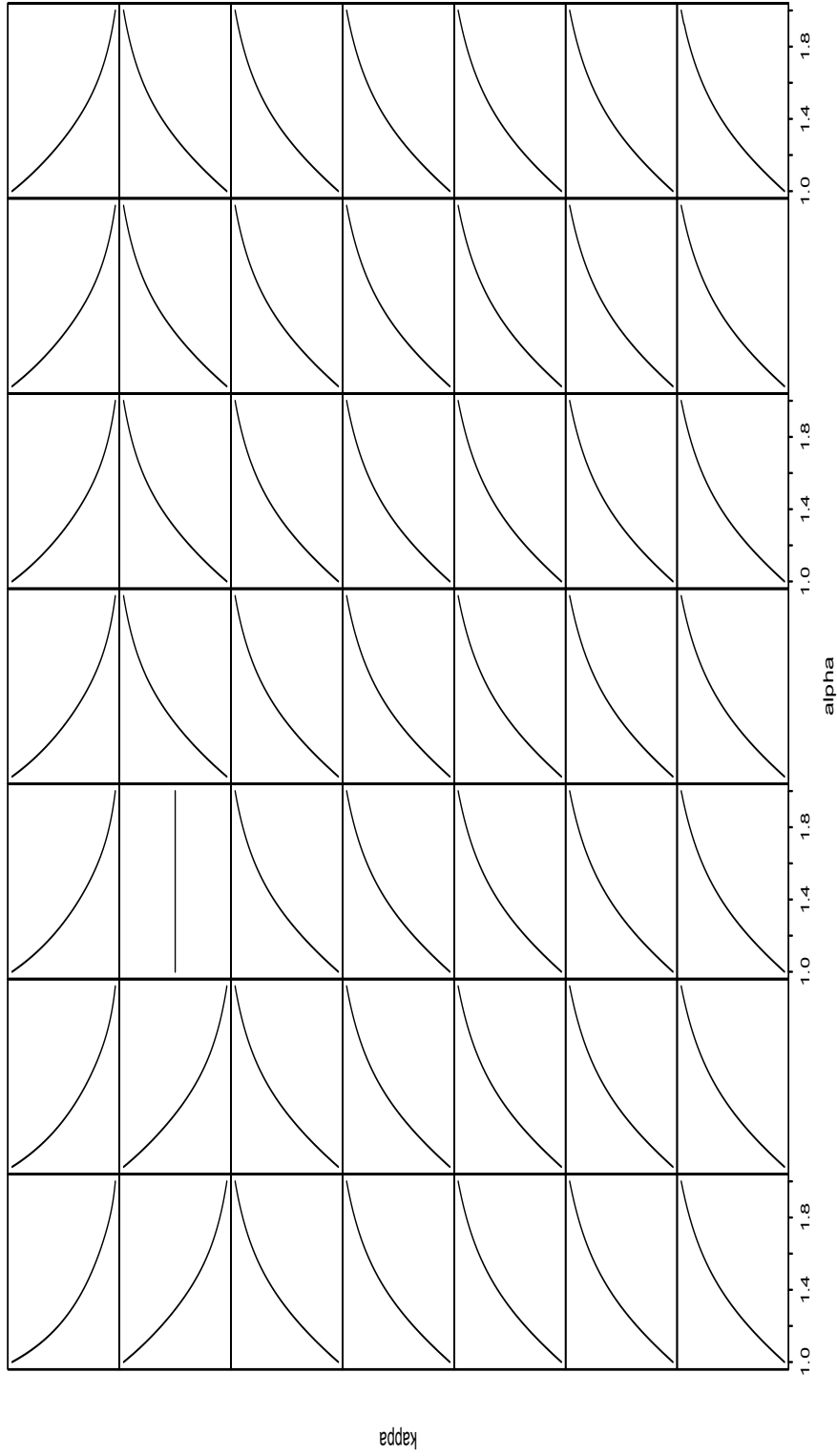
Figure C.1: each subfigure plots values of the objective function $\kappa(\epsilon) = \max\limits_{\alpha \in \{1, 1.01, \ldots, 2\}} \left\{ F_{Z_2}^{-1}(1 - \epsilon_*) F_{Z_\alpha}^{-1}(1 - \epsilon)/F_{Z_\alpha}^{-1}(1 - \epsilon_*) \right\}$ for type C ($\epsilon_* = 0.1$) against $\alpha \in \{1, 1.01, \ldots, 2\}$ at a particular tolerance level ϵ, and subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
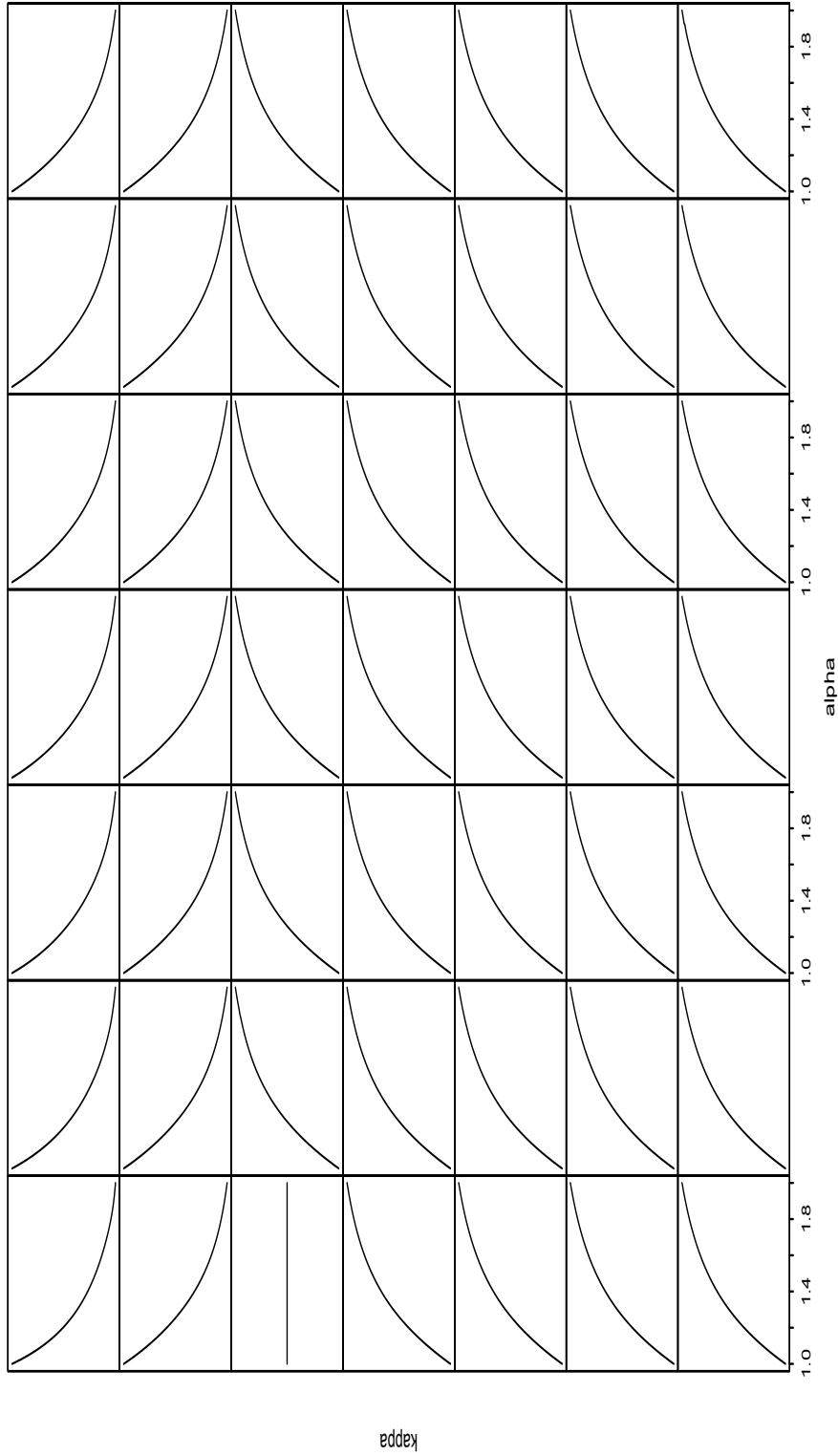
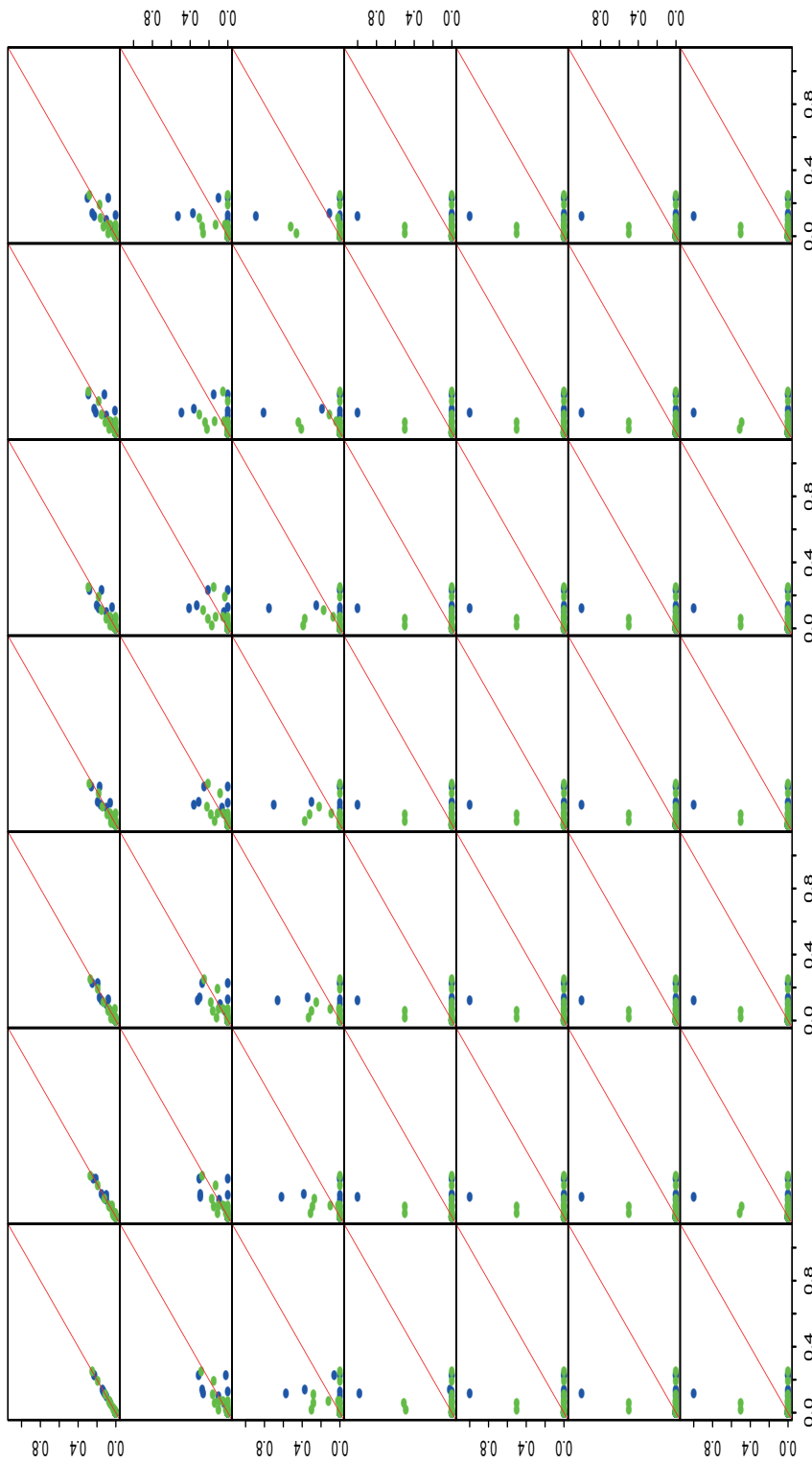Figure C.2: each subfigure plots values of the objective function $\kappa(\epsilon) = \max\limits_{\alpha \in \{1, 1.01, \ldots, 2\}} \left\{ F_{Z_2}^{-1}(1 - \epsilon_*) F_{Z_\alpha}^{-1}(1 - \epsilon)/F_{Z_\alpha}^{-1}(1 - \epsilon_*) \right\}$ for type D ($\epsilon_* = 0.15$) against $\alpha \in \{1, 1.01, \ldots, 2\}$ at a particular tolerance level ϵ, and subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
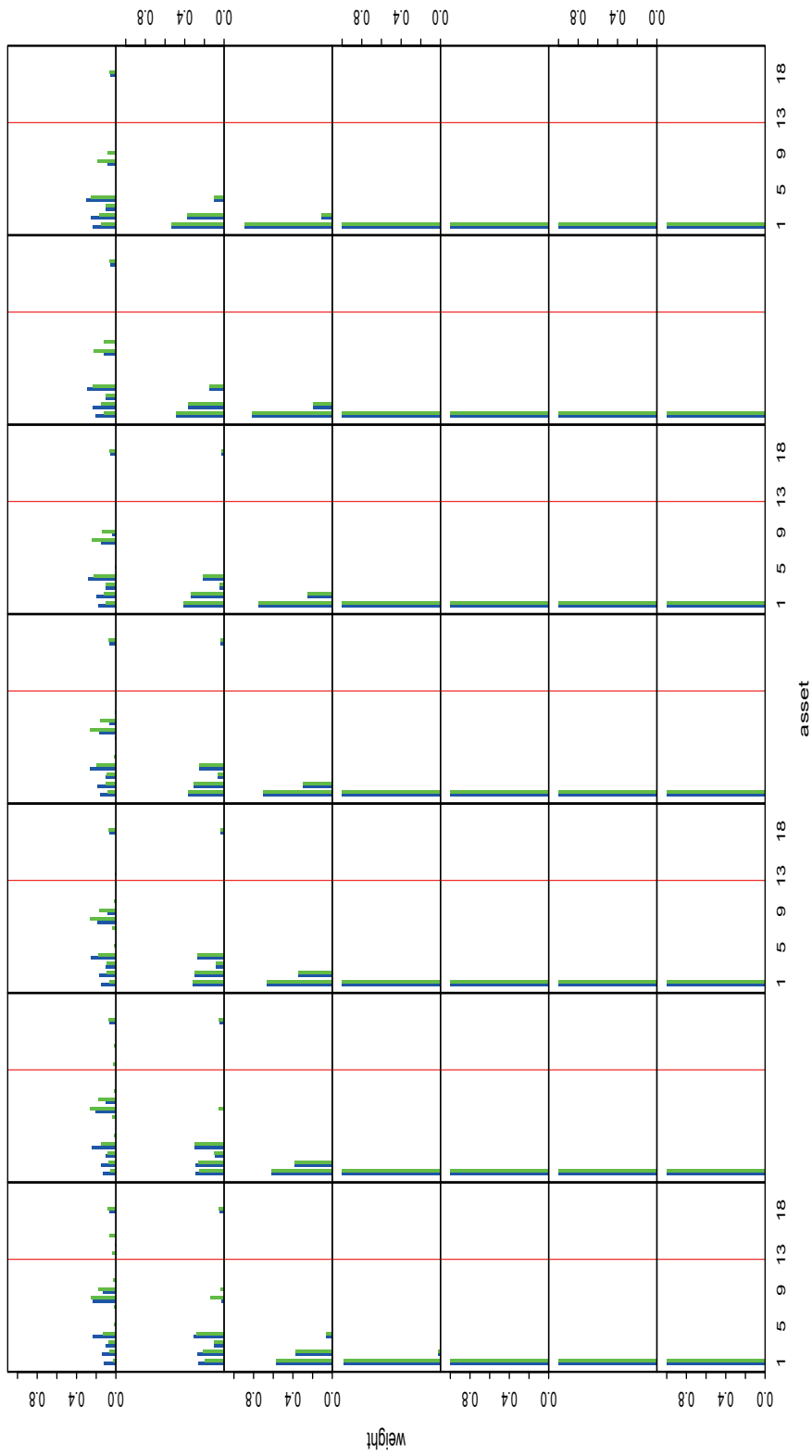
Figure C.3: each subfigure plots the optimal weights against the asset number for both the type A (blue) and C (green) problems where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
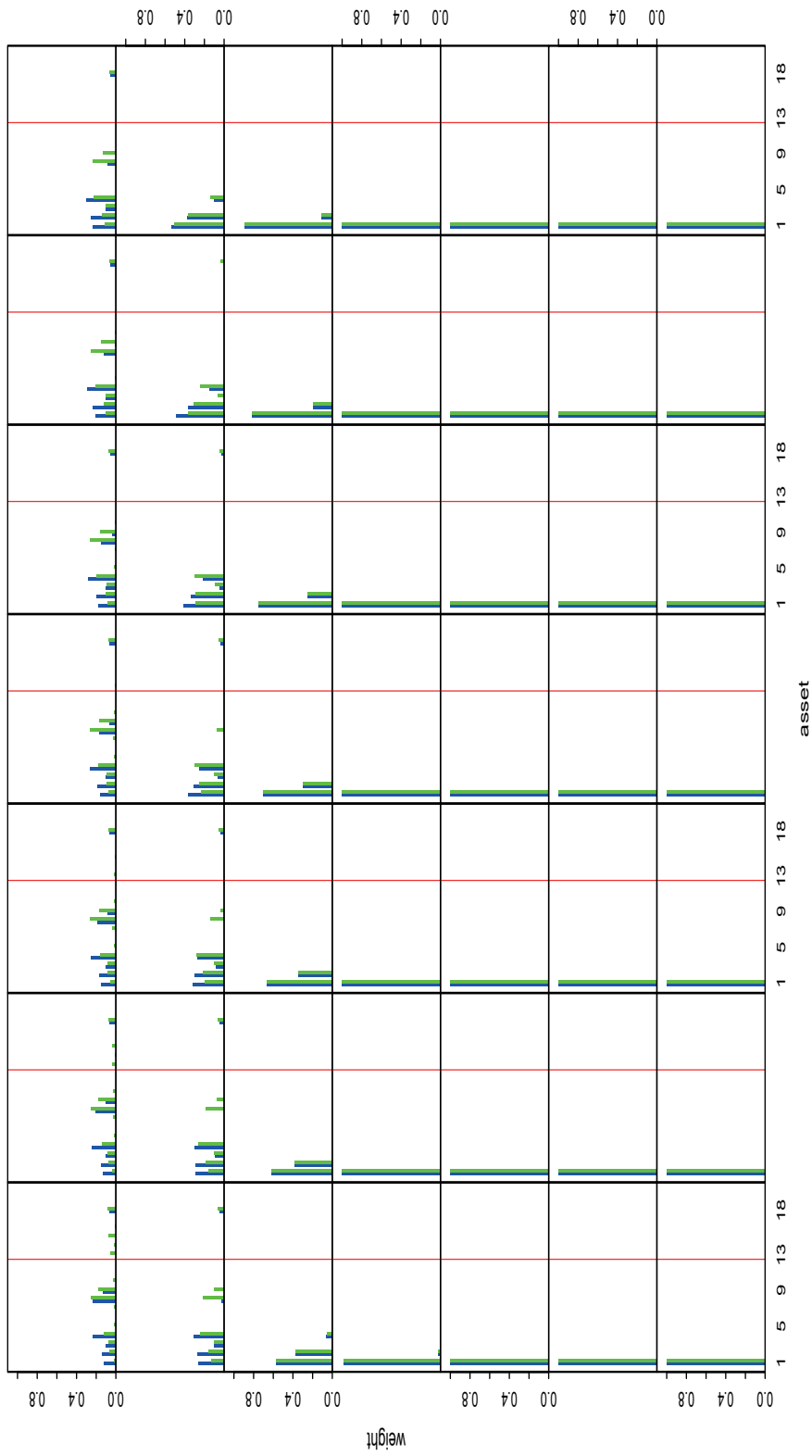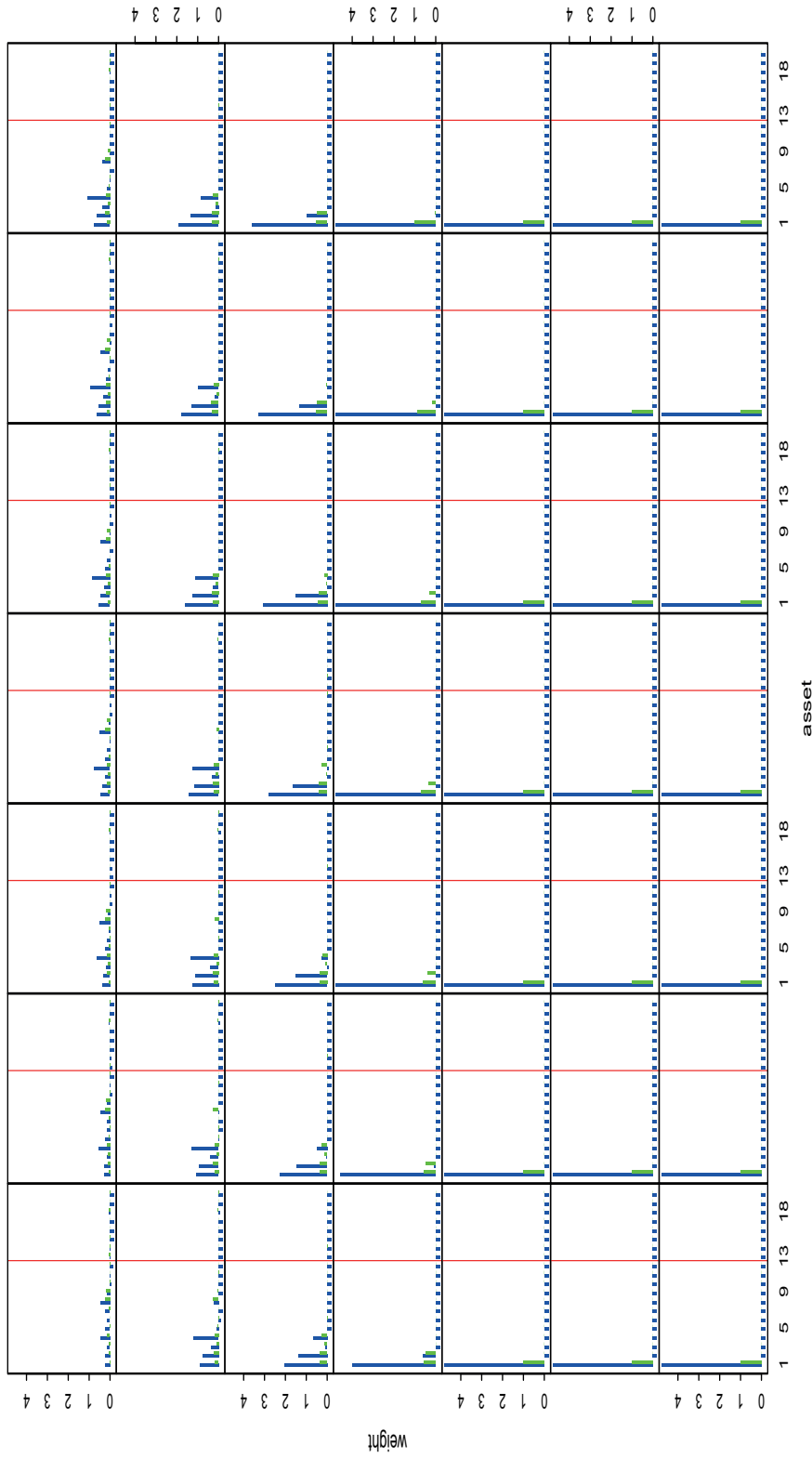
Figure C.4: each subfigure plots the optimal weights against the asset number for both the type A (blue) and D (green) problems where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
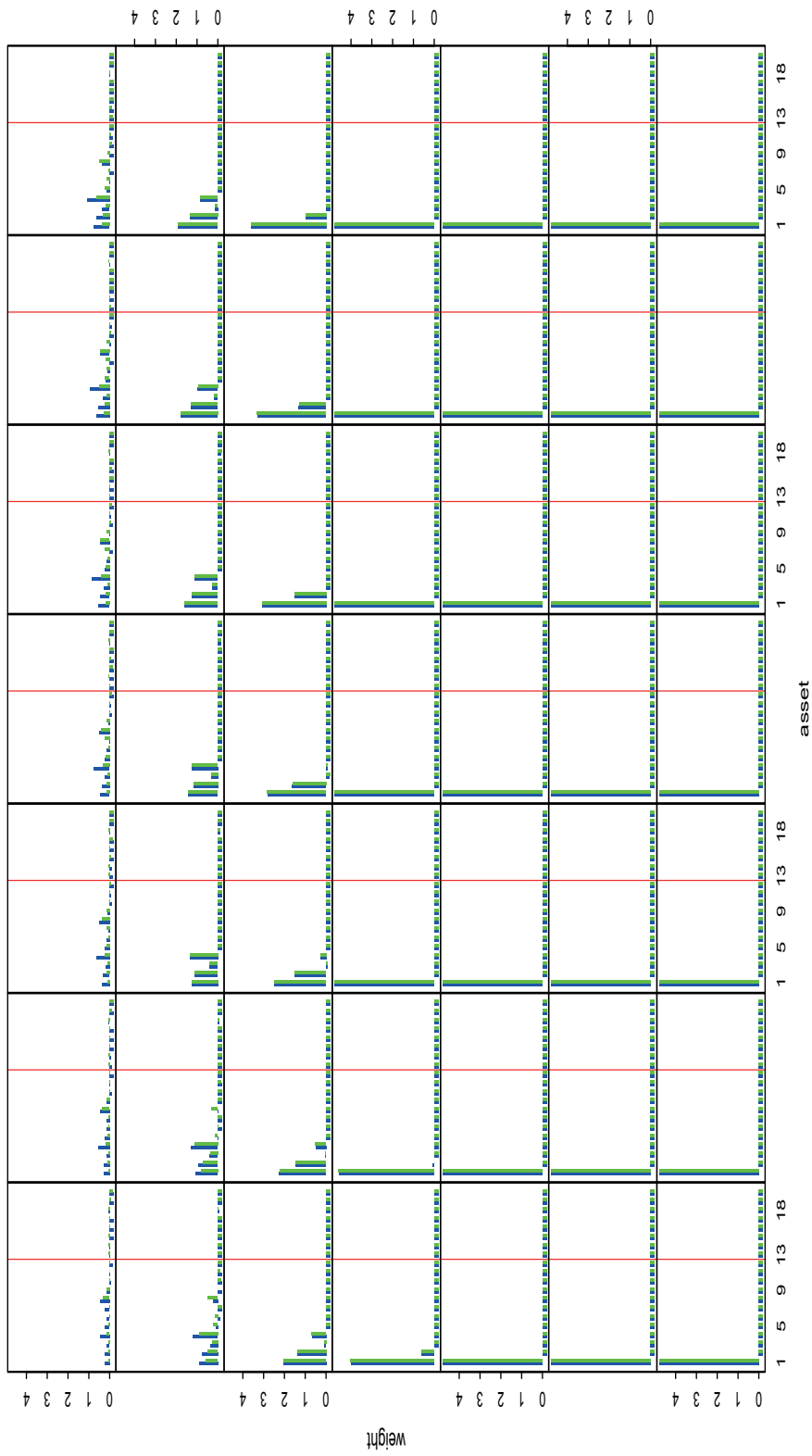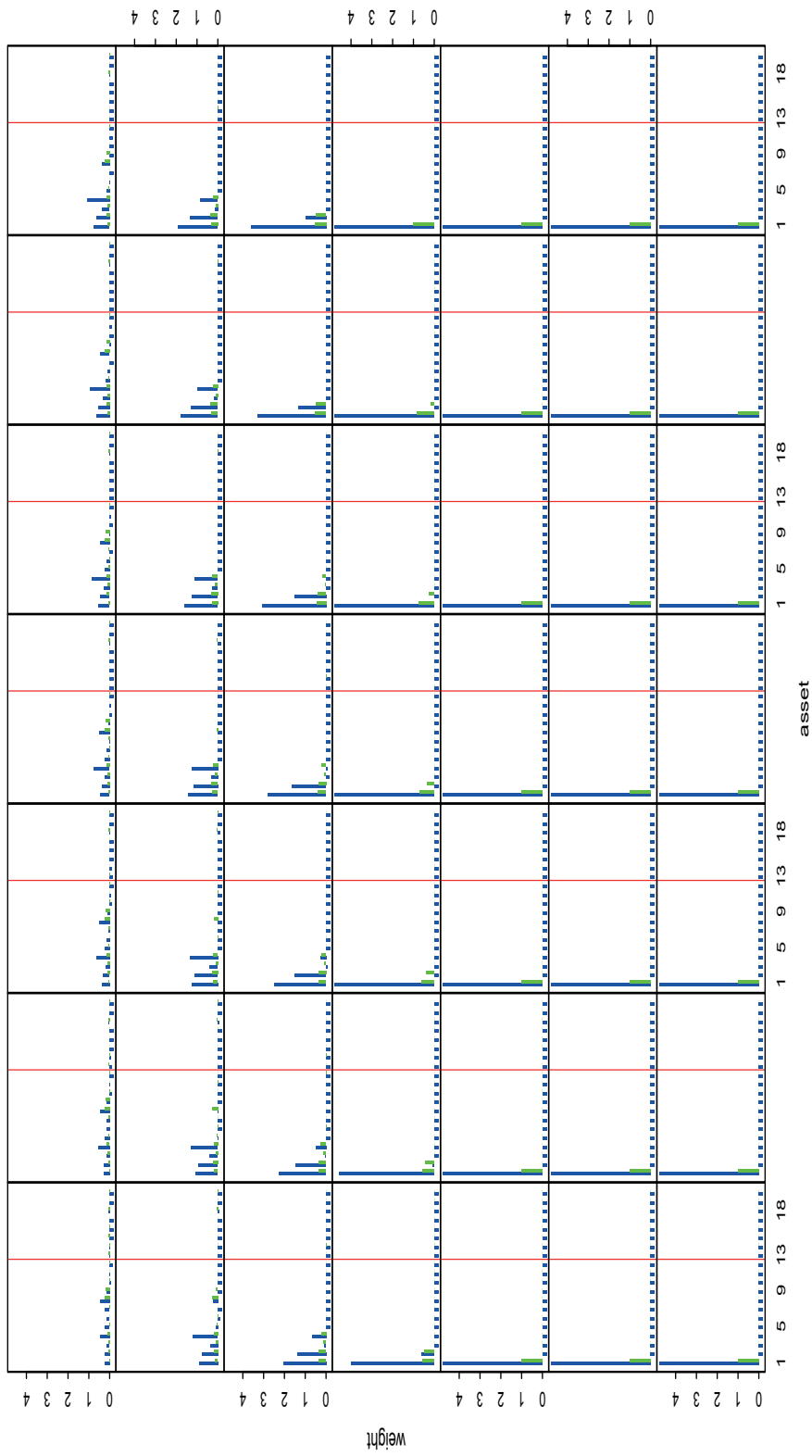
Figure C.5: each subfigure plots the optimal weights against the asset number for both the type A (blue) and C (green) problems where short-selling is allowed up to a maximum of one fifth the total wealth for each asset at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
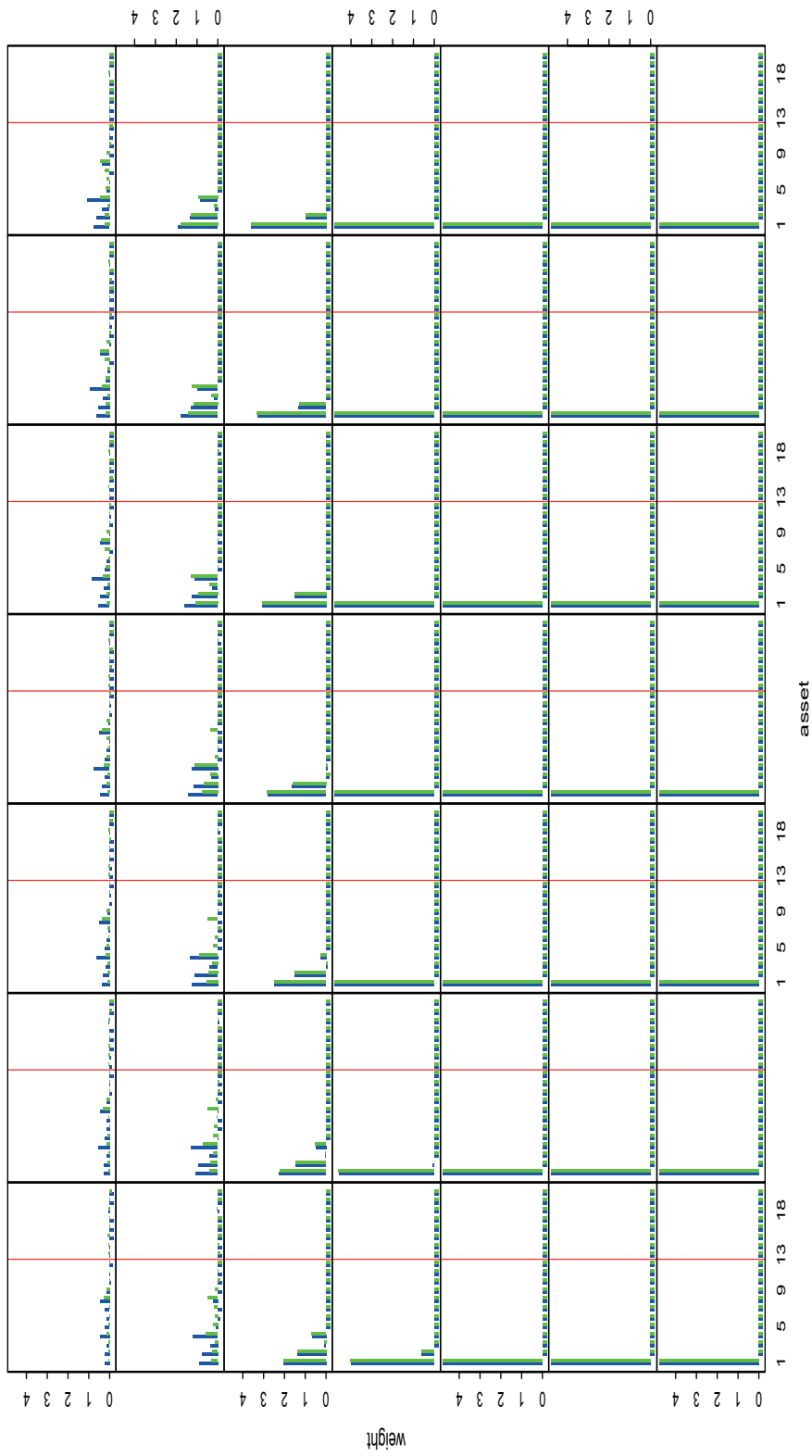
Figure C.6: each subfigure plots the optimal weights against the asset number for both the type A (blue) and D (green) problems where short-selling is allowed up to a maximum of one fifth the total wealth for each asset at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
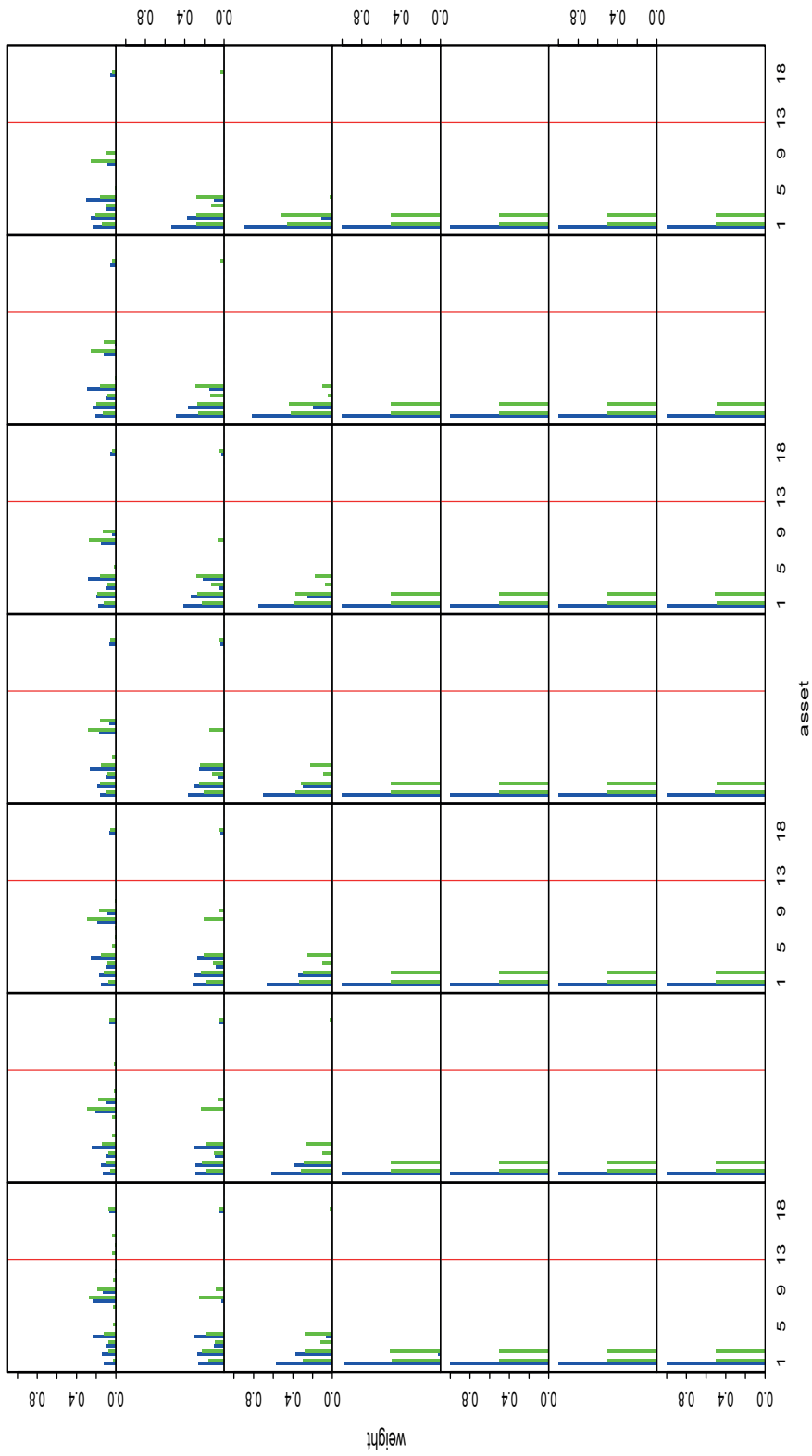
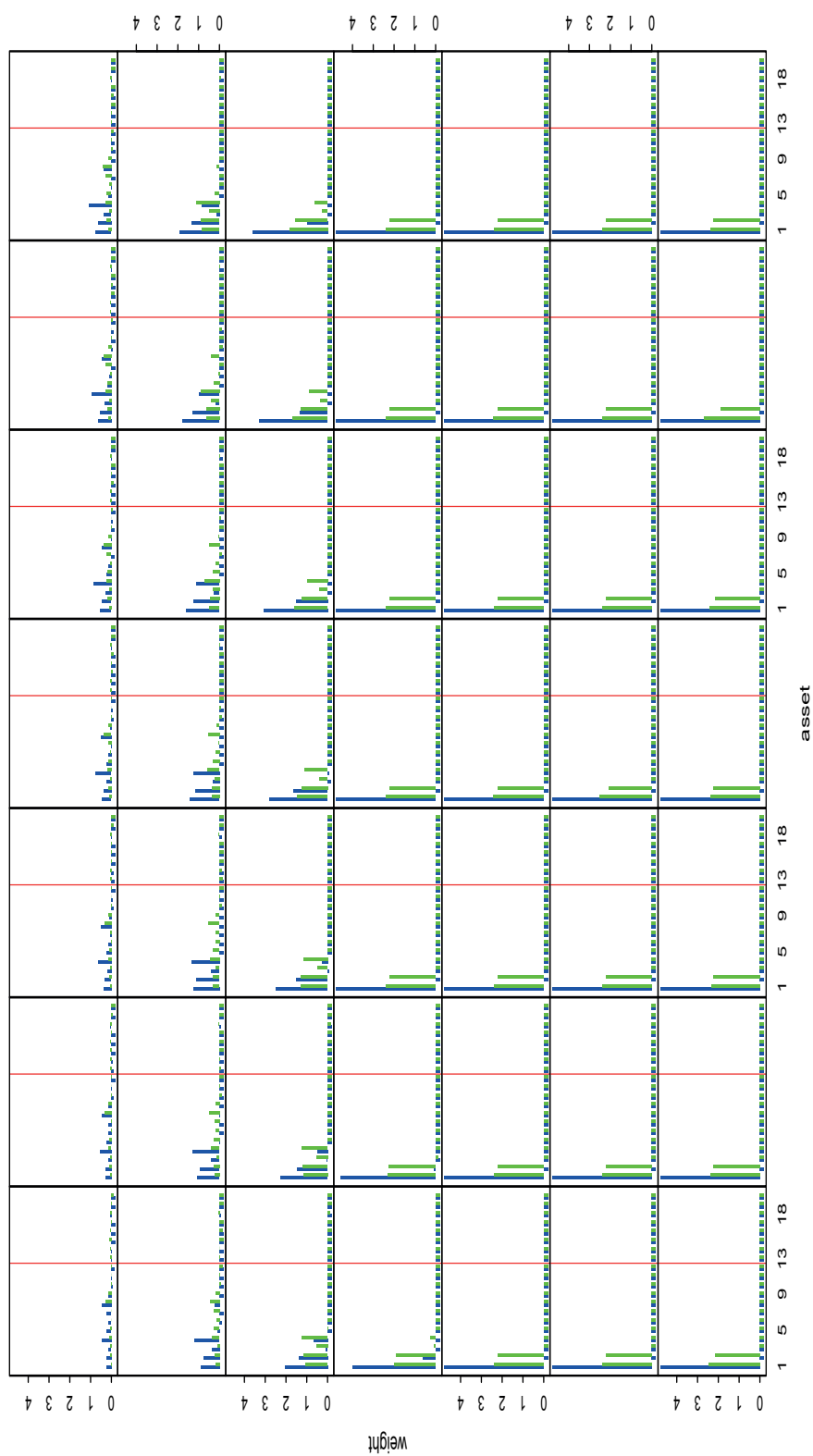Figure C.7: each subfigure plots the optimal weights of the type II problem both with added location uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
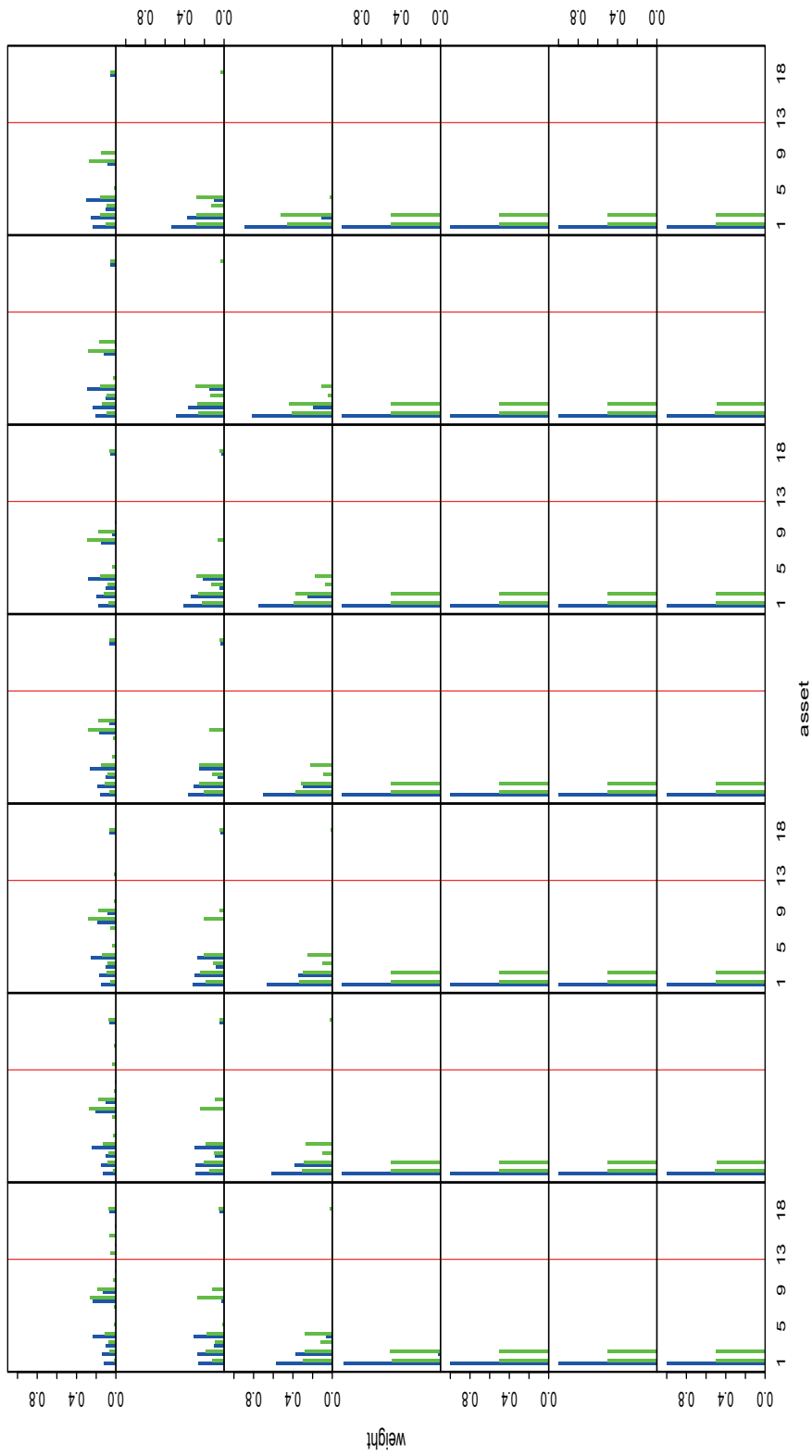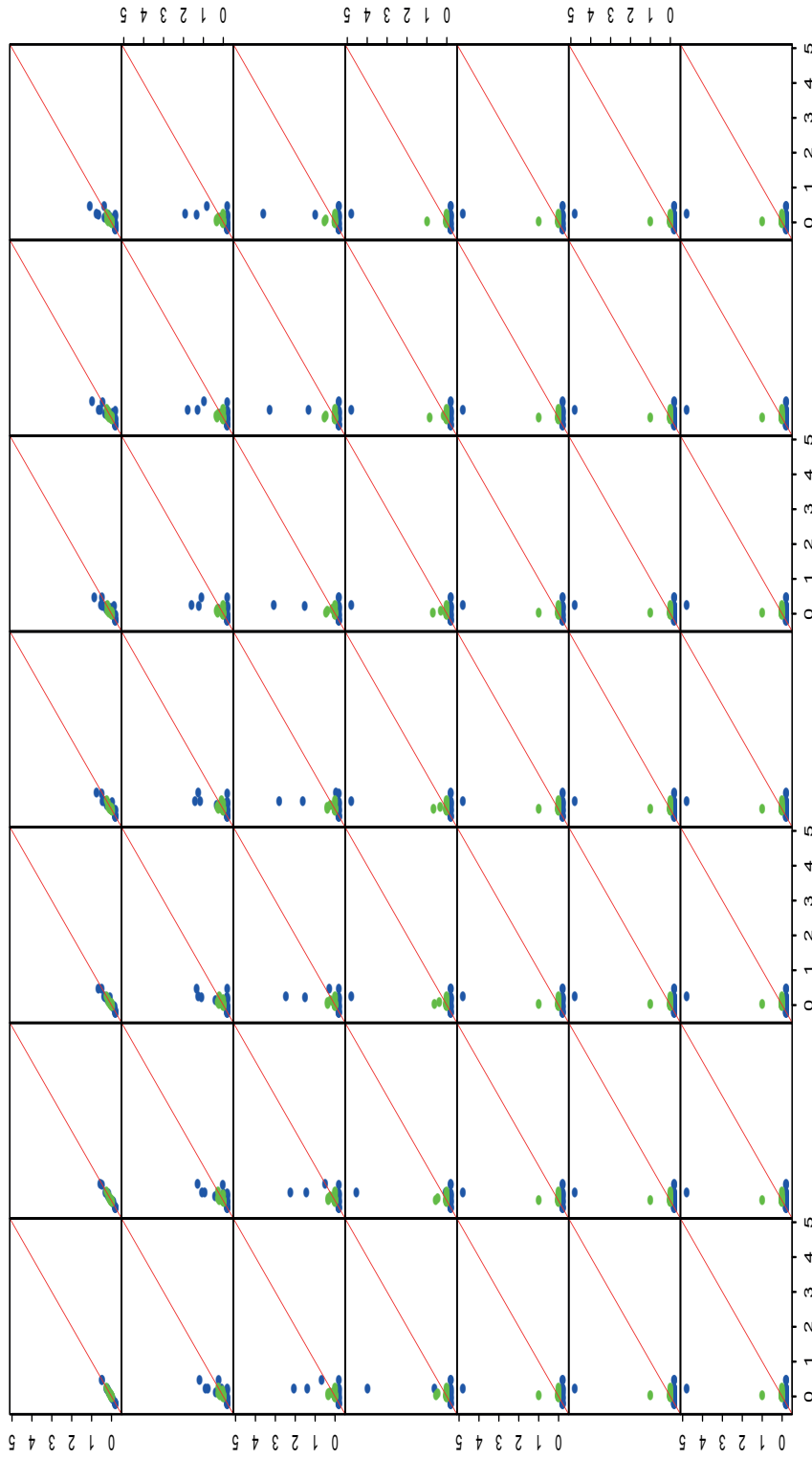
Figure C.8: each subfigure plots the optimal weights of the type III problem both with added location uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
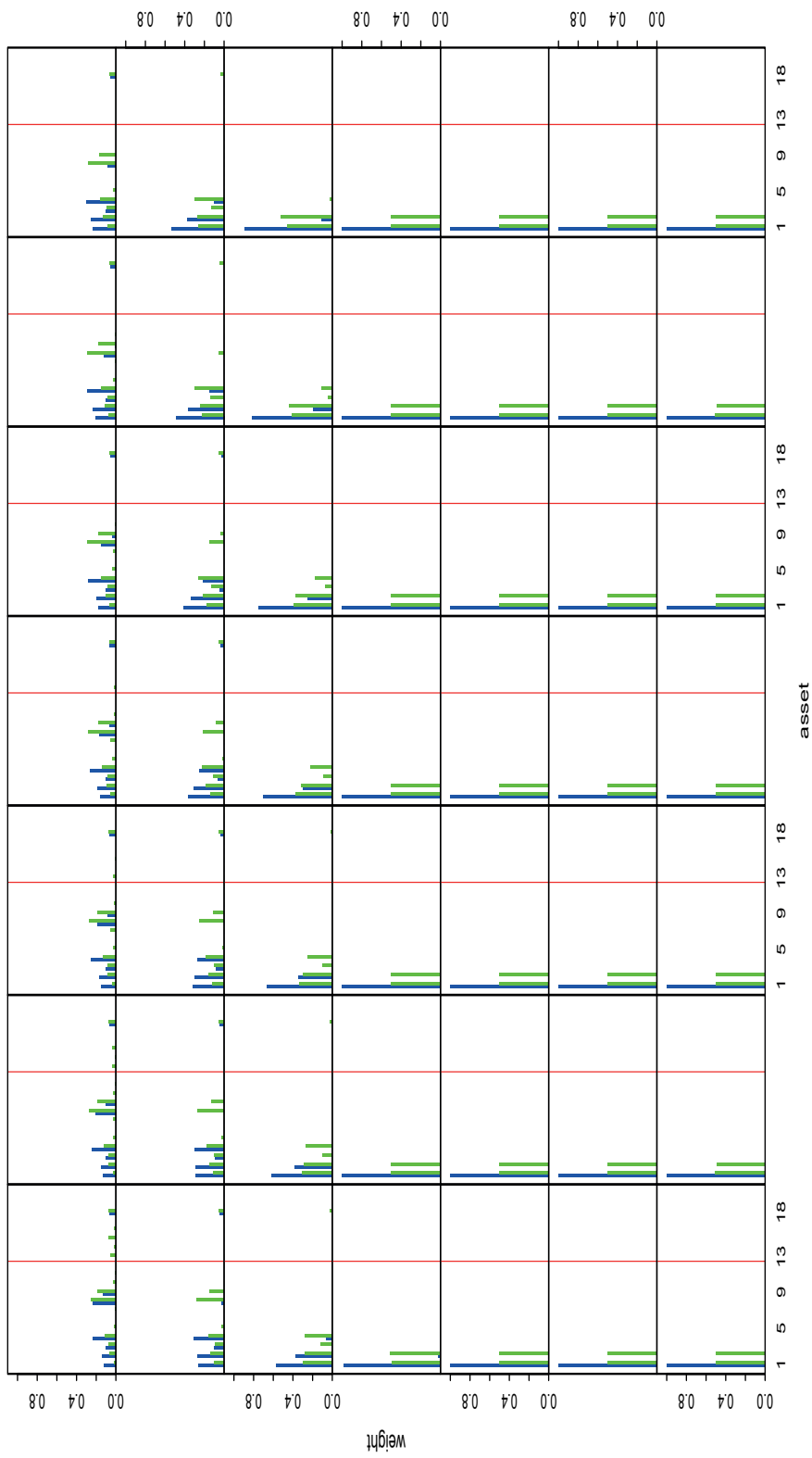
Figure C.9: each subfigure plots the optimal weights of the type IV problem both with added location uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
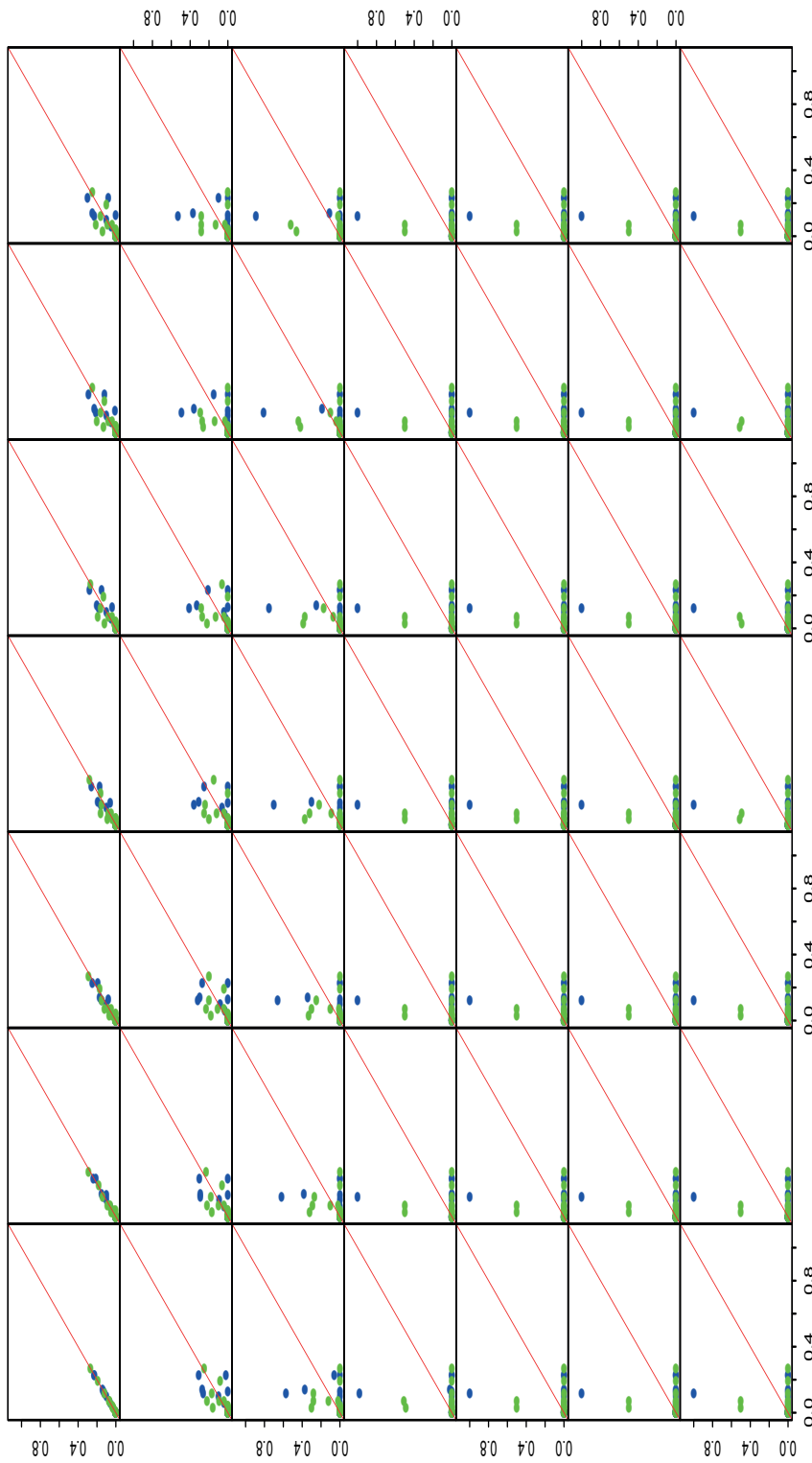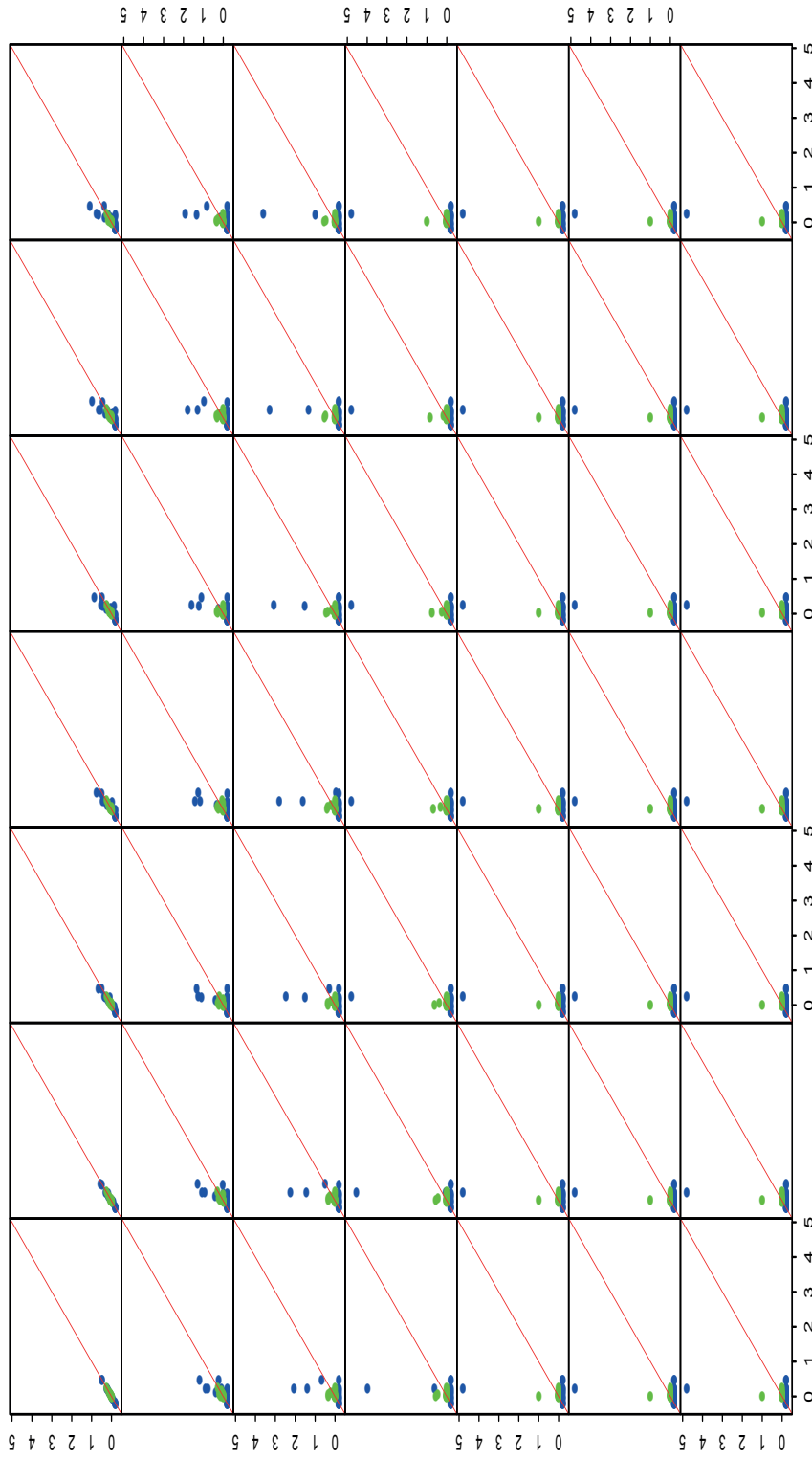
Figure C.10: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type II problem both with added location uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
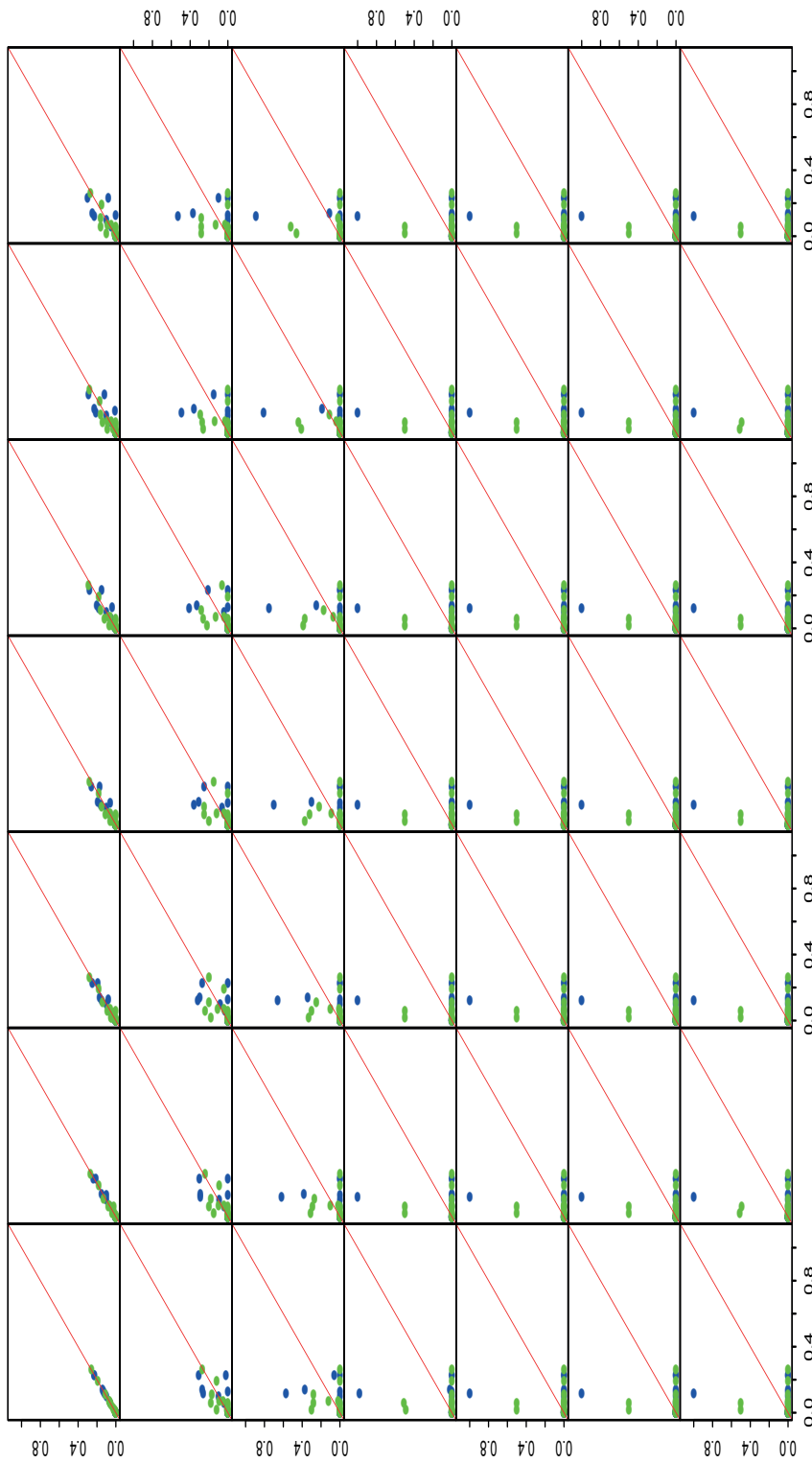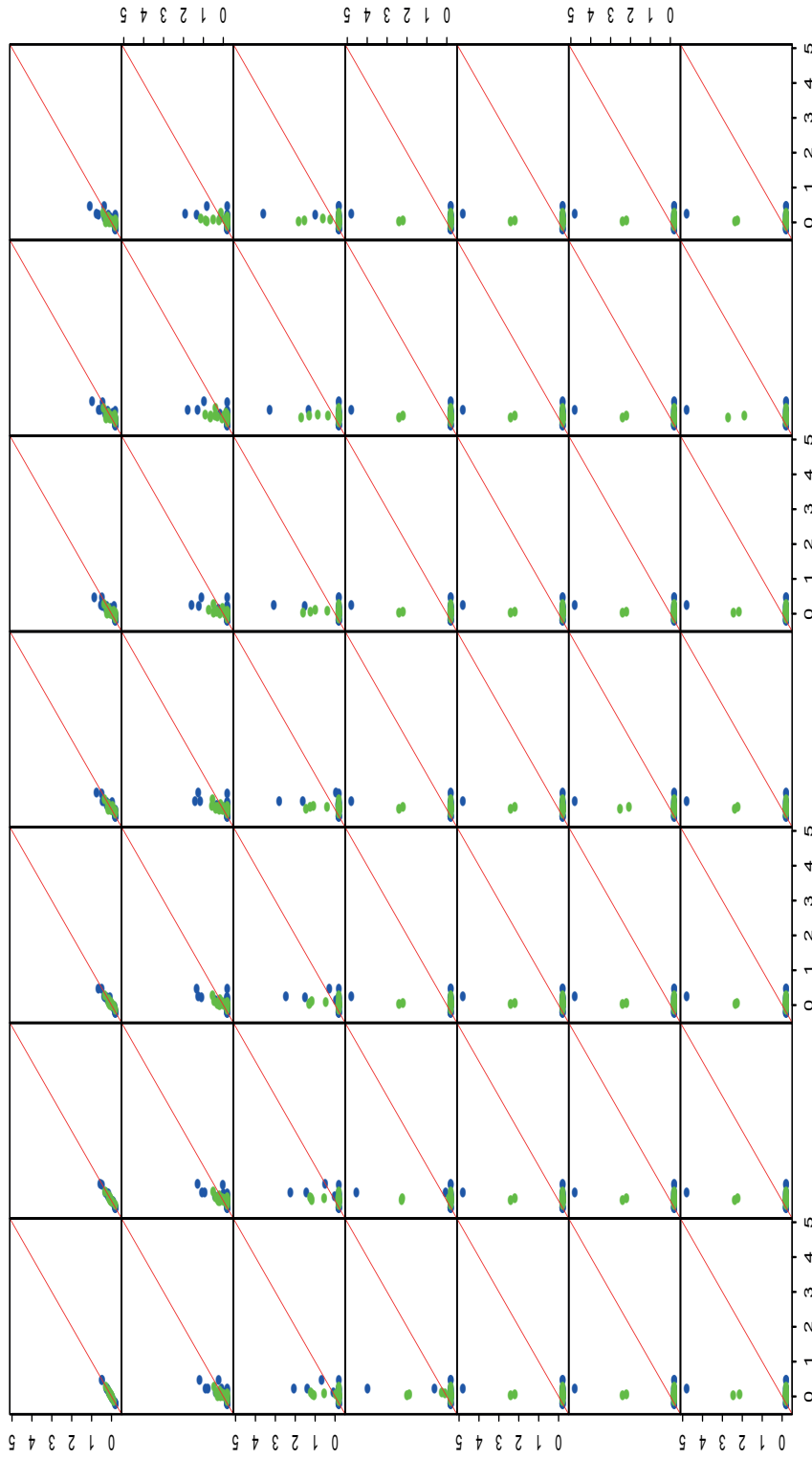
Figure C.11: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type III problem both with added location uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.12: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type IV problem both with added location uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
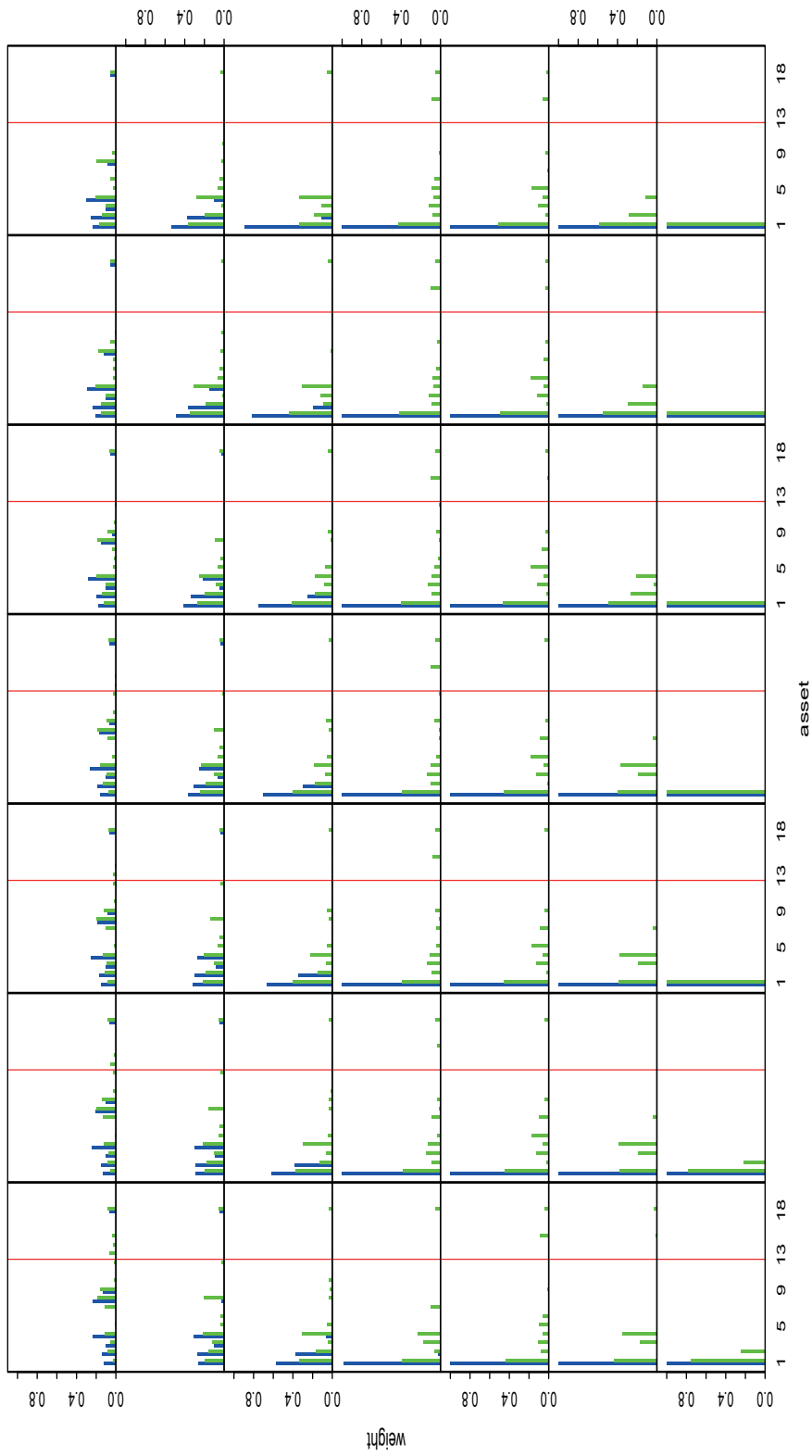
Figure C.13: each subfigure plots the optimal weights of the type II problem both with added location uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
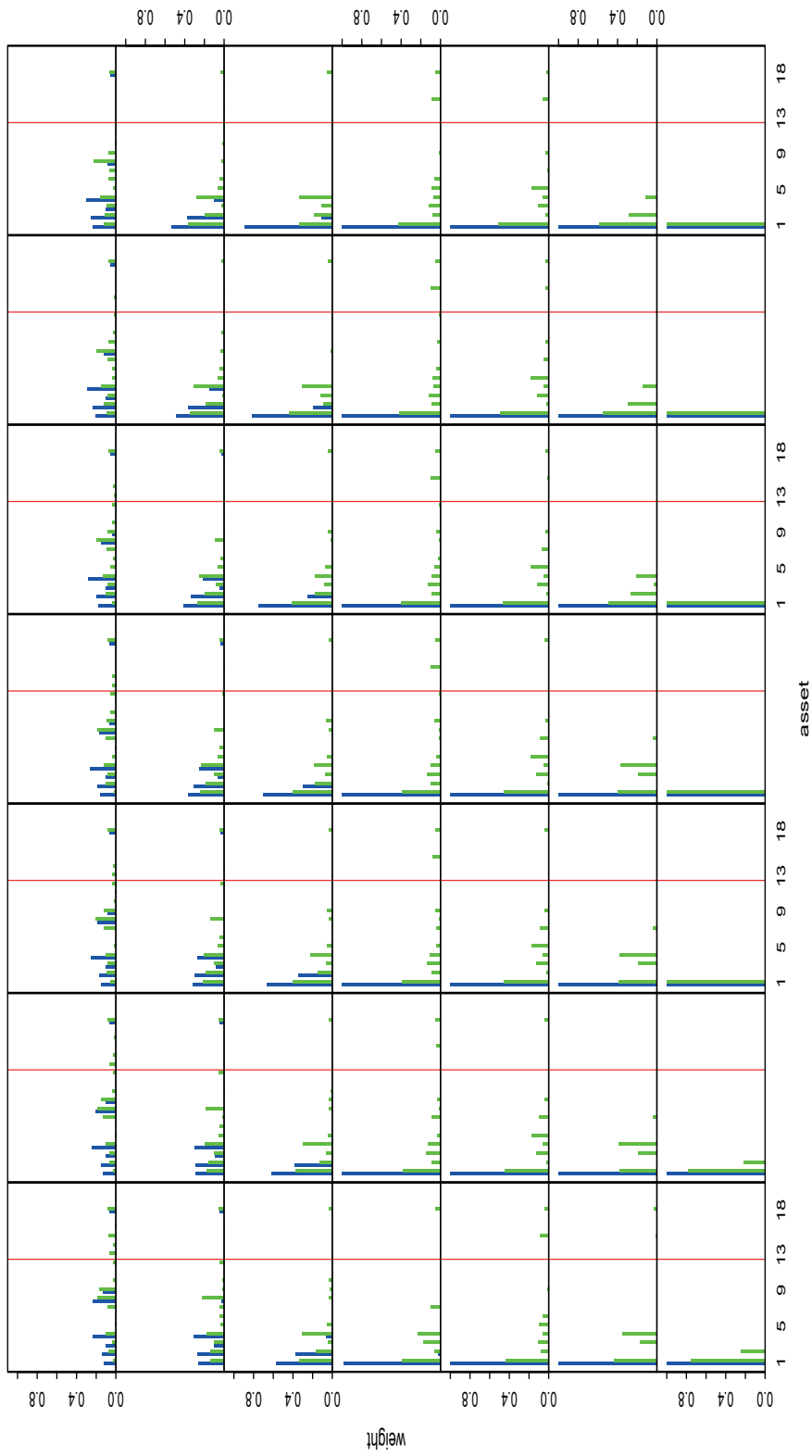
Figure C.14: each subfigure plots the optimal weights of the type III problem both with added location uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
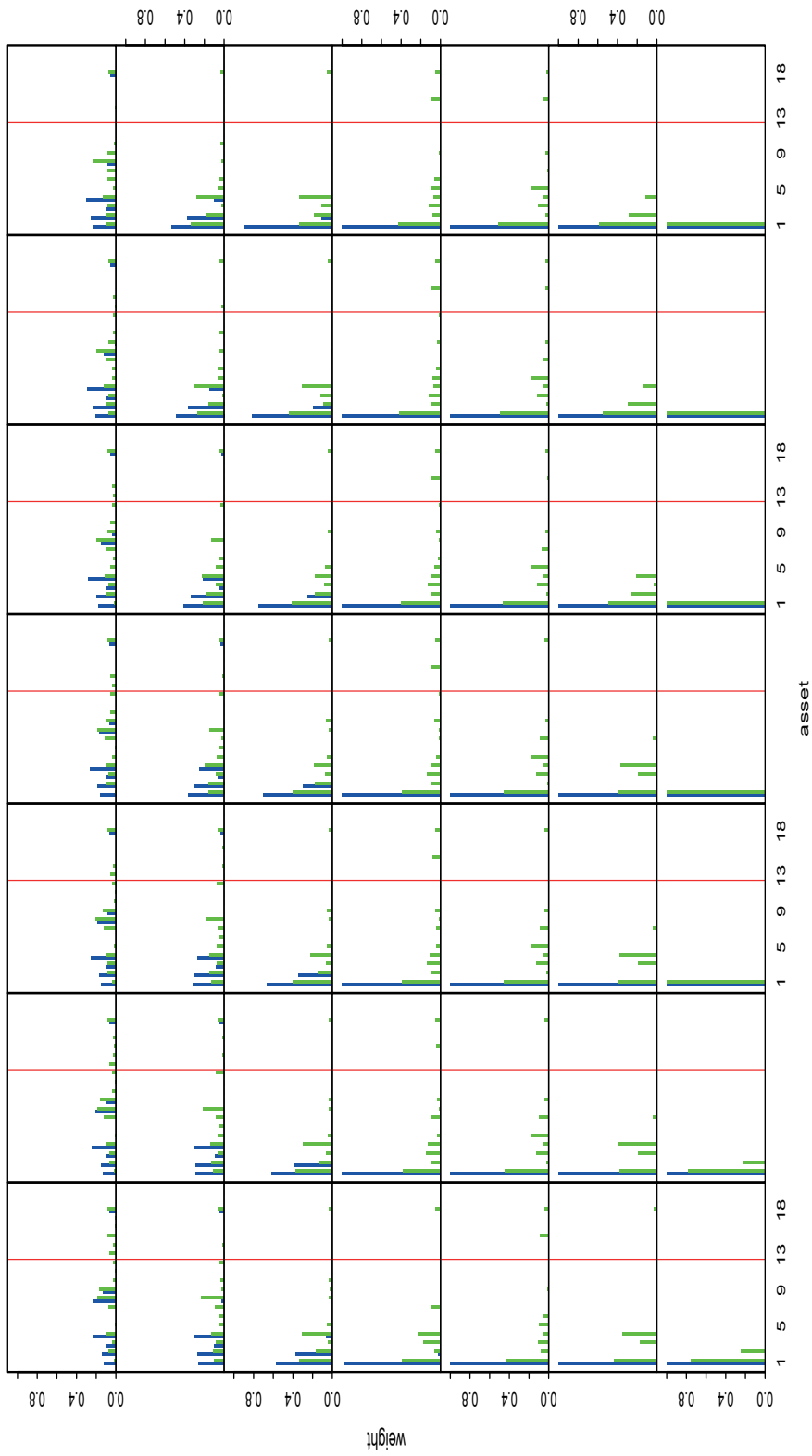
Figure C.15: each subfigure plots the optimal weights of the type IV problem both with added location uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
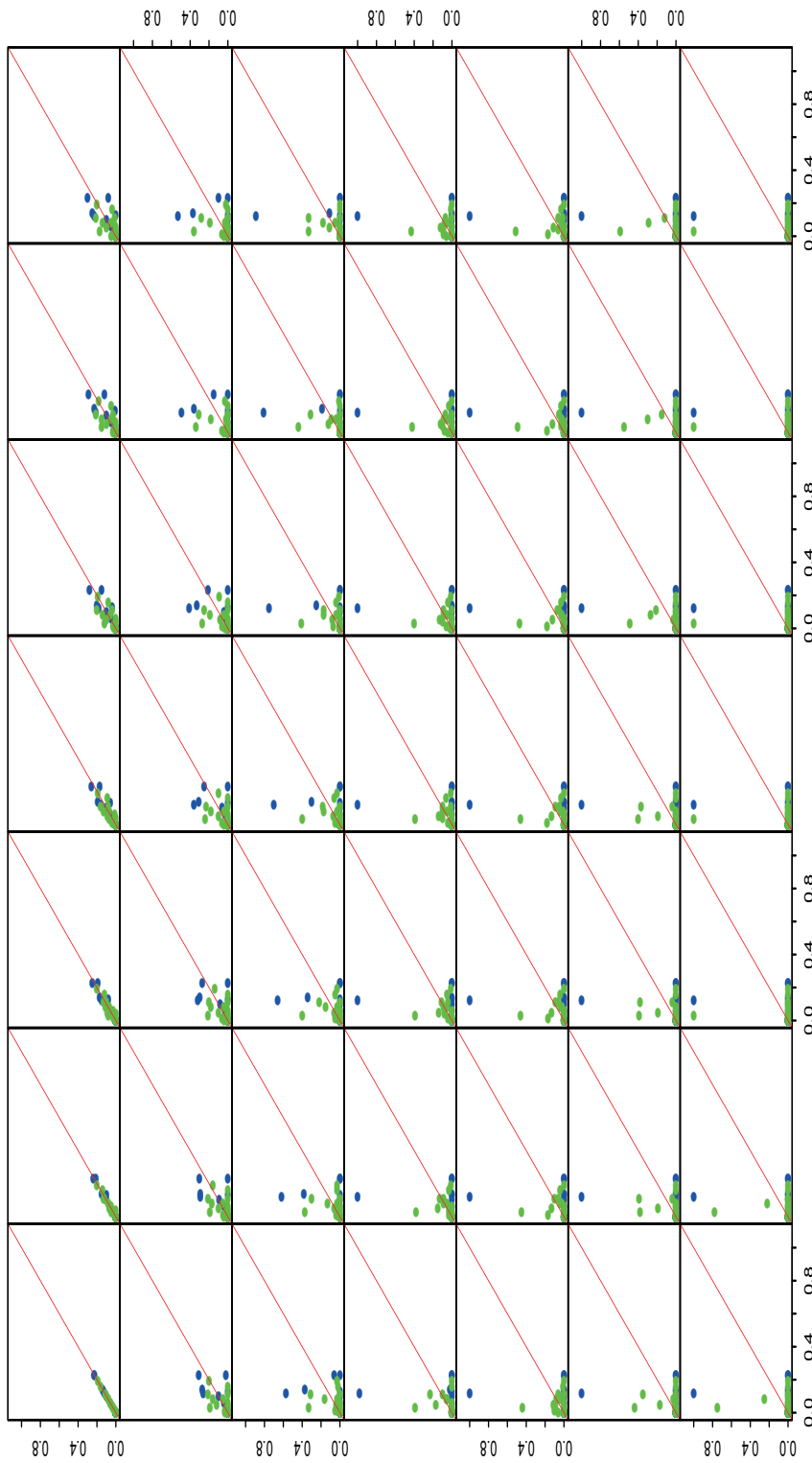
Figure C.16: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type II problem both with added location uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
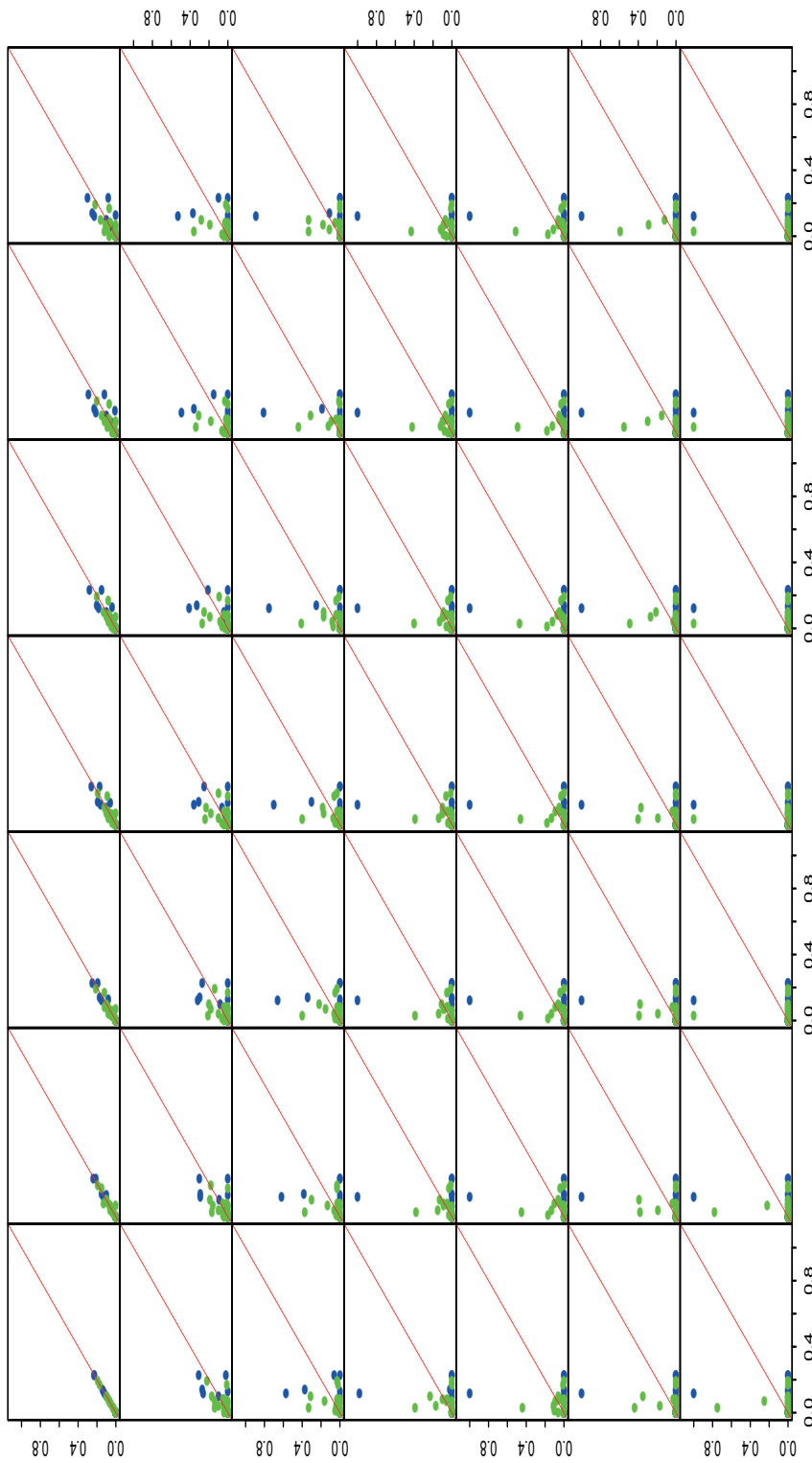
Figure C.17: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type III problem both with added location uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
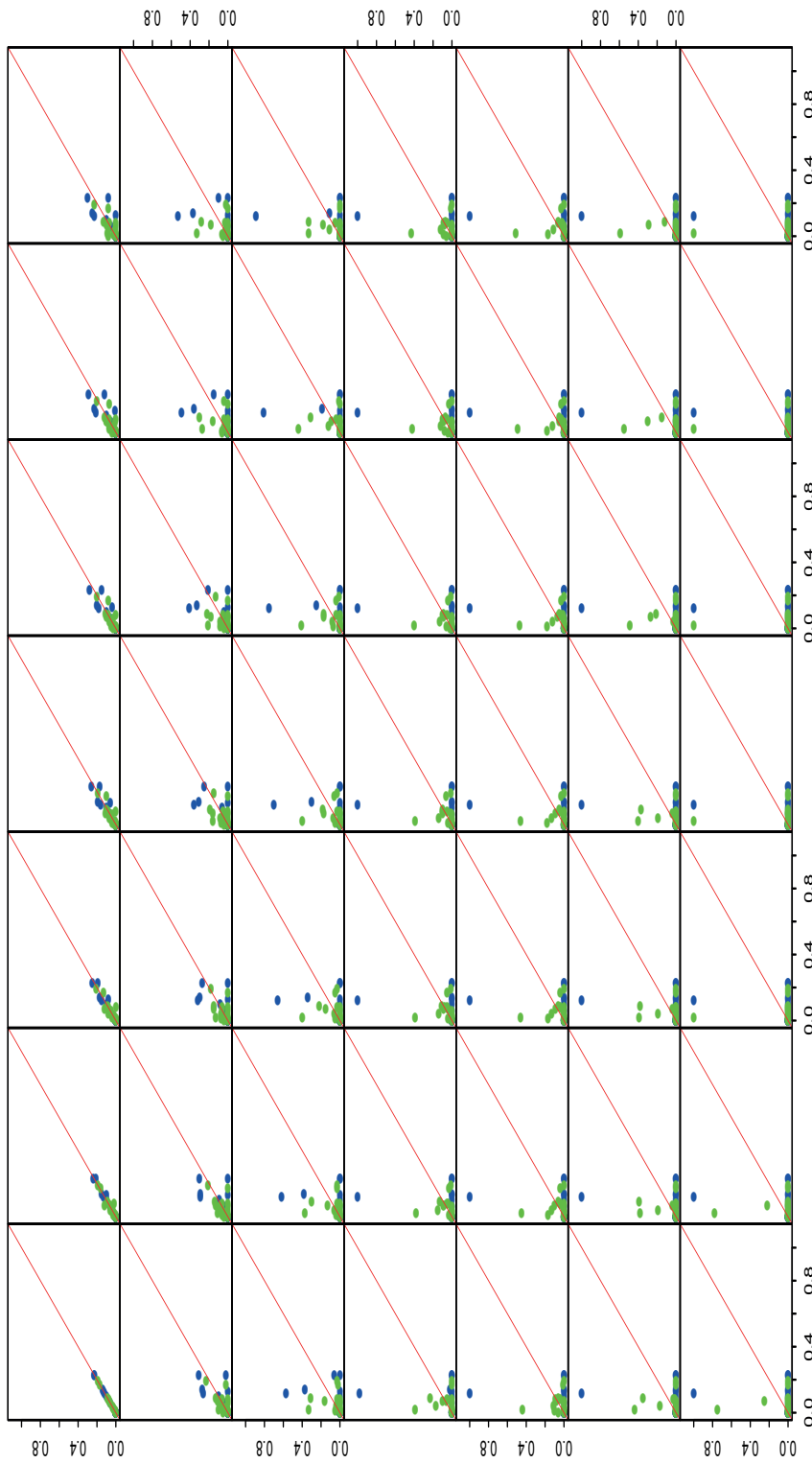
Figure C.18: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type IV problem both with added location uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.19: each subfigure plots the optimal weights of the type II problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
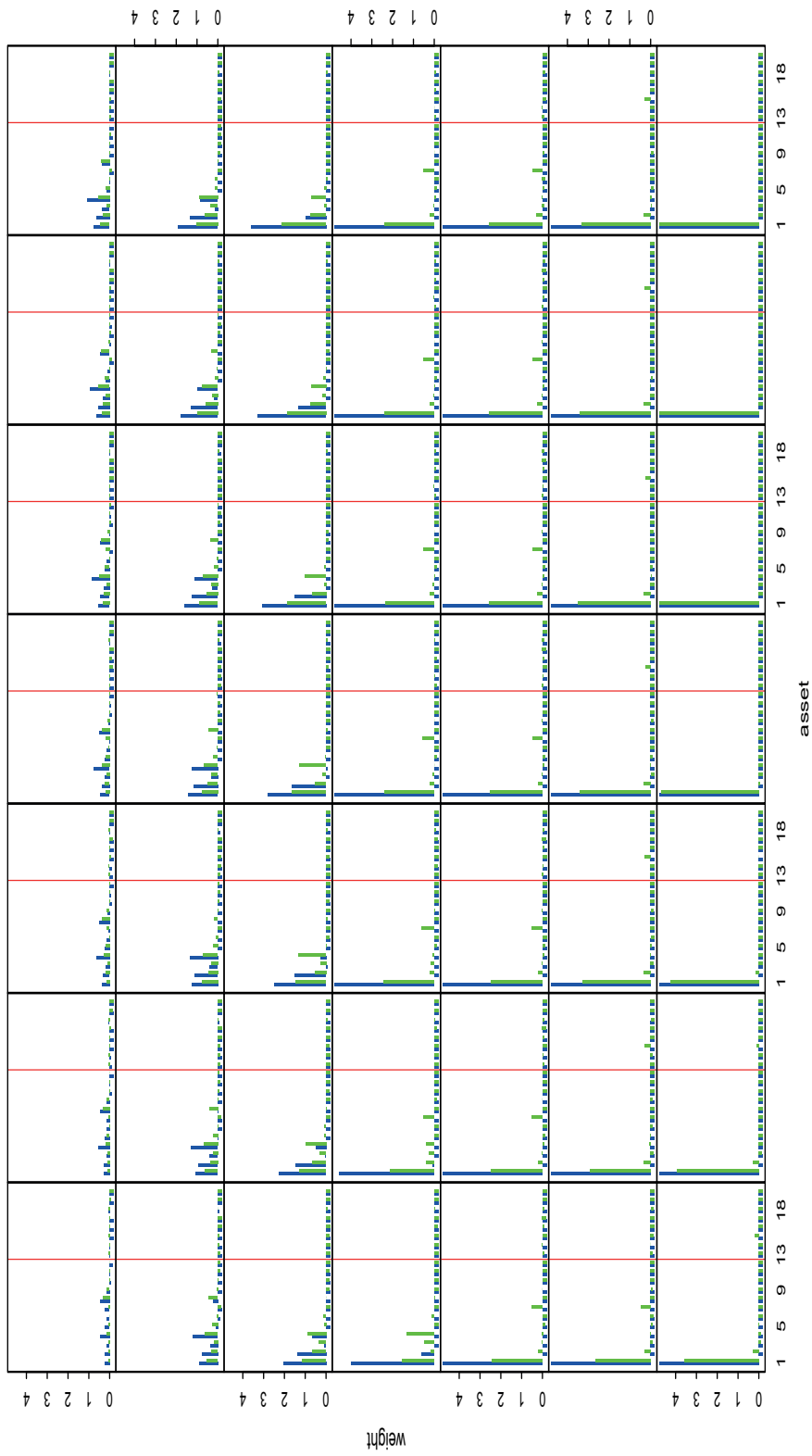
Figure C.20: each subfigure plots the optimal weights of the type III problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
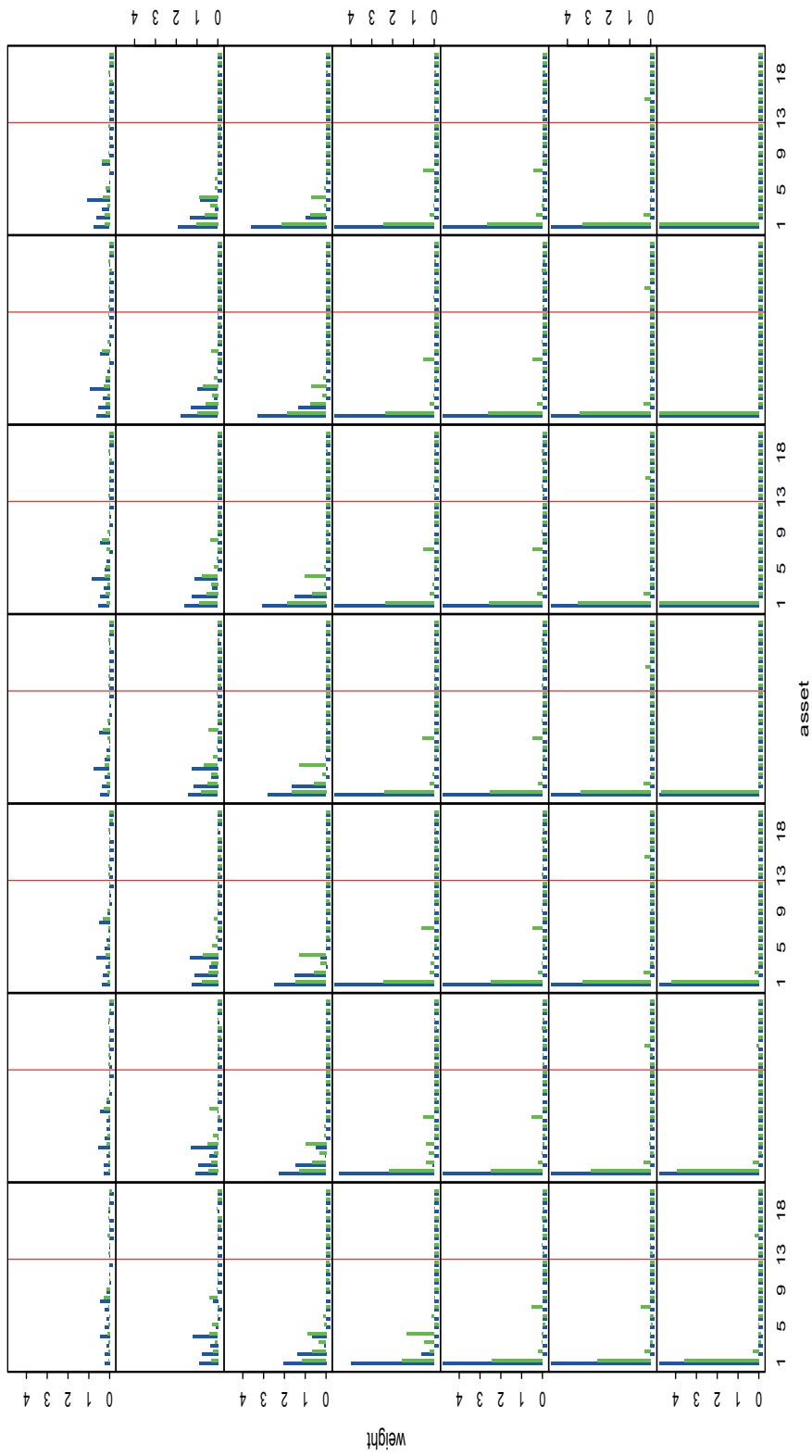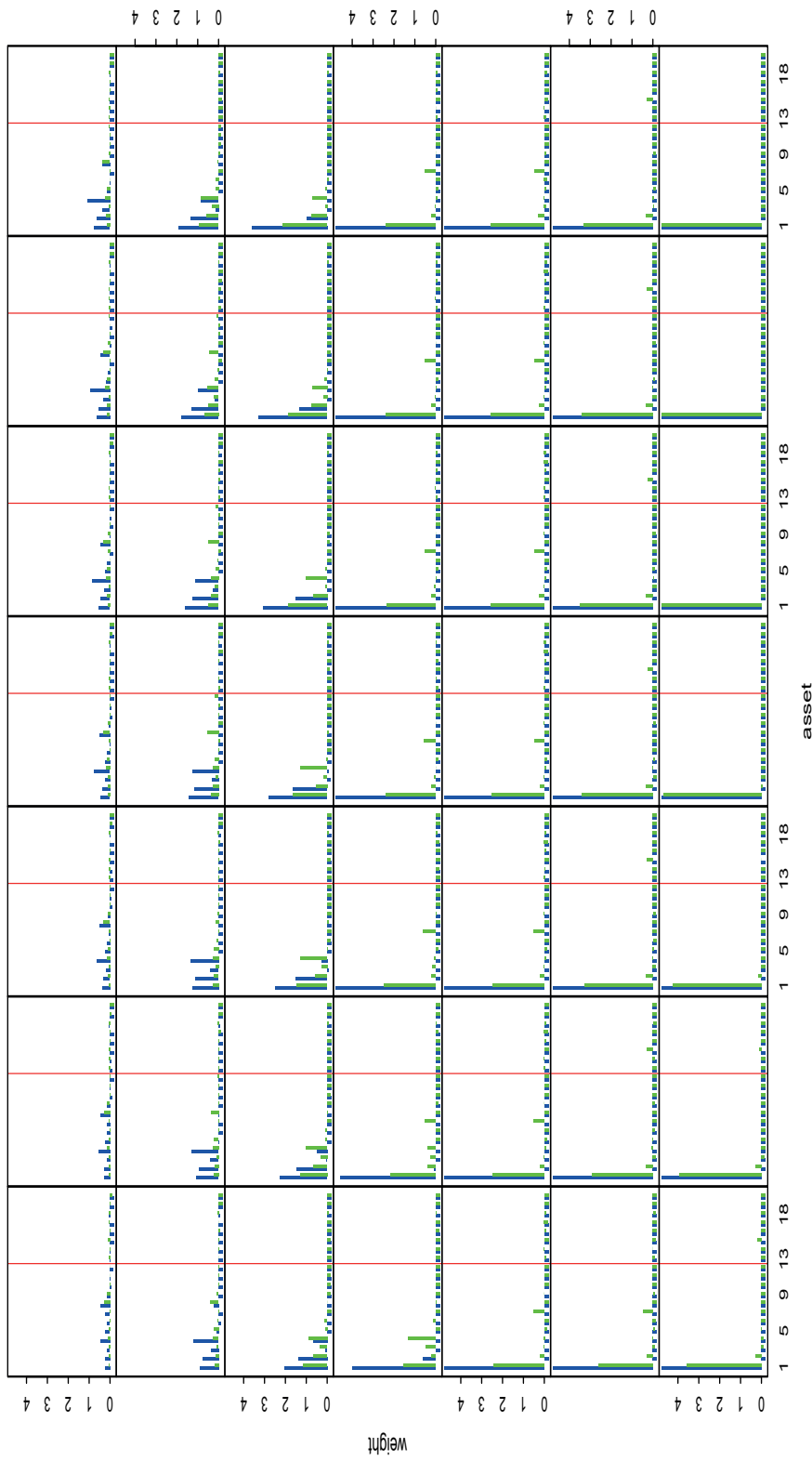
Figure C.21: each subfigure plots the optimal weights of the type IV problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.22: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type II problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \dots, 0.49$.

Figure C.23: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type III problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.24: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type IV problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.25: each subfigure plots the optimal weights of the type II problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
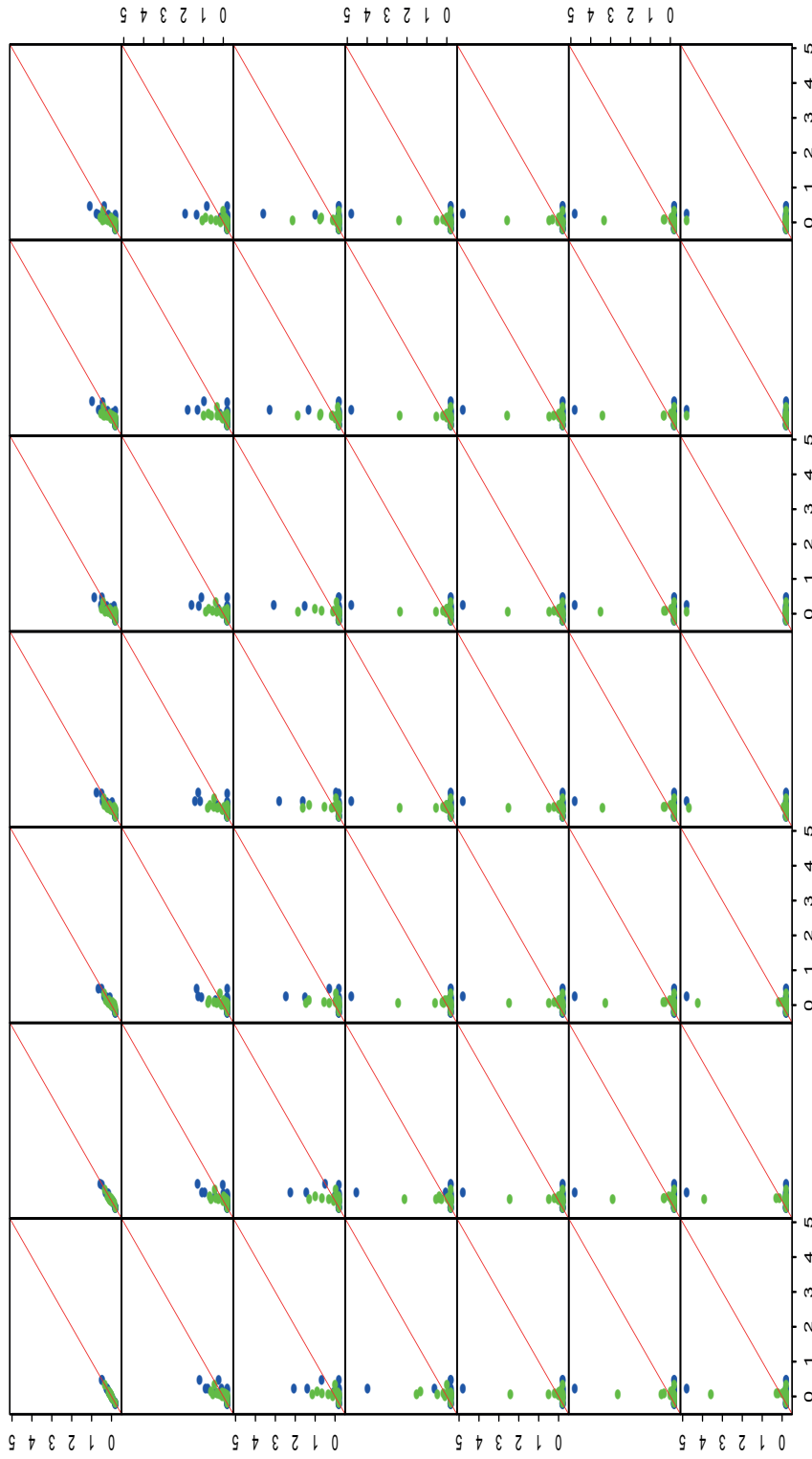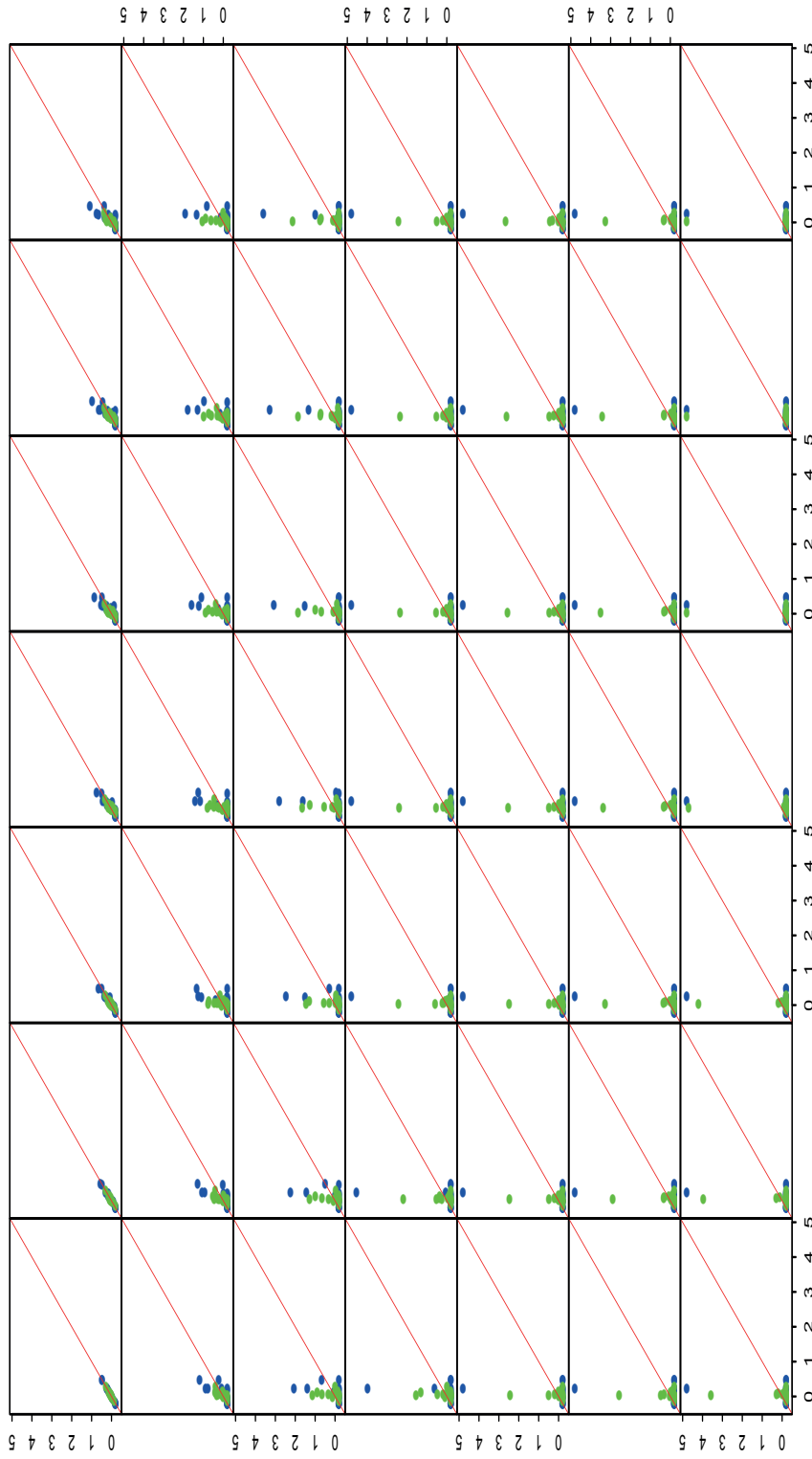
Figure C.26: each subfigure plots the optimal weights of the type III problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.27: each subfigure plots the optimal weights of the type IV problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.28: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type II problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
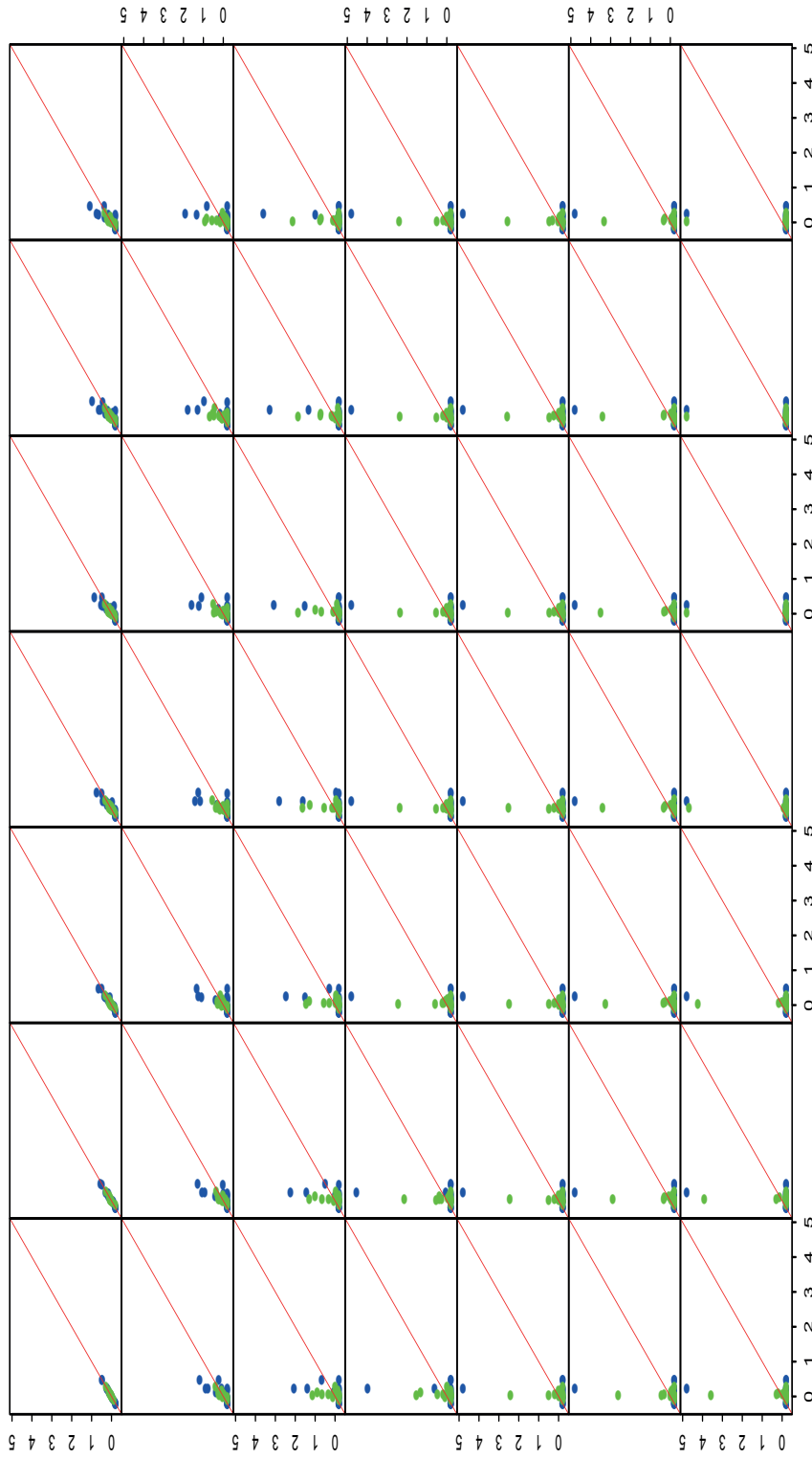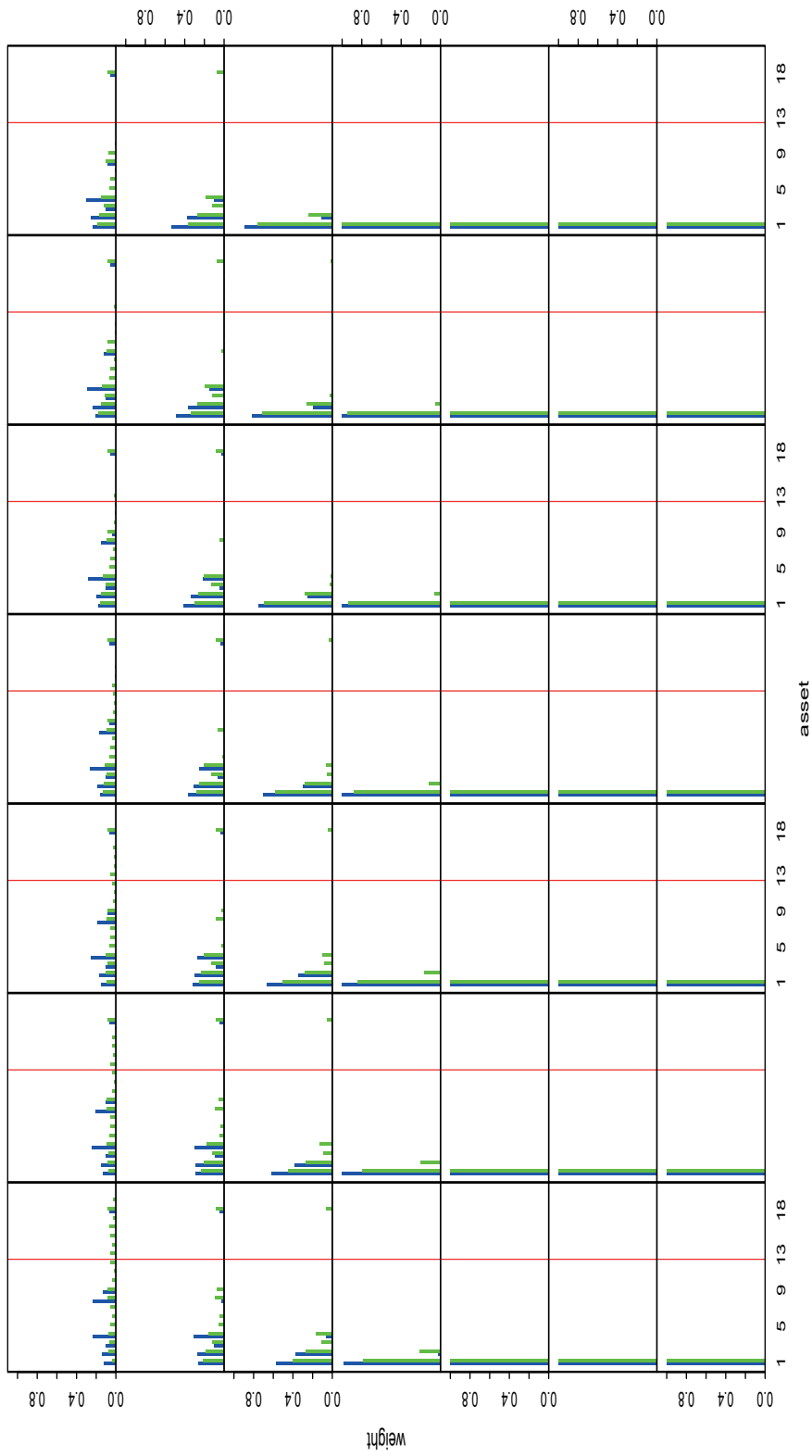
Figure C.29: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type III problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
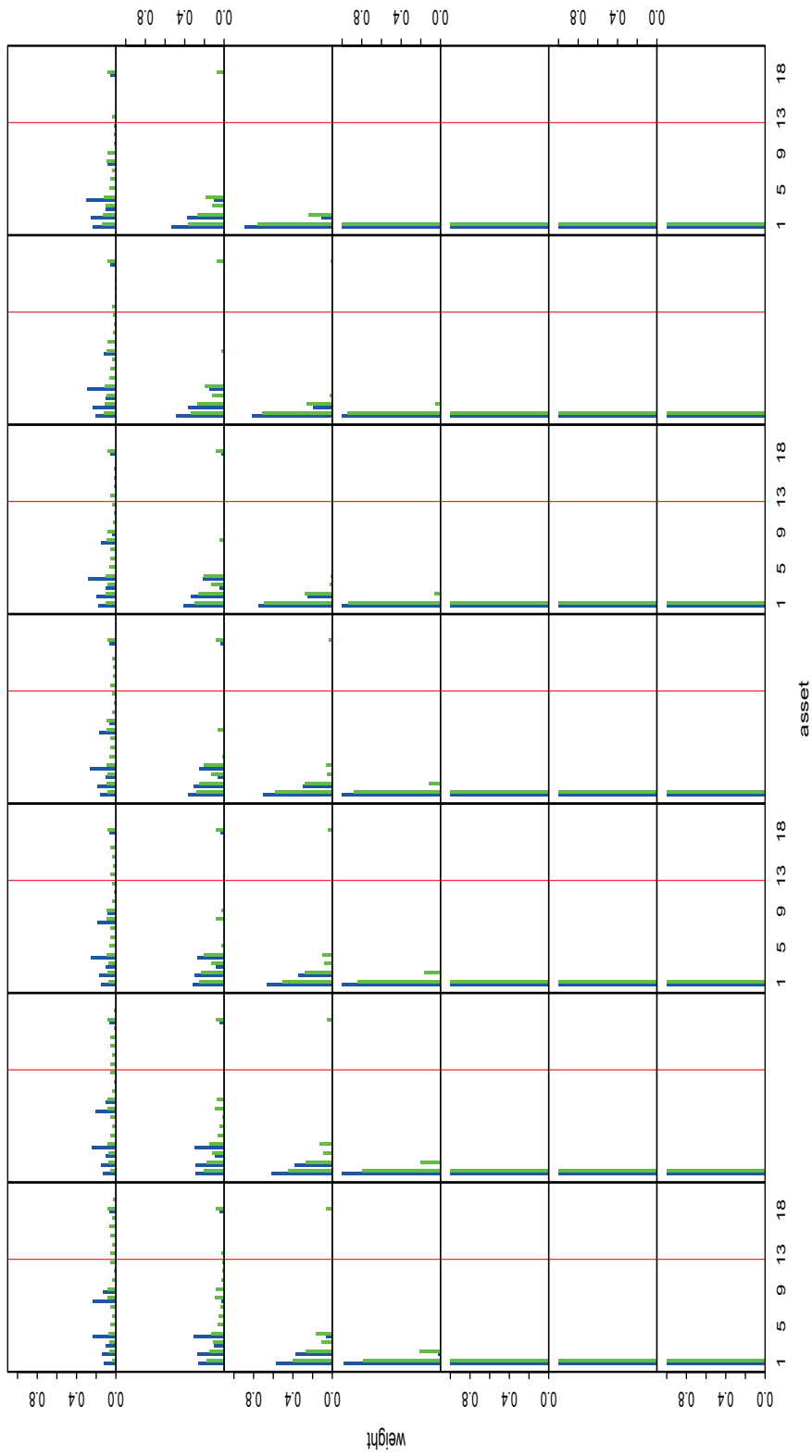
Figure C.30: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type IV problem both with added eigenvalue uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
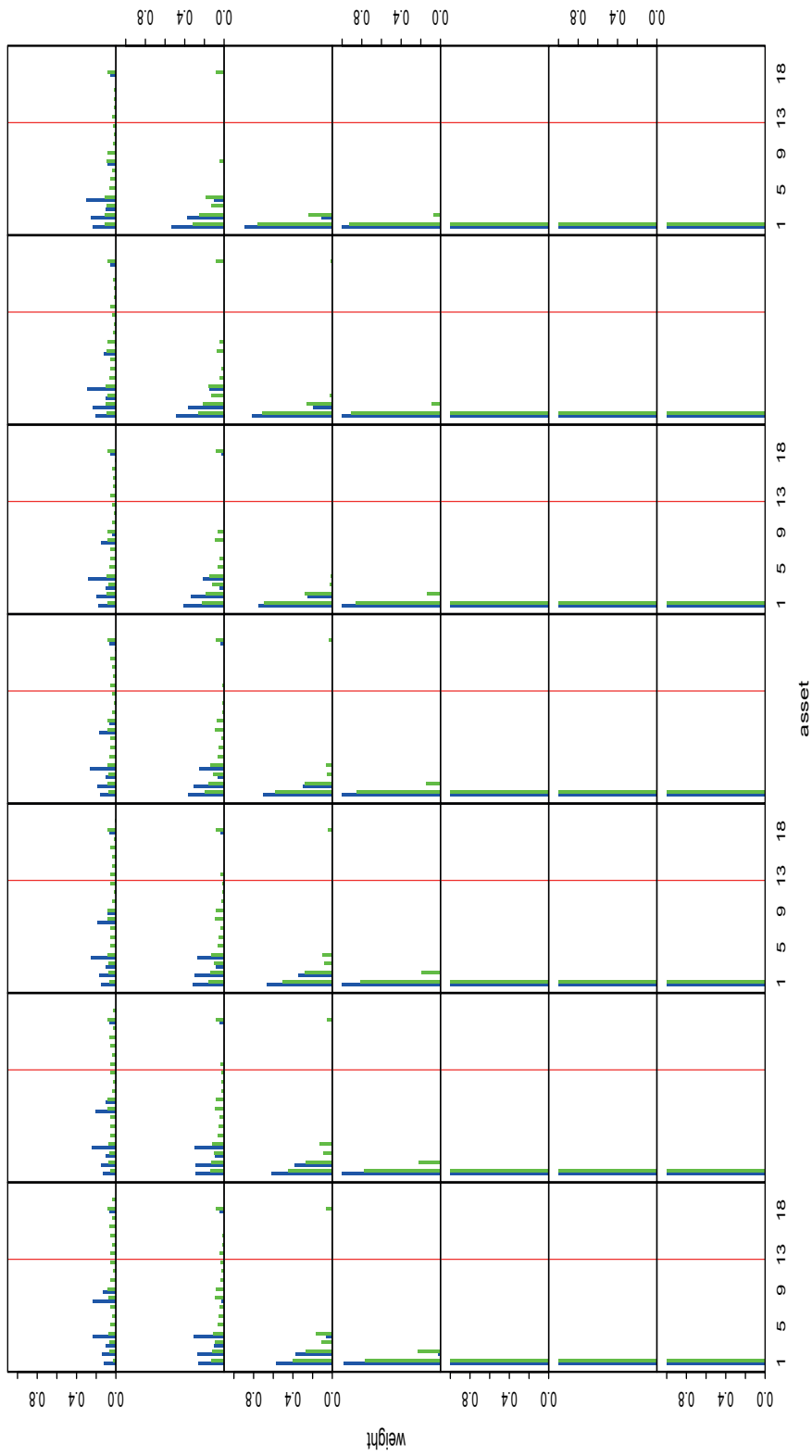
Figure C.31: each subfigure plots the optimal weights of the type II problem both with added eigenvector uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.32: each subfigure plots the optimal weights of the type III problem both with added eigenvector uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
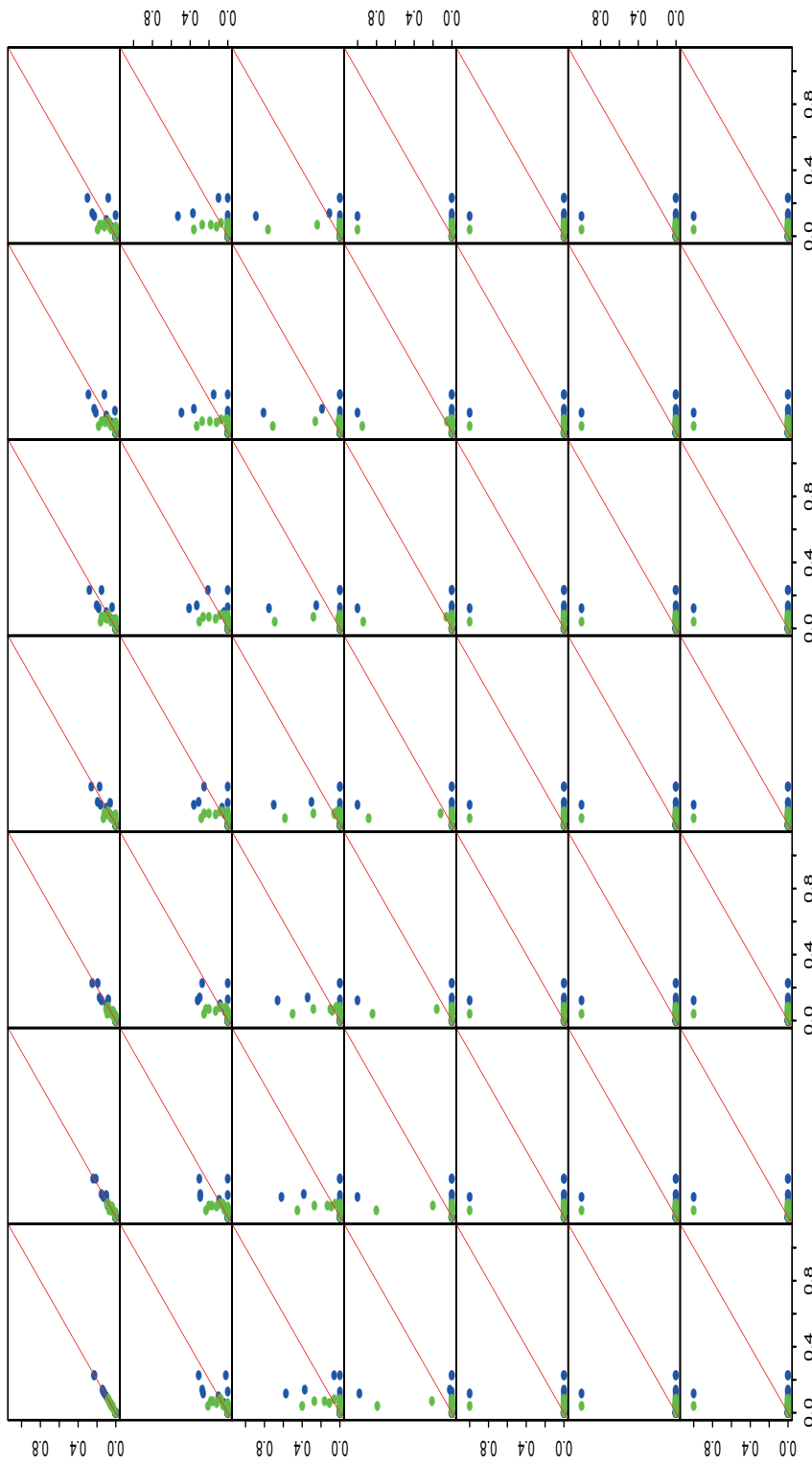
Figure C.33: each subfigure plots the optimal weights of the type IV problem both with added eigenvector uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is disallowed at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.34: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type II problem both with added eigenvector uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
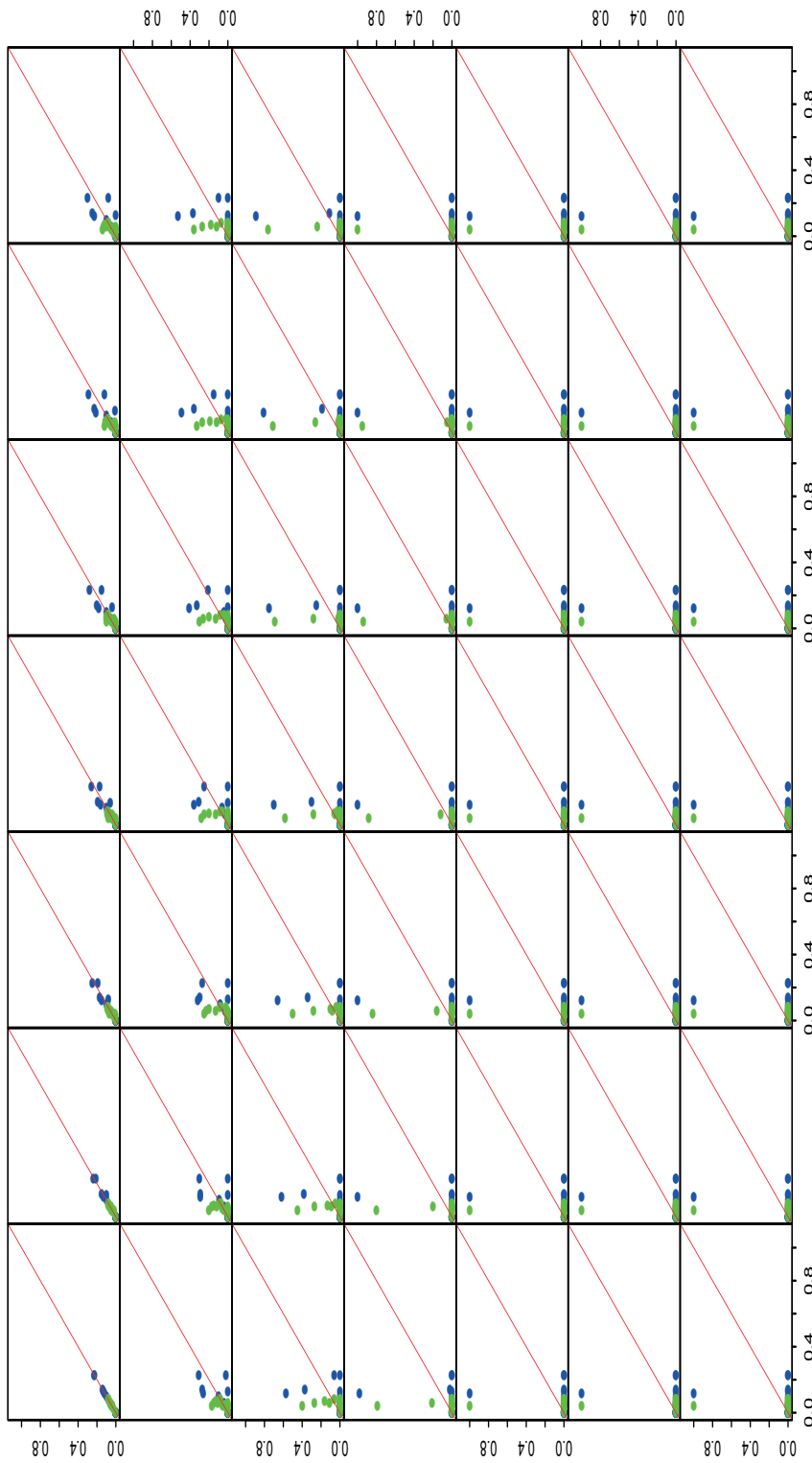
Figure C.35: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type III problem both with added eigenvector uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
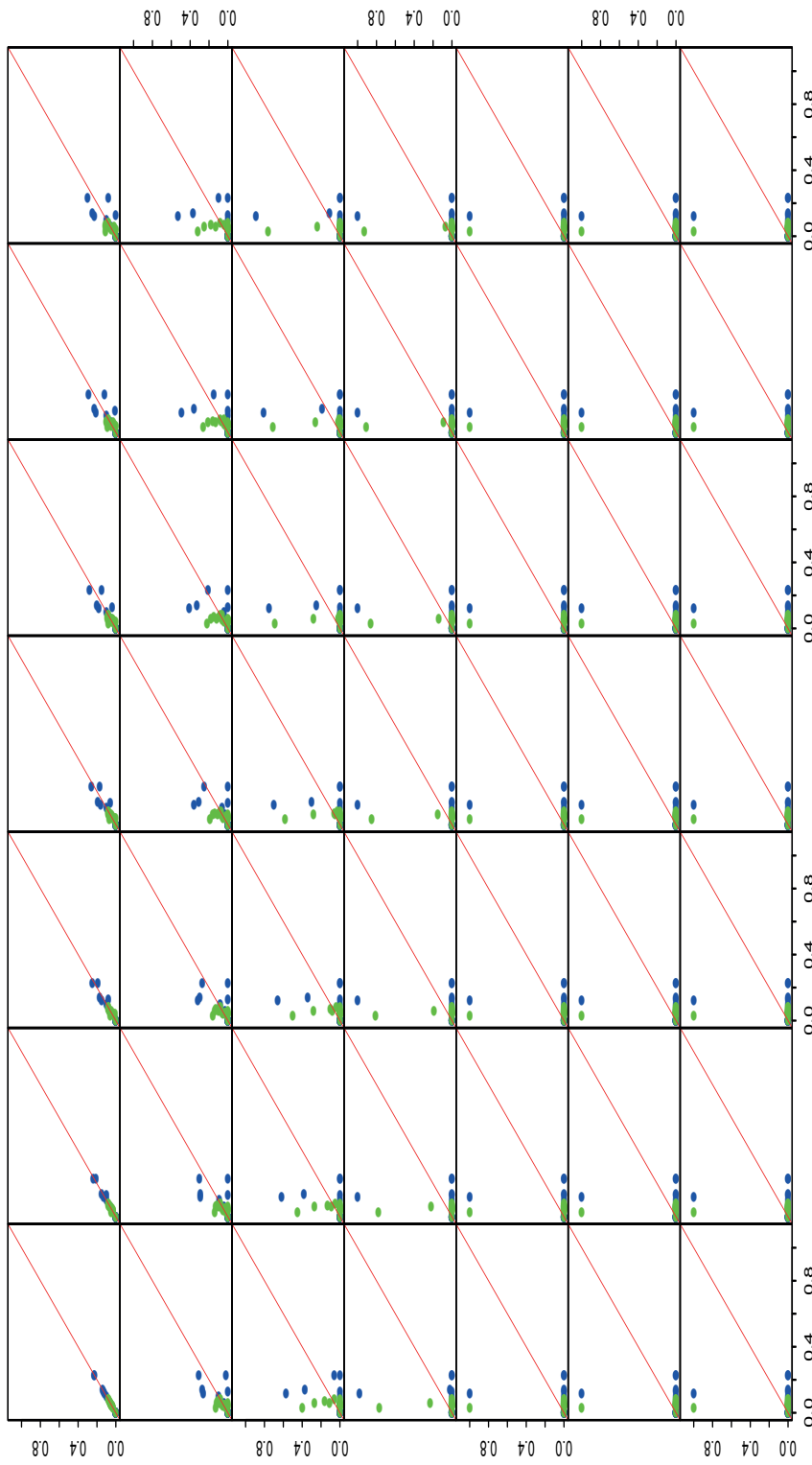
Figure C.36: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type IV problem both with added eigenvector uncertainty (green) and without added uncertainty (blue), where short-selling is disallowed; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
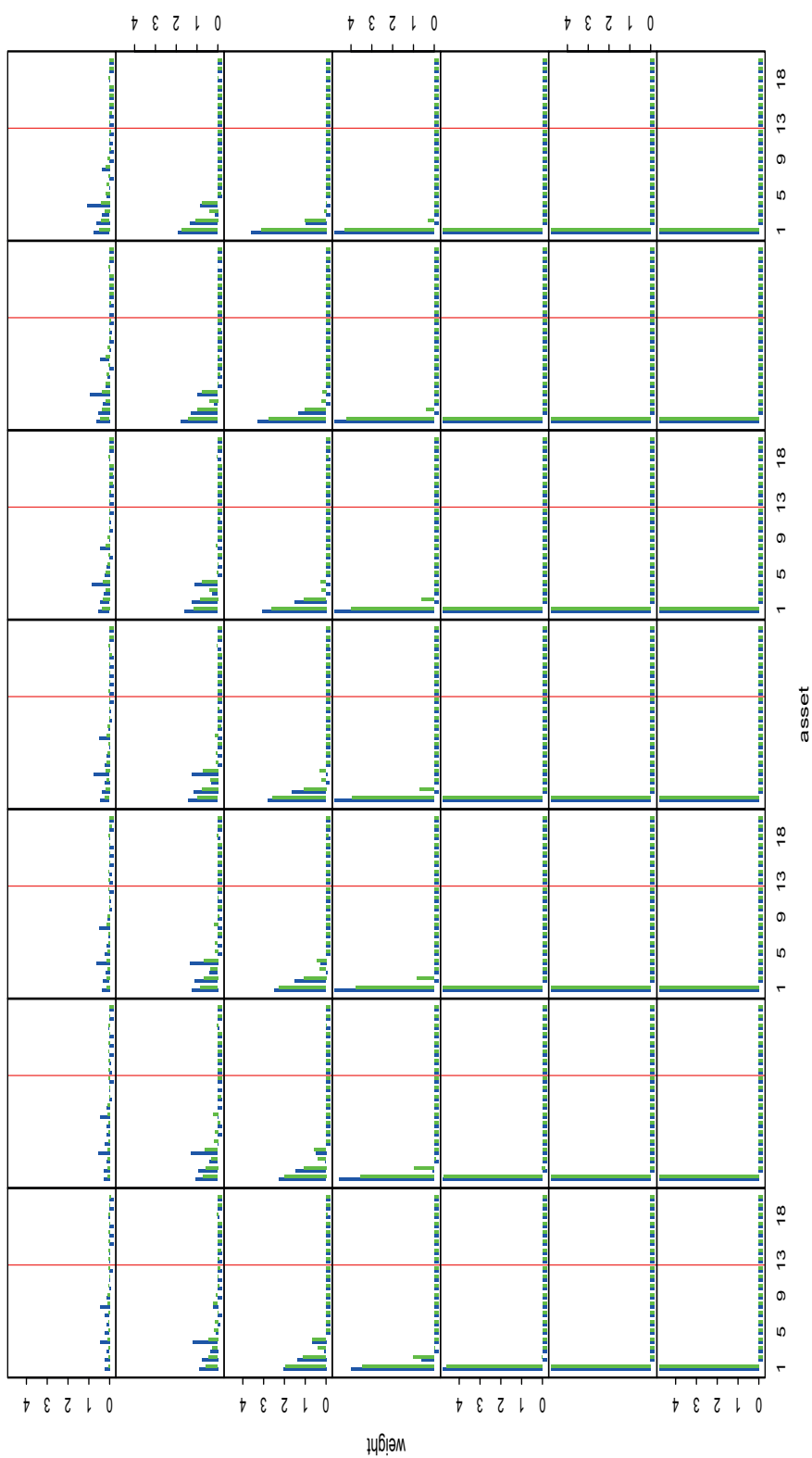
Figure C.37: each subfigure plots the optimal weights of the type II problem both with added eigenvector uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
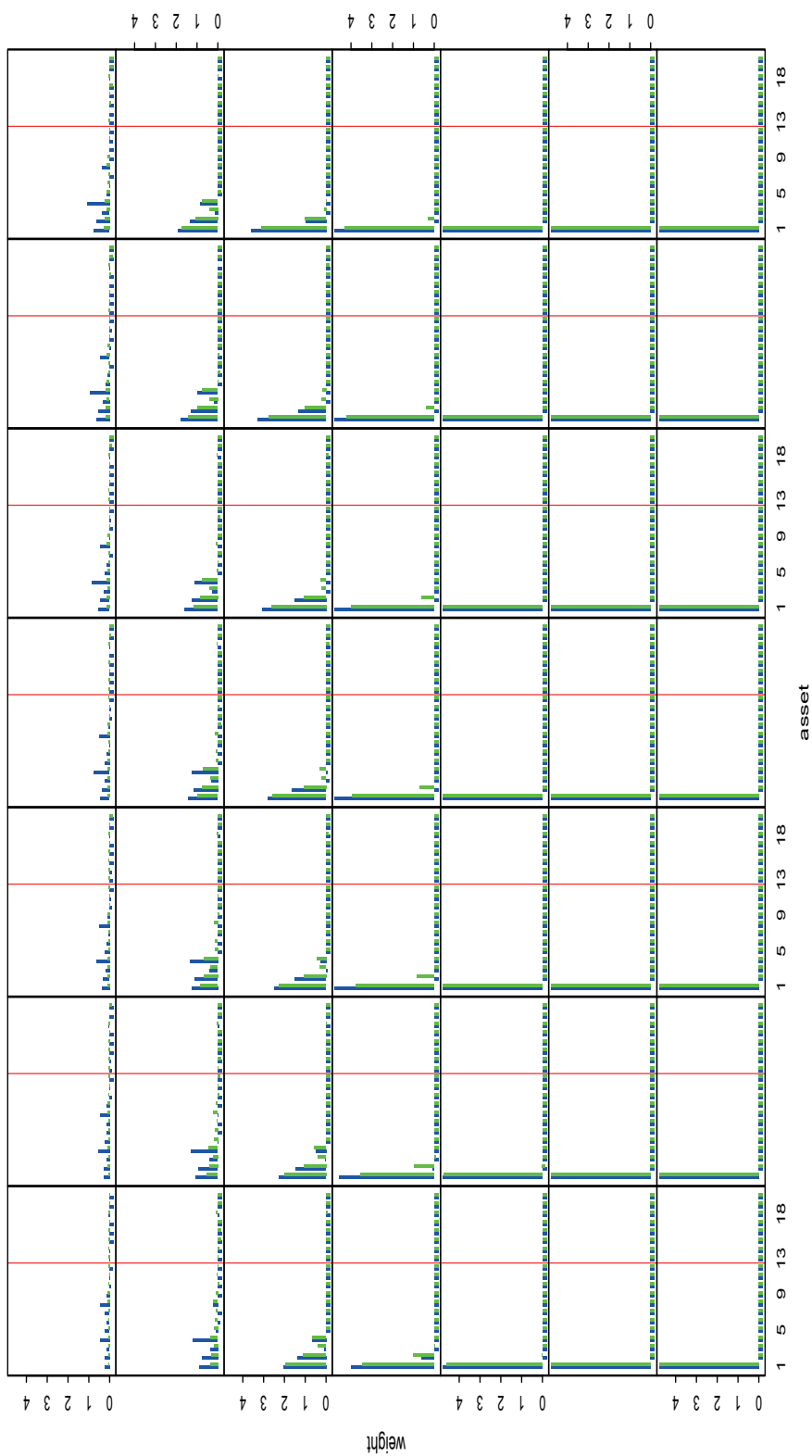
Figure C.38: each subfigure plots the optimal weights of the type III problem both with added eigenvector uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
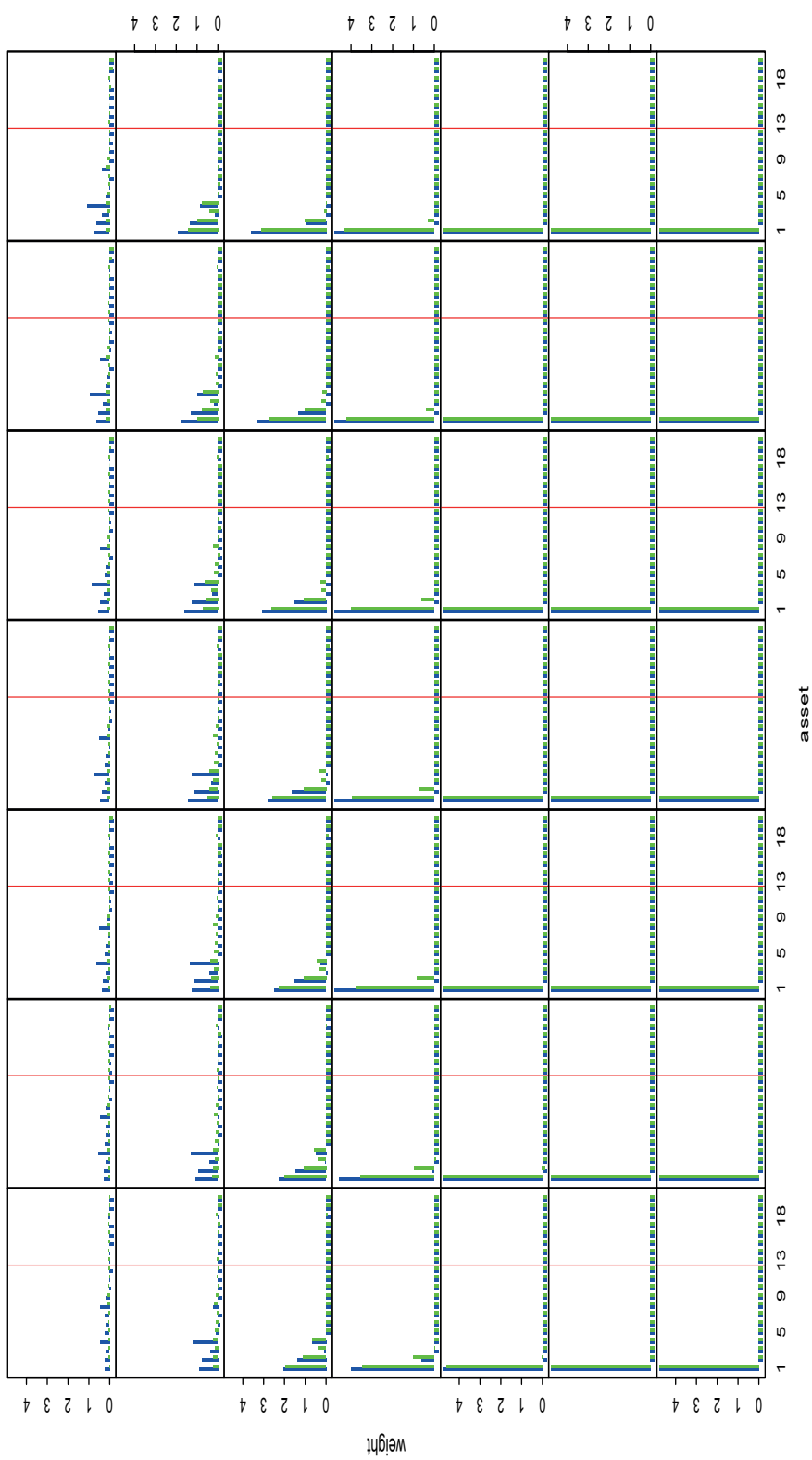
Figure C.39: each subfigure plots the optimal weights of the type IV problem both with added eigenvector uncertainty (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
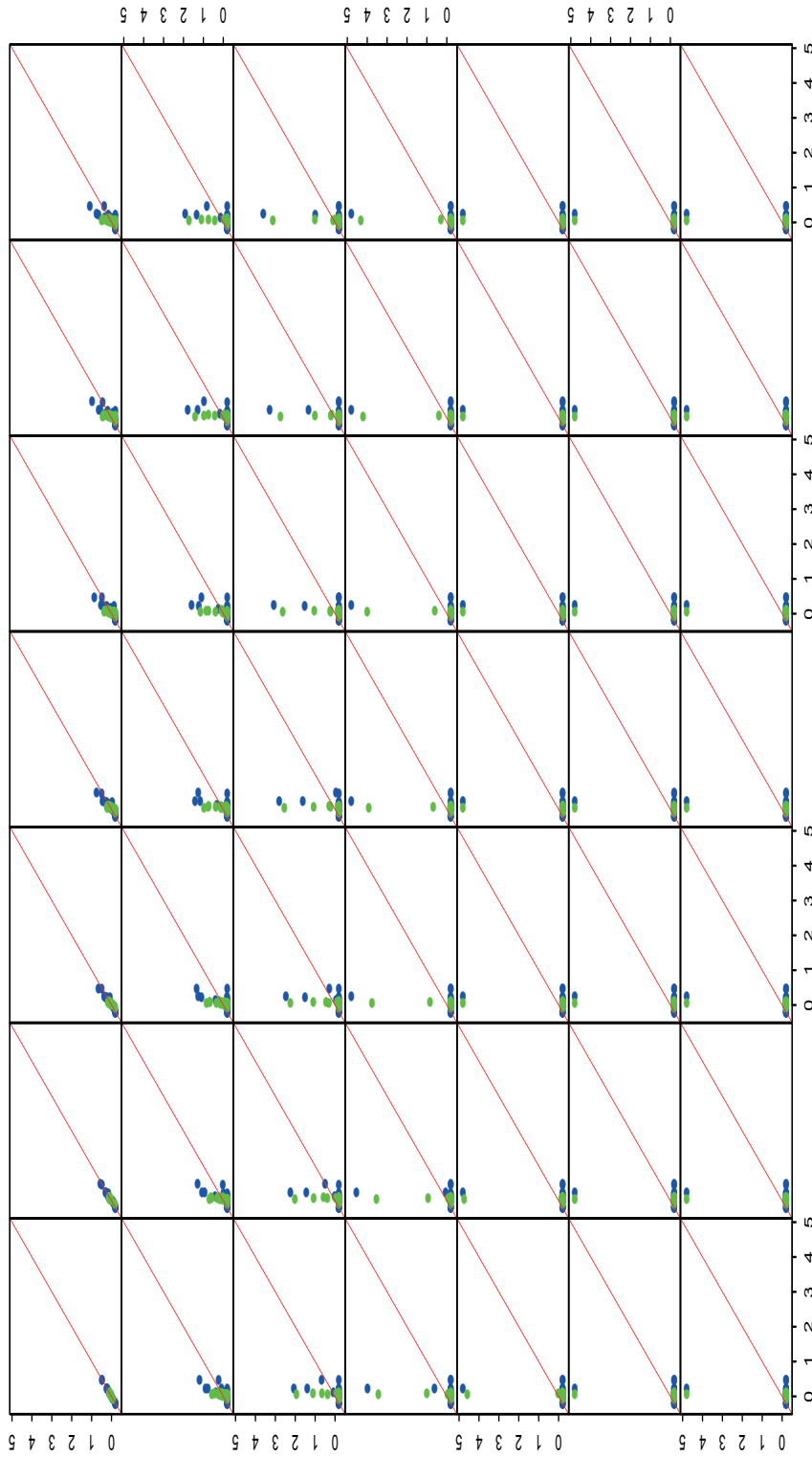
Figure C.40: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type II problem both with added eigenvector uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.41: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type III problem both with added eigenvector uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
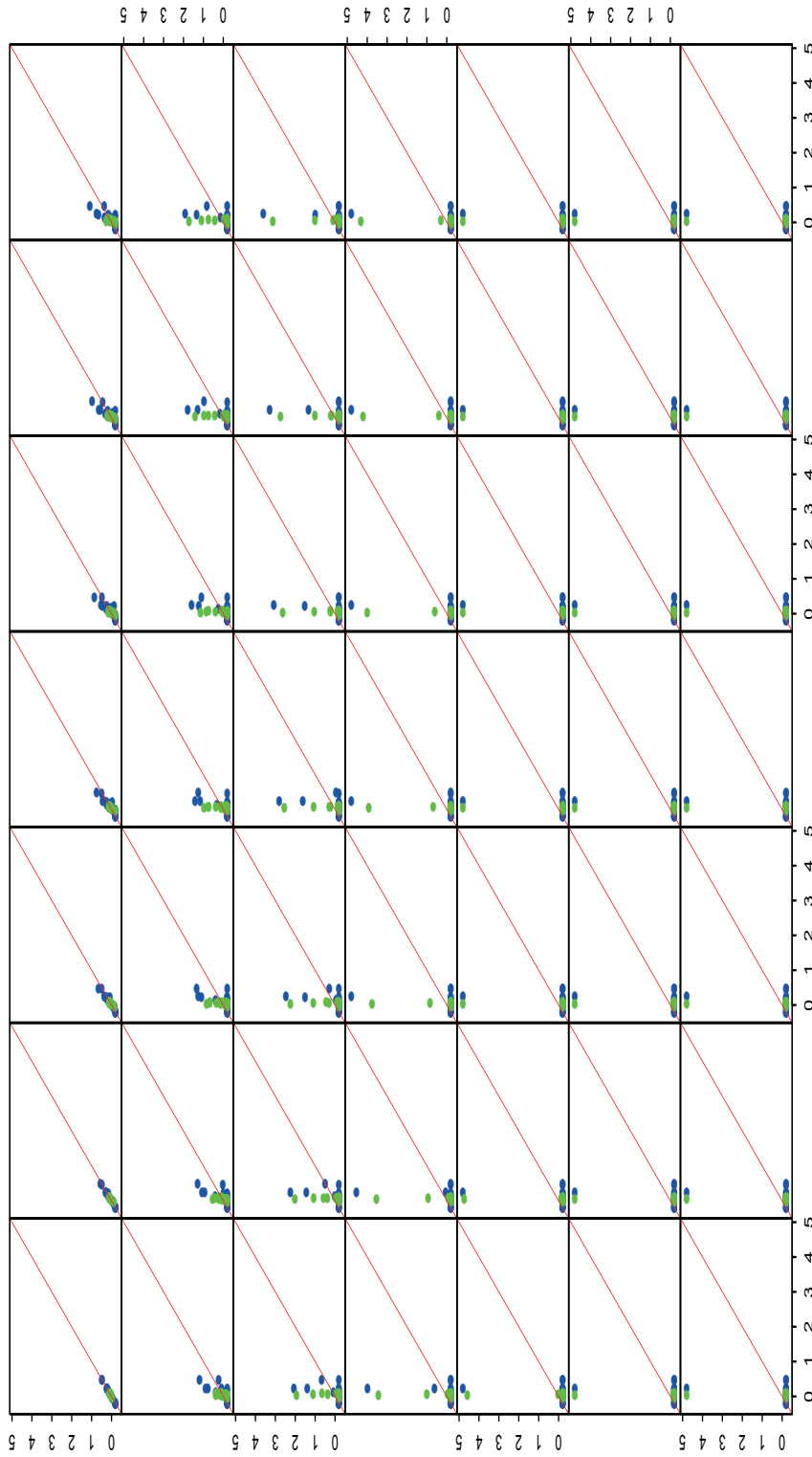
Figure C.42: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type IV problem both with added eigenvector uncertainty (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
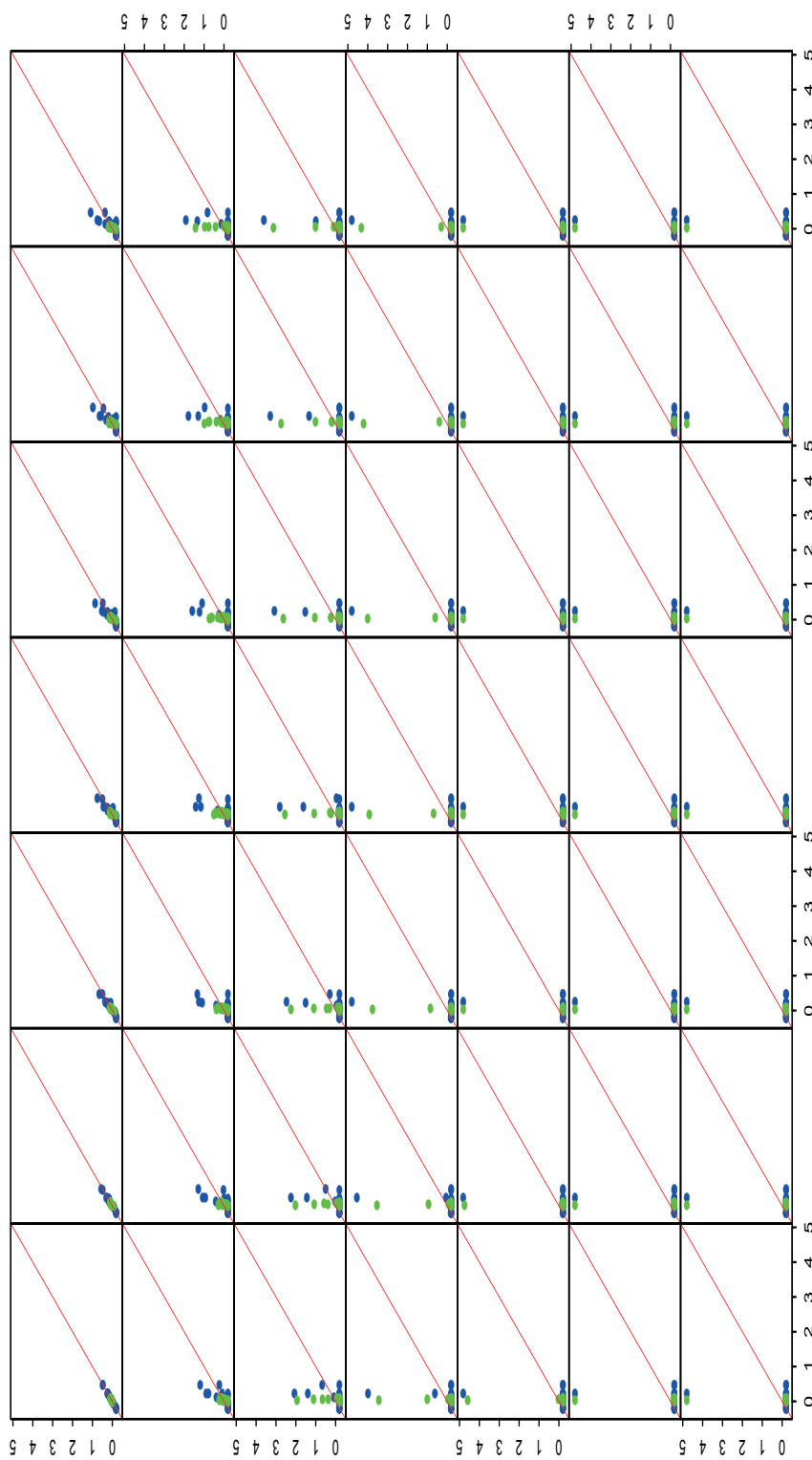
Figure C.43: each subfigure plots the optimal weights of the type II problem both with added mean, eigenvalue and eigenvector uncertainty plus trading costs (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
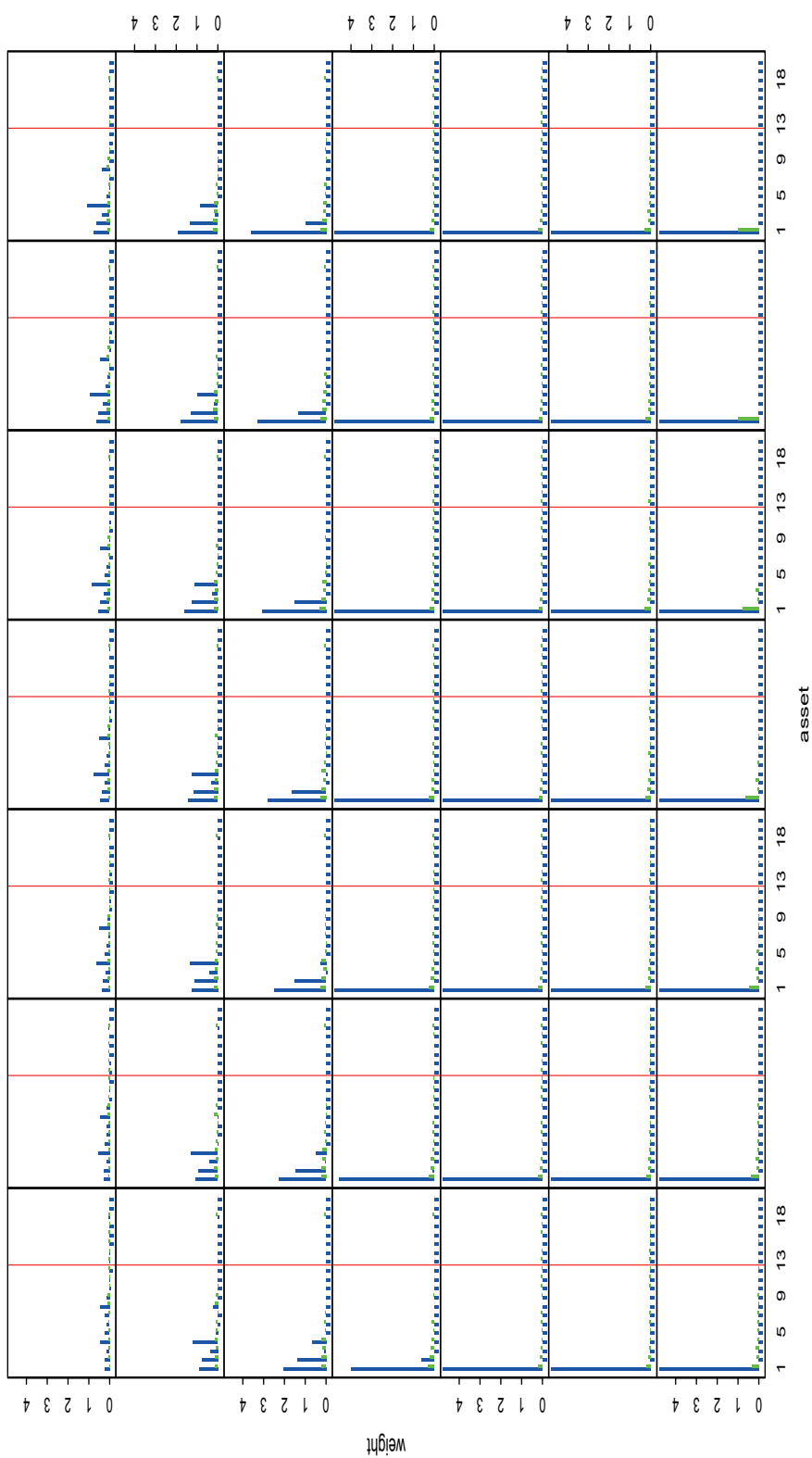
Figure C.44: each subfigure plots the optimal weights of the type III problem both with added mean, eigenvalue and eigenvector uncertainty plus trading costs (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
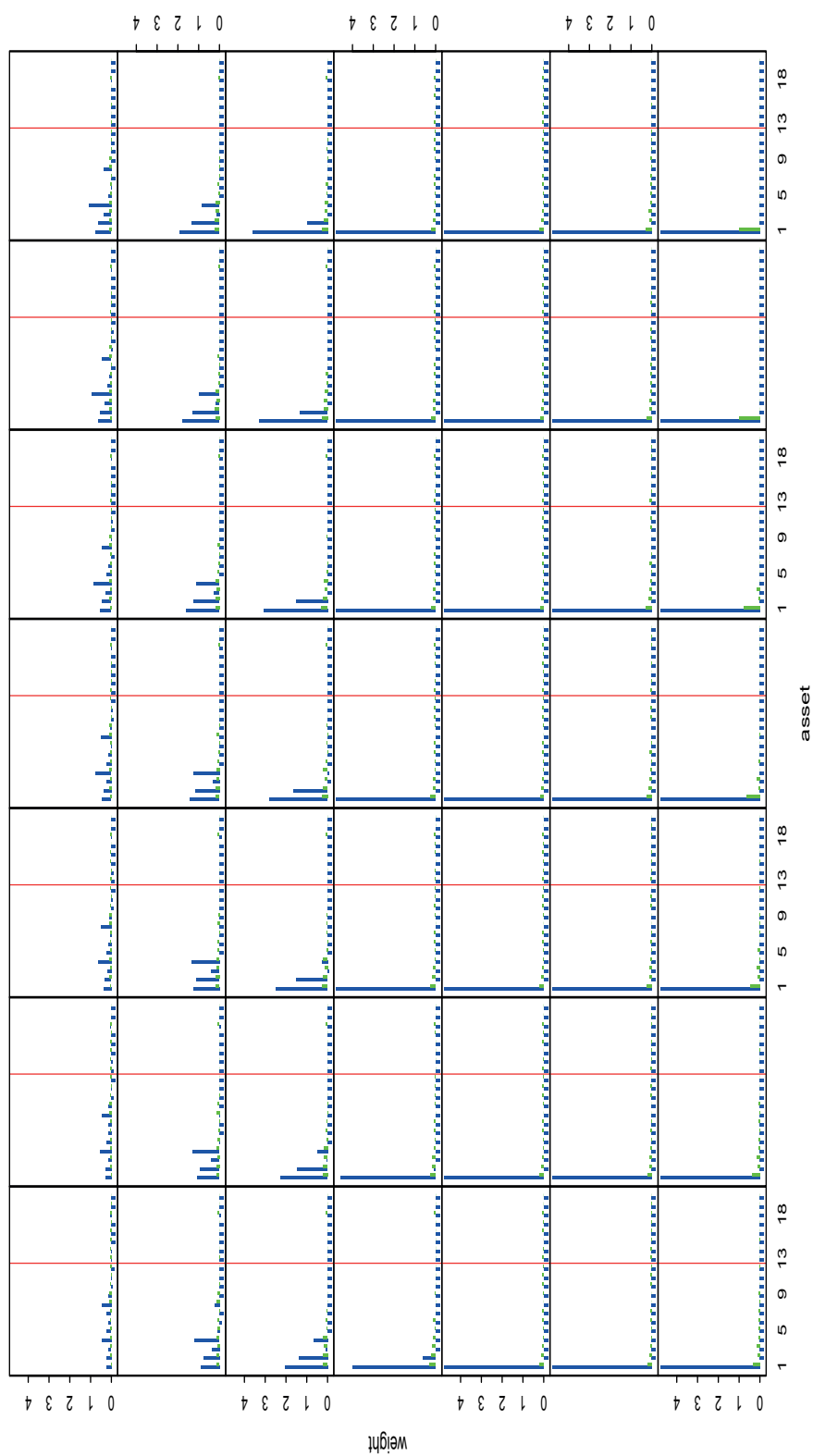
Figure C.45: each subfigure plots the optimal weights of the type IV problem both with added mean, eigenvalue and eigenvector uncertainty plus trading costs (green) and without added uncertainty (blue) against the asset number where short-selling is allowed up to a maximum of one-fifth the total wealth at a particular tolerance level ϵ; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.

Figure C.46: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type II problem both with added mean, eigenvalue and eigenvector uncertainty plus trading costs (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
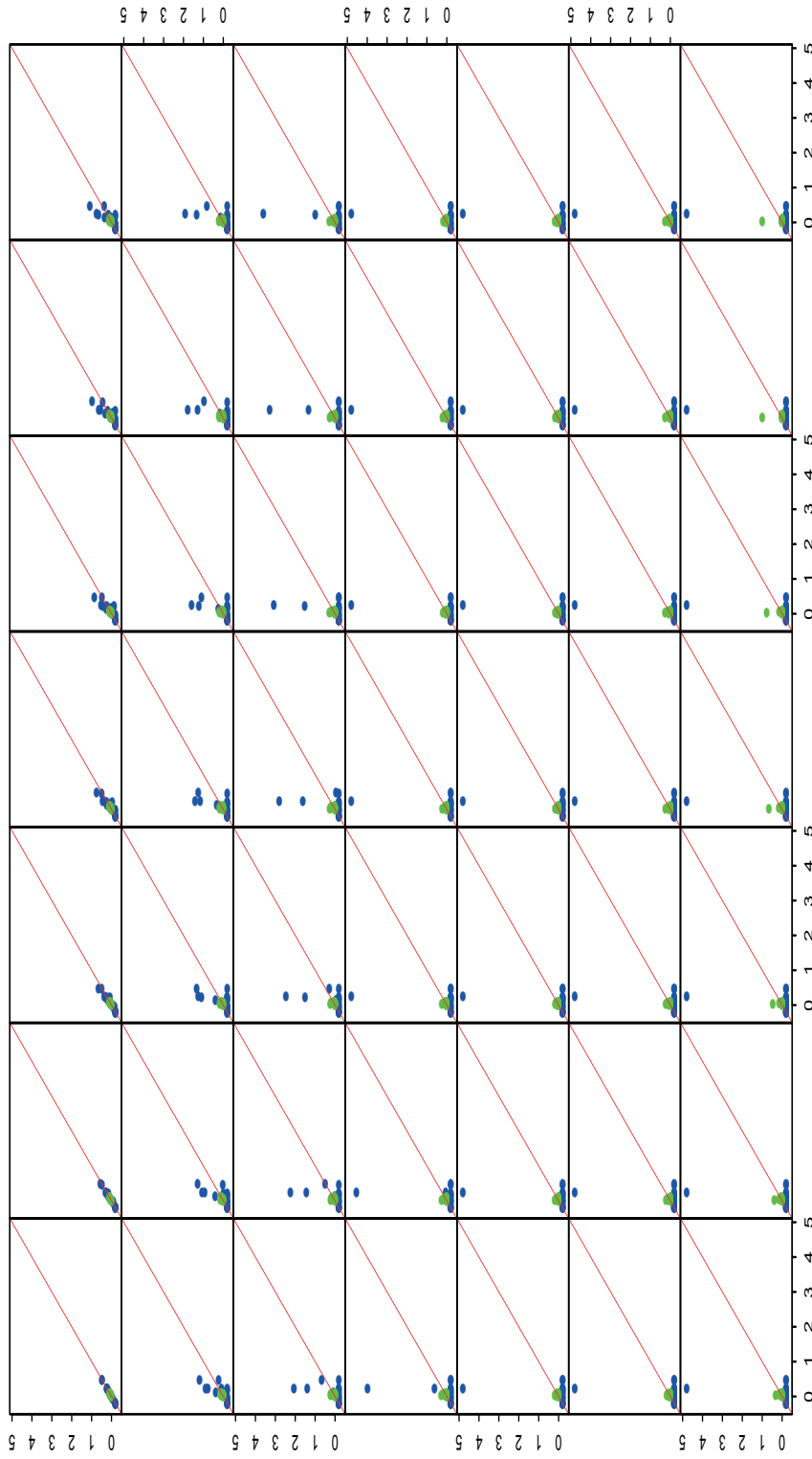
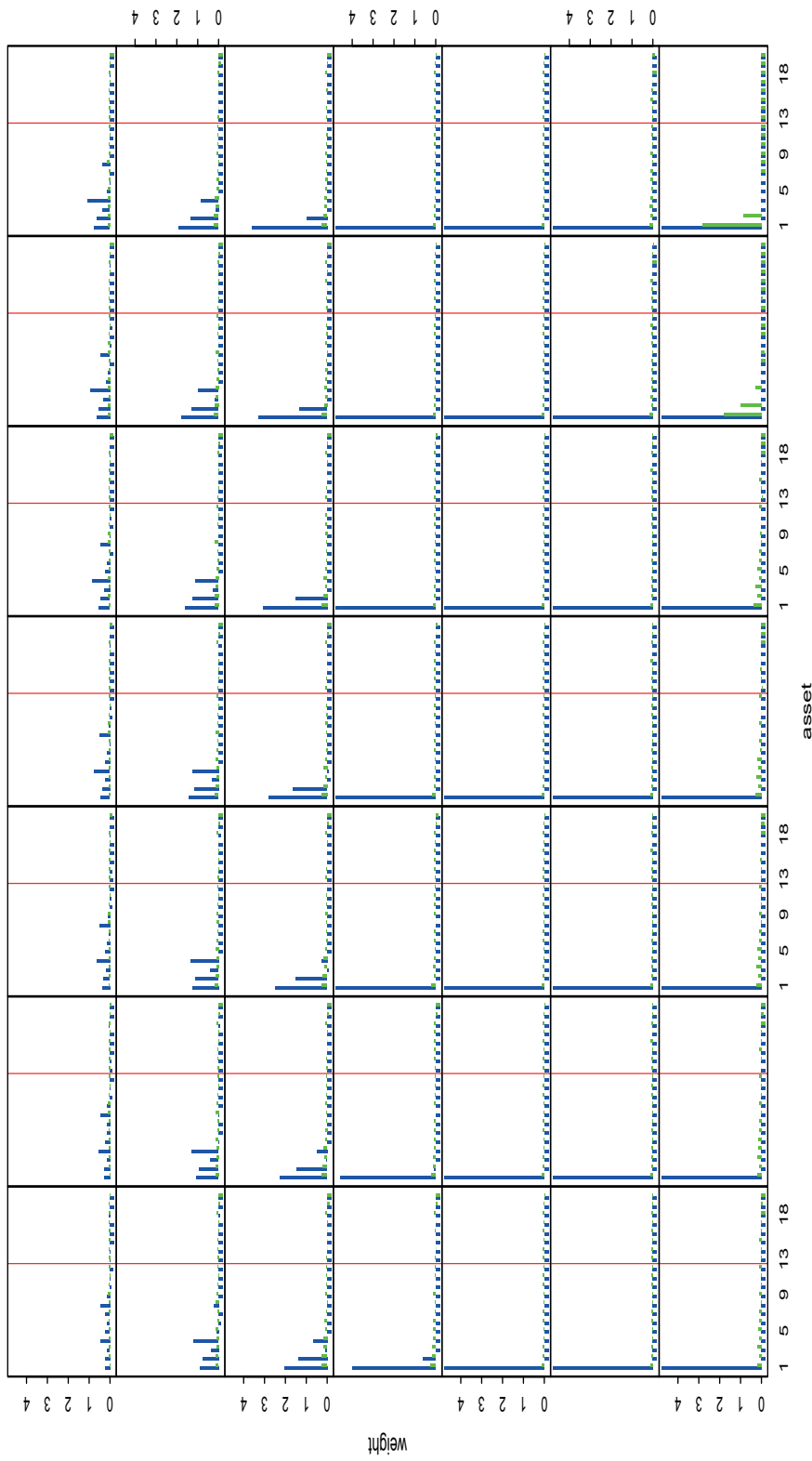Figure C.47: each subfigure plots the optimal weights at a particular tolerance level ϵ against those at tolerance level 0.01 of the type III problem both with added mean, eigenvalue and eigenvector uncertainty plus trading costs (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
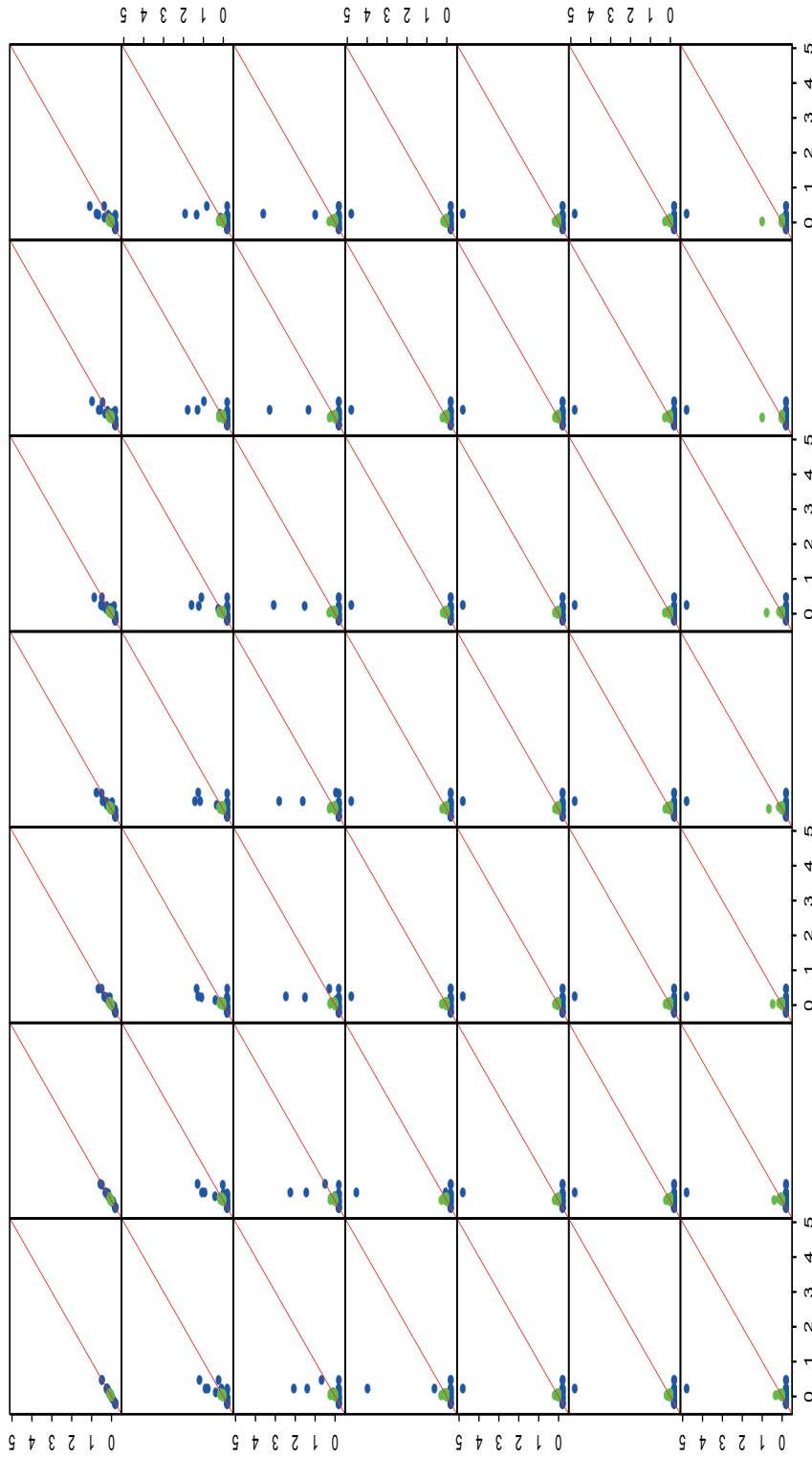
Figure C.48: each subfigure plots the optimal weights at a particular tolerance level at tolerance level 0.01 of the type IV problem both with added mean, eigenvalue and eigenvector uncertainty plus trading costs (green) and without added uncertainty (blue), where short-selling is allowed up to a maximum of one-fifth the total wealth; subfigures from left to right then top to bottom correspond respectively to $\epsilon = 0.01, 0.02, \ldots, 0.49$.
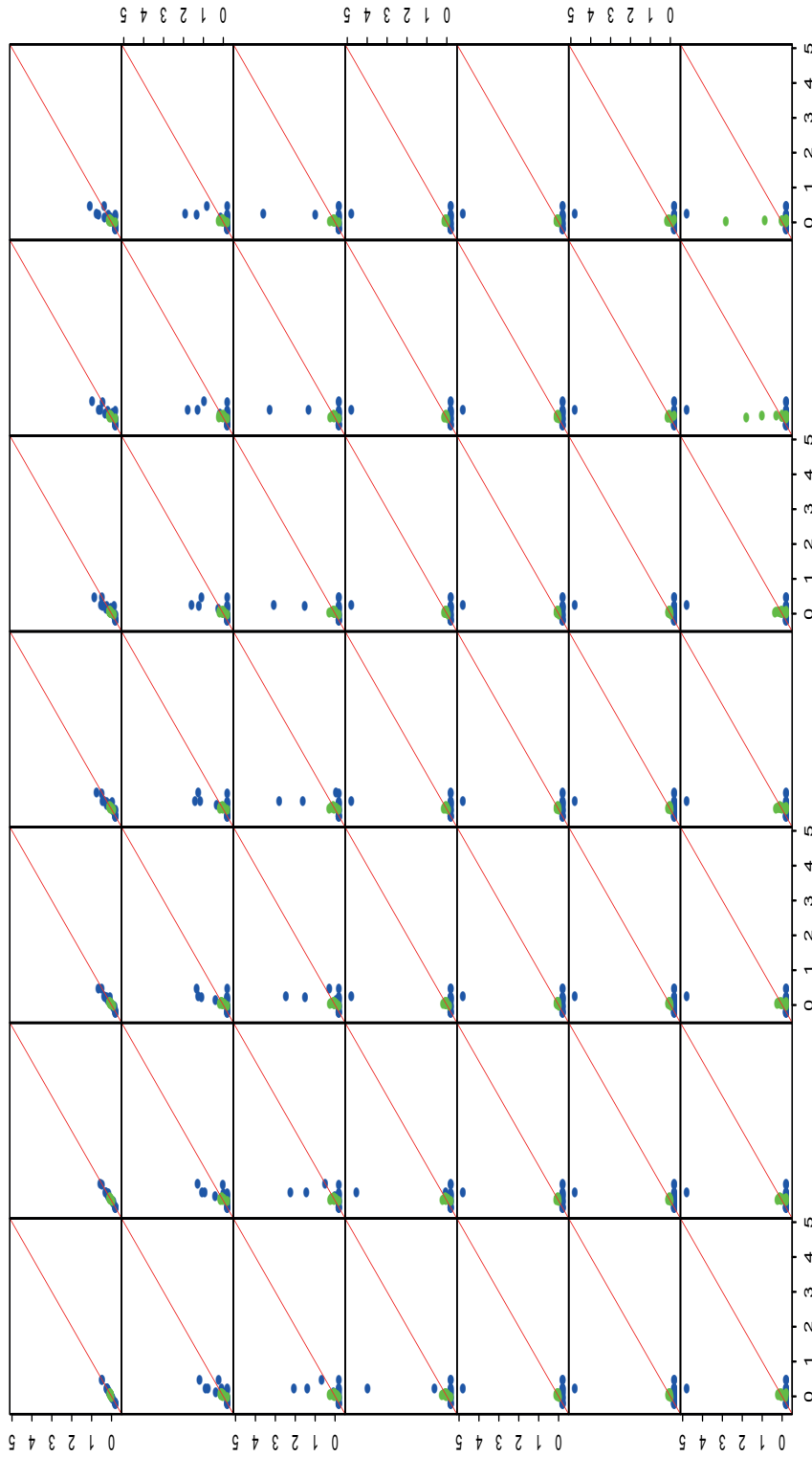
References

[1] Abdul-Hamid, H., and Nolan, J. P. Multivariate stable densities as functions of one dimensional projections. *Journal of Multivariate Analysis 67*, 1 (1998), 80 – 89.

[2] Adler, R. J. Discussion: heavy tail modeling and teletraffic data. *The Annals of Statistics 25*, 5 (1997), 1849–1852.

[3] Ahmadi-Javid, A. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications 155*, 3 (2012), 1105–1123.

[4] Akgiray, V., and Booth, G. G. The stable-law model of stock returns. *Journal of Business and Economic Statistics 6* (1988), 55–57.

[5] Alexander, C. Principles of the skew. *Risk 14*, 1 (2001a), 29–32.

[6] Alexander, C., and Lazar, E. Normal mixture GARCH(1,1): applications to exchange rate modelling. *Journal of Applied Econometrics 21* (2006), 307–336.

[7] Alexander, C., and Lazar, E. Modelling regime-specific stock price volatility. *Oxford Bulletin of Economics and Statistics 71* (2009), 761–797.

[8] Alexander, G. J., and Francis, J. C. *Portfolio Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1986.

[9] Aparicio, F. M., and Estrada, J. Empirical distributions of stock returns: European securities markets, 1990-95. *European Journal of Finance 7* (2001), 1–21.

[10] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. Coherent measures of risk. *Mathematical Finance 9*, 3, 203–228.

[11] Ausin, M. C., and Galeano, P. Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics and data Analysis 51*, 5 (2007).

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

[13] Babu, G. J. Efficient estimation of the reciprocal of the density quantile function at a point. *Statistics and Probability Letters 4* (1986), 133–139.

[14] Babu, G. J. Joint asymptotic distribution of marginal quantiles and quantile functions in samples from a multivariate population. *Journal of Multivariate Analysis 27* (1988), 15–23.

[15] Babu, G. J., and Rao, C. R. Estimation of the reciprocal of the density quantile function at a point. *Journal of Multivariate Analysis 33* (1990), 106–124.

[16] Bai, X., Russel, J., and Tiao, G. Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics 114* (2003), 349–360.

[17] Bai, X., Russel, J. R., and Tiao, G. C. Beyond Merton's utopia (I): effects of non-normality and dependence on the precision of variance using high-frequency financial data. *University of Chicago, GSB Working Paper* (July, 2001).

[18] Bakshi, G., Kapadia, N., and Madan, D. Stock return characteristics, skew laws, and the differential pricing for individual equity options. *The Review of Financial Studies 14*, 1 (2003), S29–32.

[19] Barmish, B. R., and Lagoa, C. M. The uniform distribution: a rigorous justification for the use in robustness analysis. *Mathematics of Control, Signals, and Systems 10* (1997), 203–222.

[20] Bauwens, L., Lubrano, M., and Richard, J.-F. *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, New York, 1999.

[21] BCBS. Consultative document May 2012. Fundamental review of the trading book. Basel Committee on Banking Supervison. Bank for International Settlements, Basel.

[22] BCBS. Consultative document October 2013. Fundamental review of the trading book: A revised market risk framework. Basel Committee on Banking Supervison. Bank for International Settlements, Basel.

[23] BCBS. Regulatory consistency assessment programme (RCAP) - analysis of risk-weighted assets for market risk, 2013. Basel Committee on Banking Supervison. Bank for International Settlements, Basel.

[24] Bekaert, G., and Gray, S. F. Target zones and exchange rates: An empirical investigation. *Journal of International Economics 45* (1998), 1–35.

[25] Ben-Tal, A., El Ghaoui, R., and Nemirovski, A. *Robust Optimization*. Princeton University Press, 2009.

[26] Ben-Tal, A., Margalit, T., and Nemirovski, A. The ordered subsets mirror descent optimization method with applications to tomography. *SIAM Journal on Optimization 12* (2001), 79–108.

[27] Ben-Tal, A., and Nemirovski, A. Robust convex optimization. *Mathematics of Operations Research 23*, 4 (1998), 769–805.

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

[29] Ben-Tal, A., and Nemirovski, A. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming 88*, 3 (2000), 411–424.

[30] Ben-Tal, A., and Nemirovski, A. Lectures on modern convex optimization: Analysis, algorithms, engineering applications, 2001.

[31] Ben-Tal, A., and Nemirovski, A. Robust optimization-methodology and applications. *Mathematical Programming Series B 92* (2002), 811–833.

[32] Ben-Tal, A., and Nemirovski, A. Selected topics in robust convex optimization. *Mathematical Programming Series B 112*, 1 (2008), 125–158.

[33] Ben-Tal, A., Nemirovski, A., and Roos, C. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization 13*, 2 (2002), 535–560.

[34] Bernardo, J. M., and Smith, A. F. M. *Bayesian Theory*. John Wiley and Sons, Ltd, 2000.

[35] Bertsimas, D., Pachmanova, D., and Sim, M. Robust linear optimization under general norms. *OR Letters 32*, 6 (2004), 510–516.

[36] Bertsimas, D., and Popescu, I. Optimal inequalities in probability theory: a convex optimization approach. *SIAM Journal on Optimization 15*, 3 (2005), 780–804.

[37] Bertsimas, D., Popescu, I., and Sethuraman, J. Moment problems and semidefinite programming. In *Handbook on Semidefinite Programming*, R. Saigal, H. Wolkovitz, and L. Vandenberghe, Eds. Kluwer Academic Publishers, 2000, pp. 469–509.

[38] Bertsimas, D., and Sim, M. The price of robustness. *Operations Research 32*, 1 (2004), 35–53.

[39] Black, F. Capital market equilibrium with restricted borrowing. *Journal of Business 45*, 3 (1972), 444–455.

[40] Black, F. Studies of stock market volatility changes. In *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, A. S. Association, Ed. American Statistical Association, Alexandria, VA, 1976.

[41] Boyd, S., El Ghaoui, L., Feron, E., and Balakrishnan, V. *Linear Matrix Inequalities in System and Control Theory*, vol. 15 of Studies in Applied Mathematics. SIAM, Philadelphia, 1994.

[42] Boyd, S., and Vandenberghe, L. *Convex Optimization*, vol. 8 of *Advances in Computational Management Science*. Springer, The Netherlands, 2004.

[43] Brennan, M. J. Capital market equilibrium with divergent borrowing and lending. *Journal of Financial and Quantitative Analysis 6*, 5 (1971), 1197–1205.

[44] Broda, S., Haas, M., Kraus, J., Paolella, M., and Steude, S. Stable mixture GARCH models. *Swiss Finance Institute, SFI. Research Paper Series* (2011).

[45] Broda, S. A. Modeling fat tails in stock returns: a multivariate stable-GARCH approach. *Computational Statistics 26* (2011), 1–23.

[46] Broda, S. A., Haas, M., Kraus, J., Paolella, M. S., and Steude, S. C. Stable mixture GARCH models. *Journal of Econometrics 172* (2013), 292–306.

[47] Byczkowski, T., Nolan, J. P., and Rajput, B. Approximation of multidimensional stable densities. *Journal of Multvariate Analysis 46*, 1 (1993), 13–31.

[48] Carmichael, B., Koumou, G., and Moran, K. Unifying portfolio diversification measures using Rao's quadratic entropy. Tech. rep., 2015. CIRPEE Working Paper 15-08.

[49] Cazalet, Z., Grison, P., and Roncalli, T. The smart beta indexing puzzle. *The Journal of Index Investing 5*, 1 (2014), 97–119.

[50] Chambers, J. M., Mallows, C., and Stuck, B. W. A method for simulating stable random variables. *Journal of the American Statistical Association 71* (1997), 340–344.

[51] Cheng, B. N., and Rachev, S. T. Multivariate stable future prices. *Mathematical Finance 5* (1995), 133–153.

[52] Chib, S., and Greenberg, E. Markov Chain Monte Carlo simulation methods in econometrics. *Economic Theory 12* (1996), 409–431.

[53] Choueifaty, Y., Froidure, T., and Reynier, J. Properties of the most diversified portfolio. *Journal of Investment Strategies 2*, 2 (2013), 49–70.

[54] Craven, P., and Wahba, G. Smoothing noisy data with spline functions. *Numerische Mathematik 31* (1979), 377–403.

[55] Cvitanić, J., and Karatzas, I. On dynamic measures of risk. *Finance and Stochastics 3*, 4 (1999), 451–482.

[56] Daníelsson, J. The emperor has no clothes: Limits to risk modelling. *Journal of Banking and Finance 26*, 7 (2002), 1273–1296.

[57] de Boor. *A Practical Guide to Splines.* Springer, New York, 1978.

[58] de Santis, G., and Gérard, B. International asset pricing and portfolio diversification with time-varying risk. *The Journal of Finance 52*, 5 (1997), 1881–1912.

[59] Dittmar, R. F. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance 57*, 1 (2002), 369–403.

[60] Doganoglu, T., Hartz, C., and Mittnik, S. Portfolio optimization when risk factors are conditionally varying and heavy tailed. *Computational Economics 29*, 3-4 (2007), 333–354.

[61] DuMouchel, W. H. Stable distributions in statistical inference. 1: Symmetric stable distributions compared to other symmetric long-tailed distributions. *JASA 68* (1973), 469–477.

[62] DuMouchel, W. H. Stable distributions in statistical inference. 2: Information from stably distributed samples. *JASA 70* (1975), 386–393.

[63] DuMouchel, W. H. Estimating the stable index α in order to measure tail thickness: A critique. *The Annals of Statistics 11* (1983), 1019–1031.

[64] Eilers, P. H. C., and Marx, B. D. Flexible smoothing with B-splines and penalties. *Statistical Science 11*, 2 (1996), 89–121.

[65] El Ghaoui, L. Optimization models and applications. http://livebooklabs.com/keeppies/c5a5868ce26b8125, 2015. Livebook visited July 2015.

[66] El Ghaoui, L., and Lebret, H. Robust solutions to least-squares problems with uncertain data. *SIAM Journal of Matrix Analysis and Applications 18*, 4 (1997), 1035–1064.

[67] El Ghaoui, L., Oks, M., and Oustry, F. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Oper. Res. 51*, 4 (2003), 543–556.

[68] Embrechts, P. An academic response to Basel 3.5. *Risks 2*, 1 (2014), 25–48.

[69] Embrechts, P., and Hofert, M. Practices and issues in operational risk modeling under Basel II. *Lithuanian Mathematical Journal 51*, 2.

[70] Embrechts, P., Klüppelberg, C., and Korn, R. *Modelling Extremal Events for Insurance and Finance.* Springer, 1997.

[71] Embrechts, P., Puccetti, G., and Rüschendorf, L. Model uncertainty and VaR aggregation. *Journal of Banking and Finance 37* (2013), 2750–2764.

[72] Engle, R. F. Stock volatility and the crash of '87: Discussion. *Review of Financial Studies 3* (1990), 103–106.

[73] Engle, R. F. Dynamic conditional correlation - a simple class of multivariate GARCH models, May 2000. UCSD Economics Discussion Paper No. 2000-09.

[74] Engle, R. F., and Sheppard, K. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working Paper 8554, National Bureau of Economic Research, October 2001.

[75] Fama, E. F. Mandelbrot and the stable paretian hypothesis. *The journal of business 36*, 4 (1963), 420–429.

[76] Fama, E. F. The behavior of stock market prices. *Journal of Business 38* (1965), 34–105.

[77] Fama, E. F. Portfolio analysis in a stable paretian market. *Management Science 11*, 3 (1965a), 404–419.

[78] Fama, E. F. The behavior of stock-market prices. *The journal of Business 38*, 1 (1965b), 34–105.

[79] Fofack, H., and Nolan, J. P. Extremes, modes and other characteristics of stable distributions. *Extremes 2* (1999), 39–58.

[80] Francis, J. C. Skewness and investors' decisions. *Journal of Financial and Quantitative Analysis 10*, 1 (1975), 163–174.

[81] Friend, W. E., and Westerfield, R. Co-skewness and capital asset pricing. *Journal of Finance 35* (1980), 897–914.

[82] Gabriel, V., Murat, C., and Thiele, A. Recent advances in robust optimization: An overview. *European Journal of Operational Research 235*, 3 (2014), 471–483.

[83] Gaivoronski, A. A., and Pflug, G. C. Value-at-risk in portfolio optimization: Properties and computational approach. *Journal of Risk 7*, 2 (2005), 1–31.

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

[85] Gawronski, W., and Wiessner, M. Asymptotic and inequalities for the mode of stable laws. *Statistics and Decisions 10* (1992), 183–197.

[86] Giacometti, R., Bertocchi, M. I., Rachev, S. T., and Fabozzi, F. J. Stable distributions in the BlackâĂŞLitterman approach to asset allocation. *Quantitative Finance 7*, 4 (2007), 423–433.

[87] Glosten, L. R., Jagannathan, R., and Runkle, D. E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*, 5 (1993), 1779–1801.

[88] Goffin, J.-L., Luo, Z.-Q., and Ye, Y. Complexity analysis of an interior cutting plane method for convex feasibility problems. *SIAM Journal on Optimization 6* (1996), 638–652.

[89] Goldfarb, D., and Iyengar, G. Robust portfolio selection problems. *Mathematics of Operations Research 28* (2003), 1–38.

[90] Golub, G. H., Heath, M., and Wahba, G. Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics 21*, 2 (1979), 215–223.

[91] Gu, C. Cross-validating non-gaussian data. *Journal of Computational and Graphical Statistics 1* (1992), 169–179.

[92] Gu, C. *Smoothing Spline ANOVA Models.* Springer, New York, 2002.

[93] Gu, C., and Kim, Y. J. Penalized likelihood regression: general approximation and efficient approximation. *Canadian Journal of Statistics 34*, 4 (2002), 619–628.

[94] Gu, C., and Wahba, G. Minimizing gcv/gml scores with multiple smoothing parameters via the newton method. *SIAM Journal on Scientific and Statistical Computing 12*, 2 (1991), 383–393.

[95] Haas, M., Mittnik, S., and Paolella, M. S. A new approach to markov switching GARCH models. *Journal of Financial Econometrics 2*, 4 (2004a), 493–530.

[96] Haas, M., Mittnik, S., and Paolella, M. S. Mixed normal conditional heteroscedasticity. *Journal of Financial Econometrics 2*, 2 (2004b), 211–250.

[97] Haas, M., Mittnik, S., and Paolella, M. S. Asymmetric multivariate normal mixture GARCH. *Computational Statistics and Data Analysis 53* (2009), 2129–2154.

[98] Hall, P. On unimodality and rates of convergences for stable laws. *Journal of the London Mathematical Society 2*, 30 (1984), 371–384.

[99] Hamilton, J. D. Regime switching models. In *The New Palgrave Dictionary of Economics*, S. N. Durlauf and L. E. Blume, Eds. Palgrave Macmillan, Basingstoke, 2008.

[100] Harvey, C. R., Liechty, J. C., and Liechty, M. W. Parameter uncertainty in asset allocation. In *The Oxford Handbook of Quantitative Asset Management* (2011), B. Scherer and K. Winston, Eds., Oxford University Press.

[101] Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition.* Springer-Verlag, New York, 2009.

[102] Heiss, F., and Winschel, V. Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics 144*, 1 (2008), 62 – 80.

[103] Hill, B. M. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*, 5 (1975), 1163–1174.

[104] Huisman, R., Koedijk, K. G., Kool, C. J. M., and Palm, F. The tail-fatness of FX returns reconsidered. *De Economist 150* (2002), 299–312.

[105] Jacquier, E., and Polson, N. Bayesian econometrics in finance. In *The Oxford Handbook of Bayesian Econometrics*, J. Geweke, G. Koop, and H. Van Dijk, Eds. Oxford University Press, 2011.

[106] Jagannathan, R., and Ma, T. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance 58*, 4 (2003), 1651–1638.

[107] Janicki, A., and Weron, A. *Simulation and Chaotic Behavior of α-Stable Stochastic Processes.* Marcel Dekker, New York, 1994.

[108] Jansen, D. W., and de Vries, C. G. On the frequency of large stock returns: putting booms and busts into perspective. *Review of Economics and Statistics 73* (1991), 18–24.

[109] Kacperczyk, M., and Damien, P. Asset allocation under distribution uncertainty. Tech. rep., NYU Stern School of Business, 2011.

[110] Kane, A. Skewness preference and portfolio choice. *Journal of Financial and Quantitative Analysis 17* (1977), 15–25.

[111] Kanniappan, P., and Sastry, S. M. Uniform convergence of convex optimization problems. *Journal of Mathematical Analysis and Applications 96*, 1 (1983), 1 – 12.

[112] Kim, T.-H., and White, H. On more robust estimation of skewness and kurtosis. *Finance Research Letters 1* (2004), 56–73.

[113] Kim, W. C., Kim, M. J., Kim, J. H., and Fabozzi, F. J. Robust portfolios that do not tilt factor exposure. *European Journal of Operational Research 234*, 2 (2014), 411 – 421.

[114] Kim, Y. J., and Gu, C. Smoothing spline gaussian regression: more scalable computation via efficient approximation. *Journal of the Royal Statistical Society, Series B 66* (2004), 337–356.

[115] Kirchler, M., and Huber, J. Fat tails and volatility clustering in experimental asset markets. *Journal of Economic Dynamics and Controls 31* (2007), 1844–1874.

[116] Kon, S. J. Models of stock returns: a comparison. *The Journal of Finance 39* (1984), 147–165.

[117] Koutrouvelis, I. A. Regression type estimation of the parameters of stable laws. *Journal of the American Statistical Association 75* (1980), 918–928.

[118] Koutrouvelis, I. A. An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics: Simulation and Computation 10* (1981), 17–28.

[119] Kozubowski, T. J., and Rachev, S. T. The theory of geometric stable distributions and its use in modeling financial data. *European Journal of Operational Research 74* (1994), 310–324.

[120] Kuester, K., Mittnik, S., and Paolella, M. S. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics 4* (2006), 53–89.

[121] Kurowicka, D., and Joe, H., Eds. *Vine Copula Handbook*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2011.

[122] Lancaster, P., and Šalkauskas, K. *Curve and Surface Fitting: an introduction*. Academic Press, London, 1986.

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

[124] Lin, X., and Zhang, D. Inference in generalized additive mixed models using smoothing splines. *Journal of the Royal Statistical Society, Series B 61* (1999), 381–400.

[125] Lobo, M. S., and Boyd, S. The worst-case risk of a portfolio. Tech. rep., Stanford University, 2000.

[126] Lobo, M. S., Fazel, M., and Boyd, S. Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research 152*, 1 (2007), 341–365.

[127] Löfberg, J. YALMIP : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference* (Taipei, Taiwan, 2004).

[128] Malevergne, Y., Pisarenko, V., and Sornette, D. Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance 5* (2005), 379–401.

[129] Mallows, C. L. Some comments on c_p. *Technometrics 15* (1973), 661–675.

[130] Mandelbrot, B. New methods in statistical economics. *Journal of Political Economy 71*, 5 (1963a), 421–440.

[131] Mandelbrot, B. The variation of certain speculative prices. *Journal of Business 36* (1963b), 394–419.

[132] Mandelbrot, B. The Variation of Some Other Speculative Prices. *The Journal of Business 40* (1967a), 393.

[133] Mandelbrot, B., and Taylor, H. M. On the distribution of stock price differences. *Operations research 15*, 6 (1967b), 1057–1062.

[134] Maringer, D. *Portfolio Management with Heuristic Optimization.* Cambridge University Press, 2005.

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

[136] Markowitz, H. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly 3* (1956), 111–133.

[137] Marx, B. D., and Eilers, P. H. Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis 28* (1998), 193–209.

[138] McNeil, A. J., Frey, R., and Embrechts, P., Eds. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press, 2006.

[139] Meerschaert, M. M., and Scheffler, H.-P. Chapter 15 - portfolio modeling with heavy tailed random vectors. In *Handbook of Heavy Tailed Distributions in Finance*, S. T. Rachev, Ed., vol. 1 of *Handbooks in Finance*. North-Holland, Amsterdam, 2003, pp. 595 – 640.

[140] Merton, R. C. An analytic derivation of the efficient portfolio frontier. *The Journal of Financial and Quantitative Analysis 7*, 4 (1972), 1851–1872.

[141] Mikosch, T. Modeling dependence and tails of financial time series. In *Modeling dependence and tails of financial time series*, B. Finkenstaedt and H. Rootzén, Eds. Chapman and Hall, London, 2003, pp. 185–286.

[142] Milgrom, P., and Segal, I. Envelope theorems for arbitrary choice sets. *Econometrica 70*, 2 (2002), 583–601.

[143] Mittnik, S., and Paolella, M. S. Prediction of financial downside risk with heavy tailed conditional distributions. In *Handbook of Heavy Tailed Distributions in Finance*, S. T. Rachev, Ed. Elsevier Science, Amsterdam, 2003.

[144] Mittnik, S., and Rachev, S. T. Alternative multivariate stable distributions and their applications to financial modeling. In *Stable Processes and Related Topics*, Cambanis et al., Eds. Birkhäuser, Boston, 1991.

[145] Mittnik, S., and Rachev, S. T. Modeling asset returns with alternative stable models. *Econometric Reviews 12* (1993), 261–330.

[146] Modarres, R., and Nolan, J. P. A method for simulating stable random vectors. *Computational Statistics 9* (1994), 11–19.

[147] Muirhead, R. J. *Samples from a Multivariate Normal Distribution, and the Wishart and Multivariate Beta Distributions*. John Wiley and Sons, Inc., 2008, pp. 79–120.

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

[149] Natarajan, K., Sim, M., and Uichanco, J. Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance 20*, 4 (2010), 695–731.

[150] Nemirovski, A. Polynomial time methods in convex programming. In *The Mathematics of Numerical Analysis: 1995 Ams-Siam Summer Seminar in Applied Mathematics July 17-August 11, 1995 Park City, Utah (Lectures in Applied Mathematics)*, vol. 32. AMS, Providence, 1996, pp. 543–589.

[151] Nemirovski, A. Prox-method with rate of convergence $o(1/t)$ for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization 15* (2004), 229–251.

[152] Nemirovski, A. On safe tractable approximations of chance-constraints. *European Journal of Operational Research 219* (2012), 707–718.

[153] Nolan, J. P. Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic Models 13* (1997), 759–772.

[154] Nolan, J. P. Parameterizations and modes of stable distributions. *Statistics and Probability Letters 38* (1998), 187–195.

[155] Nolan, J. P. An algorithm for evaluating stable densities in Zolotarev's (M) parameterization. *Mathematical and Computer Modelling 29*, 10âĂŞ12 (1999), 229 – 233.

[156] Nolan, J. P. Fitting data and assessing goodness of fit with stable distributions. In *Proceedings of the Conference on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics* (American University, Washington, DC, June 1999).

[157] Nolan, J. P. Maximum likelihood estimation of stable parameters. In *Lévy Processes*, O. E. Barndorff-Nielson et al., Eds. Birkhäuser, Boston, 2001.

[158] Nolan, J. P. *Stable Distributions: Models for Heavy Tailed Data.* Birkäuser, Boston, 2003.

[159] Nolan, J. P., and Panorska, A. K. Data analysis for heavy tailed multivariate samples. *Communications in Statistics - Stochastic Models 13* (1997), 687–702.

[160] Nolan, J. P., Panorska, A. K., and McCulloch, J. H. Estimation of stable spectral measures. *Mathematical and Computer Modelling 34* (2001), 1113–1122.

[161] Nolan, J. P., and Rajput, B. Calculation of multi-dimensional stable densities. *Communications in Statistics - Simulation and Computation 24*, 3 (1995), 551–566.

[162] Panorska, A. K. Generalized stable models for financial asset returns. *Journal of Computational and Applied Mathematics 70* (1996), 111–114.

[163] Peiró, A. Skewness in financial returns. *Journal of Banking and Finance 6* (1999), 847–862.

[164] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2014.

[165] Rachev, S. T., Kim, J. R., and Mittnik, S. Econometric modeling in the presence of heavy-tailed innovations. *Communications in Statistics - Stochastic Models 13* (1997), 841–886.

[166] Rachev, S. T., and Mittnik, S. *Stable Paretian Models in Finance.* Chapman and Hall, New York, 1994.

[167] Rachev, S. T., and Xin, H. Test for association of random variables in the domain of attraction of multivariate stable law. *Probability and Mathematical Statistics 14*, 1 (1993), 125–141.

[168] Rao, C., and Nayak, T. Cross entropy, dissimilarity measures, and characterizations of quadratic entropy. *IEEE Transactions on Information Theory 31*, 5 (1985), 589–593.

[169] Rao, R. C. Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology 21* (1982b), 24–43.

[170] Rockafellar, R. *Conjugate Duality and Optimization.* Society for Industrial and Applied Mathematics, 1974.

[171] Roll, R. R. A critique of the asset pricing theory's tests part I. On past and potential testability of the theory. *Journal of Financial Economics 4*, 2 (1977), 129–176.

[172] Roy, A. D. Safety first and the holding of assets. *Econometrica 20* (1952), 413–449.

[173] Rozelle, J., and Fielitz, B. Skewness in common stock returns. *Financial Review 15* (1980), 1–23.

[174] Rujeerapaiboon, N., Kuhn, D., and Wiesemann, W. Robust growth-optimal portfolios. *Management Science.*

[175] Samorodnitsky, G., and Taqqu, M. S. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, London, 1994.

[176] Schoenberg, J. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society* (1938), 522–536.

[177] Scott, R. C., and Horvath, P. A. On the direction of preference for moments of higher order than the variance. *The Journal of Finance 35*, 4 (1980), 915–919.

[178] Sharpe, W. F. A simplified model for portfolio analysis. *Management Sciences 9*, 2 (1963), 277–293.

[179] Sharpe, W. F. Capital asset prices with and without negative holdings. *The Journal of Finance 46*, 2 (1991), 489–509.

[180] Sharpe, W. F. The sharpe ratio. *Journal of Porfolio Management 21*, 1 (1994), 49–58.

[181] Silverman, B. W. Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *Journal of the Royal Statistical Society, Series B 47*, 1 (1985), 1–53.

[182] Simkowitz, M., and Beedles, W. Asymmetric stable distributed security returns. *Journal of the American Statistical Association 75* (1980), 306–312.

[183] Simon, C., and Blume, L. *Mathematics for Economists.* Norton, 1994.

[184] Solnik, B. The advantages of domestic and international diversification. In *International Capital Markets*, E. J. Elton and M. J. Gruber, Eds. North-Holland Publishing Company, Amsterdam, 1973.

[185] Stoyanov, S., Samorodnitsky, G., Rachev, S., and Ortobelli Lozza, S. Computing the portfolio conditional value-at-risk in the alpha-stable case. *Probability and Mathematical Statistics 26*, 1 (2006), 1–22.

[186] Szegö, G. Measures of risk. *Journal of Banking and Finance 26*, 7 (2002), 1253–1271.

[187] Tang, H., Chiu, K., and Lei, X. Finite mixture of ARMA-GARCH model for stock price prediction. In *3rd International Workshop on Computational Intelligence in Economics and Finance* (North Carolina, U.S.A., September 2003), pp. 1112–1119.

[188] Tobin, J. Liquidity preference as behavior towards risk. *Review of Economic Studies 26*, 1 (1958), 65–86.

[189] Tobin, J. The theory of portfolio selection. In *The Theory of Interest Rates*, F. Hahn and F. Brechling, Eds. Macmillan and Co. Ltd., London, 1965.

[190] Toh, K. C., Todd, M., and Tütüncü, R. H. SDPT3 – a MATLAB software package for semidefinite programming. *Optimization Methods and Software 11* (1999), 545–581.

[191] Tsiang, S. The rationale of the mean-standard deviation analysis, skewness preference, and the demand for money. *Journal of Financial and Quantitative Analysis 6*, 3 (1972), 354–371.

[192] Tucker, A. L., and Pond, L. The probability distribution of foreign exchange price changes: Tests of candidate processes. *Review of Economics and Statistics 70* (1988), 638–647.

[193] Uchaikin, V. V., and Zolotarev, V. M. *Chance and Stability: Stable Distributions and Their Applications*. VSP BV, Zeist, 1999.

[194] Varadhan, R. *Alabama: Constrained Nonlinear Optimization*, 2015. R package version 2015.3-1.

[195] Varian, H. Portfolio of Nobel laureates: Markowitz, Miller and Sharpe. *The Journal of Economic Perspectives 7*, 1 (1993), 159–169.

[196] Wahba, G. Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data. In *Approximation Theory III*, E. Cheney, Ed. Academic Press, London, 1980.

[197] Wahba, G. *Spline models for observational data*. SIAM, Philadelphia, 1990.

[198] Wang, Y. Smoothing spline models with correlated random errors. *Journal of the American Statistical Association 93*, 441 (1998b), 341–348.

[199] Weron, R. Levy-stable distributions revisited: tail index > 2 does not exclude the Levy-stable regime. *International Journal of Modern Physics C 12* (2001), 209–223.

[200] Woerheide, W., and Persson, D. An index of portfolio diversification. *Financial Services Review 2*, 2 (1992), 73 – 85.

[201] Wood, S. N. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society, Series B 62* (2000), 413–428.

[202] Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association 99* (2004), 673–686.

[203] Wood, S. N. *Generalized additive models: An introduction with R*. Chapman and Hall/CRC, 2006.

[204] Wuertz, D., Maechler, M., and core team members., R. *stabledist: Stable Distribution Functions*, 2013. R package version 0.6-6.

[205] Ye, K., Parpas, P., and Rustem, B. Robust portfolio optimization: a conic programming approach. *Computational Optimization and Applications 52*, 2 (2012), 463–481.

[206] Zolotarev, V. M. *One-Dimensional Stable Distributions*. American Mathematical Society, Providence, 1986.

[207] Zymler, S., Kuhn, D., and Rustem, B. Worst-case value at risk of nonlinear portfolios. *Manage. Sci. 59*, 1 (2013), 172–188.

Curriculum Vitae

Zheng Wei YAP   **March 23, 2017**
zheng.yap@epfl.ch

Personal Information

- Male, Singapore citizen

Education

-
  Swiss Federal Institute of Technology   Lausanne, Switzerland
  Ph.D. Statistics and Operations Research   *February 2013 - Present*
-
  Nanyang Technological University   Singapore
  B.Sc. Mathematical Sciences   *August 2008 - June 2012*