Portfolio Risk Minimization under Departures from Normality

by

Geoffrey J. Lauprête

B.S., Tufts University (1993)
S.M., Massachusetts Institute of Technology (1998)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2001

Signature of Author ....
Sloan School of Management
August 10, 2001

Certified by ....
Roy E. Welsch
Professor of Statistics and Management Science
Thesis Supervisor

Certified by ....
Alexander Samarov
Principal Research Scientist
Thesis Supervisor

Accepted by ....
James B. Orlin
Edward Pennell Brooks Professor of Operations Research
Codirector, Operations Research Center

Portfolio Risk Minimization under Departures from Normality

by

Geoffrey J. Lauprête

Submitted to the Sloan School of Management
on August 10, 2001, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis revisits the portfolio selection problem in cases where returns cannot be modeled as Gaussian. The emphasis is on the development of financially intuitive and statistically sound approaches to portfolio risk minimization.

When returns exhibit asymmetry, we propose using a quantile-based measure of risk which we call shortfall. Shortfall is related to Value-at-Risk and Conditional Value-at-Risk, and can be tuned to capture tail risk. We formulate the sample shortfall minimization problem as a linear program. Using results from empirical process theory, we derive a central limit theorem for the shortfall portfolio estimator. We warn about the statistical pitfalls of portfolio selection based on the minimization of rare events, which happens to be the case when shortfall is tuned to focus on extreme tail risk.

In the presence of heavy tails and tail dependence, we show that portfolios based on the minimization of alternative robust measures of risk may in fact have lower variance than those based on the minimization of sample variance. We show that minimizing the sample mean absolute deviation yields portfolios that are asymptotically more efficient than those based on the minimization of the sample variance, when returns have a multivariate Student-t distribution with degrees of freedom less than or equal to 6. This motivates our consideration of other robust measures of risk, for which we present linear and quadratic programming formulations. We carry out experiments on simulated and historical data, illustrating the fact that the efficiency gained by considering robust measures of risk may be substantial.

Finally, when the number of return observations is of the same order of magnitude as, or smaller than, the dimension of the portfolio being estimated, we investigate the applicability of regularization to sample risk minimization. We examine both L1- and L2-regularization. We interpret regularization from a Bayesian perspective, and provide an algorithm for choosing the regularization parameter. We validate the use of regularization in portfolio selection on simulated and historical data, and conclude that regularization can yield portfolios with smaller risk, and in particular smaller variance.

Thesis Supervisor: Roy E. Welsch
Title: Professor of Statistics and Management Science

Thesis Supervisor: Alexander Samarov
Title: Principal Research Scientist

Acknowledgements

I would like to use this space to thank everyone who, directly or indirectly, contributed to this thesis. Beginning with my advisors Alex Samarov and Roy Welsch, whose enthusiasm motivated many good ideas, and whose healthy skepticism allowed me to part with bad ones. I would like to thank Alex in particular for suggesting that I take a closer look at a measure of risk which he called shortfall, and for pointing me towards the right literature in financial economics and in asymptotic statistics. I would also like to thank Roy for sharing his knowledge of robust and biased regression techniques with me, and for deepening my understanding of statistics. I was able to apply many of the things I learned from him to the portfolio estimation problem.

I would next like to thank Dimitris Bertsimas, for being on my committee, and for demanding explanations that in the end made my work more coherent. I am indebted to him for long discussions and for coming up with the first algorithms I used to tackle the shortfall optimization problem.

The Operations Research Center faculty, students, and administration made my stay at MIT much more pleasant than it otherwise might have been. In particular, many thanks to Paulette Mosley and Laura Rose for being very competent and very patient administrators.

Of the people I met over the years while at MIT and in Boston, I'd like to cite Alex Ernst, Oded Asheri, Peter Kempthorne from Summa Capital, Paul Samuelson and Brendan Bradley, Richa Saxena, Dave Kane, Rob Furdak from Numeric, Rebecca O'Brien, Ted Theodosopoulos for sharing data from Fleet and for helping me with simulations, Arie Bergrun, Terri Sojot, Omar from Switzerland, and from the ORC, Jeremie, Adam, Alp, Martin, Murat, Omar, Hernan, Fernando and Nico.

I'd also like to mention Malika, Erin and Carolyn, for being stylish, witty and wise. From France, Nathalie, Larry, Denis and Laurent, for keeping in touch for all these years. Sayan, for always engaging me with a unique perspective. Marc and Henry B. from Palo Alto, for keeping me wired to the West Coast. Alain, for lots of things, but in particular for being in perfect musical agreement with me.

Amy and Bear, I know that I couldn't have had one without the other ... I owe you the world for your support.

And finally, thank you's all the way home to my parents, and to Luc and Matthieu.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Consider the problem of selecting a portfolio that has minimal risk, subject to a given expected return constraint. Under the assumption that asset returns[1] are multivariate Gaussian, with known covariance matrix, every wealth-seeking, risk-averse investor will prefer the portfolio that has minimal variance (see for example the introductory texts in financial economics, Huang and Litzenberger, 1988, and Ingersoll, 1987). Letting $\mathbf{R}$ be the random return vector in $\mathbb{R}^N$, with mean $\boldsymbol{\mu}$ and covariance matrix Σ, the preceding proposition means that each and every wealth-seeking, risk-averse investor prefers the portfolio that solves the optimization problem

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^t\Sigma\mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^t\mathbf{e} = 1 \\ & \mathbf{x}^t\boldsymbol{\mu} = r_p, \end{aligned} \tag{1.1}$$

where $\mathbf{x}^t\boldsymbol{\mu} = r_p$ is the given expected return constraint, r_p being a target return, and $\mathbf{x}^t\mathbf{e} = 1$ is the budget, or convexity, constraint.

Now suppose that the distribution of returns is unknown, except that the expected return vector $\boldsymbol{\mu}$ and the target return are r_p given[2]. Also, assume that a sample of return observations $\mathbf{R}_1, \dots, \mathbf{R}_T$ is

[1] By returns, we mean net returns, i.e. if the price of an asset is P_t at time t, and P_{t+1} at time $t+1$, then the net return of the asset over the period is $R = (P_{t+1} - P_t)/P_t$.

[2] The assumption that the expected return vector is known is not realistic. In practice, the expected return vector would be estimated, through a mix of fundamental and technical analysis. The problem of estimating the expected return vector $\boldsymbol{\mu}$ will not addressed in this thesis. In fact, to keep the problem simple, we will not even model the possible dependence between $\boldsymbol{\mu}$ and the sample of return observations. We leave this last issue for future research, and throughout this document we will assume that $\boldsymbol{\mu}$ is deterministic.

available. Then, a natural approach to selecting a portfolio with minimal risk is to solve the problem

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^t \hat{\Sigma} \mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1 \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p, \end{aligned} \tag{1.2}$$

where

$$\hat{\Sigma} = \frac{1}{T} \sum_{i=1}^{T} (\mathbf{R}_i - \bar{\mathbf{R}})(\mathbf{R}_i - \bar{\mathbf{R}})^t$$

is the sample covariance matrix, and

$$\bar{\mathbf{R}} = \frac{1}{T} \sum_{i=1}^{T} \mathbf{R}_i$$

is the sample mean. Solving Problem (1.2) can be expected to be optimal under two conditions: (i) the data are in fact multivariate Gaussian (and independent and identically distributed) and (ii) T is large compared to N. Because of the Gaussian assumption (i), the variance is the optimal measure of risk, and the optimal portfolio is the solution to (1.1). Then, if (i) and (ii) are satisfied, the sample covariance matrix is the maximum likelihood estimate of Σ, and the solution to Problem (1.2) will be the optimal estimator of the solution to Problem (1.1), according to classical asymptotic statistics (for example, by the invariance principle of maximum likelihood estimators).

But in actual financial markets, the Gaussian model may be completely unsatisfactory:

- the empirical distribution of asset returns may in fact be asymmetric. This is obvious for assets like options (see, e.g. Bookstaber and Clarke, 1984). More generally, skewness may occur due to the greater contagion and spillover of volatility effects between assets and markets in down rather than up market movements, see, e.g. King and Wadhani (1990), Hamao, Masulis, and Ng (1990), Neelakandan (1994), and Embrechts, McNeil, and Straumann (1999).
- even if one decides to model asset returns as multivariate elliptically symmetric[3], the empirical distribution of asset returns may still may have heavier tails and have more tail dependence than

[3] A multivariate random varianble $\mathbf{R}$ with mean $\boldsymbol{\mu}$ and dispersion matrix Ω has an elliptically symmetric distribution if

the Gaussian. With heavy tails, marginal return distributions tend to have tails that decay more slowly than the Gaussian - see for example Campbell, Lo and MacKinlay (1997) and Bouchaud and Potters (2000) for a discussion of heavy-tailed distributions in finance. Tail dependence[4] simply reflects the observation that the extreme return of one asset is likely to be accompanied by extreme returns in other assets, for example, in the context of a market crash or of a market surge - see again Embrechts, McNeil, and Straumann (1999) for a discussion of tail dependence and its applications in risk management, and Lindskog (2000) for evidence that stock returns may have more tail-dependence than the Gaussian. The Gaussian has zero tail-dependence, so that extreme events occur independently; but other elliptically symmetric distributions, such as the multivariate Student-t, may have positive tail dependence.

These departures from normality will cause the following two difficulties. First, asymmetry in the distribution of returns makes variance (or standard deviation) as a risk measure intuitively inadequate because it equally penalizes desirable upside and undesirable downside deviations from the mean. So even if Σ were known, Problem (1.1) would not be the right problem to solve for every wealth-seeking, risk-averse investor. This might discredit using Problem (1.2) as an approach to portfolio selection. Second, even if one is ready to assume that the distribution of returns is elliptically symmetric[5], heavy tails and tail dependence may have a negative impact on the portfolio estimation procedure (1.2) - this idea is made rigorous in Chapter 5. Intuitively, this difficulty arises because extreme returns, or outliers, are more frequent than under the Gaussian, which makes the estimation of variances and covariances

---

its probability density $f(.)$ is of the form

$$f(\mathbf{r}) = \frac{1}{\sqrt{\det(\Sigma)}} g\left((\mathbf{r}-\boldsymbol{\mu})^t \Omega^{-1} (\mathbf{r}-\boldsymbol{\mu})\right),$$

where $g(.) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$. If the covariance matrix of $\mathbf{R}$ is well-defined, then Ω is proportional to the covariance matrix. See appendix for more on elliptically symmetric distributions.

[4] A measure of tail dependence can be defined for pairs of random variables X and Y. Let the distribution functions of X and Y be F_1 and F_2 respectively. Then a coefficient of (lower) tail dependence between X and Y is

$$\lim_{\alpha \rightarrow 0^+} \Pr[Y < F_2^{-1}(\alpha) \mid X < F_1^{-1}(\alpha)] = \lambda,$$

provided a limit $\lambda \in [0,1]$ exists. If $\lambda \in (0,1]$ then X and Y are said to be asymptotically dependent (in the lower tail); if $\lambda = 0$, then they are asymptotically independent. See Embrechts, McNeil, and Straumann (1999) for more details.

[5] Notice that departures from normality within the class of elliptically symmetric distributions actually do not affect the optimality property of variance as a risk measure (assuming second-moments are well-defined): each and every wealth-seeking, risk-averse investor still prefer the portfolio that has minimal variance (for every level of target return). See Ingersoll (1987). However, this result says nothing about whether using (1.2) to choose portfolios is a good idea. In fact, we will prove that under certain circumstances, it is not.

less efficient - i.e., the variance of the estimates of variance and covariance will be higher.

In response to these difficulties, we propose in this thesis an alternative approach to portfolio selection that is based on a very simple observation. Notice that we can rewrite Problem (1.2) as

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \min_q \tfrac{1}{T} \textstyle\sum_{i=1}^{T} (\mathbf{x}^t \mathbf{R}_i - q)^2 \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1 \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p. \end{aligned} \tag{1.3}$$

Notice also that the sample variance of portfolio $\mathbf{x} \in \mathbb{R}^N$ can be expressed as

$$\hat{\sigma}^2(\mathbf{x}) = \min_q \frac{1}{T} \sum_{i=1}^{T} (\mathbf{x}^t \mathbf{R}_i - q)^2.$$

Now consider the problem that is obtained by replacing the objective function in (1.3) with the sample mean of a piecewise linear function. Specifically, consider the alternative portfolio optimization problem

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \min_q \tfrac{1}{T} \textstyle\sum_{i=1}^{T} |\mathbf{x}^t \mathbf{R}_i - q| \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1 \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p, \end{aligned} \tag{1.4}$$

and more generally the problem

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \min_q \tfrac{1}{T} \textstyle\sum_{i=1}^{T} [\rho_\alpha(\mathbf{x}^t \mathbf{R}_i - q)] \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1 \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p, \end{aligned} \tag{1.5}$$

where

$$\rho_\alpha(z) = z - \frac{1}{\alpha} z 1_{\{z<0\}}, \tag{1.6}$$

for $\alpha \in (0,1)$. For $\alpha = 50\%$, $\rho_\alpha(.)$ is equal to the absolute value (see Figure 1-1), and Problem (1.5) is

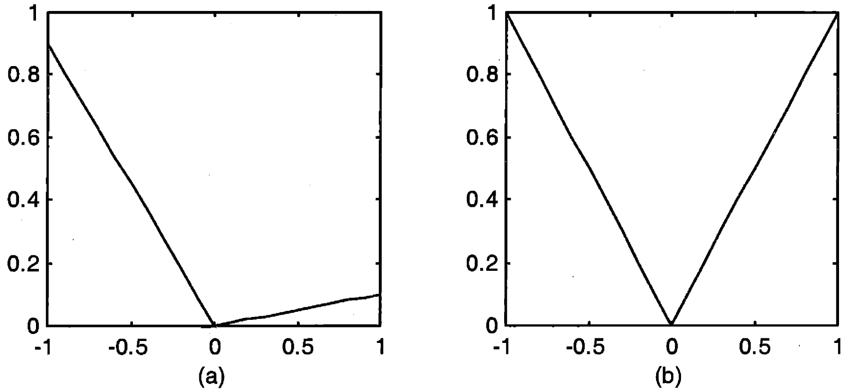

Figure 1-1: (a): $\alpha\rho_\alpha(.)$, $\alpha = 10\%$; (b): $\rho_\alpha(.)$, $\alpha = 50\%$.

equivalent to Problem (1.4). We define a corresponding measure of risk,

$$\hat{s}_\alpha(\mathbf{x}) = \min_q \frac{1}{T} \sum_{i=1}^{T} [\rho_\alpha(\mathbf{x}^t \mathbf{R}_i - q)], \tag{1.7}$$

which we call the (sample) α-shortfall, or shortfall for short, of portfolio $\mathbf{x} \in \mathbb{R}^N$. For $\alpha = 50\%$, (1.7) is the sample mean absolute deviation.

This thesis will show that the problem defined by (1.5) can serve as the basis for an alternative approach to portfolio selection that is more intuitive, and statistically more robust than (1.3), under departures from normality. We also will address the small-sample issue in portfolio selection, and will show how to amend (1.5) and (1.3) in situations when the number of return observations T is of the same order of magnitude, or smaller than, N the dimension of the portfolio being estimated. Note that we will deal exclusively with the static version of the portfolio selection problem, its multi-period extensions being outside the scope of this thesis.

1.1 Literature Review

We review here previous work on portfolio selection based on the minimization of alternative measures of risk. We begin with Markowitz himself, who explicitly suggested, in the line of research that led to his Nobel prize in economics, that alternatives to variance should play a role in portfolio selection. His proposal included a measure of risk which he called semi-variance. The semi-variance of random variable X with mean μ is defined as

$$\sigma^2_{semi} = E[(\mu - X)^-]^2,$$

where the expectation is taken with respect to the distribution of X, and where

$$(\mu - X)^- = \begin{cases} \mu - X & \text{if } X \leqslant \mu \\ 0 & \text{otherwise.} \end{cases}$$

Other similar downside measures of risk in the financial economics literature include the lower partial moments, defined as

$$LPM_i = E[(c - X)^-]^i,$$

with $i = 0, 1$, or 2, for an arbitrary c value. LPM_0 portfolio selection corresponds to Roy's (1952) "safety first" rule, which minimizes the probability of a loss, for a given target expected return.

Variance is an optimal measure of risk when returns are multivariate Gaussian, or more generally multivariate elliptically symmetric, because wealth-seeking, risk-averse investors maximize their expected utility by choosing a portfolio that has minimal variance (for every level of target return). Bawa (1975, 1978) related LPM_i portfolio selection to expected utility maximization by finding distributional conditions for one asset to be stochastically dominated[6] by another. The conclusion is that LPM_0 portfolio selection yields portfolios that are not first-order stochastically dominated, i.e. such portfolios are preferred by at least some wealth-seeking investors, among all portfolios with the same level of target return. Similarly, mean-LPM_1 portfolio selection yields portfolios that are not second-order stochastically dominated, for every level of target return, i.e. such portfolios are preferred by at least some wealth-seeking, risk-averse investors, among all portfolios with the same level of target return.. These results do not depend on the underlying return distribution being multivariate Gaussian.

In recent years, the quantile-based measures of risk Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) , have found extensive use in the financial industry. The α-level VaR (see e.g. Jorion, 1997, Dowd, 1998, and Duffie and Pan, 1997) of random variable X, which is just (minus) its α-quantile,

[6] See appendix for a review of stochastic dominance concepts.

is defined as

$$VaR = -\inf\{z \mid \Pr(X \leqslant z) \geqslant \alpha\}.$$

While VaR measures the worst return which can be expected with a certain probability, it does not address how large these returns can be expected to be in the small-probability even that returns are below the VaR. Also, it is not a convex risk measure and may therefore discourage diversification. To address this issue, CVaR, also called mean-excess function, Tail VaR, or expected shortfall (see Embrechts, McNeil, and Strauman, 1999, and Artzner, Delbaen, Eber and Heath, 1999), can be used. The α-level CVaR is defined as

$$CVaR = -E[X \mid X \leqslant q_\alpha],$$

where q_α is the α-quantile of X.

Optimization of the LPM_1 and LPM_2 measures of risk has been claimed to be difficult (see for example Grootveld and Hallerbach, 1999), because of the discontinuity due to the $(.)^-$ operator - we claim otherwise, and in fact show that these problems can be formulated as generic linear and quadratic programs. Using VaR in portfolio selection has been considered in Lemus (1999) and Lemus, Samarov, and Welsch (1999). CVaR optimization has been formulated as a linear program in Rockafellar and Uryasev (1999), who, with Tasche (2000) have examined some of the mathematical properties of CVaR.

The statistical properties of these alternatives to variance seem to have received very little attention. Grootveld and Hallerbach (1999) mention that LPM's do not seem to offer significantly different portfolios when used instead of variance in portfolio selection. Yamai and Yoshiba (2001) mention the need to develop efficient estimation techniques for VaR and CVaR.

1.2 Thesis Organization

In Chapter 2, we motivate the shortfall measure of risk by examining its population properties, i.e. we operate under the assumption that the distribution of returns is known. Shortfall is related to VaR and CVaR, mentioned earlier, and is indexed by a probability level α. When the distribution of returns is asymmetric, the parameter α may be chosen so that upside and downside returns are penalized differently, and shortfall may be preferable to variance. Also, when the distribution of returns is symmetric, shortfall is proportional to the standard deviation, meaning that shortfall portfolio selection is equivalent to variance portfolio selection. We investigate the natural connection between shortfall portfolio selection and expected utility maximization. We obtain closed form expressions for the gradient and hessian of the shortfall. We show that shortfall is a convex risk measure, giving it an important advantage over VaR. Finally, we discuss how classical mean-variance portfolio analysis results, such as two-fund separation and the concept of beta, generalize to mean-shortfall analysis.

In Chapter 3, we formulate the sample shortfall portfolio optimization problem as a linear program (LP). We also formulate the problem in a way that emphasizes its connection with the sample variance, LPM_1, LPM_2, and CVaR optimization problems. We claim that any of these problems can be formulated as either an LP or a quadratic program (QP), making them solvable with generic LP and QP solvers. We introduce the concept of regularization in sample shortfall portfolio optimization, and sample portfolio optimization in general, as a tool to enforce the uniqueness of the solution to the optimization problem, when the number of return observations T is less than N, the dimension of the portfolio that is being estimated. We revisit this important issue in Chapter 6, from a statistical perspective.

In Chapter 4, we prove a central limit theorem for the shortfall portfolio estimator, and compare it to a similar result for the variance portfolio estimator. The framework we consider is the following. Given independent and identically distributed realizations $\mathbf{R}_1, \ldots, \mathbf{R}_T$ of the random return vector $\mathbf{R} \in \mathbb{R}^N$, we consider the estimators defined by minimizing sample risk, subject to a deterministic set of linear equality constraints. We show that both the variance and shortfall portfolio estimators are consistent and asymptotically normal. Based on the form of the asymptotic covariance of the shortfall portfolio estimator, and based on a computational experiment, we warn against shortfall portfolio estimation based on small values of α.

In Chapter 5, we address the issue of robust portfolio estimation, i.e. the estimation of variance minimizing portfolios under departures from normality within the class of elliptically symmetric distributions. We develop a measure of portfolio estimator performance which we call estimation risk, and show that alternatives to the variance portfolio estimator may in fact have lower estimation risk under departures from normality, such as when returns have heavier tails and more tail dependence than the Gaussian. One alternative to the variance portfolio estimator, when returns are elliptically symmetric, is the shortfall portfolio estimator. We prove, using the results from Chapter 4, that the shortfall portfolio estimator asymptotically outperforms the variance portfolio estimator when the distribution of returns is multivariate Student-t with less than 6 degrees of freedom. Other alternative portfolio estimators include the Huber portfolio estimator, the trimean portfolio estimator, and the trimmed mean portfolio estimator. We show on artificial and real data that these alternative estimators may in practice outperform the variance portfolio estimator under departures from normality.

In Chapter 6, we revisit the topic of regularization in portfolio estimation, which we introduced in Chapter 3. However, our point of view here is statistical. We show how regularization can be motivated from both the Bayesian perspective, and from the perspective of balancing estimation error and approximation error. We provide an algorithm for choosing the regularization parameter, based on cross-validation. We then show, in examples involving simulated and historical data, that regularization may improve the finite sample performance of portfolio estimators when the the number of return observations T is less than, or of the same order of magnitude as, the dimension N of the portfolio that is being estimated.

Finally, in Chapter 7 we conclude this thesis by summarizing our contributions to the portfolio selection literature, and by offering directions for future research.

1.3 Some Notation

In this thesis, we will use the following notation.

$\mathbf{e}$: vector of ones, whose size can be determined by the context in which it is used.

$\mathbf{O}_K$: $(K \times K)$ matrix of zeros.

$\mathbf{O}$: arbitrary matrix of zeros, whose size can be determined by the context in which it is used.

$\mathbf{I}_K$: $(K \times K)$ identity matrix.

$\rightsquigarrow$: "converges in distribution to".

N : the number of assets.

T : the number of return observations in a sample.

^{t} : the transpose operator.

$o(1)$: term that converges (deterministically) as T grows.

$o_p(1)$: term that converges in probability as T grows.

Chapter 2

Population Properties of the α-Shortfall Risk Measure

Suppose that asset returns $\mathbf{R} \in \mathbb{R}^N$ have a multivariate normal distribution, or more generally a multivariate elliptically symmetric distribution, with mean $\boldsymbol{\mu}$ and covariance matrix Σ. The mean-variance optimization problem is

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^t \Sigma \mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1, \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p, \end{aligned} \tag{2.1}$$

where r_p is the target portfolio expected return. Mean-variance portfolio selection is consistent with expected utility maximization in the sense that all wealth-seeking, risk-averse investors (i.e. investors with increasing, concave utility functions[1]) will, for any target return r_p, prefer portfolios with the smallest variance, or standard deviation. More generally, investors will be able to maximize their utility by restricting their portfolio choice set to portfolios that solve (2.1) for some level of r_p. The set of portfolios that solve (2.1) for some level of r_p is called the minimum variance portfolio set.

Now suppose that returns are not elliptically symmetric. Then it is not true anymore that all wealth-seeking, risk-averse investors will, for any target return, prefer portfolios with the smallest standard deviation. In fact, there exist increasing, concave utility functions $u(.)$, and random return variables

[1] See appendix for a review of utility theory and stochastic dominance.

X and Y, such that $E[u(X)] \geqslant E[u(Y)]$ even though $E(X) = E(Y)$ and the standard deviation of X is greater than the standard deviation of Y, i.e. $\sigma_X > \sigma_Y$ (see Ingersoll, 1987). When returns are asymmetric, mean-variance portfolio selection looses its expected utility maximization properties. In addition, the standard deviation equally penalizes returns above and below the mean, making it intuitively unappealing.

Notice that, in direct analogy to (1.3), we can rewrite Problem (2.1) as

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \min_q E(\mathbf{x}^t\mathbf{R} - q)^2 \\ \text{subject to} \quad & \mathbf{x}^t\mathbf{e} = 1 \\ & \mathbf{x}^t\boldsymbol{\mu} = r_p, \end{aligned} \tag{2.2}$$

since the variance of portfolio $\mathbf{x} \in \mathbb{R}^N$ can be expressed as

$$\sigma^2(\mathbf{x}) = \min_q E(\mathbf{x}^t\mathbf{R} - q)^2.$$

In this chapter, we show that we can address some of the shortcomings of mean-variance portfolio selection that occur in the presence of return distribution asymmetry, by considering the portfolio selection problem

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \min_q E[\rho_\alpha(\mathbf{x}^t\mathbf{R} - q)] \\ \text{subject to} \quad & \mathbf{x}^t\mathbf{e} = 1 \\ & \mathbf{x}^t\boldsymbol{\mu} = r_p, \end{aligned} \tag{2.3}$$

where the objective function in (2.2) has been replaced by the expectation of the piecewise linear function defined by

$$\rho_\alpha(z) = z - \frac{1}{\alpha} z 1_{\{z<0\}}, \tag{2.4}$$

for $\alpha \in (0, 1)$. We define a corresponding measure of risk,

$$s_\alpha(\mathbf{x}) = \min_q E[\rho_\alpha(\mathbf{x}^t\mathbf{R} - q)], \tag{2.5}$$

which we call the α-shortfall, or shortfall for short, of portfolio $\mathbf{x} \in \mathbb{R}^N$. Notice that this chapter examines portfolio selection from a population perspective, i.e, assuming that the distribution of $\mathbf{R}$ is known. The optimization and statistical properties of shortfall portfolio selection from a sample perspective - i.e. assuming the distribution of $\mathbf{R}$ is unknown, but a sample of return observations is available - will be examined in later chapters.

This chapter is organized as follows. In Section 1, we provide an alternative definition of (2.5), that gives a more intuitive flavor to the shortfall measure of risk. We then examine the relationship between shortfall and other risk measures, including Value-at-Risk (VaR), and Conditional VaR (CVaR), and the standard deviation. In Section 2 we motivate shortfall portfolio selection by relating it to expected utility maximization. Via theorems of stochastic dominance, we show that optimal shortfall portfolios maximize the expected utilities of at least some wealth-seeking, risk-averse investors, for any given level of target return r_p. In Section 3, we examine some of the mathematical properties of shortfall, and in Section 4 we extend some classical mean-variance portfolio analysis results to the mean-shortfall setting. Finally, in Section 5 we present the results of a numerical experiment involving portfolio selection with asymmetrically distributed return data, and in Section 6 we summarize our findings.

2.1 Alternative Definition of the Shortfall Risk Measure, and Relation to Other Risk Measures

Let $X \in \mathbb{R}$ be a random variable with a continuous density. We have the well-known result

$$\arg\min_{q\in\mathbb{R}} E[|X-q|] = q_{0.5},$$

where $q_{0.5}$ is the median of X. More generally, we have

$$\arg\min_{q\in\mathbb{R}} E[\rho_\alpha(X-q)] = q_\alpha, \tag{2.6}$$

where q_α is the α-quantile of variable X (see van der Vaart, 1998), i.e.

$$\begin{aligned} q_\alpha &= \inf\{z \mid P(X \leqslant z) \geqslant \alpha\} \\ &= F^{-1}(\alpha), \end{aligned}$$

where F is the distribution function of X. Now suppose that random variable $\mathbf{R} \in \mathbb{R}^N$ has a continuous density and mean $\boldsymbol{\mu}$. We can then write the shortfall of portfolio $\mathbf{x} \in \mathbb{R}^N$ as

$$\begin{aligned} s_\alpha(\mathbf{x}) &= \min_q E\left[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)\right] \\ &= E\left[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q_\alpha(\mathbf{x}))\right] \\ &= E\left[(\mathbf{x}^t\mathbf{R}-q_\alpha(\mathbf{x})) - \frac{1}{\alpha}(\mathbf{x}^t\mathbf{R}-q_\alpha(\mathbf{x}))1_{\{\mathbf{x}^t\mathbf{R}\leqslant q_\alpha(\mathbf{x})\}}\right] \\ &= \mathbf{x}^t\boldsymbol{\mu} - E[\mathbf{x}^t\mathbf{R} \mid \mathbf{x}^t\mathbf{R} \leqslant q_\alpha(\mathbf{x})], \end{aligned} \tag{2.7}$$

where the second equation follows from (2.6), and the third equation follows from (2.4), and where

$$q_\alpha(\mathbf{x}) = \inf\{z \mid P(\mathbf{x}^t\mathbf{R} \leqslant z) \geqslant \alpha\}$$

is the α-quantile of portfolio $\mathbf{x}$. According to (2.7), for every level of target return $\mathbf{x}^t\boldsymbol{\mu} = r_p$, shortfall captures the tail risk below the α-quantile of the portfolio. A small α parameter, say 1%, can then serve to uncover risk in the tail, that may not be obvious if a symmetric measure like standard deviation is used.

We next show how shortfall is related to VaR, CVaR, and the standard deviation.

2.1.1 Relation to VaR and CVaR

Recall that the α-level VaR of portfolio $\mathbf{x}$ can be defined as

$$\text{VaR}(\mathbf{x}) = -q_\alpha(\mathbf{x}),$$

and the α-level CVaR of portfolio $\mathbf{x}$ can be defined as

$$\text{CVaR}(\mathbf{x}) = -E[\mathbf{x}^t\mathbf{R} \mid \mathbf{x}^t\mathbf{R} \leqslant q_\alpha(\mathbf{x})].$$

Clearly, we have

$$s_\alpha(\mathbf{x}) = \mathbf{x}^t\boldsymbol{\mu} + \text{CVaR}(\mathbf{x}).$$

Some advantages of shortfall over CVaR are as follows:

- first, the mean-adjustment makes shortfall proportional to standard deviation when returns are elliptically symmetric (see next subsection). Therefore, when returns are elliptically symmetric, shortfall portfolio optimization recovers optimal mean-variance portfolios.

- second, shortfall consistent with the traditional definition of risk as a measure of volatility independent of a location parameter, typically the mean. For example, the standard deviation is independent of the mean. Notice in particular that for $\alpha = 50\%$,

$$\begin{aligned} s_{0.5}(\mathbf{x}) &= \mathbf{x}^t\boldsymbol{\mu} - E[\mathbf{x}^t\mathbf{R} \mid \mathbf{x}^t\mathbf{R} \leqslant q_{0.5}(\mathbf{x})] \\ &= E|\mathbf{x}^t\mathbf{R} - q_{0.5}(\mathbf{x})|, \end{aligned}$$

  where $q_{0.5}(\mathbf{x})$ is the median of portfolio $\mathbf{x}$. $s_{0.5}(\mathbf{x})$ is the mean of absolute deviations from the median, which turns out to be a more robust estimation of volatility than the standard deviation when returns have heavier tails and more tail-dependence than the Gaussian. More on this in Chapter 5.

2.1.2 Relation to the Standard Deviation

We show here that when the multivariate distribution of returns is elliptically symmetric, the shortfall of a portfolio is proportional to its standard deviation, where the coefficient of proportionality only depends on α. Therefore, when the multivariate distribution of returns is elliptically symmetric, mean-variance optimization is strictly equivalent to mean-shortfall optimization for any α.

Proposition 1 *(a) If the distribution of* $\mathbf{R}$ *is multivariate normal with mean* $\boldsymbol{\mu}$ *and covariance matrix* Σ*, then*

$$s_\alpha(\mathbf{x}) = \frac{\phi(z_\alpha)}{\alpha}\sqrt{\mathbf{x}^t\Sigma\mathbf{x}},$$

where $\phi()$ *is the density of the standard normal, and* z_α *is its* α*-quantile.*

(b) If the distribution of $\mathbf{R}$ *is multivariate elliptically symmetric with mean* $\boldsymbol{\mu}$ *and covariance matrix*

Σ, *then*

$$s_\alpha(\mathbf{x}) = g(\alpha)\sqrt{\mathbf{x}^t\Sigma\mathbf{x}},$$

where $g(.)$ *is independent of* $\mathbf{x}$.

Proof. (a) When the distribution of $\mathbf{R}$ is multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix Σ, then the variable $X = \mathbf{x}^t\mathbf{R}$ has a univariate normal distribution with mean $\mu = \mathbf{x}^t\boldsymbol{\mu}$ and variance $\sigma^2 = \mathbf{x}^t\Sigma\mathbf{x}$. Therefore,

$$s_\alpha(\mathbf{x}) = \mu - E[X \mid X \leqslant q_{\alpha,X}]$$

where $q_{\alpha,X}$ is the α-quantile of variable X. Therefore,

$$\begin{aligned}
s_\alpha(\mathbf{x}) &= \mu - \frac{1}{\alpha\sigma\sqrt{2\pi}}\int_{-\infty}^{q_{\alpha,X}} x\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\
&= \frac{1}{\alpha\sigma\sqrt{2\pi}}\int_{-\infty}^{q_{\alpha,X}} (x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\
&= \frac{\sigma}{\alpha\sqrt{2\pi}}\int_{-\infty}^{z_\alpha} y\exp\left(-\frac{y^2}{2}\right)dy \\
&= \frac{\phi(z_\alpha)}{\alpha}\sigma.
\end{aligned}$$

The proof of (b) is analogous. ■

The remarkable implication of the previous proposition is that when returns are multivariate elliptically symmetric, the solutions to the mean-shortfall and mean-variance problems are identical.

2.1.3 Optimal Bounds on Shortfall

In this subsection, for given values of the mean and standard deviation of a portfolio, we obtain universal bounds on the quantile and shortfall that are best possible in the sense that there exist probability distributions that attain them. This allows us to compute bounds on the quantile and shortfall even if the distribution of returns is unknown. For ease of notation, let us write $s_\alpha := s_\alpha(\mathbf{x})$ and $q_\alpha := q_\alpha(\mathbf{x})$, and $\sigma := \sqrt{\mathbf{x}^t\Sigma\mathbf{x}}$. We use the techniques from Bertsimas and Popescu (1999) to derive these bounds.

Theorem 2 *The inequalities in Table 2.1 are valid and best possible..*

(a) Optimal Bounds on q_α given μ and σ^2	$-\sigma\sqrt{(1-\alpha)/\alpha}+\mu \leqslant q_\alpha \leqslant \sigma\sqrt{\alpha/(1-\alpha)}+\mu,$
(b) Optimal Bounds on s_α given μ, σ^2, and q_α	$\left\{\begin{array}{ll} -(q_\alpha-\mu) & \text{if } q_\alpha-\mu\leqslant 0 \\ (q_\alpha-\mu)(1-\alpha)/\alpha & \text{if } q_\alpha-\mu>0 \end{array}\right\} \leqslant s_\alpha \leqslant \sigma\sqrt{(1-\alpha)/\alpha},$
(c) Optimal Bounds on s_α given μ and σ^2	$0\leqslant s_\alpha \leqslant \sigma\sqrt{(1-\alpha)/\alpha}.$

Table 2.1: Optimal Bounds on Quantile and Shortfall

α	$\frac{\varphi(z_\alpha)}{\alpha}$	$\sqrt{\frac{1-\alpha}{\alpha}}$
0.1	1.7550	3.0000
0.05	2.0627	4.3589
0.01	2.6652	9.9499
0.005	2.8919	14.1067

Table 2.2: Calculating shortfall: normal approximation vs. worst case distribution.

Proof. See appendix. ∎

Given μ and σ^2, Table 2.2 compares the normal approximation to the shortfall, given in Proposition 1, and the worst case value given in Table 2.1. We see that when $\alpha = 10\%$, the normal approximation underestimates shortfall by up to $(3-1.755)/3 = 41\%$, and when $\alpha = 1\%$, then normal approximation underestimates shortfall by up to $(9.9499-2.6652)/9.9499 = 73\%$.

2.2 Relation to Expected Utility Maximization

Now let us consider the mean-shortfall portfolio optimization problem

$$\begin{aligned} \text{minimize} \quad & s_\alpha(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x}^t\mathbf{e} = 1, \\ & \mathbf{x}^t\boldsymbol{\mu} = r_p, \end{aligned} \tag{2.8}$$

where r_p is the target return. To relate mean-shortfall portfolio selection to expected utility optimization, we will use the following theorem, due to Levy and Kroll (1978), and which we prove analytically in the appendix. This theorem states conditions under which one portfolio is preferred to another by all investors whose utility functions are in $\mathbb{U}_2 := \{u(x) \mid u'(x) > 0, -\infty < u''(x) < 0 \; \forall x \in \mathbb{R}\}$, the class of all increasing, concave utility functions. These investors form the family of wealth-seeking, risk-averse investors. Note that this theorem fits into the large literature on stochastic dominance, which attempts to find distributional conditions under which one alternative investment will be preferred to another by all investors whose utilities belong to a certain class of functions - see Levy (1992) and Fishburn (1980) for review papers on this topic.

Theorem 3 (Levy and Kroll, 1978) *Suppose that X and Y are the random returns of two portfolios. Then X is preferred to Y by all investors in $\mathbb{U}_2$ (or X "second-order stochastically dominates" Y) if and only if*

$$E(X|X \leqslant q_{\alpha,X}) \geqslant E(Y|Y \leqslant q_{\alpha,Y}) \ \forall \alpha \in (0,1), \text{ and } > \text{ for some } \alpha,$$

where $q_{\alpha,X}$ and $q_{\alpha,Y}$ are, respectively, the α-quantiles of X and Y.

Let $\alpha \in (0,1)$ be fixed. From the alternative definition of shortfall - see Equation (2.7) - and from Levy and Kroll's theorem, it is then clear that given a target return r_p, the shortfall minimizing portfolio (or portfolios) will be preferred by at least one investor in $\mathbb{U}_2$ (otherwise, this would contradict the theorem). Another way of saying this is that the shortfall minimizing portfolio is non-dominated, in the sense that there exists no portfolio with the same mean and higher shortfall, that is preferred by all investors in $\mathbb{U}_2$. Note that this result holds for arbitrary distributions of returns. In fact, mean-variance portfolios have this property also. However, under the assumption that the distribution of returns is elliptically symmetric, they are the only such portfolios. We just saw that in general, i.e. under asymmetry, they are not.

2.3 Mathematical Properties of Shortfall

The results that we prove here are of independent interest, and are used in the next section.

2.3.1 Positive Homogeneity and Convexity

Proposition 4 *Let the random variable $\mathbf{R} \in \mathbb{R}^N$ have mean $\boldsymbol{\mu}$. Then*

(a) $s_\alpha(\mathbf{x}) \geqslant 0$ for all $\mathbf{x} \in \mathbb{R}^N$ and $\alpha \in (0,1)$. Moreover, the shortfall $s_\alpha(\mathbf{x})$ is equal to zero for some $\mathbf{x}$ and α if and only if $\mathbf{x}^t\mathbf{R}$ is a constant with probability 1.

(b) The shortfall is positively homogeneous, i.e, for all $t \geqslant 0$*,*

$$s_\alpha(t\mathbf{x}) = ts_\alpha(\mathbf{x}).$$

(c) The shortfall is convex, i.e. if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$*, and* $\lambda \in [0,1]$*, then*

$$s_\alpha(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \leqslant \lambda s_\alpha(\lambda\mathbf{x}) + (1-\lambda)s_\alpha(\mathbf{y}).$$

Proof. (a) Let $X := \mathbf{x}^t\mathbf{R}$, and $q_\alpha := q_\alpha(\mathbf{x})$. Then, conditioning on the event $X \leqslant q_\alpha$, we get

$$\begin{aligned} s_\alpha(\mathbf{x}) &= E(X) - E[X \mid X \leqslant q_\alpha] \\ &= (1-\alpha)\{E[X \mid X > q_\alpha] - E[X \mid X \leqslant q_\alpha]\}, \end{aligned}$$

so that $s_\alpha(\mathbf{x}) \geqslant 0$ and is equal to 0 if and only if $P(\mathbf{x}^t\mathbf{R}$ is constant$) = 1$. Clearly, if $P(\mathbf{x}^t\mathbf{R}$ is constant$) = 1$, then $s_\alpha(\mathbf{x}) = 0$ for all $\alpha \in (0,1)$.

(b) Clearly $q_\alpha(t\mathbf{x}) = tq_\alpha(\mathbf{x})$ for all $t \geqslant 0$, and the result follows.

(c) Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, and let $X := \mathbf{x}^t\mathbf{R}$ and $Y := \mathbf{y}^t\mathbf{R}$. Then

$$\begin{aligned} & s_\alpha(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \\ = \; & [\lambda\mathbf{x}+(1-\lambda)\mathbf{y}]^t\boldsymbol{\mu} - E[\lambda X + (1-\lambda)Y \mid \lambda X + (1-\lambda)Y \leqslant q_{\alpha,\lambda}] \\ = \; & \lambda\left(\mathbf{x}^t\boldsymbol{\mu} - E[X \mid \lambda X + (1-\lambda)Y \leqslant q_{\alpha,\lambda}]\right) + (1-\lambda)\left(\mathbf{y}^t\boldsymbol{\mu} - E[Y \mid \lambda X + (1-\lambda)Y \leqslant q_{\alpha,\lambda}]\right), \end{aligned}$$

where $q_{\alpha,\lambda}$ is the α-quantile of random variable $\lambda X + (1-\lambda)Y$. But it is obvious that

$$\begin{aligned} & E[X \mid \lambda X + (1-\lambda)Y \leqslant q_{\alpha,\lambda}] \\ & \geqslant E[X \mid \lambda X \leqslant q_\alpha(\lambda\mathbf{x})] \\ & = E[X \mid X \leqslant q_\alpha(\mathbf{x})]. \end{aligned}$$

The same reasoning applied to Y yields the result

$$\begin{aligned} & s_\alpha(\lambda \mathbf{x}+(1-\lambda)\mathbf{y}) \\ \leqslant\ & \lambda\left(\mathbf{x}^t\boldsymbol{\mu} - E[X \mid X \leqslant q_\alpha(\mathbf{x})]\right) + (1-\lambda)\left(\mathbf{y}^t\boldsymbol{\mu} - E[Y \mid Y \leqslant q_\alpha(\mathbf{y})]\right) \\ =\ & \lambda s_\alpha(\mathbf{x}) + (1-\lambda)s_\alpha(\mathbf{y}), \end{aligned}$$

and we are done. ■

Remark. Artzner, Delbaen, Eber and Heath (1999) propose four axioms which, they argue, every measure of risk should satisfy. They call measures of risk that satisfy these four axioms coherent. While Artzner et al. (1999) considered only discrete probability spaces, Delbaen (2000) extends their definitions to arbitrary probability spaces. Let X be an investment's net return. A coherent risk measure $\kappa(.)$ satisfies

(i) (Translation Invariance). For any $a \in \mathbb{R}$, $\kappa(X+a) = \kappa(X) - a$.

(ii) (Subadditivity). For any investments with net returns X and Y, $\kappa(X+Y) = \kappa(X) + \kappa(Y)$.

(iii) Positive Homogeneity). For all $t \geqslant 0$, $\kappa(tX) = t\kappa(X)$.

(iv) (Positivity). If $P(X \geqslant 0) = 1$, then $\kappa(X) \leqslant 0$.

It is easy to verify that CVaR is in fact coherent, while risk measures independent of the mean, such as the standard deviation, semi-variance, and shortfall, violate axioms (i) and (iv). Notice however that these axioms were developed in the context of setting margin requirements for certain investment strategies, and not for the purpose of portfolio selection. In the context of portfolio selection, we feel that the only important axioms are (ii) and (iii), which together imply that the risk measure is convex, and therefore that risk is diversifiable. Note that VaR, or even a centered version of VaR, is not convex - and in fact is not coherent - and that may be a reason for not using VaR in portfolio selection. Convexity also has important implications for portfolio optimization and the statistical analysis of portfolio estimation, as we will see in, respectively, Chapters 3 and 5.

2.3.2 Derivatives

The gradient of the shortfall, which we present below, can also be found in Tasche (2000) and in Scaillet (2000).

Proposition 5 *Let the random variable* $\mathbf{R} \in \mathbb{R}^N$ *have mean* $\boldsymbol{\mu}$ *and a continuous density. Then the gradient of* $s_\alpha(\mathbf{x})$ *with respect to* $\mathbf{x} \in \mathbb{R}^N$ *is*

$$\nabla_{\mathbf{x}} s_\alpha(\mathbf{x}) = \boldsymbol{\mu} - E[\mathbf{R} \mid \mathbf{x}^t\mathbf{R} \leqslant q_\alpha(\mathbf{x})], \tag{2.9}$$

and the Hessian of $s_\alpha(\mathbf{x})$ *with respect to* $\mathbf{x}$ *is*

$$\nabla^2_{\mathbf{x}} s_\alpha(\mathbf{x}) = \frac{f_{\mathbf{x}^t\mathbf{R}}(q_\alpha(\mathbf{x}))}{\alpha} Cov[\mathbf{R} \mid \mathbf{x}^t\mathbf{R} \leqslant q_\alpha(\mathbf{x})], \tag{2.10}$$

where $f_{\mathbf{x}^t\mathbf{R}}(.)$ *is the density of* $\mathbf{x}^t\mathbf{R}$ *and where* $Cov[\mathbf{R} \mid .]$ *is the conditional covariance matrix of* $\mathbf{R}$.

Proof. Consider x_i, the i^{th} element of $\mathbf{x}$. Then

$$\frac{\partial s_\alpha(\mathbf{x})}{\partial x_i} = \mu_i - \frac{1}{\alpha}\frac{\partial}{\partial x_i} E[\mathbf{x}^t\mathbf{R} \mid \mathbf{x}^t\mathbf{R} \leqslant q_\alpha(\mathbf{x})],$$

where μ_i is the i^{th} element of $\boldsymbol{\mu}$. Writing the last expectation as a bivariate integral in the variables $U = \sum_{j \neq i} x_j R_j$ and $V = R_i$ and differentiating with respect to x_i, we obtain

$$\begin{aligned}
\frac{\partial s_\alpha(\mathbf{x})}{\partial x_i} &= \mu_i - \frac{1}{\alpha}\frac{\partial}{\partial x_i} \int\int_{\mathbb{R}^2} (u + x_i v) 1_{\{u + x_i v \leqslant q_\alpha(\mathbf{x})\}} f_{U,V}(u,v)\,du dv \\
&= \mu_i - \frac{1}{\alpha}\frac{\partial}{\partial x_i} \int_{-\infty}^{\infty}\int_{-\infty}^{q_\alpha(\mathbf{x}) - x_i v} (u + x_i v) f_{U,V}(u,v)\,du dv \\
&= \mu_i - \frac{1}{\alpha}\frac{\partial}{\partial x_i} \int_{-\infty}^{\infty}\int_{-\infty}^{q_\alpha(\mathbf{x}) - x_i v} v f_{U,V}(u,v)\,du dv \\
&\quad - \frac{1}{\alpha}\frac{\partial}{\partial x_i} \int_{-\infty}^{\infty} \left(\frac{\partial q_\alpha(\mathbf{x})}{\partial x_i} - v \right) f_{U,V}(q_\alpha(\mathbf{x}) - x_i v, v)\,du dv.
\end{aligned} \tag{2.11}$$

By definition of the quantile $q_\alpha(\mathbf{x})$

$$\begin{aligned}\alpha &= \int\int_{\mathbb{R}^2} 1_{\{u+x_i v \leqslant q_\alpha(\mathbf{x})\}} f_{U,V}(u,v)dudv \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{q_\alpha(\mathbf{x})-x_i v} f_{U,V}(u,v)dudv.\end{aligned}$$

Differentiating the last expression with respect to x_i we see that the last term in (2.11) is zero. The result then follows.

The proof expression for the Hessian follows from similar arguments. ■

Remark. (2.10) also implies that the shortfall is convex.

2.4 Some Results from Classical Portfolio Analysis

In this section we recover two results from classical mean-variance portfolio analysis (see Ingersoll, 1987): two fund separation in the presence of riskless asset, and the concept of a portfolio's beta.

2.4.1 Two Fund Separation in the Presence of a Riskless Asset

One result from classical mean-variance analysis is that the set of solutions to the mean-variance problem

$$\begin{aligned}\text{minimize} \quad & \mathbf{x}^t\Sigma\mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^t\mathbf{e} = 1, \\ & \mathbf{x}^t\boldsymbol{\mu} = r_p,\end{aligned} \tag{2.12}$$

where r_p is the target portfolio expected return, form a convex curve is in reward-risk space, i.e. in mean-standard deviation space, as shown in Figure 2-1. In the presence of a riskless asset with rate of return r_f, the problem becomes

$$\begin{aligned}\text{minimize} \quad & \mathbf{x}^t\Sigma\mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^t\boldsymbol{\mu} + (1-\mathbf{x}^t\mathbf{e})r_f = r_p,\end{aligned} \tag{2.13}$$

and the solutions to (2.13) for $r_p > r_f$ (and assuming that the portfolio with minimum variance, among all solutions to (2.12) has expected return greater than r_f) span a line which passes between r_f and the tangency portfolio, marked by a circle in Figure 2-1. Any portfolio on this frontier can be constructed by combining the riskless asset with the tangency portfolio. And if the all investors choose portfolios that solve (2.13), then the tangency portfolio is the market, which is the Capital Asset Pricing Model result.

Tasche (2001) shows that this form of two-fund separation holds whenever the risk measure is convex and positively homogeneous, which is the case with shortfall, as we showed in the previous section. Therefore, the solutions to

$$\begin{aligned} \text{minimize} \quad & s_\alpha(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1, \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p, \end{aligned} \tag{2.14}$$

for all r_p form a convex curve in reward risk space, i.e. in mean-shortfall space, and in the presence of a riskless asset, the solutions to

$$\begin{aligned} \text{minimize} \quad & s_\alpha(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x}^t \boldsymbol{\mu} + (1 - \mathbf{x}^t \mathbf{e}) r_f = r_p, \end{aligned} \tag{2.15}$$

span a line which passes between the riskless asset and the tangency portfolio.

2.4.2 Shortfall Beta

We define the shortfall beta, derived by analogy with the standard beta from mean-variance analysis.

Definition 6 *Assume that* $\mathbf{R}$ *has mean* $\boldsymbol{\mu}$*, and a continuous density. Then the* α*-shortfall beta of asset* i *with respect to portfolio* $\mathbf{x} \in \mathbb{R}^N$ *is*

$$\beta_{i,\alpha}(\mathbf{x}) = \frac{1}{s_\alpha(\mathbf{x})} \frac{\partial s_\alpha(\mathbf{x})}{x_i} = \frac{\mu_i - E(R_i \mid \mathbf{x}^t \mathbf{R} \leq q_\alpha(\mathbf{x}))}{s_\alpha(\mathbf{x})}.$$

The quantity $\beta_{i,\alpha}(\mathbf{x})$ can be interpreted as the relative change in shortfall when increasing the weight of asset i. Note that, as with the standard beta from mean-variance, we have

$$\sum_{i=1}^{N} x_i \beta_{i,\alpha}(\mathbf{x}) = 1,$$

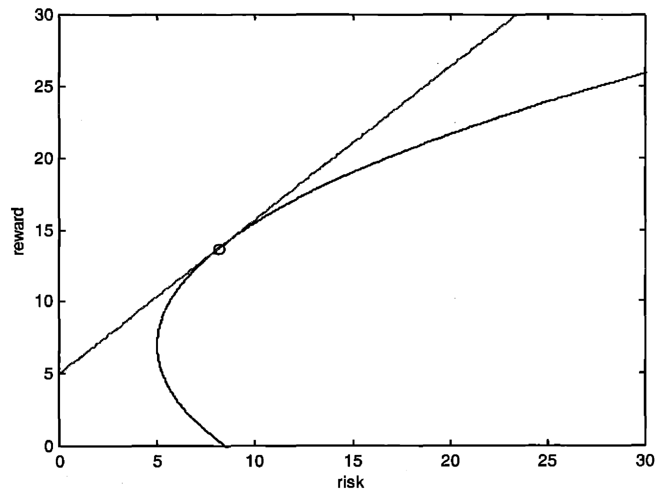

Figure 2-1: Two-fund separation: – : minimum risk frontier, -. : minimum risk frontier in presence of riskless asset, o = tangent portfolio. $r_f = 5$.

which effectively gives a decomposition of the portfolio shortfall into the individual assets' contributions, cf. Tasche (2000) for a similar decomposition in terms of VaR. The following result is similar to the classical Capital Asset Pricing Model result, relating the expected return of an optimal portfolio with the expected returns of the individual assets.

Proposition 7 *Call* $\mathbf{x}_\alpha$ *the optimal solution to 2.15. Then,* $x_{i,\alpha}$*, the ith component of* $\mathbf{x}_\alpha$*, satisfies*

$$\mu_i - r_f = \beta_{i,\alpha}(\mathbf{x}_\alpha)(r_p - r_f).$$

Proof. The Lagrangian of Problem (2.15) is $L(\mathbf{x}, \gamma) = s_\alpha(\mathbf{x}) + \gamma(r_p - (\mathbf{x}^t\boldsymbol{\mu} + (1 - \mathbf{x}^t\mathbf{e})r_f))$. Taking its partial derivatives with respect to $\mathbf{x}$ and γ, and setting them to 0 yields

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{x}} &= \boldsymbol{\mu} - E(\mathbf{R} \mid \mathbf{x}^t\mathbf{R} \leqslant q_\alpha(\mathbf{x})) + \gamma(\boldsymbol{\mu} - \mathbf{e}r_f) = 0 \qquad (2.16) \\ \frac{\partial L}{\partial \gamma} &= r_p - (\boldsymbol{\mu}^t\mathbf{x} + (1 - \mathbf{x}^t\mathbf{e})r_f) = 0. \end{aligned}$$

Now, multiplying the equation (2.16) by $\mathbf{x}^t$ yields

$$\mathbf{x}^t\boldsymbol{\mu} - E(\mathbf{x}^t\mathbf{R} \mid \mathbf{x}^t\mathbf{R} \leqslant q_\alpha(\mathbf{x})) + \gamma(\mathbf{x}^t\boldsymbol{\mu} - \mathbf{x}^t\mathbf{e}r_f) = s_\alpha(\mathbf{x}) + \gamma(r_p - r_f) = 0,$$

so that $\gamma = -s_\alpha(\mathbf{x})/(r_p - r_f)$. Plugging in the value of γ in equation (2.16) we get

$$(\mu - \mathbf{e}r_f) = \frac{\boldsymbol{\mu} - E(\mathbf{R} \mid \mathbf{x}_\alpha^t\mathbf{R} \leqslant q_\alpha(\mathbf{x}_\alpha))}{s_\alpha(\mathbf{x}_\alpha)}(r_p - r_f)$$

or

$$(\boldsymbol{\mu} - \mathbf{e}r_f) = \beta_\alpha(\mathbf{x}_\alpha)(r_p - r_f), \tag{2.17}$$

by definition of $\beta_\alpha(\mathbf{x}_\alpha)$, and we are done. ■

More interesting, however, is the shortfall beta's dependence on α. Some assets may contribute little to risk for some levels of α, but significantly at other levels of α. We illustrate this fact empirically in the next section.

2.5 Computational Experiment

In this experiment, we illustrate the fact that in the presence of asymmetrically distributed data, mean-variance and mean-shortfall optimization may yield different portfolio weights. Furthermore, we show that the risk-sensitivity of a portfolio with respect to any of its underlying assets, its *shortfall beta*, may also depend on the value of α.

Data

We generate data for three assets: (A) symmetrically distributed, (B) skewed to the left and (C) skewed to the right . The assets are designed to have the same mean and standard deviation, and to be uncorrelated with each other, which will make mean-variance optimization blind to their differences. Asset A has a normal distribution. Asset B is the distribution of a portfolio consisting in a stock with a lognormal distribution, combined with a call on 75% of the value of the stock, financed by borrowing at a riskless rate $r_f = 2.5\%$. Finally, asset C is the distribution of a portfolio consisting in a stock with a lognormal distribution, combined with a put on 75% of the value of the stock, financed by borrowing

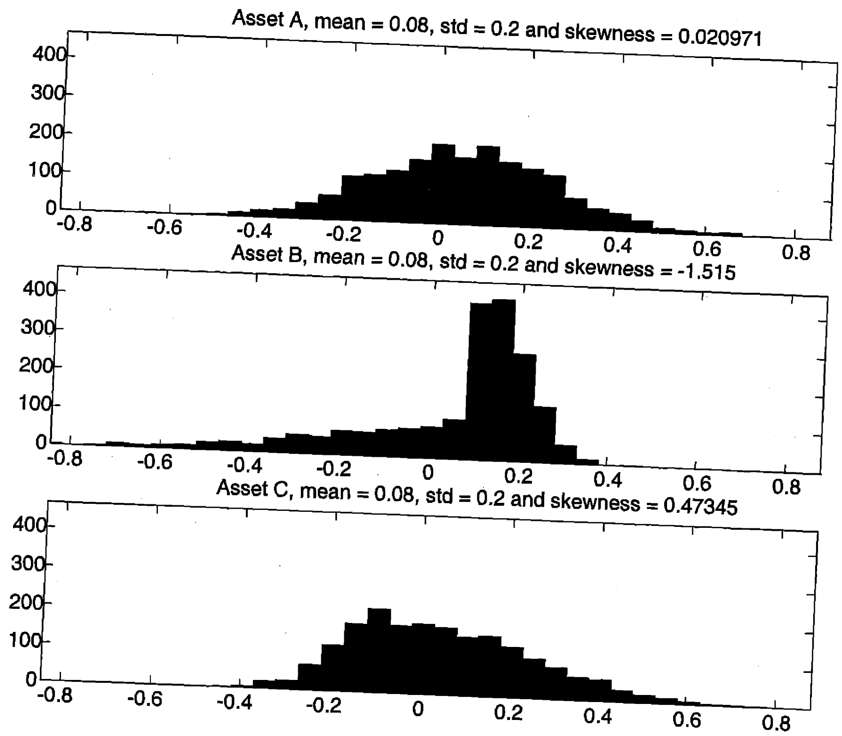

Figure 2-2: Histogram of returns for each asset, sample of size 2000.

at a continuously compounded rate of return $r_f = 2.5\%$.

The log of the underlying stock returns have normal distributions with mean equal to 8% and standard deviation equal to 20%. The price of the call and put options, used to calculate the returns of those options, were determined using the classical Black-Scholes formula, assuming a maturity of one period, and a strike price equal to the price of the asset. In each sample that we use in our experiments, the mean and standard deviation of each asset are standardized to be 8% and 20% respectively. Figure (2-2) shows histograms of the returns of each asset, for a sample of 2000 observations. The distributional asymmetry of assets B and C is clear.

Shortfall and Shortfall Beta of Fixed Weight Portfolios

Next we examine the shortfall of three different fixed weight portfolios: $x_1 = [\ 1/3 \quad 1/3 \quad 1/3\ ]$, $x_2 = [\ 0.1 \quad 0.8 \quad 0.1\ ]$ and $x_3 = [\ 0.1 \quad 0.1 \quad 0.8\ ]$ for different values of α. We repeat the following experiment 100 times:

(i) generate a sample of $T = 2000$ observations from the asymmetric multivariate distribution described above;

(ii) standardize the data so that each asset has mean and standard deviation 8% and 20% respectively;

(iii) calculate the shortfall of each portfolio x_1, x_2, and x_3, and the beta of each asset with respect to each portfolio.

In Figure (), for values of α between 2% and 50%, we plot the median shortfall, over the 100 experiments, of each portfolio. We also plot the 10% and 90% quantiles, over the 100 experiments, to give intuition about the variability of the shortfall estimates. A expected, portfolio x_2 has a higher shortfall than portfolio x_3 for values of α below 40%, reflecting that fact that portfolio x_2 is highly loaded on the negatively skewed asset B, whereas portfolio x_3 is heavily loaded on the positively skewed asset C. Note however that portfolio x_1, the equally weighted portfolio, has the lowest shortfall of all portfolios, at every value of α, a clear reflection of the power of diversification.

In Figure 2-4, we report the shortfall beta of each asset with respect to portfolio x_1. For portfolio x_1 and low values of α, asset B has the highest shortfall beta, indicating asset B is responsible for most of the shortfall of the portfolio. Asset C has the smallest shortfall beta. For portfolio x_1 and high values of α, all assets have shortfall beta about 1, indicating comparable contributions to the portfolio's shortfall. The message is that contrary to the standard beta, the shortfall beta can vary with α, indicating that an asset can have different contributions to the risk of a portfolios at different values of α.

Weights given by Mean-Variance and Mean-Shortfall Optimization

Next we propose to compare the portfolio weights obtained via mean-variance optimization and via mean-shortfall optimization on samples of our asymmetric distribution. We repeat the following experiment 100 times:

generate a sample of $T = 2000$ observations from the asymmetric multivariate distribution described above;

standardize the data so that each asset has mean and standard deviation 8% and 20% respectively;

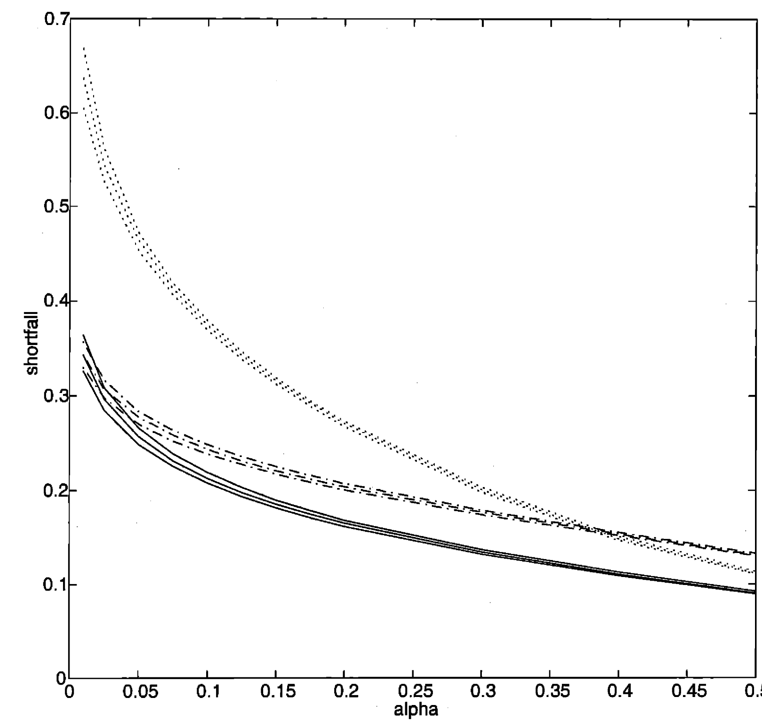

Figure 2-3: Shortfall of portfolios x_1 (—), x_2 ($\cdots$), and x_3 (-.-.). For each portfolio and α-level combination, the 10%, 50%, and 90% quantiles over 100 samples are represented.

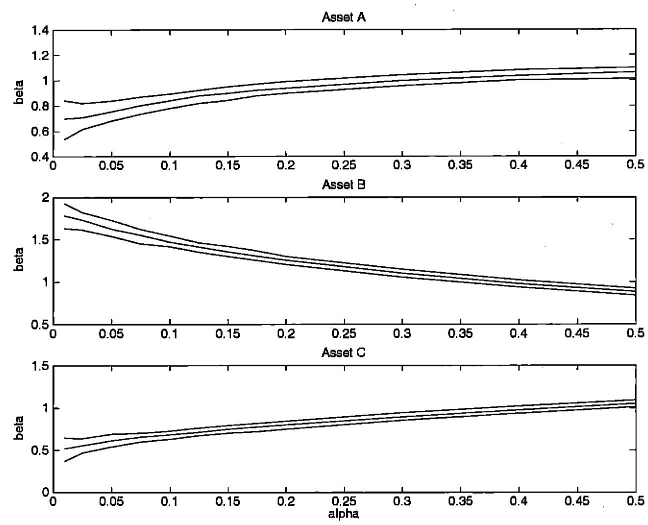

Figure 2-4: Shortfall beta of each asset with respect to portfolio x_1. For each asset, and α-level combination, the 10%, 50%, and 90% quantiles over 100 samples are represented.

do mean-variance optimization and mean-shortfall optimization for values of α between 2% and 50%. We use a target rate of return of $r_p = 8\%$, and constrain the weights to be non-negative.

Figure 2-5 shows cumulative distribution function of returns for the optimal portfolios MV, $S(0.01)$, and $S(0.10)$ on one sample of 2000 observations: the shortfall portfolios dominate in the tails, as expected, but the MV portfolio dominate in the mid-range (+/- 10%).

Figure 2-6 gives the weights assigned to each asset, for α ranging from 2% to 50%. We see that MV optimization (which is independent of α) gives equal weight to each asset, as expected. Shortfall optimization, especially for low levels of α, puts less weight on asset B, and extra weight on asset C, also as expected. The weight assigned to asset A, the symmetrically distributed variable, seems to be the same for MV and shortfall optimization.

2.6 Conclusion

In this chapter, we introduced shortfall, a quantile-based asymmetric measure of risk. We showed how it arose naturally as a measure of risk by considering distributional conditions of second order stochastic dominance. We examined its connections with other risk measures such as VaR, CVaR, and

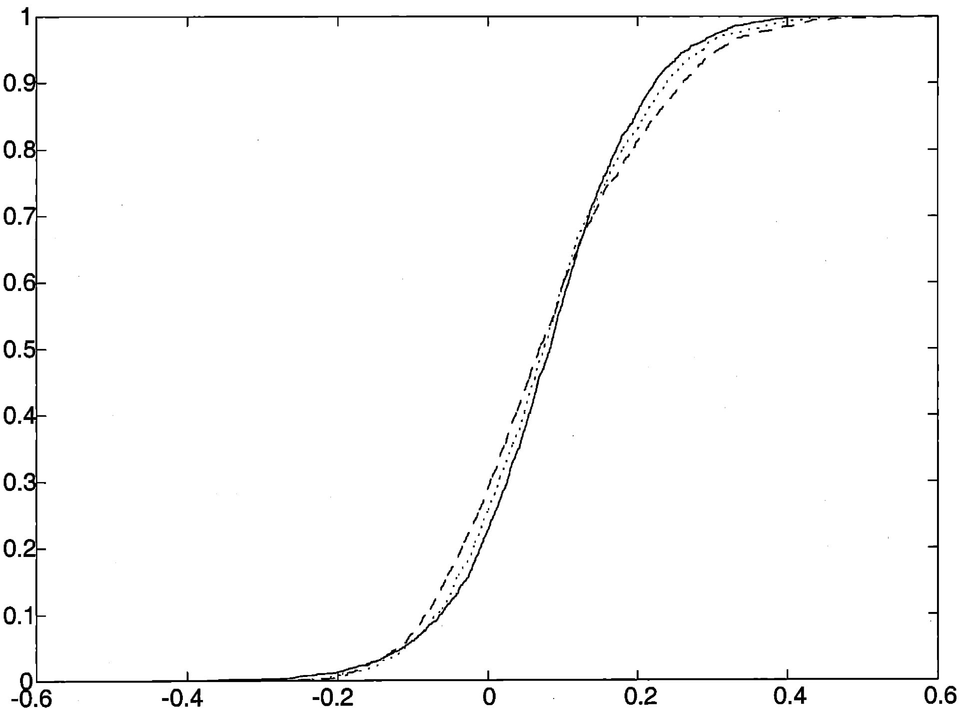

Figure 2-5: Cumulative distribution of MV (—), S(0.10) (···), and S(0.01) (-.), sample of 2000 observations.

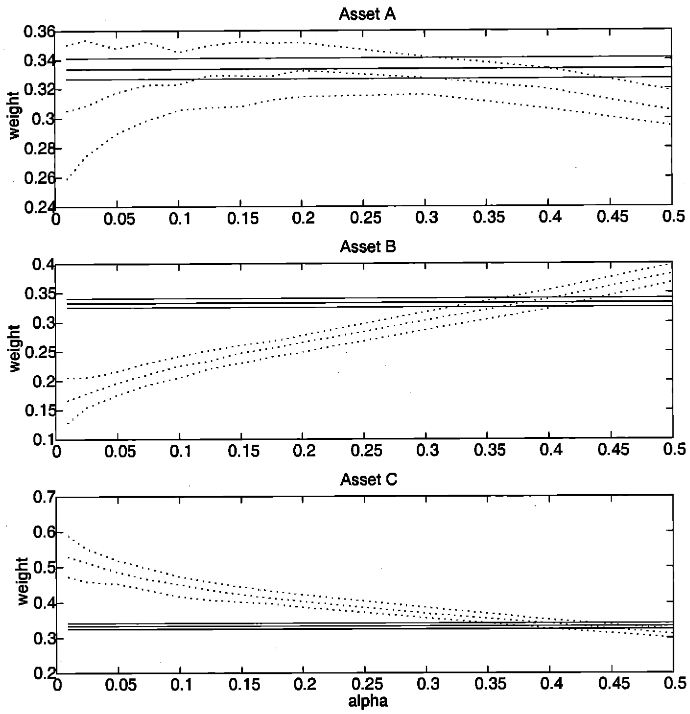

Figure 2-6: Weights for each asset:: $-$: V, $\cdots = \mathrm{s}_\alpha$. For each portfolio, asset, and α-level combination, the 10%, 50%, and 90% quantiles over 100 samples are plotted.

the standard deviation, and its mathematical properties. We also showed that two results from classical mean-variance analysis, two-fund separation with a riskless asset, and the concept of an asset's beta, extended to mean-shortfall portfolio analysis. In the next Chapters 3 and 4, we examine the sample shortfall portfolio optimization problem, and develop a central limit theorem for the shortfall estimator. In Chapter 5 we show that the shortfall estimator can outperform the variance portfolio estimator even when returns are elliptically symmetric.

Chapter 3

Sample α-Shortfall Portfolio Optimization

Given a sample of T return observations $\mathbf{R}_1, \cdots, \mathbf{R}_T$, where $\mathbf{R}_i \in \mathbb{R}^N$, $i = 1, \cdots, T$, for any $\alpha \in (0, 1)$, consider the sample α-shortfall portfolio optimization problem, introduced in Chapter 1, and defined generally as

$$\begin{array}{ll} \text{minimize}_{\mathbf{x},q} & \frac{1}{T}\sum_i^T \rho_\alpha(\mathbf{x}^t\mathbf{R}_i - q) \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{array} \tag{3.1}$$

where

$$\rho_\alpha(z) = z - \frac{1}{\alpha} z 1_{\{z<0\}}, \tag{3.2}$$

and where $\mathbf{A}$ is an $(M \times N)$ matrix with linearly independent rows and $\mathbf{b}$ is an M-dimensional vector. In this chapter, we develop a linear programming formulation of problem (3.1) and examine some of its properties. We emphasize comparisons between Problem (3.1) and the sample variance portfolio optimization problem defined as

$$\begin{array}{ll} \text{minimize} & \mathbf{x}\hat{\Sigma}\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{array} \tag{3.3}$$

where

$$\hat{\Sigma} = \frac{1}{T}\sum_{i=1}^{T}(\mathbf{R}_i - \bar{\mathbf{R}})(\mathbf{R}_i - \bar{\mathbf{R}})^t$$

is the sample covariance matrix, and

$$\bar{\mathbf{R}} = \frac{1}{T}\sum_{i=1}^{T}\mathbf{R}_i$$

is the sample covariance mean.

This chapter is organized as follows. In Section 1 we develop a linear programming (LP) formulation of Problem (3.1) with $N+1+T$ variables and $M+2T$ constraints, with $N+1$ dense columns. In Section 2 we present an alternative formulation of the sample variance portfolio optimization problem which does not make use of the covariance matrix. This formulation highlights the similarities between the shortfall and variance problems, and has computational advantages when $T < N - M$. We also provide alternative formulations for the sample CVaR, LPM_1 and LPM_2 portfolio optimization problems, which show that each of these problems is solvable using a generic LP or QP solver. In Section 3 we note that when $T < N - M$, the sample α-shortfall and the sample variance portfolio optimization problems will have a non-unique and unbounded set of solutions. We introduce regularization in sample portfolio optimization to guarantee solution uniqueness. Regularization is implemented by adding a strictly convex penalty to the objective of the sample portfolio optimization problem, where the penalty is proportional to either the L_1- or the L_2-norm of the weight vector $\mathbf{x}$. Finally, in Section 4 we present the results of some numerical experiments. First, we compare the running times of our portfolio optimization formulations against the classical formulation of the variance problem *with* covariance matrix. Then we show that the sample α-shortfall portfolio optimization problem, because it is an LP, can readily incorporate cardinality constraints. We use this fact to solve an index tracking problem using shortfall optimization.

3.1 Linear Programming Formulation of the Sample α-Shortfall Portfolio Optimization Problem

Using the definition of $\rho_\alpha(.)$ we can write Problem (3.1) as

$$\begin{array}{ll} \text{minimize}_{\mathbf{x},q} & \frac{1}{T}\sum_i^T(\mathbf{x}^t\mathbf{R}_i - q) - \frac{1}{T}\sum_i^T \frac{1}{\alpha}(\mathbf{x}^t\mathbf{R}_i - q)1_{\{\mathbf{x}^t\mathbf{R}_i \leqslant q\}} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{array}$$

which can then be rewritten (using the usual LP formulation tricks), letting

$$\bar{\mathbf{R}} = \frac{1}{T}\sum_{i=1}^{T}\mathbf{R}_i,$$

as

$$\begin{array}{ll} \min_{\mathbf{x},q,\mathbf{z}} & \mathbf{x}^t\bar{\mathbf{R}} - q + \frac{1}{\alpha T}\sum_{i=1}^T z_i \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & z_i \geqslant q - \mathbf{x}^t\mathbf{R}_i,\ i = 1, \dots, T, \\ & z_i \geqslant 0,\ i = 1, \dots, T. \end{array} \tag{3.4}$$

which has $N+1+T$ variables and $M+2T$ constraints, with $N+1$ dense columns. Note that Uryasev and Rockafellar (1999) have independently derived a formulation similar to the one described above, in the context of CVaR portfolio optimization - see Section 2.

We now notice an interesting geometric fact. Let $(\mathbf{x}_\alpha, q_\alpha)$ be a solution to (3.4), and let $\mathbf{P}_\alpha$ be the convex-hull of all solutions to (3.4). Let $\mathcal{T} = \{1, \dots, T\}$ and $\mathcal{H}$ denote the set of $K = N - M$ element subsets of $\mathcal{T}$. Elements h of $\mathcal{H}$ have relative complement $\bar{h} = \mathcal{T} - h$. Let $\mathbf{R}(h)$ denote the $(K \times N)$ matrix with rows $\mathbf{R}_i$, $i \in h$.

Proposition 8 *Every optimal basic feasible solution to (3.4) has the form*

$$(\mathbf{x}_\alpha^t, q_\alpha)^t = \begin{bmatrix} \mathbf{A} & \mathbf{0}_M \\ \mathbf{R}(h) & -\mathbf{e}_{N-M} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{b} \\ \mathbf{0}_{N-M} \end{bmatrix},$$

for some $h \in \mathcal{H}$. Also, $\mathbf{P}_\alpha$ is the convex-hull of all such solutions.

Proof. Follows from the definition of basic feasible solution and the formulation (3.4). See Bertsimas and Tsitsiklis (1997) for a discussion of basic feasible solutions in linear programming. ■

The preceding lemma says that every optimal basic feasible solution $\mathbf{x}_\alpha$ (and its associated α-quantile q_α) determines a hyperplane $P = \{\mathbf{R} \in \mathbb{R}^N \mid \mathbf{x}_\alpha^t \mathbf{R} = q_\alpha\}$ that passes through at least the $N - M$ points in the sample $\{\mathbf{R}_1, \dots, \mathbf{R}_T\}$.

3.1.1 Alternative Derivation of the α-Shortfall Optimization Linear Programming Formulation

It is interesting to note that we can derive the formulation (3.4) by starting with the alternative definition of shortfall presented in chapter 2, namely

$$s_\alpha(\mathbf{x}) = \mathbf{x}^t \boldsymbol{\mu} - E[\mathbf{x}^t \mathbf{R} \mid \mathbf{x}^t \mathbf{R} \leqslant q_\alpha(\mathbf{x})],$$

where

$$q_\alpha(\mathbf{x}) = \inf\{z \mid P(\mathbf{x}^t \mathbf{R} \leqslant z) \geqslant \alpha\}$$

is the α-quantile of portfolio $\mathbf{x}$. The sample α-shortfall of portfolio $\mathbf{x}$, can be defined in direct analogy to its population counterpart, namely

$$\hat{s}_\alpha(\mathbf{x}) = \mathbf{x}^t \bar{\mathbf{R}} - \frac{1}{\alpha T} \sum_{i=1}^{T} (\mathbf{x}^t \mathbf{R}_i) 1_{\{\mathbf{x}^t \mathbf{R}_i \leqslant \hat{q}_\alpha(\mathbf{x})\}}, \tag{3.5}$$

where $\bar{\mathbf{R}} = \left(\sum_{i=1}^{T} \mathbf{R}_i\right)/T$ is the sample mean, and where $\hat{q}_\alpha(\mathbf{x})$ is the sample α-quantile of portfolio $\mathbf{x}$ defined as

$$\hat{q}_\alpha(\mathbf{x}) = \inf\{z \in \mathbb{R} \mid \frac{1}{T} \sum_{i=1}^{T} 1_{\{\mathbf{x}^t \mathbf{R}_i \leq z\}} \geqslant \alpha\}.$$

Using (3.5), we can formulate the shortfall optimization problem as

$$\begin{aligned} \text{minimize} \quad & \hat{s}_\alpha(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.6}$$

A naive formulation the sample α-shortfall portfolio optimization problem follows from the observation that (3.6) can be rewritten as the following linear program (assuming that $K := \alpha T$ is integer):

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^t\bar{\mathbf{R}} - z \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \tfrac{1}{K}\textstyle\sum_{i\in S}(\mathbf{x}^t\mathbf{R}_i) \geqslant z, \end{aligned} \tag{3.7}$$

where S ranges over all K-element subsets of $\{1, \cdots, T\}$. Of course, the number of constraints is exponential in the number of observations T, so unless N and T are small, the problem as formulated in (3.7) cannot be directly input into a computer (however, see the Appendix for a description of an algorithm which solves (3.7) by sequentially generating the constraints of the problem, and which terminates in polynomial time).

The next theorem shows that formulations (3.7) (which does not involve q) and (3.4) are equivalent.

Theorem 9 *Assume that αT is integer. Then the problems (3.7) and (3.4) have the same optimal solution.*

Proof. Let $K := \alpha T$. Given a vector $\mathbf{v} \in \mathbb{R}^T$, observe that the optimal solution of the linear program

$$\begin{aligned} \text{minimize}_{\mathbf{z}} \quad & \textstyle\sum_{i=1}^T v_i u_i \\ \text{subject to} \quad & \textstyle\sum_{i=1}^T u_i = K, \\ & 0 \leqslant u_i \leqslant 1, \; i = 1, \ldots, T \end{aligned} \tag{3.8}$$

is equal to the sum of the K smallest components of the vector $\mathbf{v}$, i.e. it is equal to $\sum_{i=1}^K v_{(i)}$, where $v_{(1)} \leqslant \cdots \leqslant v_{(T)}$ are the components of $\mathbf{v}$ ordered from smallest to largest. By strong duality, the optimal solution of problem (3.8) is the same as the optimal solution of the dual problem

$$\begin{aligned} \text{maximize}_{q,\mathbf{y}} \quad & Kq + \textstyle\sum_{i=1}^{T} y_i \\ \text{subject to} \quad & t + y_i \geqslant v_i,\ i = 1, \dots, T \\ & y_i \leqslant 0,\ i = 1, \dots, T. \end{aligned} \tag{3.9}$$

Now notice that Problem (3.7) can be written as

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \mathbf{x}^t\bar{\mathbf{R}} - \tfrac{1}{K}\min_{\mathbf{u}} \textstyle\sum_{i=1}^{T} (\mathbf{x}^t\mathbf{R}_i)u_i \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \textstyle\sum_{i=1}^{T} u_i = K, \\ & 0 \leqslant u_i \leqslant 1,\ i = 1, \dots, T. \end{aligned}$$

Using (3.9), we can rewrite the last problem as

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \mathbf{x}^t\bar{\mathbf{R}} - \tfrac{1}{K}\max_{q,\mathbf{y}}(Kq + \textstyle\sum_{i=1}^{T} y_i) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & q + y_i \geqslant \mathbf{x}^t\mathbf{R}_i,\ i = 1, \dots, T \\ & y_i \leqslant 0,\ i = 1, \dots, T. \end{aligned}$$

Using the fact that $\max(\theta) = -\min(-\theta)$, we obtain,

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \mathbf{x}^t\bar{\mathbf{R}} - q - \tfrac{1}{K}\textstyle\sum_{i=1}^{T} y_i \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & q + y_i \geqslant \mathbf{x}^t\mathbf{R}_i,\ i = 1, \dots, T \\ & y_i \leqslant 0,\ i = 1, \dots, T. \end{aligned}$$

and the conclusion follows by letting $z_i = -y_i$. ■

3.2 Alternative Formulations of the Sample Variance, CVar, LPM_1, and LPM_2 Portfolio Optimization Problems

Before we present alternative formulations of other portfolio optimization problems, let us notice that for any $\alpha \in (0,1)$, the sample α-shortfall portfolio optimization

$$\begin{array}{ll} \text{minimize} & \frac{1}{T}\sum_i^T \rho_\alpha(\mathbf{x}^t\mathbf{R}_i - q) \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array}$$

problem can be also be written as - using definition (3.2) -

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, q\in\mathbb{R}, t^+\in\mathbb{R}^T, t^+\in\mathbb{R}^T} & \frac{1}{T}\sum_i^T t_i^+ + \frac{1}{T}\sum_i^T(\frac{1}{\alpha}-1)t_i^- \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x}^t\mathbf{R}_i - q = t_i^+ - t_i^-,\ i = 1,\dots,T, \\ & \mathbf{t}^+, \mathbf{t}^- \geqslant 0, \end{array} \tag{3.10}$$

In particular, the sample 50%-shortfall problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{T}\sum_i^T |\mathbf{x}^t\mathbf{R}_i - q| \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array},$$

can be rewritten as

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, q\in\mathbb{R}, t^+\in\mathbb{R}^T, t^+\in\mathbb{R}^T} & \frac{1}{T}\sum_i^T t_i^+ + \frac{1}{T}\sum_i^T t_i^- \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x}^t\mathbf{R}_i - q = t_i^+ - t_i^-,\ i = 1,\dots,T, \\ & \mathbf{t}^+, \mathbf{t}^- \geqslant 0. \end{array}$$

3.2.1 Alternative Formulation of the Sample Variance Portfolio Optimization Problem

From the definitions of the sample covariance matrix and the sample mean, the sample variance problem (3.3) can be rewritten as

$$\begin{array}{ll} \text{minimize} & \frac{1}{T}\sum_i^T (\mathbf{x}^t\mathbf{R}_i - q)^2 \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array}$$

It is then easy to see that the problem can be rewritten without a covariance matrix, as follows:

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, q\in\mathbb{R}, t^+\in\mathbb{R}^T, t^+\in\mathbb{R}^T} & \frac{1}{T}\sum_i^T (t_i^+)^2 + \frac{1}{T}\sum_i^T (t_i^-)^2 \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x}^t\mathbf{R}_i - q = t_i^+ - t_i^-,\ i = 1, \ldots, T, \\ & \mathbf{t}^+, \mathbf{t}^- \geqslant 0, \end{array} \tag{3.11}$$

which is a QP with $N + 1 + 2T$ variables, and $M + 3T$ constraints, with $N + 1$ dense columns.

3.2.2 Alternative Formulation of the Sample CVaR Portfolio Optimization Problem

The sample CVaR (with parameter $\alpha \in (0, 1)$) optimization problem is

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N} & -\frac{1}{\alpha T}\sum_i^T (\mathbf{x}^t\mathbf{R}_i) 1_{\{\mathbf{x}^t\mathbf{R}_i \leq \hat{q}_\alpha(\mathbf{x})\}} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array}$$

The following formulation of the sample CVaR optimization problem is due to Uryasev and Rockafellar (1999):

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, q\in\mathbb{R}} & -q - \frac{1}{\alpha T}\sum_i^T (\mathbf{x}^t\mathbf{R}_i - q) 1_{\{\mathbf{x}^t\mathbf{R}_i \leq q\}} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{array}$$

which can be reformulated as

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, q\in\mathbb{R}, t^+\in\mathbb{R}^T, t^+\in\mathbb{R}^T} & -q + \frac{1}{\alpha T}\sum_i^T t_i^- \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x}^t\mathbf{R}_i - q = t_i^+ - t_i^-,\ i = 1, \dots, T, \\ & \mathbf{t}^+, \mathbf{t}^- \geqslant 0, \end{array} \tag{3.12}$$

which is an LP with $N + 1 + 2T$ variables, and $M + 3T$ constraints, with $N + 1$ dense columns.

3.2.3 Alternative Formulation of the Sample LPM$_1$ Portfolio Optimization Problem

The sample LPM$_1$ (with parameter c) optimization problem is

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N} & \frac{1}{T}\sum_i^T (c - \mathbf{x}^t\mathbf{R}_i) 1_{\{\mathbf{x}^t\mathbf{R}_i \leqslant c\}} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array}$$

which can be reformulated as

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, t^+\in\mathbb{R}^T, t^+\in\mathbb{R}^T} & \frac{1}{T}\sum_i^T t_i^- \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x}^t\mathbf{R}_i - c = t_i^+ - t_i^-,\ i = 1, \dots, T, \\ & \mathbf{t}^+, \mathbf{t}^- \geqslant 0, \end{array} \tag{3.13}$$

which is an LP with $N + 1 + 2T$ variables, and $M + 3T$ constraints, with $N + 1$ dense columns.

3.2.4 Alternative Formulation of the Sample LPM$_2$ Portfolio Optimization Problem

The sample LPM$_2$ (with parameter c) optimization problem is

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N} & \frac{1}{T}\sum_i^T (c - \mathbf{x}^t\mathbf{R}_i)^2 1_{\{\mathbf{x}^t\mathbf{R}_i \leqslant c\}} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array}$$

which can be reformulated as

$$
\begin{array}{ll}
\text{minimize}_{\mathbf{x}\in\mathbb{R}^N, t^+\in\mathbb{R}^T, t^+\in\mathbb{R}^T} & \frac{1}{T}\sum_i^T (t_i^-)^2 \\
\text{subject to} & \mathbf{Ax} = \mathbf{b} \\
 & \mathbf{x}^t\mathbf{R}_i - c = t_i^+ - t_i^-, \; i = 1, \ldots, T, \\
 & \mathbf{t}^+, \mathbf{t}^- \geqslant 0,
\end{array} \tag{3.14}
$$

which is a QP with $N + 1 + 2T$ variables, and $M + 3T$ constraints, with $N + 1$ dense columns.

3.3 Non-Uniqueness and Unboundedness of the Sample Solution Set and Regularization

In this section, we propose a solution to the problem of non-uniqueness and unboundedness which characterizes the sample shortfall and variance portfolio optimization problems when $T \leqslant (N - M)$ for example. Note that $T \leqslant (N - M)$ is realistic, and corresponds to any sample portfolio optimization problem where the number of assets is large, but the number of observations is limited. For example, consider a sample portfolio optimization problem involving 2000 stocks, with monthly returns going back only 10 years - see Ledoit (1995) for an example of sample variance portfolio optimization with 10 years of monthly return observations on a relatively large set of assets, consisting of all stocks in the CRSP universe.

For the α-shortfall problem, $T \leqslant (N - M)$ implies that there exists no basic feasible solution. For the variance problem, $T \leqslant (N - M)$ implies the covariance matrix has rank less than N. This will of course create computational difficulties for any LP and QP solver. More disturbing, however, is the fact that for the sample portfolio optimization problems we are considering, the optimal objective function will be 0 (the optimal solution has 0 risk), and that the optimal feasible set will be unbounded.

Proposition 10 *Assume that $T \leqslant (N - M)$. Then the sample α-shortfall and the sample variance portfolio optimization problems have an unbounded solution set for which the sample risk is* 0.

Proof. Assume without loss of generality that the vectors $\mathbf{R}_1, \ldots, \mathbf{R}_T$ are independent, and independent of the rows of **A**. Then the matrix

$$M = \begin{bmatrix} \mathbf{A} \\ \mathbf{R}_1^t \\ \vdots \\ \mathbf{R}_T^t \end{bmatrix},$$

has $T + M$ linearly independent columns. So the system

$$M\mathbf{x} = \begin{bmatrix} \mathbf{b} \\ q\mathbf{e} \end{bmatrix} \tag{3.15}$$

has at least one solution, for arbitrary q, as does the system

$$M\mathbf{x} = \mathbf{0}. \tag{3.16}$$

Therefore, any portfolio optimization of the form

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, q\in\mathbb{R}} & \frac{1}{T}\sum_i^T f(\mathbf{x}^t\mathbf{R}_i - q) \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{array}$$

- where $f(.) \geqslant 0$ and $f(0) = 0$ - has at least one feasible solution with 0 sample risk - since there exists $\mathbf{x}$ which satisfies (3.15) - and the solution set is unbounded - since there exists at least one feasible direction satisfying (3.16). ■

The problem of the non-uniqueness and unboundedness of the optimal solution set can be solved by regularization methods, as we describe now[1]. Notice that we can enforce uniqueness in our sample portfolio optimization problems by adding a strictly convex penalty to the objective function. This is exactly what regularization does. Specifically, we will consider two types of penalty.

- L_1-norm penalty $||\mathbf{x}||_1^1 = \sum_{i=1}^N |x_i|$ where $|.|$ stands for the absolute value, and x_i, $i = 1, \ldots, N$ are the elements of $\mathbf{x}$;
- L_2-norm penalty $||\mathbf{x}||_2^2 = \sum_{i=1}^N (x_i)^2$.

[1]Note that regularization is also motivated, but from a statistical perspective, in Chapter V.

Now, consider the sample α-shortfall portfolio optimization problem. The L_1-norm regularized version of the problem is

$$\begin{array}{ll} \text{minimize}_{\mathbf{x},q} & \hat{s}_\alpha(\mathbf{x}) + \lambda||\mathbf{x} - \mathbf{x}_m||_1^1 \\ \text{subject to} & \mathbf{Ax} = \mathbf{b}. \end{array} \tag{3.17}$$

where λ can be an arbitrary positive real, and $\mathbf{x}_m \in \mathbb{R}^N$ [2]. This can be written as an LP using the usual trick with absolute values:

$$\begin{array}{ll} \text{minimize}_{\mathbf{x},q} & \hat{s}_\alpha(\mathbf{x}) + \lambda \sum_{i=1}^N z_i \\ \text{subject to} & \mathbf{Ax} = \mathbf{b} \\ & z_i \geqslant x_i - x_{i,m},\ i = 1, \ldots, N \\ & z_i \geqslant -x_i + x_{i,m},\ i = 1, \ldots, N. \end{array}$$

The L_2-norm regularized version of the problem is

$$\begin{array}{ll} \text{minimize}_{\mathbf{x},q} & \hat{s}_\alpha(\mathbf{x}) + \lambda||\mathbf{x} - \mathbf{x}_m||_2^2 \\ \text{subject to} & \mathbf{Ax} = \mathbf{b}. \end{array} \tag{3.18}$$

which can be written as a QP. In fact, the sample α-shortfall and variance portfolio optimization problems with L_1-norm or L_2-norm regularization can each be solved as either an LP or a QP, using the usual algorithms for either problem formulation. In addition, with the added penalty, the objective functions will be strictly convex, guaranteeing a unique solution.

3.4 Computational Experiments

We implemented our portfolio optimization formulations in CPLEX, which is a solver for LPs and QPs. All data manipulations described below (except for the optimizations) were done in Matlab, and CPLEX was called from Matlab using executable "mex" files (for the optimizations). These files allow us to pass data to the CPLEX solver, which then returns the optimal solution after optimization. We ran our experiments on a Dell Precision 410 with a 400 Mhz Dual Pentium II, and with 256MB of RAM.

In what follows we report the running times of portfolio optimization problems and formulations,

[2] Again, see Chapter 5 for motivation and suggestions concerning the choice of λ and $\mathbf{x}_m$

for datasets of different sizes. We also consider an application of the α-shortfall portfolio optimization problem which leverages the fact that the problem can be formulated as an LP. Specifically, we consider the index tracking problem, and show that this hard QP with integrality constraints can be solved by considering an easier LP formulation with cardinality constraints.

3.4.1 Comparing the Running Times of Sample Portfolio Optimization Problems and Formulations

We will compare the running times of the following problems and formulations:

- **shortfall**: formulation (3.4) for the sample α-shortfall portfolio optimization problem, implemented with $\alpha = 50\%$;
- **variance-A:** formulation (3.3) for the sample variance portfolio problem implemented with covariance matrix;
- **variance-B**: formulation (3.11) for the sample variance portfolio optimization problem implemented without covariance matrix calculation.

We solve these problems for datasets of size $(T \times N)$, where $N = 100, 200, 500, 1000, 2000$, and $T = 120, 260, 520$. Notice that the values of T which we choose represent, respectively, the number of months in ten years, the number of trading days in one year, and the number of weeks in ten years. For each combination of N and T, the dataset consists in T observations from the distribution $N(\mathbf{0},\Sigma)$, where Σ is an $(N \times N)$ covariance matrix with ones on the diagonal, and 0.5 off-diagonal elements.

The following tables reports the running times to optimality for each problem. For $T < N$, we do not report the results of the "unregularized" optimization problems, for the reasons that we described in Section 4 above. Problems are regularized using $\mathbf{x}_m = \mathbf{0}$ and $\lambda = 1$, where $\mathbf{x}_m$ and λ are defined as in (3.17) and (3.18) . L1-regularization is only implemented for shortfall. For variance-A and variance-B, and for L2-regularized shortfall, the CPLEX barrier optimizer (based on interior point methods) was used. For shortfall and shortfall-L1, the dual simplex solver was used, as it gave shorter running times.

The following insights emerge from this experiment.

1. for values of N and T less than 500, the variance-A is solvable in CPLEX about an order of magnitude faster than the other two problems, with or without regularization.

		N=100	N=200	N=500
variance-A	T=120	0.1		
	T=260	0.1	0.5	
	T=520	0.1	0.5	11.1
shortfall	T=120	0.4		
	T=260	2.4	4.8	
	T=520	9.6	24.6	43.5
variance-B	T=120	0.3		
	T=260	1.5	3.3	
	T=520	7.2	14.6	36.1

Table 3.1: Sample Portfolio Optimization Problems (unregularized)

		N=100	N=200	N=500	N=1000	N=2000
variance-A	T=120	0.1	0.5	9.6	92.3	606.2
	T=260	0.1	0.5	9.5	78.7	606.3
	T=520	0.1	0.5	9.4	92.0	717.0
shortfall	T=120	0.6	1.0	2.6	5.4	10.1
	T=260	2,7	5.9	15.0	26.0	46.7
	T=520	17.6	27.5	61.4	120.0	176.7
variance-B	T=120	0.5	1.0	2.4	4.8	10.1
	T=260	2.4	5.5	12.6	21.0	37.0
	T=520	13.2	25.3	51.4	80.3	138.7

Table 3.2: Sample Portfolio Optimization Problems (L2-regularized)

		N=100	N=200	N=500	N=1000	N=2000
shortfall	T=120	0.6	1.8	10.5	35.0	129.5
	T=260	2.2	6.5	30.4	102.7	354.9
	T=520	8.6	22.5	101.1	310.5	1036.8

Table 3.3: Sample Portfolio Optimization Problems (L1-regularized)

2. for $N = 2000$, and after L2-regularization, the situation is inverted, with shortfall and variance-B vastly outperforming variance (by a factor of 60 for $N = 2000, T = 120$, for example). To see why this is happening, notice that in the variance-A problem, the factor determining running time is the size of the quadratic matrix, which in this case is (2000×2000), and dense. In the other two problems, the only dense part is the constraint matrix, which is only $(T \times 2000)$. This constraint matrix is smaller than the corresponding dense quadratic matrix, and the problems run faster.

3. L1-regularization is at least five-times more expensive than L2-regularization, for large N, such as $N = 1000$ or 2000.

We conclude this experiment by noting that we did not notice a strong dependence of running time on α in the case of shortfall. In regularized problems, the value of the λ parameter may affect running times, but we have not examined this issue. We also note that while we do not report the average running time over Monte Carlo repetitions of our experiment, we did run a small Monte Carlo (10 repetitions), and the insights described above do not change.

3.4.2 Solving the Index-Tracking Problem with Cardinality Constraints Using Sample Shortfall Portfolio Optimization

In this experiment, we consider the index-tracking problem, which we define as follows. Given N assets, whose return vector $\mathbf{R}$ is a random variable with mean $\boldsymbol{\mu}$ and covariance matrix Σ, and given a portfolio $\mathbf{x}_b$ - b for "benchmark" - of these assets, find a portfolio $\mathbf{x} \geqslant \mathbf{0}$ of $K < N$ assets which has minimum tracking error, where tracking error is defined as

$$\sqrt{E[(\mathbf{x}^t\mathbf{R} - \mathbf{x}_b^t\mathbf{R})^2]} = \sqrt{(\mathbf{x} - \mathbf{x}_b)^t\Sigma^*(\mathbf{x} - \mathbf{x}_b)}, \tag{3.19}$$

where $\Sigma^* := \boldsymbol{\mu}\boldsymbol{\mu}^t + \Sigma$. We also add the following twist to the problem: the nonzero elements in $\mathbf{x}$ must lie in the interval $[a, b]$.

The index-tracking problem can be formulated as a QP with cardinality constraints. To this end, we introduce binary variables

$$y_i = \begin{cases} 1 & \text{if asset } i \text{ is included in } \mathbf{x}, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the problem can be written as

$$
\begin{aligned}
\text{minimize}_{\mathbf{x},q} \quad & (\mathbf{x}-\mathbf{x}_b)^t\Sigma^*(\mathbf{x}-\mathbf{x}_b) \\
\text{subject to} \quad & \mathbf{x}^t\mathbf{e} = 1 \\
& \mathbf{x}^t\mathbf{y} = K \\
& ay_i \leqslant x_i \leqslant by_i,\ i=1,\dots,N, \\
& y_i \in \{0,1\},\ i=1,\dots,N, \\
& \mathbf{x} \geqslant \mathbf{0}.
\end{aligned}
\tag{3.20}
$$

This QP with cardinality constraints is a hard problem for which there exists no commercially available solvers.

To solve (3.20), we use the fact that when returns are normal, the shortfall of a portfolio is proportional to its standard deviation. Technically speaking, the exact relationship - see Chapter 1 - is

$$
s_\alpha(\mathbf{x}-\mathbf{x}_b) = \frac{\phi(z_\alpha)}{\alpha}\sqrt{(\mathbf{x}-\mathbf{x}_b)^t\Sigma^*(\mathbf{x}-\mathbf{x}_b)}
$$

where $\phi(.)$ is the density of the standard normal, and z_α is the α-quantile of the standard normal, and where $s_\alpha(\mathbf{x}-\mathbf{x}_b)$ is the α-shortfall of portfolio $\mathbf{x}-\mathbf{x}_b$, assuming returns are normal with covariance matrix Σ^*. This shows that Problem (3.20) is equivalent to

$$
\begin{aligned}
\text{minimize}_{\mathbf{x},q} \quad & s_\alpha(\mathbf{x}-\mathbf{x}_b) \\
\text{subject to} \quad & \mathbf{x}^t\mathbf{e} = 1 \\
& \mathbf{x}^t\mathbf{y} = K \\
& ay_i \leqslant x_i \leqslant by_i,\ i=1,\dots,N, \\
& y_i \in \{0,1\},\ i=1,\dots,N, \\
& \mathbf{x} \geqslant \mathbf{0}.
\end{aligned}
$$

Now this last problem can be approximately solved as an LP, by generating data $\mathbf{R}_i^*$, $i=1,\dots,T$ from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ^*, and solving the related sample shortfall portfolio optimization problem, i.e., by solving

$$
\begin{aligned}
\text{minimize}_{\mathbf{x},q} \quad & \mathbf{x}^t\bar{\mathbf{R}} - q + \tfrac{1}{T\alpha}\textstyle\sum_i^T z_i \\
\text{subject to} \quad & \mathbf{x}^t\mathbf{e} = 1 \\
& \mathbf{x}^t\mathbf{y} = K \\
& a y_i \leqslant x_i \leqslant b y_i, \; i = 1, \dots, N, \\
& z_i \geqslant q - (\mathbf{x} - \mathbf{x}_b)^t \mathbf{R}_i^*, \; i = 1, \dots, T, \\
& y_i \in \{0, 1\}, \; i = 1, \dots, N, \\
& \mathbf{x}, \mathbf{z} \geqslant \mathbf{0}.
\end{aligned} \tag{3.21}
$$

where we have used the formulation of Problem (3.4) above. Problem (3.21) is a mixed integer linear program, for which there is a large literature as well as several commercially available solvers.

To illustrate our approach to the index-tracking problem, we run the following experiments for $N = 96$, using a covariance matrix Σ and a mean vector $\boldsymbol{\mu}$ which were calculated from SP100 data obtained in the CRSP database[3]. We generated random samples of size $T = 100, 200, 500, 1000$ from a distribution with mean $\mathbf{0}$ and covariance Σ^*, with Σ^* as defined after (3.19). We let $\mathbf{x}_b = \mathbf{e}/N$, an equally weighted index or benchmark. We let $a = 0.3\%$, and $b = 10\%$. We finally let $\alpha = 50\%$. We solve Problem (3.21) using the mixed integer solver in CPLEX, and we report in Table (3.4) the tracking error, and in Table (3.5) the running time, both as a function of K and T. Tracking error is calculated using (3.19). We run the mixed integer programming algorithm in CPLEX until five feasible solutions have been found. The reason for stopping the algorithm after five solutions is that running times to exact optimality can be excessive for problems of the size which we are considering. Moreover, the best solution found among the first five solutions is typically either the optimal or very close to the problem.

The following insights emerge from this experiment:

1. the proposed approach successfully solves the index tracking problem, for a given $\boldsymbol{\mu}$ and Σ, within reasonable running times.

[3] $\boldsymbol{\mu}$ was calculated as the sample mean, multiplied by 21, of 96 stocks, for the period January 02, 1996, to December 31, 1999. Σ was calculated as the sample covariance matrix, multiplied by 21, of those same stocks over the same period. The 96 stocks were selected using the following criteria: they were in the SP100 on December 31, 1999, and had daily return data for the entire four year period under consideration. The data come from the CRSP database, and were obtained using the Wharton Research Database Service (WRDS).

K=	96	90	80	70	60	50	30
T=100	0.0%	0.46%	0.54%	0.66%	0.81%	0.90%	1.1%
T=200	0.0%	0.38%	0.48%	0.63%	0.72%	0.93%	1.1%
T=500	0.0%	0.29%	0.46%	0.59%	0.73%	0.79%	1.0%
T=1000	0.0%	0.27%	0.37%	0.54%	0.60%	0.78%	0.90%

Table 3.4: Tracking Error (per month) as a function of K and T.

K=	96	90	80	70	60	50	30
T=100	0.7	1.7	3.2	5.0	7.4	9.3	12.5
T=200	0.8	7.0	11.7	18.0	21.8	27.1	32.4
T=500	0.9	32.8	59.5	81.6	102.7	121.9	135.8
T=1000	1.2	128.2	206.6	313.3	372.9	436.4	519.3

Table 3.5: Running Time (in seconds) as a function of K and T.

2. as expected, the tracking error increases as K decreases, since it becomes increasingly more difficult to track the index with a smaller number of assets.

3. the running times are monotonically increasing as K decreases and T increases.

3.5 Conclusion

In this chapter we developed an LP formulation of the sample α-shortfall portfolio optimization problem, and alternative formulations of the variance, CVaR, LPM_1 and LPM_2 portfolio optimization problems. The alternative formulation of the variance portfolio optimization problem circumvents the need to calculate the covariance matrix. We noted the need to regularize our problems when $T < N - M$. When N is large and T is less than $N - M$, we showed via simulation that our formulations may have significantly lower running times than the classical formulation of the variance portfolio optimization problem, which involves the covariance matrix. This improvement in running time occurs because when $T < N - M$, our formulations do not require the manipulation of a dense $(N \times N)$ covariance matrix. Finally, we showed that we could solve the index-tracking problem, a QP with cardinality constraints, by solving a related sample shortfall optimization problem, due to the linearity of the latter problem.

Chapter 4

Central Limit Theorem for the α-Shortfall Portfolio Estimator

Let $\mathbf{R}$ be a random return vector in $\mathbb{R}^N$ with mean $\boldsymbol{\mu}$, positive definite covariance matrix Σ, and a continuous density. Let $\mathbf{A}$ be an $(M \times N)$ matrix with linearly independent rows and let $\mathbf{b}$ be an M-dimensional vector. Assume $\mathbf{b} \neq \mathbf{0}$ so the next problem is non-trivial. The optimal shortfall portfolio, for $\alpha \in (0,1)$, over the set $\mathbf{Ax} = \mathbf{b}$ is the vector $\mathbf{x}_\alpha$ which along with its quantile $q_\alpha(\mathbf{x}_\alpha) := \inf\{z | \Pr(\mathbf{x}^t\mathbf{R} \leqslant z) \geqslant \alpha\}$ solves the problem

$$(\mathbf{x}_\alpha, q_\alpha(\mathbf{x}_\alpha)) = \arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q \in \mathbb{R}}} E[\rho_\alpha(\mathbf{x}^t\mathbf{R} - q)],$$

where

$$\rho_\alpha(z) := \begin{cases} z & \text{if } z \geqslant 0 \\ (1 - \frac{1}{\alpha})z & \text{if } z < 0. \end{cases}$$

Now suppose that we do not know the distribution of $\mathbf{R}$, but that we are given independent and identically distributed realizations $\mathbf{R}_1, \dots, \mathbf{R}_T$ of the random return vector $\mathbf{R}$. Given these realizations, we want to estimate $\mathbf{x}_\alpha$. We define the shortfall portfolio estimator is defined as the vector $\hat{\mathbf{x}}_\alpha$ which, along with its sample α-quantile $\hat{q}_\alpha$, solves

$$(\hat{\mathbf{x}}_\alpha,\hat{q}_\alpha) = \arg\min_{\substack{\mathbf{A}\mathbf{x}=\mathbf{b}\\ q\in\mathbb{R}}} \frac{1}{T}\sum_{i=1}^{T}[\rho_\alpha(\mathbf{x}^t\mathbf{R}_i - q)], \tag{4.1}$$

This chapter establishes the asymptotic normality of $\hat{\mathbf{x}}_\alpha$, proving that $\hat{\mathbf{x}}_\alpha$ converges to $\mathbf{x}_\alpha$ at the usual $\sqrt{T}$ rate.

For completeness, we also establish the asymptotic normality of the the variance portfolio estimator $\hat{\mathbf{x}}_V$ which solves

$$\hat{\mathbf{x}}_V = \arg\min_{\mathbf{A}\mathbf{x}=\mathbf{b}} \mathbf{x}^t\hat{\Sigma}\mathbf{x},$$

where $\hat{\Sigma} = \hat{\Sigma}= \sum_{i=1}^{T}(\mathbf{R}_i - \bar{\mathbf{R}})(\mathbf{R}_i - \bar{\mathbf{R}})^t/T$ is the sample covariance matrix, and $\bar{\mathbf{R}} = \sum_{i=1}^{T}\mathbf{R}_i/T$ is the sample mean, showing that $\hat{\mathbf{x}}_V$ converges to

$$\mathbf{x}_V = \arg\min_{\mathbf{A}\mathbf{x}=\mathbf{b}} \mathbf{x}^t\Sigma\mathbf{x},$$

at the $\sqrt{T}$ rate. Notice that $\hat{\mathbf{x}}_V$, along with its sample mean $\hat{q}_V$, solves

$$(\hat{\mathbf{x}}_V,\hat{q}_V) = \arg\min_{\substack{\mathbf{A}\mathbf{x}=\mathbf{b}\\ q\in\mathbb{R}}} \frac{1}{T}\sum_{i=1}^{T}(\mathbf{x}^t\mathbf{R}_i - q)^2. \tag{4.2}$$

This last fact will be used in the proofs to follow.

This chapter is organized as follows. In Section 1, we start by proving the asymptotic normality of the variance portfolio estimator. We start with the variance estimator because proving its asymptotic normality is easier, and will serve as a blue-print for the proof of the asymptotic normality of the shortfall estimator. In Section 2, we derive the asymptotic normality of the shortfall portfolio estimator, using some results from empirical process theory. In Section 3 we present the results of a computational experiment which illustrates that shortfall portfolio estimation with small values of α may be dangerous. Finally in Section 4 we offer some concluding remarks.

4.1 Asymptotic Normality of the Variance Portfolio Estimator

Establishing asymptotic normality is easier for $\hat{\mathbf{x}}_V$ than for $\hat{\mathbf{x}}_\alpha$, so we start with the variance portfolio estimator. The ideas that will be developed here are similar to those used in the shortfall proof, to follow in the next section. But there, the details are more involved because the function $\rho_\alpha(.)$ is not differentiable at 0. An alternative, somewhat more direct, proof for $\hat{\mathbf{x}}_V$ is provided in the appendix.

4.1.1 Assumption (A)

We start by making an assumption on the distribution of $\mathbf{R}$.

(A): $\mathbf{R}$, with mean $\boldsymbol{\mu}$ and covariance matrix Σ, has a continuous density, and $\mathbf{R}$ has finite fourth moments.

The assumption of finite fourth moments is necessary for the existence of the asymptotic covariance matrix of $\hat{\mathbf{x}}_V$, as we will see below.

4.1.2 Notation

To simplify notation, we introduce the following. Let $\mathbf{W} = (\mathbf{R}^t, -1)^t$, and let P be the distribution of $\mathbf{W}$, $E(.)$ be the expectation with respect to $\mathbf{W}$. Define $\mathbf{y}_V = (\mathbf{x}_V, q_V)^t, \hat{\mathbf{y}}_V = (\hat{\mathbf{x}}_V^t, \hat{q}_V)$, and

$$\begin{aligned}\mathcal{Y} &= \left\{(\mathbf{x}^t, q)^t \mid \mathbf{A}\mathbf{x} = \mathbf{b},\ q \in \mathbb{R}\right\} \\ &= \left\{\mathbf{y} \in \mathbb{R}^{N+1} \mid \mathbf{A}_0\mathbf{y} = \mathbf{b}\right\},\end{aligned}$$

with $\mathbf{A}_0 = [\mathbf{A}\ \mathbf{0}]$, and $\mathbf{0}$ is an N-dimensional vector of zeros. Note that

$$\begin{aligned}\mathbf{y}_V &= \arg\min_{\mathbf{y}\in\mathcal{Y}} E[(\mathbf{y}^t\mathbf{W})^2] \\ &= \arg\min_{\mathbf{y}\in\mathcal{Y}} \mathbf{y}^t E[\mathbf{W}\mathbf{W}^t]\mathbf{y} \\ &= \arg\min_{\mathbf{y}\in\mathcal{Y}} \mathbf{y}^t \Psi \mathbf{y} \qquad (4.3)\end{aligned}$$

where $\Psi = E[\mathbf{W}\mathbf{W}^t]$, and

$$\begin{aligned}\hat{\mathbf{y}}_V &= \arg\min_{\mathbf{y}\in\mathcal{Y}} \left[\frac{1}{T}\sum_{i=1}^{T}(\mathbf{y}^t\mathbf{w}_i)^2\right] \\ &= \arg\min_{\mathbf{y}\in\mathcal{Y}} \mathbf{y}^t \left[\frac{1}{T}\sum_{i=1}^{T}(\mathbf{w}_i\mathbf{w}_i^t)\right]\mathbf{y} \\ &= \arg\min_{\mathbf{y}\in\mathcal{Y}} \mathbf{y}^t\hat{\Psi}\mathbf{y}, \end{aligned} \tag{4.4}$$

where $\hat{\Psi} = \left[\frac{1}{T}\sum_{i=1}^{T}(\mathbf{w}_i\mathbf{w}_i^t)\right]$. Define $\mathcal{Z} = \{\mathbf{z} \in \mathbb{R}^{N+1} \mid \mathbf{A}_0\mathbf{z} = \mathbf{0}\}$, the space of first order feasible variations.

4.1.3 Outline of Proof

We outline below the proof of the asymptotic normality of $\hat{\mathbf{x}}_V$.

Step1: Writing the Lagrangean of the problem

$$\begin{aligned} &\text{minimize} && \mathbf{y}^t\hat{\Psi}\mathbf{y} \\ &\text{subject to} && \mathbf{A}_0\mathbf{y} = \mathbf{0}, \end{aligned}$$

we can get a closed form expression for $\hat{\mathbf{y}}_V$, which depends on the product of a random matrix (converging in probability to a positive definite matrix) multiplied by a random vector (converging in distribution by virtue of the CLT). Then, using Slutzky's Lemma (see below) and some linear algebra, one gets $\sqrt{T}(\hat{\mathbf{y}}_V - \mathbf{y}_V) \rightsquigarrow N(\mathbf{0}, L_V^* Q_V^* L_V^*)$, for some L_V^* and Q_V^* expressible in closed form. This is Theorem 13.

Step 2: Using linear algebra, from the asymptotic normality of $\hat{\mathbf{y}}_V$ one gets the asymptotic normality of $\hat{\mathbf{x}}_V$, $\sqrt{T}(\hat{\mathbf{x}}_V - \mathbf{x}_V) \rightsquigarrow N(\mathbf{0}, L_V Q_V L_V)$, for some L_V and Q_V expressible in closed form, which is Corollary 14.

4.1.4 Uniqueness of the Variance Portfolio Estimator

We will need the following lemma in the proofs to follow.

Lemma 11 $\mathbf{y}_V$ *is uniquely defined, and* $\hat{\mathbf{y}}_V$ *are uniquely defined with probability one.*

Proof. Notice that

$$\Psi = \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} + \begin{pmatrix} \boldsymbol{\mu} \\ -1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}^t & -1 \end{pmatrix},$$

which shows that Ψ is positive definite. Also, with probability one, $\hat{\Sigma}$ is positive definite, by the assumption that $\mathbf{R}$ has a continuous density. Therefore,

$$\hat{\Psi} = \begin{bmatrix} \hat{\Sigma} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} + \begin{pmatrix} \bar{\mathbf{R}} \\ -1 \end{pmatrix} \begin{pmatrix} \bar{\mathbf{R}}^t & -1 \end{pmatrix},$$

where $\bar{\mathbf{R}}$ is the sample mean. Therefore, $\hat{\Psi}$ is positive definite. From the fact that $\mathcal{Y}$ is a convex set, and from the *strict* convexity of the objective functions in (4.3) and (4.4), we conclude that both $\mathbf{y}_V$ and $\hat{\mathbf{y}}_V$ are the uniquely defined. ■

4.1.5 Proof of Asymptotic Normality

We will need the following lemma, whose proof can be found in van der Vaart (1998), Lemma 2.8.

Lemma 12 (Slutsky's Lemma) *Let $\hat{\mathbf{y}}$ and $\mathbf{Y}$ be random vectors in $\mathbb{R}^N$ such that $\hat{\mathbf{y}} \rightsquigarrow \mathbf{Y}$. Let $\hat{M}$ be a random $N \times N$ matrix that converges in probability to a constant invertible matrix M. Then, assuming $\hat{M}$ is invertible, $\hat{M}^{-1}\hat{\mathbf{y}} \rightsquigarrow M^{-1}\mathbf{Y}$.*

The next theorem proves the asymptotic normality of $\hat{\mathbf{y}}_V$.

Theorem 13 *Let $\hat{\mathbf{y}}_V$ be defined as in (4.4). Suppose assumption (A) holds. Then*

$$\sqrt{T}(\hat{\mathbf{y}}_V - \mathbf{y_V}) \rightsquigarrow N(\mathbf{0}, L_V^* Q_V^* L_V^*)$$

where

$$Q_V^* = Cov(\mathbf{W}\mathbf{W}^t \mathbf{y}_V)$$

and where L_V^* is the upper-left $(N+1)$-dimensional corner of M_V^{-1} where

$$M_V = \begin{bmatrix} \Psi & \mathbf{A}_0^t \\ \mathbf{A}_0 & \mathbf{0}_M \end{bmatrix}$$

is nonsingular.

Proof. Notice that, along with a unique vector of Langrange multipliers $\boldsymbol{\lambda}_V \in \mathbb{R}^M$, $\mathbf{y}_V$ uniquely solves the system of linear equations

$$\begin{cases} \Psi \mathbf{y} & +\mathbf{A}_0^t \boldsymbol{\lambda} & = \mathbf{0} \\ \mathbf{A}_0 \mathbf{y} & +\mathbf{0}_M \boldsymbol{\lambda} & = \mathbf{b}. \end{cases} \tag{4.5}$$

Let

$$M_V = \begin{bmatrix} \Psi & \mathbf{A}_0^t \\ \mathbf{A}_0 & \mathbf{0}_M \end{bmatrix}.$$

Then M_V is invertible by uniqueness of $\mathbf{y}_V$ and $\boldsymbol{\lambda}_V$ in (4.5). Notice also that, along with a unique set of Langrange multipliers $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^M$, with probability one $\hat{\mathbf{y}}$ uniquely solves

$$\begin{cases} \hat{\Psi} \mathbf{y} & +\mathbf{A}_0^t \boldsymbol{\lambda} & = \mathbf{0} \\ \mathbf{A}_0 \mathbf{y} & +\mathbf{0}_M \boldsymbol{\lambda} & = \mathbf{b}, \end{cases} \tag{4.6}$$

and let

$$\hat{M}_V = \begin{bmatrix} \hat{\Psi} & \mathbf{A}_0^t \\ \mathbf{A}_0 & \mathbf{0}_M \end{bmatrix}.$$

With probability one, $\hat{M}_V$ is invertible by uniqueness of $\hat{\mathbf{y}}_V$ and $\hat{\boldsymbol{\lambda}}_V$ in (4.6).

Using (4.6), write

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix} = \hat{M}_V \begin{bmatrix} \mathbf{y}_V \\ \boldsymbol{\lambda}_V \end{bmatrix} + \hat{M}_V \begin{bmatrix} \hat{\mathbf{y}}_V - \mathbf{y}_V \\ \hat{\boldsymbol{\lambda}}_V - \boldsymbol{\lambda}_V \end{bmatrix} \quad (4.7)$$

$$= \begin{bmatrix} \hat{\Psi}\mathbf{y}_V + \mathbf{A}_0^t \boldsymbol{\lambda}_V \\ \mathbf{b} \end{bmatrix} + \hat{M}_V \begin{bmatrix} \hat{\mathbf{y}}_V - \mathbf{y}_V \\ \hat{\boldsymbol{\lambda}}_V - \boldsymbol{\lambda}_V \end{bmatrix}.$$

Since $\hat{M}_V$ is invertible with probability one, we can rewrite (4.7) (after multiplying both sides by $\sqrt{T}$) as

$$\sqrt{T} \begin{bmatrix} \hat{\mathbf{y}}_V - \mathbf{y}_V \\ \hat{\boldsymbol{\lambda}}_V - \boldsymbol{\lambda}_V \end{bmatrix} = -\sqrt{T}\hat{M}_V^{-1} \left\{ \begin{bmatrix} \hat{\Psi}\mathbf{y}_V + \mathbf{A}_0^t \boldsymbol{\lambda}_V \\ \mathbf{b} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix} \right\}$$

$$= -(M_V + o_P(1))^{-1} \begin{bmatrix} \sqrt{T}(\hat{\Psi}\mathbf{y}_V + \mathbf{A}_0^t \boldsymbol{\lambda}_V) \\ \mathbf{0} \end{bmatrix}. \quad (4.8)$$

Notice that

$$\begin{aligned} \sqrt{T}(\hat{\Psi}\mathbf{y}_V + \mathbf{A}_0^t \boldsymbol{\lambda}_V) &= \sqrt{T}((\hat{\Psi}-\Psi)\mathbf{y}_V + \Psi\mathbf{y}_V + \mathbf{A}_0^t \boldsymbol{\lambda}_V) \\ &= \sqrt{T}[(\hat{\Psi}-\Psi)\mathbf{y}_V] \text{ by (4.5)}, \\ &= \sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T} \mathbf{w}_i \mathbf{w}_i^t \mathbf{y}_V - E(\mathbf{W}\mathbf{W}^t \mathbf{y}_V)\right). \end{aligned}$$

Therefore, by the classical central limit theorem, $\sqrt{T}(\hat{\Psi}\mathbf{y}_V + \mathbf{A}_0^t \boldsymbol{\lambda}_V)$ converges in distribution to a normal vector with mean $\mathbf{0}$ and covariance

$$Q_V^* = Cov(\mathbf{W}\mathbf{W}^t \mathbf{y}_V)$$

So by Slutsky's Lemma, and from (4.8)

$$\sqrt{T} \begin{bmatrix} \hat{\mathbf{y}}_V - \mathbf{y}_V \\ \hat{\boldsymbol{\lambda}}_V - \boldsymbol{\lambda}_V \end{bmatrix} \rightsquigarrow \mathbf{U}_V,$$

where $\mathbf{U}_V$ is normal with mean $\mathbf{0}$ and covariance

$$Cov(\mathbf{U}_V) = M_V^{-1} \begin{bmatrix} Q_V^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_M \end{bmatrix} M_V^{-1}.$$

Now write

$$M_V^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

where B_{11} has dimension $(N+1) \times (N+1)$ so that

$$\begin{aligned} Cov(\mathbf{U}_V) &= \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} Q_V & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_M \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \\ &= \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} Q_V B_{11} & Q_V B_{12} \\ \mathbf{0} & \mathbf{0}_M \end{bmatrix} \\ &= \begin{bmatrix} B_{11} Q_V B_{11} & B_{11} Q_V B_{12} \\ B_{21} Q_V B_{11} & B_{21} Q_V B_{12} \end{bmatrix}. \end{aligned}$$

It follows that the asymptotic covariance of $\sqrt{T}(\hat{\mathbf{y}} - \mathbf{y}_V)$ is asymptotically normal with mean $\mathbf{0}$ and covariance $B_{11} Q_V B_{11}$. The conclusion then follows with $L_V^* := B_{11}$. ■

Focusing on $\hat{\mathbf{x}}_V$ instead of $\hat{\mathbf{y}}_V$ (remember that $\hat{\mathbf{y}}_V = (\hat{\mathbf{x}}_V^t, \hat{q}_V)^t$) , we get the following corollary, which proves the asymptotic normality of the variance portfolio estimator, and gives its asymptotic covariance matrix in closed form.

Corollary 14 *Let $\hat{\mathbf{x}}_V$ be defined as in (4.2). Suppose assumption (A) holds. Then*

$$\sqrt{T}(\hat{\mathbf{x}}_V - \mathbf{x}_V) \rightsquigarrow N(\mathbf{0}, L_V Q_V L_V),$$

where

$$\begin{aligned} Q_V &= Cov[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R}^{\mathbf{t}} \mathbf{x}_{\mathbf{V}} - q_V)] \\ &= Cov[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t \mathbf{x}_V], \end{aligned}$$

and

$$L_V = \Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}.$$

Proof. We know from Theorem 13 that

$$\sqrt{T}(\hat{\mathbf{y}}_V - \mathbf{y}_V) \rightsquigarrow N(\mathbf{0}, L_V^* Q_V^* L_V^*)$$

where

$$Q_V^* = Cov(\mathbf{W}\mathbf{W}^t\mathbf{y}_V)$$

and where L_V^* is the upper-left $(N+1)$-dimensional corner of M_V^{-1}, where

$$M_V = \begin{bmatrix} \Psi & \mathbf{A}_0^t \\ \mathbf{A}_0 & \mathbf{0}_M \end{bmatrix}$$

Write

$$L_V^* = \begin{bmatrix} L_V & \mathbf{f} \\ \mathbf{f}^t & d \end{bmatrix}$$

where L_V is the upper-left N-dimensional corner of M_V^{-1}, and where $\mathbf{f} \in \mathbb{R}^N$, $d \in \mathbb{R}$. Let $\Gamma = E(\mathbf{R}\mathbf{R}^t)$. Then since (i) Γ is positive definite, by virtue of Σ being positive definite, and (ii) M_V is invertible from Theorem 13, we have from Proposition 51 in the appendix

$$\begin{aligned} L_V &= \Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}, && (4.9) \\ \mathbf{f} &= \left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right]\boldsymbol{\mu} \\ &= L_V\boldsymbol{\mu}. && (4.10) \end{aligned}$$

We now partition Q_V^* as

$$\begin{aligned}
Q_V^* &= Cov(\mathbf{W}\mathbf{W}^t\mathbf{y}_V) \\
&= E\left[\mathbf{W}\mathbf{W}^t(\mathbf{W}^t\mathbf{y}_V)^2\right] - E\left[\mathbf{W}(\mathbf{W}^t\mathbf{y}_V)\right]E\left[\mathbf{W}^t(\mathbf{W}^t\mathbf{y}_V)\right] \\
&= \begin{bmatrix} E[\mathbf{R}\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)^2] & E[-\mathbf{R}(\mathbf{R}^t\mathbf{x}_V - q_V)^2] \\ E[-\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)^2] & E[(\mathbf{R}^t\mathbf{x}_V - q_V)^2] \end{bmatrix} \\
&\quad - \begin{bmatrix} E[\mathbf{R}(\mathbf{R}^t\mathbf{x}_V - q_V)] \\ 0 \end{bmatrix} \begin{bmatrix} E[\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)] & 0 \end{bmatrix}, \text{ by definition of } \mathbf{y}_V \text{ and } \mathbf{W}, \\
&= \begin{bmatrix} E[\mathbf{R}\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)^2] - E[\mathbf{R}(\mathbf{R}^t\mathbf{x}_V - q_V)][\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)] & E[-\mathbf{R}(\mathbf{R}^t\mathbf{x}_V - q_V)^2] \\ E[-\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)^2] & E[(\mathbf{R}^t\mathbf{x}_V - q_V)^2] \end{bmatrix} \\
:&= \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^t & Q_{22} \end{bmatrix}. \qquad (4.11)
\end{aligned}$$

Notice that by definition of $\hat{\mathbf{y}}_V$ we have

$$\hat{\mathbf{x}}_V - \mathbf{x}_V = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \end{bmatrix} (\hat{\mathbf{y}}_V - \mathbf{y}_V)$$

where here $\mathbf{0} \in \mathbb{R}^N$. It follows that $\sqrt{T}(\hat{\mathbf{x}}_V - \mathbf{x}_V)$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix

$$\begin{aligned}
& \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \end{bmatrix} L_V^* Q_V^* L_V^* \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}^t \end{bmatrix} \\
= & \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \end{bmatrix} \begin{bmatrix} L_V & \mathbf{f} \\ \mathbf{f}^t & d \end{bmatrix} Q_V^* \begin{bmatrix} L_V & \mathbf{f} \\ \mathbf{f}^t & d \end{bmatrix} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}^t \end{bmatrix} \\
= & \begin{bmatrix} L_V & \mathbf{f} \end{bmatrix} Q_V^* \begin{bmatrix} L_V \\ \mathbf{f}^t \end{bmatrix} \\
= & \begin{bmatrix} L_V & \mathbf{f} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^t & Q_{22} \end{bmatrix} \begin{bmatrix} L_V \\ \mathbf{f}^t \end{bmatrix} \\
= & L_V \begin{bmatrix} \mathbf{I} & \boldsymbol{\mu} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^t & Q_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \boldsymbol{\mu}^t \end{bmatrix} L_V \\
= & L_V \left[Q_{11} + \boldsymbol{\mu} Q_{12}^t + Q_{12} \boldsymbol{\mu}^t + \boldsymbol{\mu} Q_{22} \boldsymbol{\mu}^t \right] L_V,
\end{aligned} \tag{4.12}$$

where the second to last equation follows from (4.9) and (4.10). Focusing on the middle term in (4.12), and using (4.11), we write

$$\begin{aligned}
& \left[Q_{11} + \boldsymbol{\mu} Q_{12}^t + Q_{12} \boldsymbol{\mu}^t + \boldsymbol{\mu} Q_{22} \boldsymbol{\mu}^t \right] \\
= & E[\mathbf{R}\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)^2] - E[\mathbf{R}(\mathbf{R}^t\mathbf{x}_V - q_V)][\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)] \\
& + \boldsymbol{\mu} E[-\mathbf{R}^t(\mathbf{R}^t\mathbf{x}_V - q_V)^2] + E[-\mathbf{R}(\mathbf{R}^t\mathbf{x}_V - q_V)^2]\boldsymbol{\mu}^t \\
& + \boldsymbol{\mu} E[(\mathbf{R}^t\mathbf{x}_V - q_V)^2]\boldsymbol{\mu}^t \\
= & E[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t (\mathbf{R}^t\mathbf{x}_V - q_V)^2] - E\left[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R}^t\mathbf{x}_V - q_V)\right] E\left[(\mathbf{R} - \boldsymbol{\mu})^t (\mathbf{R}^t\mathbf{x}_V - q_V)\right] \\
= & Cov\left[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R}^t\mathbf{x}_V - q_V)\right] := Q_V
\end{aligned}$$

But notice that by definition q_V, $q_V = \boldsymbol{\mu}^t\mathbf{x}_V$ so that.

$$\begin{aligned}
Q_V &= Cov\left[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R}^t\mathbf{x}_V - q_V)\right] \\
&= Cov\left[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})\mathbf{x}_V\right].
\end{aligned}$$

Using Q_V in (4.12) yields the following simple expression for the asymptotic covariance of $\sqrt{T}(\hat{\mathbf{x}}_V - \mathbf{x}_V)$:

$$L_V Q_V L_V.$$

■

4.2 Asymptotic Normality of the Shortfall Portfolio Estimator

Our approach to proving the asymptotic normality of the shortfall portfolio estimator is inspired by the work of van de Geer (1990) and Pollard (1991) on the asymptotic properties of the least absolute deviations (LAD) linear regression estimator. Using the Asymptotic Equicontinuity Criterion stated below, a result from empirical process theory, we show that the minimizer of the empirical function (4.1) approaches in probability the solution to a quadratic optimization problem. We use the closed form solution to this QP, which is a random variable, to derive the asymptotic distribution of the shortfall portfolio estimator. Notice that classical theorems such as Huber's M-estimator consistency and asymptotic normality theorems (see for example Huber, 1981) do not apply directly to our estimator, because of the non-differentiability of the ρ_α function.

4.2.1 Assumption (B)

We start by making an assumption about the distribution of $\mathbf{R}$.

(B): $\mathbf{R}$, with mean $\boldsymbol{\mu}$ and covariance matrix Σ, has a continuous density, and for every $\mathbf{x} \in \mathbb{R}^N$ and $q \in \mathbb{R}$ such that $f_{\mathbf{x}^t\mathbf{R}}(q) > 0$, the density $f_{\mathbf{R},\mathbf{x}^t\mathbf{R}}(\mathbf{r}, q)$ is well-defined, $Cov[\mathbf{R} \mid \mathbf{x}^t\mathbf{R} - q = 0]$ is well defined, and the rank of $Cov[\mathbf{R} \mid \mathbf{x}^t\mathbf{R} - q = 0]$ is $N - 1$.

Obviously, the rank of $Cov[\mathbf{R} \mid \mathbf{x}^t\mathbf{R} - q = 0]$ is at most $N-1$. Assumption (B) will hold for example when $Cov[\mathbf{R} \mid \mathbf{x}^t\mathbf{R} - q = 0]$ is well defined for every $\mathbf{x} \in \mathbb{R}^N$ and $q \in \mathbb{R}$ such that $f_{\mathbf{x}^t\mathbf{R}}(q) > 0$, and when the density of $\mathbf{R}$ is continuous and has a support which is closed in $\mathbb{R}^N$.

4.2.2 Notation

We introduce some notation. Let $\mathbf{W} = (\mathbf{R}^t, -1)^t$, and let P be the distribution of $\mathbf{W}$, $E(.)$ be the expectation with respect to $\mathbf{W}$. Let $P_T = (1/T)\sum_{i=1}^T 1_{\mathbf{W}_i}$, where $1_{\mathbf{W}_i}$ is the point mass at $\mathbf{W}_i := (\mathbf{R}_i^t, -1)^t$, and let $\int g dP_T = (1/T)\sum_{i=1}^T g(\mathbf{W}_i)$ for any function $\mathbb{R}^{N+1} \to \mathbb{R}$. For any function g :

$\mathbb{R}^{N+1} \rightarrow \mathbb{R}$, let

$$||g||_{L_2} := ||g||_{L_2(P)} = E[g^2(\mathbf{W})].$$

Define $\mathbf{y}_\alpha = (\mathbf{x}_\alpha, q_\alpha)^t, \hat{\mathbf{y}} = (\hat{\mathbf{x}}_\alpha^t, \hat{q}_\alpha)$, and

$$\mathcal{Y} = \left\{\mathbf{y} \in \mathbb{R}^{N+1} \mid \mathbf{A}_0\mathbf{y} = \mathbf{b}\right\},$$

with $\mathbf{A}_0 = [\mathbf{A}\ \mathbf{0}]$. Note that

$$\mathbf{y}_\alpha = \arg\min_{\mathbf{y}\in\mathcal{Y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})] \tag{4.13}$$

As in the previous section, define $\mathcal{Z} = \{\mathbf{z} \in \mathbb{R}^{N+1} \mid \mathbf{A}_0\mathbf{z} = \mathbf{0}\}$, the space of first order feasible variations.

We will need the following result, in the form of a lemma.

Lemma 15 *The gradient and Hessian of $E[\rho_\alpha(\mathbf{y}^t\mathbf{W})] = E[\rho_\alpha(\mathbf{x}^t\mathbf{R} - q)]$ are well-defined and can be written as*

$$\begin{aligned} \nabla_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})] &= \begin{bmatrix} E(\mathbf{R}) - \frac{1}{\alpha}E(\mathbf{R}1_{\{\mathbf{x}^t\mathbf{R}-q\leqslant 0\}}) \\ -1 + \frac{1}{\alpha}\Pr(\mathbf{x}^t\mathbf{R} - q \leqslant 0) \end{bmatrix} \\ &= E(\mathbf{W}) - \frac{1}{\alpha}E(\mathbf{W}1_{\{\mathbf{y}^t\mathbf{W}\leqslant 0\}}), \end{aligned} \tag{4.14}$$

and

$$\begin{aligned} \nabla^2_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})] &= f_{\mathbf{x}^t\mathbf{R}}(q)\frac{1}{\alpha}\begin{bmatrix} E[\mathbf{R}\mathbf{R}^t \mid \mathbf{x}^t\mathbf{R}-q = 0] & E[\mathbf{R} \mid \mathbf{x}^t\mathbf{R}-q = 0] \\ E[\mathbf{R}^t \mid \mathbf{x}^t\mathbf{R}-q = 0] & 1 \end{bmatrix} \\ &= f_{\mathbf{y}^t\mathbf{W}}(0)\frac{1}{\alpha}E[\mathbf{W}\mathbf{W}^t \mid \mathbf{y}^t\mathbf{W} = 0], \end{aligned} \tag{4.15}$$

where $f_{\mathbf{x}^t\mathbf{R}}(.)$ is the density of variable $\mathbf{x}^t\mathbf{R}$, and $f_{\mathbf{y}^t\mathbf{W}}(.)$ is the density of variable $\mathbf{y}^t\mathbf{W}$.

Proof. See the appendix. ■

4.2.3 Outline of Proof

We outline below the proof of the asymptotic normality of $\hat{\mathbf{x}}_\alpha$.

Step1: Using the Asymptotic Equicontinuity Lemma (defined below), a result about the convergence of empirical processes indexed by functions that form a polynomial class (i.e. a class that is "not too complex"), we show that the empirical process

$$\begin{aligned} l_T(\mathbf{z}) &= \alpha\sum_{i=1}^{T}\left[\rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W}_i - T^{-1/2}\mathbf{z}^t\mathbf{W}_i) - \rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W}_i)\right] \\ &= \mathbf{z}^t\mathbf{U}_T + \frac{1}{2}\mathbf{z}^t\Sigma_\alpha\mathbf{z} + o_p(1), \end{aligned}$$

uniformly over $||z|| \leqslant K$. The matrix Σ_α is deterministic and expressible in closed form. $\mathbf{U}_T$ converges to a known distribution by the classical CLT for i.i.d. vectors. This is Lemma 21. Note that the minimizer $\hat{\mathbf{z}}$ of $l_T(.)$ is expressible as $\hat{\mathbf{z}} = \sqrt{T}(\hat{\mathbf{y}}_\alpha - \mathbf{y}_\alpha)$.

Step 2: Write the Lagrangean corresponding to

$$\begin{aligned} &\text{minimize} \quad \mathbf{z}^t\mathbf{U}_T + \tfrac{1}{2}\mathbf{z}^t\Sigma_\alpha\mathbf{z} \\ &\text{subject to} \quad \mathbf{A}_0\mathbf{z} = \mathbf{0}, \end{aligned}$$

to get a closed form solution $\vec{\mathbf{z}}$ for the problem. By a convexity argument and some linear algebra, we show that $\hat{\mathbf{z}}$ converges to $\vec{\mathbf{z}}$ whose asymptotic distribution just follows by the classical CLT for i.i.d. vectors. This shows that $\sqrt{T}(\hat{\mathbf{y}}_\alpha - \mathbf{y}_\alpha) \rightsquigarrow N(\mathbf{0}, L_\alpha^* Q_\alpha^* L_\alpha^*)$, for some L_α^* and Q_α^* expressible in closed form. This is Theorem 22.

Step 3: Using linear algebra, from the asymptotic normality of $\hat{\mathbf{y}}_\alpha$ one gets the asymptotic normality of $\hat{\mathbf{x}}_\alpha$, $\sqrt{T}(\hat{\mathbf{x}}_\alpha - \mathbf{x}_\alpha) \rightsquigarrow N(\mathbf{0}, L_\alpha Q_\alpha L_\alpha)$ for some L_α and Q_α expressible in closed form, which is Corollary 23.

4.2.4 Uniqueness of the Shortfall Portfolio Estimator

In the following proofs, we will need the results below.

Lemma 16 *Under assumption (B), the following statements are true:*

(a) For all $\mathbf{z} \in \mathcal{Z}$,

$$E[\rho_\alpha((\mathbf{y}_\alpha + \mathbf{z})^t\mathbf{W}) - \rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W})] = \frac{1}{2}\mathbf{z}^t[\nabla^2_\mathbf{y} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}]\mathbf{z} + o(||\mathbf{z}||^2),$$

(b) For all $\mathbf{z} \in \mathcal{Z}$,

$$\mathbf{z}^t[\nabla^2_\mathbf{y} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}]\mathbf{z} > \bar{\lambda}_\alpha||\mathbf{z}||^2$$

for some $\bar{\lambda}_\alpha > 0$.

(c) There exists an $\epsilon_\alpha > 0$ *and a* $\gamma_\alpha > 0$ *such that for all* $\mathbf{z} \in \mathcal{Z}$, $||\mathbf{z}|| < \epsilon_\alpha$,

$$E[\rho_\alpha((\mathbf{y}_\alpha + \mathbf{z})^t\mathbf{W}) - \rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W})] > \gamma_\alpha||\mathbf{z}||^2.$$

Therefore $\mathbf{z} = \mathbf{0}$ *is the unique minimizer of* $E[\rho_\alpha((\mathbf{y}_\alpha + \mathbf{z})^t W) - \rho_\alpha(\mathbf{y}_\alpha^t W)]$ *over* $\mathcal{Z}$, *and* $\mathbf{y}_\alpha$ *is the unique minimizer of the shortfall* $E[\rho_\alpha(\mathbf{y}^t W)]$ *over* $\mathcal{Y}$.

Proof. See the appendix. ■

Remark 17 *Let* $\Sigma_\alpha = \alpha\nabla^2_\mathbf{y} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}$. *From the preceding Lemma 16 (b), we see that*

$$\begin{aligned} &\textit{minimize} \quad \mathbf{z}^t\Sigma_\alpha\mathbf{z} \\ &\textit{subject to} \quad \mathbf{A}_0\mathbf{z} = \mathbf{0} \end{aligned}$$

has a unique solution, $\mathbf{0} \in \mathbb{R}^{N+1}$. *Writing the Lagrangean* $L(\mathbf{z}, \boldsymbol{\lambda}) = \mathbf{z}^t\Sigma_\alpha\mathbf{z} + \boldsymbol{\lambda}^t(\mathbf{A}_0\mathbf{z})$ *of the above problem and taking its derivative with respect to* $\mathbf{z}$ *we get*

$$\begin{aligned} \nabla_{(\mathbf{z},\boldsymbol{\lambda})} L(.,.) &= \begin{bmatrix} \Sigma_\alpha\mathbf{z} + \mathbf{A}_0^t\boldsymbol{\lambda} \\ \mathbf{A}_0\mathbf{z} \end{bmatrix} \\ &= M_\alpha \begin{bmatrix} \mathbf{z} \\ \boldsymbol{\lambda} \end{bmatrix}, \end{aligned}$$

with

$$M_\alpha = \begin{bmatrix} \Sigma_\alpha & \mathbf{A}_0^t \\ \mathbf{A}_0 & \mathbf{0} \end{bmatrix}$$

By uniqueness of the optimal solution $\mathbf{0}$*, and by uniqueness of its Lagrange multipliers* $\boldsymbol{\lambda}$*, the system*

$$\nabla_{(\mathbf{z},\boldsymbol{\lambda})} L(.,.) = \mathbf{0}$$

must have a unique solution. This implies that M_α *is invertible.*

4.2.5 Asymptotic Equicontinuity

We will need some results from empirical process theory - see Pollard (1985) and Pollard (1984, Ch. II). We start with the following definitions.

Definition 18 *A class of sets* $\mathcal{D}$ *is said to be a polynomial class (also called a Vapnik-Cervonenkis class) if there exists a polynomial* $p(.)$ *with the property: for every finite set* A *there are at most* $p(card(A))$ *distinct subsets of the form* $A \cap D$ *with* D *in* $\mathcal{D}$*. (Here* $card(A)$ *is the cardinality of* A*).*

Definition 19 *The graph of a function* $g : \mathbb{R}^{N+1} \rightarrow \mathbb{R}$ *is defined as the set*

$$\left\{(\mathbf{w},t) \in \mathbb{R}^{N+2} \mid 0 \leqslant t \leqslant g(\mathbf{w}) \text{ or } 0 \geqslant t \geqslant g(\mathbf{w})\right\}.$$

A collection $\mathcal{G}$ of functions is said to be a polynomial class if the graphs of all the functions in $\mathcal{G}$ form a polynomial class of sets. Define the class of functions

$$\mathcal{G} = \{g_{\mathbf{z},\mathbf{v}} : \mathbb{R}^{N+1} \rightarrow \mathbb{R} \mid g_{\mathbf{z},\mathbf{v}}(\mathbf{w}) = \mathbf{z}^t \mathbf{w} 1_{\{\mathbf{y}_\circ^t \mathbf{w} \leqslant \mathbf{v}^t \mathbf{w}\}}, \mathbf{z}, \mathbf{v} \in \mathbb{R}^{N+1}\}.$$

Notice that the graph of $g_{\mathbf{z},\mathbf{v}}$ for $\mathbf{z},\mathbf{v} \in \mathbb{R}^{N+1}$ can be written as

$$\begin{aligned}\{(\mathbf{w},t) \;\in\; & \mathbb{R}^{N+2} \mid 0 \leqslant t \leqslant \mathbf{z}^t\mathbf{w}1_{\{\mathbf{y}_\alpha^t\mathbf{w}\leqslant\mathbf{v}^t\mathbf{w}\}} \text{ or } 0 \geqslant t \geqslant \mathbf{z}^t\mathbf{w}1_{\{\mathbf{y}_\alpha^t\mathbf{w}\leqslant\mathbf{v}^t\mathbf{w}\}}\} \\ = \; & \{\{0 \leqslant t\} \cap \{0 \leqslant -t + \mathbf{z}^t\mathbf{w}\} \cap \{0 \leqslant (\mathbf{v} - \mathbf{y}_\alpha)^t\mathbf{w}\} \\ & \cup \{\{0 \geqslant t\} \cap \{0 \geqslant -t + \mathbf{z}^t\mathbf{w}\} \cap \{0 \leqslant (\mathbf{v} - \mathbf{y}_\alpha)^t\mathbf{w}\} \\ & \cup \{\{t = 0\} \cap \{0 > (\mathbf{v} - \mathbf{y}_\alpha)^t\mathbf{w}\}\,.\end{aligned}$$

Therefore, the graph of $g_{\mathbf{z},\mathbf{v}}$ is the union of three sets, each of which is the intersection of three halfspaces. From Lemma 15 in Pollard (1984, Ch. II), we know that a class of sets expressible as finite unions and intersections of halfspaces is a polynomial class. So the collection of functions $\mathcal{G}$ as defined above is a polynomial class.

For any $g \in \mathcal{G}$, define the empirical process

$$\nu_T(g) = T^{1/2} \int gd(P_T - P), \;\; g \in \mathcal{G}.$$

Also define the envelope of $\mathcal{G}$ as any measurable function G such that $|g| < G$ almost surely for all $g \in \mathcal{G}$. We will need the following lemma for polynomial classes (see Lemma 15 in Pollard, 1984, Ch. VII).

Lemma 20 (Aymptotic Equicontinuity) *Suppose that $\mathcal{G}$ is a polynomial class with envelope G satisfying $||G||_{L_2} < \infty$. Then the asymptotic equicontinuity criterion holds: for all $\delta > 0$, there is a T_δ such that*

$$\Pr(\sup_{\substack{g,\tilde{g}\in\mathcal{G} \\ ||g-\tilde{g}||_{L_2}<\delta}} |\nu_T(g) - \nu_T(\tilde{g})| > \delta) < \delta$$

for all $T > T_\delta$.

4.2.6 Proof of the Asymptotic Normality of the Shortfall Portfolio Estimator

Consider the process

$$l_T(\mathbf{z}) = \alpha \sum_{i=1}^{T} \left[\rho_\alpha(\mathbf{y}_\alpha^t \mathbf{W}_i - T^{-1/2}\mathbf{z}^t \mathbf{W}_i) - \rho_\alpha(\mathbf{y}_\alpha^t \mathbf{W}_i) \right], \tag{4.16}$$

indexed by $\mathbf{z} \in \mathcal{Z}$, the space of first order feasible variations. We show in the next lemma that $l_T(\mathbf{z})$ converges in probability to a random quadratic function of $\mathbf{z}$, uniformly over $||\mathbf{z}|| \leqslant K$, for K fixed.

Lemma 21 *Let $K > 0$ be fixed. Under assumption (B), for $\mathbf{z} \in \mathcal{Z}$,*

$$l_T(\mathbf{z}) = \mathbf{z}^t \mathbf{U}_T + \frac{1}{2}\mathbf{z}^t \Sigma_\alpha \mathbf{z} + o_P(1),$$

uniformly over all $||\mathbf{z}|| \leqslant K$, where

$$\Sigma_\alpha = f_{\mathbf{y}_\alpha^t \mathbf{W}}(0) E(\mathbf{W}\mathbf{W}^t | \mathbf{y}_\alpha^t \mathbf{W} = 0),$$

and where $\mathbf{U}_T$ does not depend on $\mathbf{z}$,

$$\mathbf{U}_T \rightsquigarrow N(0, Q_\alpha^*),$$

with

$$Q_\alpha^* = Cov\left[(1-\alpha)\mathbf{W} 1_{\{\mathbf{y}_\alpha \mathbf{W} \leqslant 0\}} - \alpha \mathbf{W} 1_{\{\mathbf{y}_\alpha \mathbf{W} > 0\}}) \right].$$

Proof. We expand (4.16) using the definition of ρ_α to get

$$\begin{aligned} l_T(\mathbf{z}) &= \sum_{i=1}^{T}\Big[\alpha(\mathbf{y}_\alpha\mathbf{W}_i - T^{-1/2}\mathbf{z}^t\mathbf{W}_i)1_{\{\mathbf{y}_\alpha\mathbf{W}_i - T^{-1/2}\mathbf{z}^t\mathbf{W}_i\geqslant 0\}} \\ &\quad + (1-\alpha)(T^{-1/2}\mathbf{z}^t\mathbf{W}_i - \mathbf{y}_\alpha\mathbf{W}_i)1_{\{\mathbf{y}_\alpha\mathbf{W}_i - T^{-1/2}\mathbf{z}^t\mathbf{W}_i\leqslant 0\}}\Big] \\ &\quad - \sum_{i=1}^{T}\left[\alpha(\mathbf{y}_\alpha\mathbf{W}_i)1_{\{\mathbf{y}_\alpha\mathbf{W}_i\geqslant 0\}} + (1-\alpha)(-\mathbf{y}_\alpha\mathbf{W}_i)1_{\{\mathbf{y}_\alpha\mathbf{W}_i\leqslant 0\}}\right] \\ &= (1-\alpha)\sum_{i=1}^{T} T^{-1/2}\mathbf{z}^t\mathbf{W}_i 1_{\{\mathbf{y}_\alpha\mathbf{W}_i\leqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}} - \alpha\sum_{i=1}^{T} T^{-1/2}\mathbf{z}^t\mathbf{W}_i 1_{\{\mathbf{y}_\alpha\mathbf{W}_i\geqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}} \\ &\quad - \sum_{i=1}^{T}\mathbf{y}_\alpha\mathbf{W}_i 1_{\{0\leqslant\mathbf{y}_\alpha\mathbf{W}_i\leqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}} + \sum_{i=1}^{T}\mathbf{y}_\alpha\mathbf{W}_i 1_{\{0\geqslant\mathbf{y}_\alpha\mathbf{W}_i\geqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}}, \end{aligned} \tag{4.17}$$

where the second equality follows after some algebra. Consider the first term in expression (4.17). Let $\mathcal{G}(K) = \{g_{\mathbf{z},\mathbf{v}} : \mathbb{R}^{N+1}\to\mathbb{R} \mid g_{\mathbf{z},\mathbf{v}}(\mathbf{w}) = \mathbf{z}^t\mathbf{w}1_{\{\mathbf{y}_o^t\mathbf{w}\leqslant\mathbf{v}^t\mathbf{w}\}}, \mathbf{z},\mathbf{v}\in\mathbb{R}^{N+1},\ ||\mathbf{z}|| < K\} \subset\mathcal{G}$. $\mathcal{G}\subset\mathcal{G}(K)$, $\mathcal{G}(K)$ is still a polynomial class, and since

$$K||\mathbf{W}|| \geqslant \sup_{g\in\mathcal{G}(K)} g(\mathbf{W})$$

its envelope $G(.)$ satisfies

$$E[G(\mathbf{W})^2] \leqslant E[(K||\mathbf{W}||)^2] = K^2\sigma^2 < \infty,$$

where $\sigma^2 := E[||\mathbf{W}||^2] < \infty$ by definition of $\mathbf{W}$ and assumption (B). Moreover,

$$\begin{aligned} ||g_{\mathbf{z},T^{-1/2}\mathbf{z}} - g_{\mathbf{z},0}||_{L_2}^2 &= E(\mathbf{z}^t\mathbf{W}(1_{\{\mathbf{y}_\alpha\mathbf{W}\leqslant T^{-1/2}\mathbf{z}^t\mathbf{W}\}} - 1_{\{\mathbf{y}_\alpha\mathbf{W}\leqslant 0\}}))^2 \\ &\leqslant E(||\mathbf{z}||^2||\mathbf{W}||^2 1_{\{0\leqslant|\mathbf{y}_\alpha\mathbf{W}|\leqslant T^{-1/2}|\mathbf{z}^t\mathbf{W}|\}}) \\ &\leqslant E(K^2||\mathbf{W}||^2 1_{\{0\leqslant|\mathbf{y}_\alpha\mathbf{W}|\leqslant T^{-1/2}|\mathbf{z}^t\mathbf{W}|\}}) \\ &\leqslant K^2E(||\mathbf{W}||^2)E(1_{\{0\leqslant|\mathbf{y}_\alpha\mathbf{W}|\leqslant T^{-1/2}K||\mathbf{W}||\}}). \end{aligned}$$

so $||g_{\mathbf{z},T^{-1/2}\mathbf{z}} - g_{\mathbf{z},0}||_{L_2} = o(1)$ uniformly in $||\mathbf{z}|| < K$, by ndomiated convergence. Therefore, by the Asymptotic Equicontinuity Criterion,

$$\nu_T(g_{\mathbf{z},T^{-1/2}\mathbf{z}}) = \nu_T(g_{\mathbf{z},0}) + o_P(1),$$

uniformly in $||\mathbf{z}|| < K$. We can therefore write the first term in (4.17) as

$$(1-\alpha)\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\mathbf{z}^t\mathbf{W}_i 1_{\{\mathbf{y}_\alpha \mathbf{W}_i \leqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}} = (1-\alpha)\nu_T(g_{\mathbf{z},0}) + (1-\alpha)T^{1/2}\int g_{\mathbf{z},T^{-1/2}\mathbf{z}}dP + o_P(1), \tag{4.18}$$

uniformly in $||\mathbf{z}|| < K$. Define

$$\mathbf{U}_{1,T} = \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\left[\mathbf{W}_i 1_{\{\mathbf{y}_\alpha^t \mathbf{W} \leqslant 0\}} - E(\mathbf{W}_i 1_{\{\mathbf{y}_\alpha^t \mathbf{W} \leqslant 0\}})\right].$$

Then (4.18) becomes

$$(1-\alpha)\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\mathbf{z}^t\mathbf{W}_i 1_{\{\mathbf{y}_\alpha^t \mathbf{W}_i \leqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}} = (1-\alpha)\mathbf{z}^t\mathbf{U}_{1,T} + (1-\alpha)T^{1/2}\int (g_{\mathbf{z},T^{-1/2}\mathbf{z}})dP + o_P(1), \tag{4.19}$$

uniformly in $||\mathbf{z}|| < K$.

Similarly, for the second expression in (4.17), we find

$$\alpha\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\mathbf{z}^t\mathbf{W}_i 1_{\{\mathbf{y}_\alpha^t \mathbf{W}_i \geqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}} = \alpha\mathbf{z}^t\mathbf{U}_{2,T} + \alpha T^{1/2}\int (\bar{g}_{\mathbf{z},T^{-1/2}\mathbf{z}})dP + o_P(1), \tag{4.20}$$

uniformly in $||\mathbf{z}|| < K$, where

$$\mathbf{U}_{2,T} = \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\left[\mathbf{W}_i 1_{\{\mathbf{y}_\alpha^t \mathbf{W} > 0\}} - E(\mathbf{W}_i 1_{\{\mathbf{y}_\alpha^t \mathbf{W} > 0\}})\right],$$

and where $\bar{g}_{\mathbf{z},T^{-1/2}\mathbf{z}} \in \bar{\mathcal{G}}(K) = \{\bar{g}_{\mathbf{z},\mathbf{v}} : \mathbb{R}^{N+1} \rightarrow \mathbb{R} | \bar{g}_{\mathbf{z},\mathbf{v}}(\mathbf{w}) = \mathbf{z}^t\mathbf{w}1_{\{\mathbf{y}_\alpha^t \mathbf{w} > \mathbf{v}^t\mathbf{w}\}}, \mathbf{z},\mathbf{v} \in \mathbb{R}^{N+1}, ||\mathbf{z}|| < K\}$.

Consider the third term in (4.17). Define for $\mathbf{w} \in \mathbb{R}^{N+1}$,

$$h_{T^{-1/2}\mathbf{z},T^{1/2}}(\mathbf{w}) = T^{1/2}(\mathbf{y}_\alpha^t \mathbf{w})1_{\{0 \leqslant \mathbf{y}_\alpha^t \mathbf{w} \leqslant T^{-1/2}\mathbf{z}^t\mathbf{w}\}}.$$

Let $\mathcal{H}(K) = \{h_{T^{-1/2}\mathbf{z},T^{1/2}} : \mathbb{R}^{N+1} \rightarrow \mathbb{R} | h_{T^{-1/2}\mathbf{z},T^{1/2}}(\mathbf{w}) = T^{1/2}(\mathbf{y}_\alpha^t \mathbf{w})1_{\{0 \leqslant \mathbf{y}_\alpha^t \mathbf{w} \leqslant T^{-1/2}\mathbf{z}^t\mathbf{w}\}}, \mathbf{z} \in \mathbb{R}^{N+1}, ||\mathbf{z}|| < K\}$. Then $\mathcal{H}$ is a VC-graph class, and its envelope H satisfies

$$\begin{aligned} ||H||_{L_2}^2 &\leqslant TE[(\mathbf{y}_\alpha^t \mathbf{w})^2 1_{\{0 \leqslant \mathbf{y}_\alpha^t \mathbf{w} \leqslant T^{-1/2}K||\mathbf{w}||\}}] \\ &\leqslant E(K||\mathbf{W}||)^2 = K^2\sigma^2 < \infty, \end{aligned}$$

where as before $\sigma^2 = E(||\mathbf{W}||^2)$. Also, $||h_{T^{-1/2}\mathbf{z},T^{1/2}}||_{L_2} = o(1)$ uniformly in $||\mathbf{z}|| \leqslant K$ by dominated convergence, so $\nu_T(h_{T^{-1/2}\mathbf{z},T^{1/2}}) = o_P(1)$ by the Asymptotic Equicontinuity Criterion. Then

$$\begin{aligned} \sum_{i=1}^T \mathbf{y}_\alpha^t \mathbf{W}_i 1_{\{0 \leqslant \mathbf{y}_\alpha^t \mathbf{W}_i \leqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}} &= \nu_T(h_{T^{-1/2}\mathbf{z},T^{1/2}}) + T^{1/2}\int h_{T^{-1/2}\mathbf{z},T^{1/2}}dP \\ &= T^{1/2}\int h_{T^{-1/2}\mathbf{z},T^{1/2}}dP + o_P(1), \end{aligned} \tag{4.21}$$

uniformly in $||\mathbf{z}|| \leqslant K$, and similarly, the fourth term in (4.17) can be written as

$$\begin{aligned} \sum_{i=1}^T \mathbf{y}_\alpha^t \mathbf{W}_i 1_{\{0 \geqslant \mathbf{y}_\alpha^t \mathbf{W}_i \geqslant T^{-1/2}\mathbf{z}^t\mathbf{W}_i\}} &= \nu_T(\bar{h}_{T^{-1/2}\mathbf{z},T^{1/2}}) + T^{1/2}\int \bar{h}_{T^{-1/2}\mathbf{z},T^{1/2}}dP \\ &= T^{1/2}\int \bar{h}_{T^{-1/2}\mathbf{z},T^{1/2}}dP + o_P(1), \end{aligned} \tag{4.22}$$

uniformly in $||\mathbf{z}|| \leqslant K$ where

$$\bar{h}_{T^{-1/2}\mathbf{z},T^{1/2}}(\mathbf{w}) = T^{1/2}(\mathbf{y}_\alpha^t \mathbf{w})1_{\{0 \geqslant \mathbf{y}_\alpha^t \mathbf{w} \geqslant T^{-1/2}\mathbf{z}^t\mathbf{w}\}}.$$

Using (4.19), (4.20), (4.21) and (4.22), we can write $l_T(\mathbf{z})$ as

$$\begin{aligned} l_T(\mathbf{z}) &= \mathbf{z}^t[(1-\alpha)\mathbf{U}_{1,T} - \alpha\mathbf{U}_{2,T}] \\ &\quad +(1-\alpha)T^{1/2}\int (g_{\mathbf{z},T^{-1/2}\mathbf{z}})dP + \alpha T^{1/2}\int (\bar{g}_{\mathbf{z},T^{-1/2}\mathbf{z}})dP \\ &\quad -T^{1/2}\int h_{T^{-1/2}\mathbf{z},T^{1/2}}dP + T^{1/2}\int \bar{h}_{T^{-1/2}\mathbf{z},T^{1/2}}dP + o_P(1), \end{aligned}$$

uniformly in $||\mathbf{z}|| \leqslant K$. But notice that for $\mathbf{z} \in \mathcal{Z}$

$$\begin{aligned}
& (1-\alpha)T^{1/2}\int (g_{\mathbf{z},T^{-1/2}\mathbf{z}})dP + \alpha T^{1/2}\int (\bar{g}_{\mathbf{z},T^{-1/2}\mathbf{z}})dP \\
& +T^{1/2}\int h_{T^{-1/2}\mathbf{z},T^{1/2}}dP + T^{1/2}\int \bar{h}_{T^{-1/2}\mathbf{z},T^{1/2}}dP \\
= & \int T^{1/2}\mathbf{z}^t\mathbf{w}\left[(1-\alpha)1_{\{\mathbf{y}_\alpha^t\mathbf{w}\leqslant T^{-1/2}\mathbf{z}^t\mathbf{w}\}} - \alpha 1_{\{\mathbf{y}_\alpha^t\mathbf{w}>T^{-1/2}\mathbf{z}^t\mathbf{w}\}}\right]dP \\
& + \int T\mathbf{y}_\alpha^t\mathbf{w}[-1_{\{0\leqslant\mathbf{y}_\alpha^t\mathbf{w}\leqslant T^{-1/2}\mathbf{z}^t\mathbf{w}\}} + 1_{\{0\geqslant\mathbf{y}_\alpha^t\mathbf{w}\geqslant T^{-1/2}\mathbf{z}^t\mathbf{w}\}}]dP \\
= & \alpha TE[\rho_\alpha((\mathbf{y}_\alpha^t - T^{-1/2}\mathbf{z})^t\mathbf{W}) - \rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W})] \\
= & \alpha T\left[\frac{1}{2}(T^{-1/2}\mathbf{z})^t[\nabla_{\mathbf{y}}^2 E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha^t}](T^{-1/2}\mathbf{z}) + o(||T^{-1/2}\mathbf{z}||^2)\right] \text{ by Lemma 16,} \\
= & \frac{1}{2}\mathbf{z}^t[\alpha\nabla_{\mathbf{y}}^2 E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha^t}]\mathbf{z} + o(1),
\end{aligned}$$

uniformly over $||\mathbf{z}|| \leqslant K$. Therefore,

$$l_T(\mathbf{z}) = \mathbf{z}^t[(1-\alpha)\mathbf{U}_{1,T} - \alpha\mathbf{U}_{2,T}] + \frac{1}{2}\mathbf{z}^t[E(\mathbf{W}\mathbf{W}^t|\mathbf{y}_\alpha^t\mathbf{W}=0)]\mathbf{z}.f_{\mathbf{y}_\alpha^t\mathbf{W}}(0)+o_P(1),$$

uniformly over $||\mathbf{z}|| \leqslant K$, where we have used the expression for $\nabla_{\mathbf{y}}^2 E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]$ found in Lemma 15. The result follows by defining $\mathbf{U}_T = (1-\alpha)\mathbf{U}_{1,T} - \alpha\mathbf{U}_{2,T}$. ■

We now use the preceding lemma to get the asymptotic distribution of the shortfall portfolio estimator and its associated sample quantile.

Theorem 22 *Under assumption (B),*

$$\sqrt{T}(\hat{\mathbf{y}}_\alpha - \mathbf{y}_\alpha) \rightsquigarrow N(\mathbf{0}, L_\alpha^* Q_\alpha^* L_\alpha^*),$$

where

$$Q_\alpha^* = Cov\left[(1-\alpha)\mathbf{W}1_{\{\mathbf{y}_\alpha^t\mathbf{W}\leqslant 0\}} - \alpha\mathbf{W}1_{\{\mathbf{y}_\alpha^t\mathbf{W}>0\}})\right],$$

and where L^*_α be the matrix consisting of the upper-left $(N+1)$-dimensional corner of M_α^{-1}, where

$$M_\alpha = \begin{bmatrix} E(\mathbf{W}\mathbf{W}^t|\mathbf{y}_\alpha^t\mathbf{W}=0).f_{\mathbf{y}_\alpha^t\mathbf{W}}(0) & \mathbf{A}_0^t \\ \mathbf{A}_0 & \mathbf{0}_N \end{bmatrix},$$

which is invertible.

Proof. Let $\Sigma_\alpha = E(\mathbf{W}\mathbf{W}^t|\mathbf{y}_\alpha^t\mathbf{W}=0).f_{\mathbf{y}_\alpha^t\mathbf{W}}(0)$ and $\mathbf{U}_T \rightsquigarrow N(\mathbf{0}, Q^*_\alpha)$ as in the Lemma 21, and consider the minimizer of

$$\Phi_T(\mathbf{z}) = \mathbf{z}^t\mathbf{U}_T + \frac{1}{2}\mathbf{z}^t\Sigma_\alpha\mathbf{z},$$

over $\mathbf{z} \in \mathcal{Z}$. Writing the Lagrangean of the last expression and taking its derivative, we see that the minimizer of $\Phi_T(\mathbf{z})$ must satisfy the following set of linear equations:

$$\begin{cases} \mathbf{U}_T + \Sigma_\alpha\mathbf{z} \quad + \mathbf{A}_0^t\boldsymbol{\lambda} & = \mathbf{0} \\ \mathbf{A}_0\mathbf{z} & = \mathbf{0}, \end{cases}$$

or

$$M_\alpha \begin{bmatrix} \mathbf{y} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\mathbf{U}_T \\ \mathbf{0}_M \end{bmatrix},$$

where $\boldsymbol{\lambda} \in \mathbb{R}^M$, and where

$$M_\alpha = \begin{bmatrix} \Sigma_\alpha & \mathbf{A}_0^t \\ \mathbf{A}_0 & \mathbf{0} \end{bmatrix},$$

which is invertible (see the remark after Lemma 16). Write the inverse of M_α as the partitioned matrix

$$M_\alpha^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^t & B_{22} \end{bmatrix}$$

Let V_α be the matrix consisting of the upper-left $(N+1)$-dimensional corner of M_α^{-1}. Then, the minimizer of $\Phi_T(\mathbf{z})$ over $\mathbf{z} \in \mathcal{Z}$ is

$$\vec{\mathbf{z}} = -B_{11}\mathbf{U}_T.$$

We next show that $\hat{\mathbf{z}}$, the minimizer of $l_T(\mathbf{z})$, converges to $\vec{\mathbf{z}}$ in probability.

Let $\delta > 0$, and let $B_\delta = B(\vec{\mathbf{z}}, \delta)$ denote the closed ball in $\mathbb{R}^{N+1}$ with center $\vec{\mathbf{z}}$ and radius δ. Write

$$l_T(\mathbf{z}) = \mathbf{z}^t\mathbf{U}_T + \frac{1}{2}\mathbf{z}^t\Sigma_\alpha\mathbf{z} + r(\mathbf{z}).$$

Because $\mathbf{U}_T$ converges in distribution, $\mathbf{U}_T$ is stochastically bounded, and we can choose a compact set K such that $B_\delta \subset K$ with probability arbitrarily close to one. Over this set K, by Lemma 21, we have

$$\triangle_T = \sup_{\mathbf{z}\in\mathcal{Z}\cap K} |r(\mathbf{z})| \to 0 \text{ in probability.}$$

Consider any $\mathbf{z} \in \mathcal{Z} \cap B_\delta^c$, where B_δ^c is the complement of B_δ in $\mathbb{R}^{N+1}$, and suppose $\mathbf{z} = \vec{\mathbf{z}}+\beta\mathbf{v}$ with $\mathbf{v} \in\mathbb{R}^{N+1}$ and $||\mathbf{v}|| = 1$. Let $\mathbf{z}_\varepsilon = \vec{\mathbf{z}}+\delta\mathbf{v}$, and notice that $\mathbf{z}_\varepsilon$ is on the boundary ∂B_δ of B_δ. By the convexity of $l_T(.)$ we have

$$\frac{\delta}{\beta}l_T(\mathbf{z}) + (1-\frac{\delta}{\beta})l_T(\vec{\mathbf{z}}) \geqslant l_T(\mathbf{z}_\varepsilon),$$

so that

$$l_T(\mathbf{z}) - l_T(\vec{\mathbf{z}}) \geqslant \frac{\beta}{\delta}[l_T(\mathbf{z}_\varepsilon) - l_T(\vec{\mathbf{z}})].$$

Using the definitions of $l_T(.)$ and $\mathbf{z}_\varepsilon$ we have

$$\begin{aligned} l_T(\mathbf{z}_\varepsilon) - l_T(\vec{\mathbf{z}}) &\geqslant \mathbf{z}_\varepsilon^t\mathbf{U}_T + \frac{1}{2}\mathbf{z}_\varepsilon^t\Sigma_\alpha\mathbf{z}_\varepsilon - \left[(\vec{\mathbf{z}})^t\mathbf{U}_T + \frac{1}{2}(\vec{\mathbf{z}})^t\Sigma_\alpha\vec{\mathbf{z}}\right] - 2\triangle_T \\ &= \delta\mathbf{v}^t\mathbf{U}_T + \delta\mathbf{v}^t\Sigma_\alpha\vec{\mathbf{z}} + \frac{1}{2}\delta^2\mathbf{v}^t\Sigma_\alpha\mathbf{v} - 2\triangle_T \\ &= \delta\mathbf{v}^t\mathbf{U}_T + \delta\mathbf{v}^t\Sigma_\alpha(-B_{11}\mathbf{U}_T)+\frac{1}{2}\delta^2\mathbf{v}^t\Sigma_\alpha\mathbf{v} - 2\triangle_T, \end{aligned} \tag{4.23}$$

where the last line follows from the definition of $\vec{\mathbf{z}}$. Using the expressions for M_α and its inverse, we can write

$$\begin{aligned} M_\alpha M_\alpha^{-1} &= \begin{bmatrix} \Sigma_\alpha & \mathbf{A}_0^t \\ \mathbf{A}_0 & \mathbf{0} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^t & B_{22} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_\alpha B_{11} + \mathbf{A}_0^t B_{12}^t & \Sigma_\alpha B_{12} + \mathbf{A}_0^t B_{22} \\ \mathbf{A}_0 B_{11} & \mathbf{A}_0 B_{12} \end{bmatrix} \\ &= \mathbf{I}. \end{aligned}$$

It follows that $\mathbf{v}^t = \mathbf{v}^t(\Sigma_\alpha B_{11} + \mathbf{A}_0^t B_{12}^t) = \mathbf{v}^t(\Sigma_\alpha B_{11})$ since $\mathbf{v} \in \mathcal{Z}$, by definition of $\mathcal{Z}$. Using this last fact in (4.23) yields

$$\begin{aligned} l_T(\mathbf{z}_\epsilon) - l_T(\vec{\mathbf{z}}) &\geqslant \delta \mathbf{v}^t \mathbf{U}_T - \delta \mathbf{v}^t \mathbf{U}_T + \frac{1}{2}\delta^2 \mathbf{v}^t \Sigma_\alpha \mathbf{v} - 2\triangle_T \\ &= \frac{1}{2}\delta^2 \mathbf{v}^t \Sigma_\alpha \mathbf{v} - 2\triangle_T \\ &\geqslant \delta^2 \gamma ||\mathbf{v}||^2 - 2\triangle_T \\ &= \delta^2\gamma - 2\triangle_T, \end{aligned}$$

where $\gamma > 0$ by Lemma 16. The last expression does not depend on $\mathbf{z}$. Now

$$\inf_{\mathbf{z} \in \mathcal{Z} \cap B_\delta^c} [l_T(\mathbf{z}) - l_T(\vec{\mathbf{z}})] \geqslant \frac{\beta}{\delta}[\delta^2\gamma - 2\triangle_T].$$

Since $\delta^2\gamma - 2\triangle_T > 0$ with probability approaching one, the minimum of $l_T(.)$ cannot occur outside of B_δ. This shows that the event $||\hat{\mathbf{z}} - \vec{\mathbf{z}}|| > \boldsymbol{\delta}$, where $\hat{\mathbf{z}}$ is the minimum of $l_T(.)$, has a probability that converges to 0. Since $\mathbf{U}_T \rightsquigarrow N(\mathbf{0}, Q_\alpha^*)$ and $\vec{\mathbf{z}} = -B_{11}\mathbf{U}_T$ we know that

$$\hat{\mathbf{z}} \rightsquigarrow N(\mathbf{0}, B_{11} Q_\alpha^* B_{11}).$$

Finally, note that $\hat{\mathbf{z}} = \sqrt{T}(\hat{\mathbf{y}}_\alpha - \mathbf{y}_\alpha)$, by definition of $l_T(.)$, which yields the desired convergence in distribution of $\hat{\mathbf{y}}_\alpha$, with $L_\alpha^* = B_{11}$. ∎

Focusing on $\hat{\mathbf{x}}_\alpha$ instead of $\hat{\mathbf{y}}_\alpha$ (remember that $\hat{\mathbf{y}}_\alpha = (\hat{\mathbf{x}}_\alpha^t, \hat{q}_\alpha)^t$) , we get the following corollary.

Corollary 23 *Under assumption (B),*

$$\sqrt{T}(\hat{\mathbf{x}}_\alpha - \mathbf{x}_\alpha) \rightsquigarrow N(\mathbf{0}, L_\alpha Q_\alpha L_\alpha),$$

where

$$Q_\alpha = Cov\left[(\mathbf{R}-\boldsymbol{\tau})\left(1_{\{\mathbf{x}_\alpha^t\mathbf{R}\leqslant q_\alpha\}} - \alpha\right)\right],$$

and where

$$L_\alpha = \frac{1}{f_{\mathbf{x}_\alpha^t\mathbf{R}}(q_\alpha)}\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\tau}\boldsymbol{\tau}^t\Omega\right),$$

with

$$\begin{aligned}\Omega &= \left[\Gamma^{-1} - \Gamma^{-1}\mathbf{A}^t\left(\mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Gamma^{-1}\right],\\ \delta &= \boldsymbol{\tau}^t\Gamma^{-1}\mathbf{A}^t\left(\mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Gamma^{-1}\boldsymbol{\tau},\\ \boldsymbol{\tau} &= E(\mathbf{R}|\mathbf{x}_\alpha^t\mathbf{R} = q_\alpha),\ \textit{and}\\ \Gamma &= E(\mathbf{R}\mathbf{R}^t|\mathbf{x}_\alpha^t\mathbf{R} = q_\alpha).\end{aligned}$$

Proof. We know from the previous theorem that

$$\sqrt{T}(\hat{\mathbf{y}}_\alpha - \mathbf{y}_\alpha) \rightsquigarrow N(\mathbf{0}, L_\alpha^* Q_\alpha^* L_\alpha^*)$$

where

$$Q_\alpha^* = Cov\left((1-\alpha)\mathbf{W}1_{\{\mathbf{y}_\alpha^t\mathbf{W}\leqslant 0\}} - \alpha\mathbf{W}1_{\{\mathbf{y}_\alpha^t\mathbf{W}>0\}}\right)$$

and where L_α^* is the upper-left $(N+1)$-dimensional corner of M_α^{-1}. Write

$$L_\alpha^* = \begin{bmatrix} L_\alpha & \mathbf{f} \\ \mathbf{f}^t & d \end{bmatrix}$$

where L_α is the upper-left N-dimensional corner of M_α^{-1}, and where $\mathbf{f} \in \mathbb{R}^N$, $d \in \mathbb{R}$. Let $\Gamma = E(\mathbf{R}\mathbf{R}^t \mid \mathbf{x}_\alpha^t\mathbf{R} = q_\alpha)$ and $\boldsymbol{\tau} = E(\mathbf{R}|\mathbf{x}_\alpha^t\mathbf{R} = q_\alpha)$. M_α can be rewritten as

$$M_\alpha = f_{\mathbf{x}_\alpha^t\mathbf{R}}(q_\alpha) \begin{bmatrix} \Gamma & -\tau & \frac{1}{f_{\mathbf{x}_\alpha^t\mathbf{R}}(q_\alpha)}\mathbf{A}^t \\ -\tau^t & 1 & \mathbf{0} \\ \frac{1}{f_{\mathbf{x}_\alpha\mathbf{R}}(q_\alpha)}\mathbf{A} & \mathbf{0} & \mathbf{0}_M \end{bmatrix}.$$

Since (i) Γ is positive definite, by virtue of $Cov(\mathbf{R}|\mathbf{x}_\alpha^t\mathbf{R} = q_\alpha)$ having rank $N-1$, and (ii) M_α is invertible from the remark after Lemma 16, we can get a closed form expression for M_α^{-1} from Proposition 51, and in particular we can get closed form expressions for L_α and $\mathbf{f}$ as follows (noticing that $Cov(\mathbf{R} \mid \mathbf{x}_\alpha^t\mathbf{R} = q_\alpha)$ is singular):

$$\begin{aligned} L_\alpha &= \frac{1}{f_{\mathbf{x}_\alpha^t\mathbf{R}}(q_\alpha)}\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\tau}\boldsymbol{\tau}^t\Omega\right) && (4.24) \\ \mathbf{f} &= \frac{1}{f_{\mathbf{x}_\alpha^t\mathbf{R}}(q_\alpha)}\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\tau}\boldsymbol{\tau}^t\Omega\right) \\ &= L_\alpha\boldsymbol{\tau} && (4.25) \end{aligned}$$

where

$$\Omega = \left[\Gamma^{-1} - \Gamma^{-1}\mathbf{A}^t\left(\mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Gamma^{-1}\right]$$

and

$$\delta = \boldsymbol{\tau}^t\Gamma^{-1}\mathbf{A}^t\left(\mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Gamma^{-1}\boldsymbol{\tau}.$$

We also partition Q_α^* as follows. Notice that, by definition of $\mathbf{W}$,

$$\begin{aligned} (1-\alpha)\mathbf{W}1_{\{\mathbf{y}_\alpha^t\mathbf{W}\leqslant 0\}} - \alpha\mathbf{W}1_{\{\mathbf{y}_\alpha^t\mathbf{W}>0\}} &= \begin{bmatrix} (1-\alpha)\mathbf{R}1_{\{\mathbf{x}_\alpha^t\mathbf{R}\leqslant q_\alpha\}} - \alpha\mathbf{R}1_{\{\mathbf{x}_\alpha^t\mathbf{R}>q_\alpha\}} \\ -(1-\alpha)1_{\{\mathbf{x}_\alpha^t\mathbf{R}\leqslant q_\alpha\}} + \alpha 1_{\{\mathbf{x}_\alpha^t\mathbf{R}>q_\alpha\}} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}(1_{\{\mathbf{x}_\alpha^t\mathbf{R}\leqslant q_\alpha\}} - \alpha) \\ -(1_{\{\mathbf{x}_\alpha^t\mathbf{R}\leqslant q_\alpha\}} - \alpha) \end{bmatrix}. \end{aligned}$$

Then

$$
\begin{aligned}
& Q^*_\alpha \\
= \; & Cov((1-\alpha)\mathbf{W}1_{\{\mathbf{y}^t_\alpha \mathbf{W}\leqslant 0\}} - \alpha \mathbf{W}1_{\{\mathbf{y}^t_\alpha \mathbf{W}>0\}}) \\
= \; & E\left\{ \begin{bmatrix} \mathbf{R}\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right) \\ -(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha) \end{bmatrix} \begin{bmatrix} \mathbf{R}^t\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right) & -(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha) \end{bmatrix} \right\} \\
& -E\begin{bmatrix} \mathbf{R}\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right) \\ -(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha) \end{bmatrix} E\begin{bmatrix} \mathbf{R}^t(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha) & -(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha) \end{bmatrix} \\
= \; & \begin{bmatrix} E[\mathbf{R}\mathbf{R}^t\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)^2] & E[-\mathbf{R}(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha)^2] \\ E[-\mathbf{R}^t\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)^2] & E[\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)^2] \end{bmatrix} \\
& - \begin{bmatrix} E\left[\mathbf{R}\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)\right] \\ 0 \end{bmatrix} \begin{bmatrix} E[\mathbf{R}^t\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)] & 0 \end{bmatrix} \\
= \; & \begin{bmatrix} \begin{array}{l} E[\mathbf{R}\mathbf{R}^t\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)^2] \\ \quad - E[\mathbf{R}\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)][\mathbf{R}^t\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)] \end{array} & E[-\mathbf{R}\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)^2] \\ E[-\mathbf{R}^t\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)^2] & E[\left(1_{\{\mathbf{x}^t_\alpha \mathbf{R}\leqslant q_\alpha\}} - \alpha\right)^2] \end{bmatrix} \\
= \; & \begin{bmatrix} Q_{11} & Q_{12} \\ Q^t_{12} & Q_{22} \end{bmatrix}.
\end{aligned}
$$

Notice that by definition of $\hat{\mathbf{y}}_\alpha$ we have

$$\hat{\mathbf{x}}_\alpha - \mathbf{x}_\alpha = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \end{bmatrix} (\hat{\mathbf{y}}_\alpha - \mathbf{y}_\alpha)$$

where here $\mathbf{0} \in \mathbb{R}^N$. It follows that $\sqrt{T}(\hat{\mathbf{x}}_\alpha - \mathbf{x}_\alpha)$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix

$$
\begin{aligned}
& \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \end{bmatrix} L^*_\alpha Q^*_\alpha L^*_\alpha \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}^t \end{bmatrix} \\
= \; & \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \end{bmatrix} \begin{bmatrix} L_\alpha & \mathbf{f} \\ \mathbf{f}^t & d \end{bmatrix} Q^*_V \begin{bmatrix} L_\alpha & \mathbf{f} \\ \mathbf{f}^t & d \end{bmatrix} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}^t \end{bmatrix} \\
= \; & \begin{bmatrix} L_\alpha & \mathbf{f} \end{bmatrix} Q^*_\alpha \begin{bmatrix} L_\alpha \\ \mathbf{f}^t \end{bmatrix} \\
= \; & \begin{bmatrix} L_\alpha & \mathbf{f} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^t & Q_{22} \end{bmatrix} \begin{bmatrix} L_\alpha \\ \mathbf{f}^t \end{bmatrix} \\
= \; & L_\alpha \begin{bmatrix} \mathbf{I} & \boldsymbol{\tau} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^t & Q_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \boldsymbol{\tau}^t \end{bmatrix} L_\alpha \\
= \; & L_\alpha \left[Q_{11} + \boldsymbol{\tau} Q_{12}^t + Q_{12} \boldsymbol{\tau}^t + \boldsymbol{\tau} Q_{22} \boldsymbol{\tau}^t \right] L_\alpha, \qquad (4.26)
\end{aligned}
$$

where the second to last equation follows from (4.24) and (4.25). Focusing on the middle term in (4.26) we write

$$
\begin{aligned}
& Q_{11} + \boldsymbol{\tau} Q_{12}^t + Q_{12} \boldsymbol{\tau}^t + \boldsymbol{\tau} Q_{22} \boldsymbol{\tau}^t \\
= \; & E[\mathbf{R}\mathbf{R}^t \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)^2] - E[\mathbf{R} \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)][\mathbf{R}^t \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)] \\
& \boldsymbol{\tau} E[-\mathbf{R}^t \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)^2] + E[-\mathbf{R} \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)^2] \boldsymbol{\tau}^t \\
& + \boldsymbol{\tau} E[\left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)^2] \boldsymbol{\tau}^t \\
= \; & E\left[(\mathbf{R} - \boldsymbol{\tau})(\mathbf{R} - \boldsymbol{\tau})^t \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)^2\right] \\
& - E\left[(\mathbf{R} - \boldsymbol{\tau}) \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] E\left[(\mathbf{R} - \boldsymbol{\tau})^t \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] \\
= \; & Cov\left[(\mathbf{R} - \boldsymbol{\tau}) \left(1_{\{\mathbf{x}_\alpha^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] := Q_\alpha
\end{aligned}
$$

Using this last expression in (4.26) yields the following simple expression for the asymptotic covariance of $\sqrt{T}(\hat{\mathbf{x}}_\alpha - \mathbf{x}_\alpha)$:

$$L_\alpha Q_\alpha L_\alpha.$$

■

Note that the expression for the asymptotic covariance of the shortfall estimator contains the term

$$\left(\frac{1}{f_{\mathbf{x}_\alpha^t \mathbf{R}}(q_\alpha)}\right)^2,$$

which suggests that for low values of α, for example $\alpha = 1\%$, the portfolio estimator may be badly estimated, compared to $\alpha = 50\%$, since we would expect the density $f_{\mathbf{x}_\alpha^t \mathbf{R}}(q_\alpha)$ to be relatively small in the tails. Intuitively, this can be explained by noticing that for low values of α, we are relying on rare events to build a portfolio. This may not be a good idea unless there are many observations, meaning large T.

4.3 Computational Experiments

4.3.1 Simulated Data

In this experiment we highlight, on simulated data, the pitfalls of using small values of α in shortfall portfolio estimation. We repeat the following experiment 500 times, with $N = 10$, with T in the range 25 to 500, and for four values of α (50%, 10%, 5%, 1%):

1. generate a sample of size T from a multivariate Gaussian with mean 0 and covariance Σ equal to an equicorrelated matrix with off-diagonal elements equal to 0.5. Solve the problem

$$\begin{array}{ll} \text{minimize} & \hat{s}_\alpha(\mathbf{x}) \\ \text{subject to} & \mathbf{e}^t\mathbf{x} = 1, \end{array}$$

where $\hat{s}_\alpha(\mathbf{x})$ is the sample shortfall of portfolio $\mathbf{x}$. Call the result $\hat{\mathbf{x}}_\alpha$.

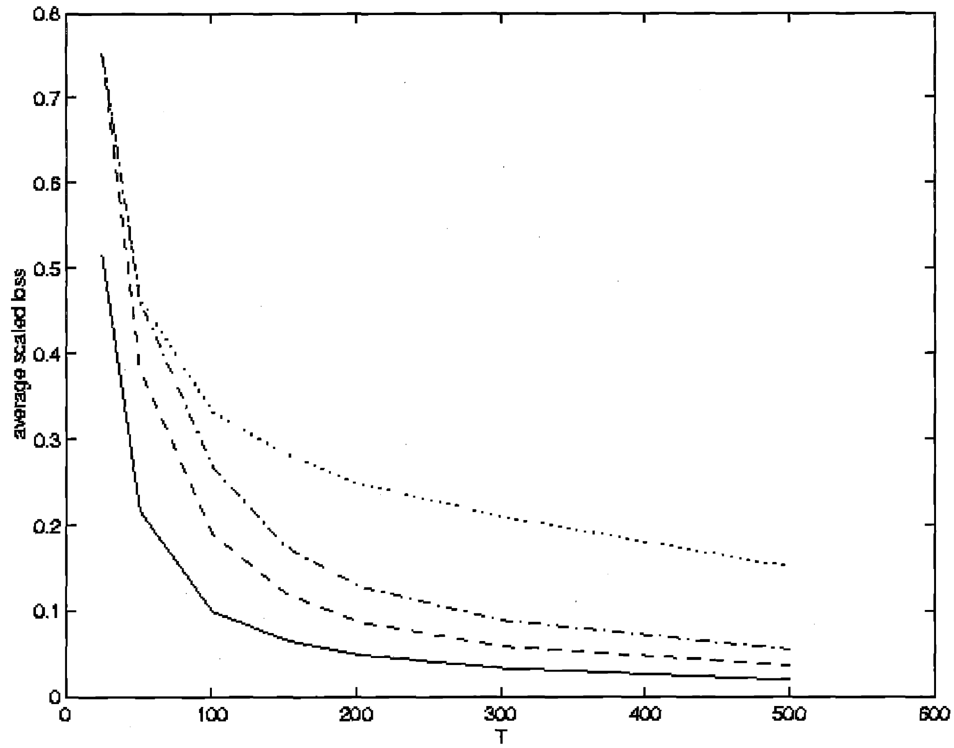

Figure 4-1: Average scaled loss over 500 Monte Carlo replications. $-$: α =50%; - -: α =10%; -. : α =5%; $\cdots$: α =1%.

2. calculate the **scaled loss** defined by

$$\begin{aligned} l &= \frac{s_\alpha(\hat{\mathbf{x}}_\alpha) - s_\alpha(\mathbf{x}_\alpha)}{s_\alpha(\mathbf{x}_\alpha)} \\ &= \frac{\sqrt{\hat{\mathbf{x}}_\alpha^t \Sigma \hat{\mathbf{x}}_\alpha} - \sqrt{\mathbf{x}_\alpha^t \Sigma \mathbf{x}_\alpha}}{\sqrt{\mathbf{x}_\alpha^t \Sigma \mathbf{x}_\alpha}}, \end{aligned}$$

where the last equation follows from Proposition 1 in Chapter 2. The scaled loss can be interpreted as the relative extra risk, in the form of shortfall, incurred by the estimation process.

Figure 4-1 shows the average scaled loss over 500 Monte Carlo replications, for various values of α and T. The message is clear. The estimation process deteriorates for small values of α, for all values of T.

4.3.2 Historical Data

In this experiment, we again show how shortfall portfolio estimation using small values of α may fail to deliver. We use the biotech data described in Chapter 5, which consists of four years of daily returns on 16 stocks. On average, the stocks are positively skewed, with an average coefficient of skewness of about 0.5. We run a backtest in which we use a rolling window of 5 months of return data, rebalancing every month, and estimating the sample variance portfolio V, and the sample shortfall portfolios S_α ($\alpha = 50\%, 25\%, 10\%$) using the constraints

$$\begin{aligned} \mathbf{x}^t\mathbf{e} &= 1 \\ \mathbf{x}^t\bar{\mathbf{R}}_{t-1} &= r_p, \end{aligned}$$

where r_p in the range 160%/260 (which corresponds to an annualized target return of 160%) to 0, where $\bar{\mathbf{R}}_{t-1}$ is the vector of mean returns calculated over the last five months. Five months was chosen arbitrarily, and corresponds to 100 trading days. Figure 4-2 shows the resulting minimum variance frontiers of the ex-post returns, for each estimator. The overall performance of the shortfall estimator peaks at $\alpha = 50\%$, and decreases for smaller values of α, with the risk-reward opportunities worsening for $\alpha = 25\%$ and then for 10%.

4.4 Conclusion

This chapter proved central limit theorems that showed that the shortfall portfolio estimator, as well as the variance estimator, converge towards their population counterparts at the usual $\sqrt{T}$ rate. We ended this chapter with a note of caution, as we argued that shortfall portfolios for low values of α may be more difficult to estimate than shortfall portfolios for moderate values of α, such as 50%. Chapter 5 uses the results from this chapter to prove that when return are generated by an elliptically symmetric distribution, the shortfall estimator may outperform the variance estimator.

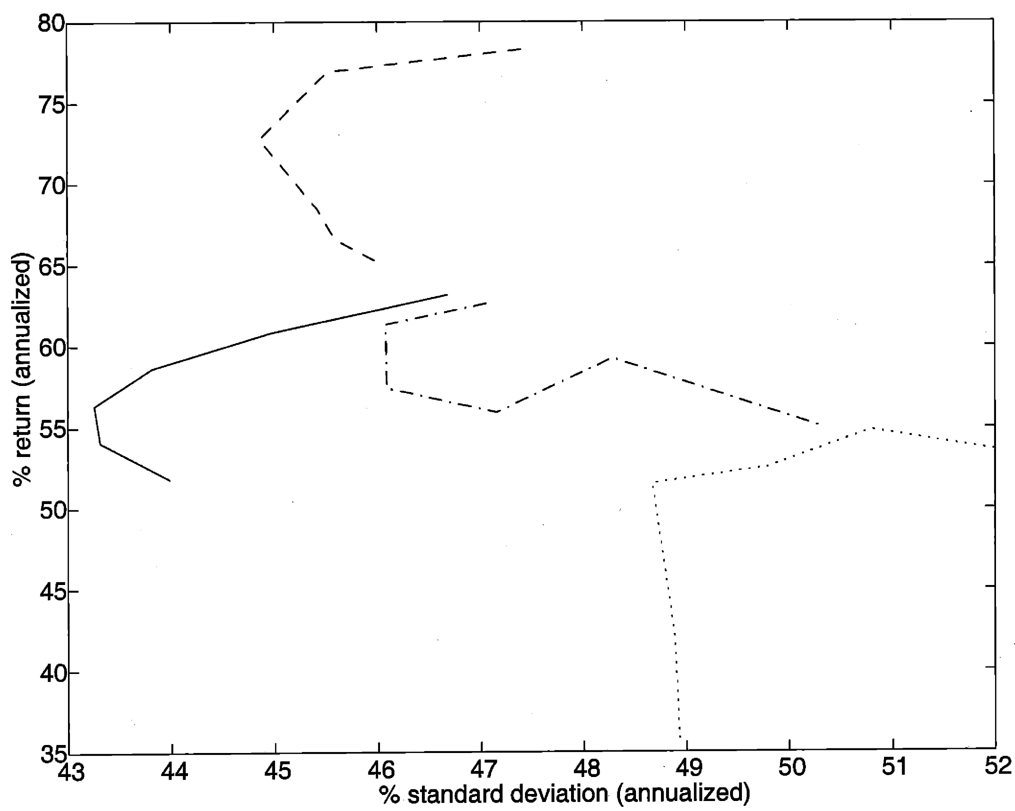

Figure 4-2: Minimum variance frontier of realized returns: – = V; - - = $S_{0.5}$; -. = $S_{0.25}$; ... = $S_{0.1}$.

Chapter 5

Robust Portfolio Estimation

Let $\mathbf{R}$ be a random return vector in $\mathbb{R}^N$ with mean $\boldsymbol{\mu}$, positive definite covariance matrix Σ, and a continuous density. Consider the problem of selecting a risk minimizing portfolio, subject to a given expected return constraint. Assuming that $\mathbf{R}$ has a multivariate Gaussian distribution, or more generally, a multivariate elliptically symmetric distribution, every wealth-seeking, risk-averse investor will prefer will prefer the portfolio that has minimal variance, i.e. every wealth-seeking, risk-averse investor will prefer the portfolio that solves

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^t \Sigma \mathbf{x} \\ \text{subject to} \quad & \mathbf{e}^t \mathbf{x} = 1 \\ & \boldsymbol{\mu}^t \mathbf{x} = r_p, \end{aligned}$$

where $r_p \in \mathbb{R}$ is the target return.

Now suppose that we do not know the exact distribution of $\mathbf{R}$, but that we are given independent and identically distributed realizations $\mathbf{R}_1, \ldots, \mathbf{R}_T$ of the random return vector $\mathbf{R}$. Suppose also that $\mathbf{R}$ has a multivariate elliptically symmetric, but not **necessarily Gaussian**, distribution Then, given these realizations, every risk-averse investor wants to estimate the minimal variance portfolio defined generally as

$$\mathbf{x}_V = \arg \min_{\mathbf{A}\mathbf{x}=\mathbf{b}} \mathbf{x}^t \Sigma \mathbf{x}, \tag{5.1}$$

where the set $\mathbf{A}\mathbf{x} = \mathbf{b}$ corresponds to the constraints $\mathbf{e}^t \mathbf{x} = 1$ and $\boldsymbol{\mu}^t \mathbf{x} = r_p$. A natural estimator of $\mathbf{x}_V$

is

$$\hat{\mathbf{x}}_V = \arg\min_{\mathbf{Ax}=\mathbf{b}} \mathbf{x}^t \hat{\Sigma} \mathbf{x}, \tag{5.2}$$

where $\hat{\Sigma}$ is the sample covariance matrix. We saw in Chapter 3 that we can rewrite (5.2) as

$$(\hat{\mathbf{x}}_V, \hat{q}_V) = \arg\min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q\in\mathbb{R}}} \frac{1}{T}\sum_{i=1}^{T}(\mathbf{x}^t\mathbf{R}_i - q)^2, \tag{5.3}$$

where $\hat{q}_V$ is the sample mean of portfolio $\hat{\mathbf{x}}_V$. When returns are multivariate Gaussian, the term $\mathbf{x}^t\mathbf{R} - q$ is Gaussian, and $(\hat{\mathbf{x}}_V, \hat{q}_V)$ may be interpreted as maximum likelihood estimators of the weight $\mathbf{x}$ and the location parameter q. Therefore, under normality, we expect $(\hat{\mathbf{x}}_V, \hat{q}_V)$ to be asymptotically efficient, a desirable property.

However, even if one is willing to assume that the distribution of returns belongs to the multivariate elliptically symmetric class, there is empirical evidence that returns may perhaps not be satisfactorily modelled as Gaussian. The following two departures from normality, inconsistent with the Gaussian assumption, have been documented. The first is heavy tails, meaning that marginal return distributions tend to have tails that decay more slowly than the Gaussian - see for example Campbell, Lo and MacKinlay (1997) and Bouchaud and Potters (2000) for a discussion of heavy-tailed distributions in finance, and evidence that stock returns sampled daily, or at higher frequencies, exhibit heavy tails.

The other documented departure from normality is tail dependence, which reflects the observation that the extreme return of one stock is likely to be accompanied by extreme returns in other stocks, for example, in the context of a market crash or of a market surge - see Embrechts, McNeil, and Straumann (1999) for a discussion of tail-dependence and its applications in risk management, and Lindskog (2000) for evidence that stock returns may have more tail dependence than the Gaussian. The Gaussian has zero tail dependence, so that extreme events occur independently; but other elliptically symmetric distributions, such as the multivariate Student-t, may have positive tail dependence. The practical implication of tail-dependence is that portfolios of assets with heavy tails also have heavy tails.

In this chapter, we argue that departures from normality in the form heavy-tailedness and tail dependence will degrade the performance of the variance portfolio estimator (5.3). We introduce a family of alternative "robust" portfolio estimators whose performance degrades less quickly than the variance portfolio estimator under departures from normality, at the price of lower performance under the Gaussian. The term "performance' is made precise in Section 1. Note that the multivariate Student-t distribution plays a critical role in our analysis, because it exhibits both marginal heavy tails, as well as tail dependence, and also because it is very tractable from both a computational and a calculation point of view.

The robust portfolio estimators which we introduce are listed below:

1. the least absolute deviation (LAD) portfolio estimator;

2. the Huber portfolio estimator;

3. the trimean portfolio estimator;

4. the trimmed mean portfolio estimator.

The LAD portfolio estimator is defined as

$$(\hat{\mathbf{x}}_{0.5}, \hat{q}_{0.5}) = \arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q \in \mathbb{R}}} \frac{1}{T} \sum_{i=1}^{T} |\mathbf{x}^t \mathbf{R}_i - q|, \tag{5.4}$$

where $\hat{q}_{0.5}$ is the sample median of portfolio $\hat{\mathbf{x}}_{0.5}$, and where $|.|$ stands for the absolute value. Note that the LAD portfolio estimator is the sample 50%-shortfall portfolio estimator, defined in Chapter 2. The trimean and trimmed mean portfolio estimators are based on underlying shortfall portfolio optimizations, as will be seen below. The Huber portfolio estimator can be seen as a hybrid between the variance and 50%-shortfall portfolio estimator. These "robust" portfolio estimators are defined in direct analogy to their counterparts in the location and regression problems in the robust statistics literature - see for example Huber (1981), and Basset and Koenker (1978) for an excellent introduction to the topic of robustness. We will see that their implementation turns out to be computationally efficient.

This chapter is organized as follows. In Section 1, we define a measure of portfolio estimator performance which we call estimation risk, and which we claim has an intuitive financial interpretation. We also introduce asymptotic estimation risk, as an approximation to estimation risk when the number of observations T is large. Then in Section 2 we evaluate the asymptotic estimation risk of $\hat{\mathbf{x}}_V$ and $\hat{\mathbf{x}}_{0.5}$ under the assumption that returns are multivariate normal, or Gaussian, and we show that $\hat{\mathbf{x}}_V$ has lower asymptotic estimation risk than $\hat{\mathbf{x}}_{0.5}$. However, under the assumption that returns have a multivariate Student-t distribution with degrees of freedom equal to or less than six, we show in Section 3 that $\hat{\mathbf{x}}_V$ has higher asymptotic estimation risk than $\hat{\mathbf{x}}_{0.5}$. The fact that $\hat{\mathbf{x}}_{0.5}$ may outperform $\hat{\mathbf{x}}_V$ under certain departures from normality with the class of elliptically symmetric distributions leads us to consider, in Section 4, other robust portfolio estimators. In Section 5 we empirically investigate the performance of $\hat{\mathbf{x}}_V$ and our robust portfolio estimators, on simulated and historical return data.

5.1 Estimation Risk: a Measure of Portfolio Estimator Performance

As argued in the introduction to this chapter, under the assumption that returns have a multivariate elliptically symmetric distribution, each and every wealth-seeking, risk-averse investor wants to estimate $\mathbf{x}_V$, defined by (5.1). Let $\hat{\mathbf{x}}$ be an arbitrary portfolio estimator satisfying the given deterministic constraints $\mathbf{Ax} = \mathbf{b}$. We will measure the performance of estimator $\hat{\mathbf{x}}$ with a measure which we will call estimation risk, defined below.

Definition 24 *The estimation risk of portfolio estimator $\hat{\mathbf{x}}$, satisfying $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$, is*

$$L_{\hat{\mathbf{x}}} = E(\hat{\mathbf{x}}^t \Sigma \hat{\mathbf{x}}) - \mathbf{x}_V^t \Sigma \mathbf{x}_V. \tag{5.5}$$

In (6.5), the expectation is taken with respect to the sample $\mathbf{R}_1, \ldots, \mathbf{R}_T$, $\hat{\mathbf{x}}$ being a function of the sample. In words, $L_{\hat{\mathbf{x}}}$ measures how much extra risk, in the form of variance, we incur with respect to the population optimal solution $\mathbf{x}_V$, because we do not know the distribution of $\mathbf{R}$, and we estimate $\mathbf{x}_V$ by $\hat{\mathbf{x}}$. Notice that

$$\begin{aligned} L_{\hat{\mathbf{x}}} &= E[(\hat{\mathbf{x}} - \mathbf{x}_V + \mathbf{x}_V)^t \Sigma (\hat{\mathbf{x}} - \mathbf{x}_V + \mathbf{x}_V)] - \mathbf{x}_V^t \Sigma \mathbf{x}_V \\ &= E[(\hat{\mathbf{x}} - \mathbf{x}_V)^t \Sigma (\hat{\mathbf{x}} - \mathbf{x}_V)] - 2\mathbf{x}_V{}^t \Sigma E[(\hat{\mathbf{x}} - \mathbf{x}_V)] + \mathbf{x}_V^t \Sigma \mathbf{x}_V - \mathbf{x}_V^t \Sigma \mathbf{x}_V \\ &= E[tr\left(\Sigma(\hat{\mathbf{x}} - \mathbf{x}_V)[(\hat{\mathbf{x}} - \mathbf{x}_V)^t\right)] - 2\mathbf{x}_V{}^t \Sigma E[(\hat{\mathbf{x}} - \mathbf{x}_V)], \end{aligned}$$

where $tr\,(.)$ stands for the trace. We expect that, asymptotically at least, we can approximate the estimation risk $L_{\hat{\mathbf{x}}}$ by a measure which we call asymptotic estimation risk, defined below.

Definition 25 *Assume that $\hat{\mathbf{x}}$ is an asymptotically unbiased estimator of $\mathbf{x}_V$. The asymptotic estima-*

tion risk of portfolio estimator $\hat{\mathbf{x}}$*,*

$$\bar{L}_{\hat{\mathbf{x}}} = E[tr\,(\Sigma Cov(\hat{\mathbf{x}}))],$$

where $Cov(\hat{\mathbf{x}})$ *is the asymptotic covariance of* $\hat{\mathbf{x}}$.

For example, in Chapter 4, we showed that both $\hat{\mathbf{x}}_V$ and $\hat{\mathbf{x}}_{0.5}$ are asymptotically normal with mean $\mathbf{x}_V$ and $\mathbf{x}_{0.5}$ respectively, and asymptotic covariances $Cov(\hat{\mathbf{x}}_V)$ and $Cov(\hat{\mathbf{x}}_{0.5})$ respectively - we used $\mathbf{x}_{0.5}$ to refer to the portfolio that, along with its median $q_{0.5}$, solves

$$(\mathbf{x}_{0.5}, q_{0.5}) = \arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q \in \mathbb{R}}} E[|\mathbf{x}^t\mathbf{R} - q|],$$

Since $\mathbf{x}_{0.5} = \mathbf{x}_V$ from Proposition 1 in Chapter 2, we see that both $\hat{\mathbf{x}}_V$ and $\hat{\mathbf{x}}_{0.5}$ are estimating the same portfolio $\mathbf{x}_V$. Now let $\bar{L}_V$ denote the asymptotic estimation risk of $\hat{\mathbf{x}}_V$, and let $\bar{L}_{0.5}$ denote the asymptotic estimation risk of $\hat{\mathbf{x}}_{0.5}$. In the sections that follow, we will show that under the assumption of that $\mathbf{R}$ has a multivariate Gaussian distribution, $\bar{L}_V < \bar{L}_{0.5}$. However, when $\mathbf{R}$ has a multivariate Student t-distribution with 6 degrees of freedom on less, $\bar{L}_{0.5} < \bar{L}_V$. This means that in certain heavy-tailed and tail-dependent conditions, $\hat{\mathbf{x}}_{0.5}$ may outperform $\hat{\mathbf{x}}_V$, at least asymptotically. This will also be our motivation for considering other robust portfolio estimators.

5.2 Asymptotic Estimation Risk Under the Multivariate Gaussian Distribution

We calculate closed form expressions for the asymptotic estimation risk of $\hat{\mathbf{x}}_V$ and $\hat{\mathbf{x}}_{0.5}$ under the assumption that $\mathbf{R}$ has a multivariate Gaussian distribution. Below, we will use the fact that

$$\mathbf{x}_V = \arg \min_{\mathbf{Ax}=\mathbf{b}} \mathbf{x}^t\Sigma\mathbf{x}.$$

can be expressed in closed form as

$$\mathbf{x}_V = \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}, \tag{5.6}$$

where the last expression follows immediately by writing the Lagrangean of

$$\begin{array}{ll} \text{minimize} & \mathbf{x}^t\Sigma\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array}$$

We also collect the following two corollaries from Chapter 4, for easier reference.

Corollary 26 *Suppose that* $\mathbf{R}$*, with mean* $\boldsymbol{\mu}$ *and covariance matrix* Σ*, has a continuous density, and suppose that* $\mathbf{R}$ *has finite fourth moments. Then*

$$\sqrt{T}(\hat{\mathbf{x}}_V - \mathbf{x}_V) \rightsquigarrow N(\mathbf{0}, L_V Q_V L_V),$$

where

$$\begin{aligned} Q_V &= Cov[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}^{\mathbf{t}}\mathbf{x_V} - q_V)] \\ &= Cov[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^t\mathbf{x}_V], \end{aligned}$$

and

$$L_V = \Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}.$$

And as we mentioned in the introduction to this chapter, the least absolute deviation portfolio estimator is the 50%-shortfall estimator, so we will use the following result concerning the asymptotic distribution of shortfall portfolio estimators.

Corollary 27 *Suppose that* $\mathbf{R}$*, with mean* $\boldsymbol{\mu}$ *and covariance matrix* Σ*, has a continuous density, and for every* $\mathbf{x} \in \mathbb{R}^N$ *and* $q \in \mathbb{R}$ *such that* $f_{\mathbf{x}^t\mathbf{R}}(q) > 0$*, the density* $f_{\mathbf{R},\mathbf{x}^t\mathbf{R}}(\mathbf{r}, q)$ *is well-defined,* $Cov[\mathbf{R} \mid \mathbf{x}^t\mathbf{R} - q = 0]$ *is well defined, and the rank of* $Cov[\mathbf{R} \mid \mathbf{x}^t\mathbf{R} - q = 0]$ *is* $N-1$*. Then*

$$\sqrt{T}(\hat{\mathbf{x}}_\alpha - \mathbf{x}_\alpha) \rightsquigarrow N(\mathbf{0}, L_\alpha Q_\alpha L_\alpha),$$

where

$$Q_\alpha = Cov\left[(\mathbf{R}-\boldsymbol{\tau})\left(1_{\{\mathbf{x}_\alpha^t\mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right],$$

and where

$$L_\alpha = \frac{1}{f_{\mathbf{x}_\alpha^t \mathbf{R}}(q_\alpha)} \left(\Omega + \frac{1}{\delta} \Omega \boldsymbol{\tau} \boldsymbol{\tau}^t \Omega \right),$$

with

$$\begin{aligned}
\Omega &= \left[\Gamma^{-1} - \Gamma^{-1} \mathbf{A}^t \left(\mathbf{A} \Gamma^{-1} \mathbf{A}^t \right)^{-1} \mathbf{A} \Gamma^{-1} \right], \\
\delta &= \boldsymbol{\tau}^t \Gamma^{-1} \mathbf{A}^t \left(\mathbf{A} \Gamma^{-1} \mathbf{A}^t \right)^{-1} \mathbf{A} \Gamma^{-1} \boldsymbol{\tau}, \\
\boldsymbol{\tau} &= E(\mathbf{R} | \mathbf{x}_\alpha^t \mathbf{R} = q_\alpha), \ \ \textit{and} \\
\Gamma &= E(\mathbf{R}\mathbf{R}^t | \mathbf{x}_\alpha^t \mathbf{R} = q_\alpha).
\end{aligned}$$

We use these corollaries in the following propositions.

5.2.1 Variance Portfolio Estimator

Theorem 28 *When* $\mathbf{R}$ *has a multivariate Gaussian distribution, the asymptotic risk of* $\hat{\mathbf{x}}_V$ *is*

$$\begin{aligned}
\bar{L}_V &= tr(\Sigma \frac{1}{T} L_V Q_V L_V) \\
&= \frac{1}{T} \left(\mathbf{x}_V^t \Sigma \mathbf{x}_V \right) [N - M].
\end{aligned}$$

where $\mathbf{x}_V^t \Sigma \mathbf{x}_V$ *is the variance of the optimal portfolio* $\mathbf{x}_V$ *.*

Proof. Let $\tilde{R} := (\mathbf{R} - \boldsymbol{\mu})^t \mathbf{x}_V$. Then

$$\begin{aligned}
& Q_V \\
=\ & Cov((\mathbf{R} - \boldsymbol{\mu})\tilde{R}) \\
=\ & Cov_{\tilde{R}} \left(E\left[(\mathbf{R} - \boldsymbol{\mu})\tilde{R} \mid \tilde{R} \right] \right) + E_{\tilde{R}} \left(Cov\left[(\mathbf{R} - \boldsymbol{\mu})\tilde{R} \mid \tilde{R} \right] \right), \text{ by decomposition of variance,} \\
=\ & Cov_{\tilde{R}} \left(\tilde{R} E\left[(\mathbf{R} - \boldsymbol{\mu}) \mid \tilde{R} \right] \right) + E_{\tilde{R}} \left(\tilde{R}^2 Cov\left[(\mathbf{R} - \boldsymbol{\mu}) \mid \tilde{R} \right] \right). \qquad (5.7)
\end{aligned}$$

Now $\left[\begin{array}{cc} (\mathbf{R}-\boldsymbol{\mu})^t & \tilde{R} \end{array} \right]^t$ has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\left[\begin{array}{cc} \Sigma & \Sigma \mathbf{x}_V \\ \mathbf{x}_V^t \Sigma & \mathbf{x}_V^t \Sigma \mathbf{x}_V \end{array} \right],$$

by definition of $\tilde{R}$. Then from Proposition 43 in the Appendix we have

$$E\left[(\mathbf{R}-\boldsymbol{\mu}) \mid \tilde{R}\right] = \mathbf{0} + \frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \tilde{R},$$

and

$$Cov\left[(\mathbf{R}-\boldsymbol{\mu}) \mid \tilde{R}\right] = \Sigma - \frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \mathbf{x}_V^t \Sigma.$$

Using the last two expressions in (5.7) we get

$$\begin{aligned} Q_V &= Cov_{\tilde{R}}\left(\frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \tilde{R}^2\right) + E_{\tilde{R}}\left(\left[\Sigma - \frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \mathbf{x}_V^t \Sigma\right] \tilde{R}^2\right) \\ &= E\left(\left[\frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \tilde{R}^2\right]\left[\frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \tilde{R}^2\right]^t\right) \\ &\quad - E\left(\left[\frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \tilde{R}^2\right]\right) E\left(\left[\frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \tilde{R}^2\right]^t\right) \\ &\quad + E_{\tilde{R}}\left(\left[\Sigma - \frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \mathbf{x}_V^t \Sigma\right] \tilde{R}^2\right) \\ &= \frac{1}{\left(\mathbf{x}_V^t \Sigma \mathbf{x}_V\right)^2} \Sigma \mathbf{x}_V \mathbf{x}_V^t \Sigma \left(E(\tilde{R}^4) - [E(\tilde{R}^2)]^2\right) + \left[\Sigma - \frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V \mathbf{x}_V^t \Sigma\right] E(\tilde{R}^2). \qquad (5.8) \end{aligned}$$

Since $\tilde{R}$ has a normal distribution with variance $\mathbf{x}_V^t \Sigma \mathbf{x}_V$, we have $E(\tilde{R}^2) = \mathbf{x}_V^t \Sigma \mathbf{x}_V$ and $E(\tilde{R}^4) = 3\left(\mathbf{x}_V^t \Sigma \mathbf{x}_V\right)^2$ - since the moment generating function of the standard normal distribution is $e^{-t/2}$, it follows that the fourth moment is equal to 3. Then, plugging these last results in (5.8) yields

$$Q_V = \Sigma\left(\mathbf{x}_V^t \Sigma \mathbf{x}_V\right) + \Sigma \mathbf{x}_V \mathbf{x}_V^t \Sigma.$$

We can now simplify the expression for the asymptotic covariance of $\sqrt{T}(\hat{\mathbf{x}}_V - \mathbf{x}_V)$, as given in Corollary

(26), as follows.

$$
\begin{aligned}
& L_V Q_V L_V \\
= & \left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right] \\
& \left(\Sigma\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right) + \Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\right)\left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right] \\
= & \left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)\left[\mathbf{I} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\right]\left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right] \\
& + \left[\mathbf{I} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\right]\mathbf{x}_V\mathbf{x}_V^t\left[\mathbf{I} - \mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right]. \qquad (5.9)
\end{aligned}
$$

But notice that we must have

$$
\begin{aligned}
\left[\mathbf{I} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\right]\mathbf{x}_V &= \mathbf{x}_V - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b} \\
&= \mathbf{x}_V - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b} \\
&= \mathbf{0},
\end{aligned}
$$

where the first and last equations follow from (5.6). Then, (5.9) becomes

$$
\begin{aligned}
L_V Q_V L_V &= \left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)\left[\mathbf{I} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\right]\left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right] \\
&= \left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)\left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right].
\end{aligned}
$$

We now can get a very simple expression for the asymptotic statistical risk $\bar{L}_V$ of $\hat{\mathbf{x}}_V$.

$$
\begin{aligned}
\bar{L}_V &= E[tr\left(\Sigma Cov(\hat{\mathbf{x}}_V)\right)] \\
&= tr\left(\frac{1}{T}\Sigma\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)\left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right]\right) \\
&= \frac{1}{T}\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right) tr\left(\mathbf{I}_N - \mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right) \\
&= \frac{1}{T}\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)\left[tr\left(\mathbf{I}_N\right) - tr\left((\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)\right] \\
&= \frac{1}{T}\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)[N - M]. \qquad (5.10)
\end{aligned}
$$

■

5.2.2 Least Absolute Deviation Portfolio Estimator

First, let us prove two lemmas which will be used in the next theorem.

Lemma 29 *Assume that the distribution of* $\mathbf{R}$ *is continuous. Then the optimal* 50%*-shortfall portfolio* $\mathbf{x}_{0.5}$*, and its median* $q_{0.5}$*, defined as*

$$(\mathbf{x}_{0.5}, q_{0.5}) = \arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q \in \mathbb{R}}} |\mathbf{x}^t \mathbf{R} - q|, \tag{5.11}$$

satisfy

$$E\left[\mathbf{R}\left(1_{\{\mathbf{x}_{0.5}^t \mathbf{R} \leqslant q_{0.5}\}} - \frac{1}{2}\right)\right] = \mathbf{A}^t \lambda,$$

for some $\lambda \in \mathbb{R}^M$.

Proof. The proof follows by writing the Lagrangean of (5.11), and taking its derivative with respect to $\mathbf{x}$ equal to zero. ■

Lemma 30 *Let* $\Gamma = E(\mathbf{R}\mathbf{R}^t \mid \mathbf{x}_V^t \mathbf{R} = q_{0.5})$. *Then* $\boldsymbol{\mu}^t \Gamma^{-1} \boldsymbol{\mu} = 1$

Proof. This follows by singularity of $Cov\left[\mathbf{R} \mid \tilde{R} = \mathbf{x}_V^t \boldsymbol{\mu}\right]$. ■

Now under the assumption that $\mathbf{R}$ is multivariate Gaussian, we know that $\mathbf{x}_{0.5} = \mathbf{x}_V$. We use this fact in the following Proposition.

Theorem 31 *When* $\mathbf{R}$ *has a multivariate Gaussian distribution, the asymptotic estimation risk of* $\hat{\mathbf{x}}_{0.5}$ *is*

$$\begin{aligned}
\bar{L}_{0.5} &= tr(\Sigma \frac{1}{T} L_{0.5} Q_{0.5} L_{0.5}) \\
&= \left(\frac{1}{f_{\mathbf{x}_V^t \mathbf{R}}(q_{0.5})}\right)^2 \frac{1}{4T} \left(\mathbf{x}_V^t \Sigma \mathbf{x}_V\right) [N - M] \\
&= \frac{\pi}{2T} \left(\mathbf{x}_V^t \Sigma \mathbf{x}_V\right) [N - M],
\end{aligned}$$

where $f_{\mathbf{x}_V^t \mathbf{R}}(.)$ *is the density of random variable* $\mathbf{x}_V^t \mathbf{R}$*, and where* $\mathbf{x}_V^t \Sigma \mathbf{x}_V$ *is the variance of the optimal portfolio* $\mathbf{x}_V$.

Proof. In what follows, we let $\alpha = 50\%$. Let $R_V := \mathbf{x}_V^t \mathbf{R}$. Notice that R_V is univariate Gaussian, and for $\alpha = 50\%$, $q_\alpha = \mathbf{x}_V^t \boldsymbol{\mu}$. Now $\left[\begin{array}{cc} \mathbf{R}^t & R_V \end{array} \right]$ has a multivariate Gaussian distribution with mean $\left[\begin{array}{cc} \boldsymbol{\mu} & \mathbf{x}_V^t \boldsymbol{\mu} \end{array} \right]$ and covariance matrix

$$\left[\begin{array}{cc} \Sigma & \Sigma \mathbf{x}_V \\ \mathbf{x}_V^t \Sigma & \mathbf{x}_V^t \Sigma \mathbf{x}_V \end{array} \right],$$

by definition of R_V. Then from Proposition 43 in the appendix we have

$$\begin{aligned} \boldsymbol{\tau} &= E\left[\mathbf{R} \mid \mathbf{x}_V^t \mathbf{R} = q_\alpha\right] = E\left[\mathbf{R} \mid R_V = \mathbf{x}_V^t \boldsymbol{\mu}\right] \\ &= \boldsymbol{\mu} + \frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V (\mathbf{x}_V^t \boldsymbol{\mu} - \mathbf{x}_V^t \boldsymbol{\mu}) \\ &= \boldsymbol{\mu}. \end{aligned} \tag{5.12}$$

This shows in particular that

$$\begin{aligned} Q_\alpha &= Cov\left[(\mathbf{R} - \boldsymbol{\tau})\left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] \\ &= Cov\left[(\mathbf{R} - \boldsymbol{\mu})\left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] \text{ by (5.12)}, \\ &= E\left[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t \left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)^2\right] \\ &\quad - E\left[(\mathbf{R} - \boldsymbol{\mu})\left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] E\left[(\mathbf{R} - \boldsymbol{\mu})^t \left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] \\ &= E\left[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t \left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)^2\right] \\ &\quad - E\left[\mathbf{R}\left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] E\left[\mathbf{R}^t \left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right], \end{aligned}$$

where the second to last equation follows from the definition of the covariance, and the last equation follows because $E\left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right) = 0$ by definition of the quantile q_α. But notice that $\left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)^2 = 1/4$ because $\alpha = 0.5$. Also, from Lemma 29,

$$\begin{aligned} & E\left[\mathbf{R}\left(1_{\{\mathbf{x}_V^t \mathbf{R} \leqslant q_\alpha\}} - \alpha\right)\right] \\ = \; & \mathbf{A}^t \lambda_\alpha, \end{aligned}$$

for some $\lambda_\alpha \in \mathbb{R}^M$. Therefore,

$$\begin{aligned} Q_\alpha &= \frac{1}{4}E\left[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^t\right] - \mathbf{A}^t\lambda_\alpha\lambda_\alpha^t\mathbf{A} \\ &= \frac{\Sigma}{4} - \mathbf{A}^t\lambda_\alpha\lambda_\alpha^t\mathbf{A}. \end{aligned} \tag{5.13}$$

Now notice that from Proposition 43 in the Appendix we also have

$$\begin{aligned} Cov\left[\mathbf{R} \mid \tilde{R} = \mathbf{x}_V^t\boldsymbol{\mu}\right] &= \Sigma - \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma \\ &= \Sigma - \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\Sigma, \text{ from (5.6)}, \\ &= \Sigma - \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A} \end{aligned}$$

so that

$$\begin{aligned} \Sigma &= Cov\left[\mathbf{R} \mid \tilde{R} = \mathbf{x}_V^t\boldsymbol{\mu}\right] + \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A} \\ &= E\left[\mathbf{R}\mathbf{R}^t \mid \tilde{R} = \mathbf{x}_V^t\boldsymbol{\mu}\right] - E\left[\mathbf{R} \mid \tilde{R} = \mathbf{x}_V^t\boldsymbol{\mu}\right]E\left[\mathbf{R}^t \mid \tilde{R} = \mathbf{x}_V^t\boldsymbol{\mu}\right] \\ &\quad + \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A} \\ &= \Gamma - \boldsymbol{\mu}\boldsymbol{\mu}^t + \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}, \end{aligned} \tag{5.14}$$

where the last equation follows from (5.12) and by definition of Γ. Using (5.13) and (5.14), we get

$$\begin{aligned} Q_\alpha &= \frac{1}{4}\left(\Gamma - \boldsymbol{\mu}\boldsymbol{\mu}^t + \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\right) - \mathbf{A}^t\lambda_\alpha\lambda_\alpha^t\mathbf{A} \\ &= \frac{1}{4}\left(\Gamma - \boldsymbol{\mu}\boldsymbol{\mu}^t\right) + \mathbf{A}^t\left(\frac{1}{4}\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1} - \lambda_\alpha\lambda_\alpha^t\right)\mathbf{A} \\ &= \frac{1}{4}\left(\Gamma - \boldsymbol{\mu}\boldsymbol{\mu}^t\right) + \mathbf{A}^t\Xi\mathbf{A}, \end{aligned} \tag{5.15}$$

with

$$\Xi = \frac{1}{4}\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1} - \lambda_\alpha\lambda_\alpha^t.$$

We can now simplify the expression for the asymptotic estimation risk of $\hat{\mathbf{x}}_\alpha$. This asymptotic estimation risk can be expressed as

$$\begin{aligned}
&\bar{L}_\alpha \\
&= E[tr\,(\Sigma Cov(\hat{\mathbf{x}}_\alpha))] \\
&= E\left[tr\left(\Sigma \frac{1}{T} L_\alpha Q_\alpha L_\alpha\right)\right] \\
&= \left(\frac{1}{f_{\mathbf{x}_V^t \mathbf{R}}(q_\alpha)}\right)^2 \frac{1}{T} tr\left(\Sigma\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) Q_\alpha \left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\right). \qquad (5.16)
\end{aligned}$$

where we have used the definitions of $Cov(\hat{\mathbf{x}}_\alpha)$, as given in Corollary (27), and $\bar{L}_\alpha$. But notice that

$$\begin{aligned}
&Q_\alpha\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) \\
&= \left(\frac{1}{4}\left(\Gamma - \boldsymbol{\mu}\boldsymbol{\mu}^t\right) + \mathbf{A}^t \Xi \mathbf{A}\right)\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) \text{ by (5.15),} \\
&= \frac{1}{4}\left(\Gamma - \boldsymbol{\mu}\boldsymbol{\mu}^t\right)\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) \text{ since } \mathbf{A}\Omega = \mathbf{0}, \\
&= \frac{1}{4}\left(\Gamma\Omega - \boldsymbol{\mu}\boldsymbol{\mu}^t\Omega + \frac{1}{\delta}\Gamma\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega - \frac{1}{\delta}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) \\
&= \frac{1}{4}\left(\Gamma\Omega - \left(\mathbf{I} - \frac{1}{\delta}\Gamma\Omega + \frac{1}{\delta}\boldsymbol{\mu}^t\Omega\boldsymbol{\mu}\mathbf{I}\right)\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) \\
&= \frac{1}{4}\left(\mathbf{I} - \mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1} - \left(\mathbf{I} - \frac{1}{\delta}(\mathbf{I} - \mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}) + \frac{1}{\delta}(1-\delta)\mathbf{I}\right)\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) \\
&= \frac{1}{4}\left(\mathbf{I} - \mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1} - \frac{1}{\delta}\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right). \qquad (5.17)
\end{aligned}$$

where the second to last equation follows by definition of Ω and because

$$\begin{aligned}
\boldsymbol{\mu}^t\Omega\boldsymbol{\mu} &= \boldsymbol{\mu}^t\left[\Gamma^{-1} - \Gamma^{-1}\mathbf{A}^t\left(\mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Gamma^{-1}\right]\boldsymbol{\mu} \\
&= \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\left(\mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Gamma^{-1} \\
&= 1 - \delta,
\end{aligned}$$

by definition of δ and because $\boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}$ =1 from Lemma 30. Similarly,

$$\begin{aligned}
&\Sigma\left(\Omega+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\\
=&\left(\Gamma-\boldsymbol{\mu}\boldsymbol{\mu}^t+\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\right)\\
&\left(\Omega+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\\
=&\left(\Gamma-\boldsymbol{\mu}\boldsymbol{\mu}^t\right)\left(\Omega+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\ \text{ since } \mathbf{A}\Omega=\mathbf{0},\\
=&\left(\mathbf{I}-\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}-\frac{1}{\delta}\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right), \qquad (5.18)
\end{aligned}$$

where the last equation is obtained by following the steps immediately preceding (5.17). Therefore, using (5.17) and (5.18) we get

$$\begin{aligned}
&\Sigma\left(\Omega+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)Q_\alpha\left(\Omega+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\\
=&\frac{1}{4}\left(\mathbf{I}_N-\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}-\frac{1}{\delta}\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\\
&\left(\mathbf{I}_N-\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}-\frac{1}{\delta}\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\\
=&\frac{1}{4}\left(\mathbf{I}_N-\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}-\frac{1}{\delta}\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right), \qquad (5.19)
\end{aligned}$$

where the last equation follows after some algebra, using the fact that $\mathbf{A}\Omega=\mathbf{0}$.

We now can get a very simple expression for the asymptotic estimation risk $\bar{L}_{0.5}$ of $\hat{\mathbf{x}}_{0.5}$. By (5.16) and (5.19) we have

$$
\begin{aligned}
& \bar{L}_{0.5} \\
= & \left(\frac{1}{f_{\mathbf{x}_V^t\mathbf{R}}(q_\alpha)}\right)^2 \frac{1}{4T} tr\left(\Sigma\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) Q_\alpha \left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\right) \\
= & \left(\frac{1}{f_{\mathbf{x}_V^t\mathbf{R}}(q_\alpha)}\right)^2 \frac{1}{4T} tr\left(\mathbf{I} - \mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1} - \frac{1}{\delta}\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) \\
= & \left(\frac{1}{f_{\mathbf{x}_V^t\mathbf{R}}(q_\alpha)}\right)^2 \\
& \frac{1}{4T}\left(tr(\mathbf{I}_N) - tr\left((\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right) - \frac{1}{\delta} tr\left(\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\right) \\
= & \left(\frac{1}{f_{\mathbf{x}_V^t\mathbf{R}}(q_\alpha)}\right)^2 \frac{1}{4T}(N - M)
\end{aligned}
$$

where the last equation follows because

$$
\begin{aligned}
& tr\left(\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) \\
= & tr((\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}^t\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\mathbf{A}^t) \\
= & 0, \text{ since } \mathbf{A}\Omega = \mathbf{0}.
\end{aligned}
$$

Now,

$$
\begin{aligned}
\left(\frac{1}{f_{\mathbf{x}_V^t\mathbf{R}}(q_\alpha)}\right)^2 & = \left(\frac{1}{\sqrt{2\pi(\mathbf{x}_V^t\Sigma\mathbf{x}_V)}} \exp\left(-\frac{(q_\alpha - \mathbf{x}_V^t\boldsymbol{\mu})}{2(\mathbf{x}_V^t\Sigma\mathbf{x}_V)}\right)\right)^{-2} \\
& = 2\pi(\mathbf{x}_V^t\Sigma\mathbf{x}_V),
\end{aligned}
$$

since $q_\alpha = \mathbf{x}_V^t\boldsymbol{\mu}$, and since $\mathbf{x}_V^t\mathbf{R}$ has a Gaussian distribution. The result follows. ■

5.3 Asymptotic Estimation Risk Under the Multivariate Student-t Distribution

We calculate closed form expressions for the asymptotic estimation risks of $\hat{\mathbf{x}}_V$ and $\hat{\mathbf{x}}_\alpha$ under the assumption that $\mathbf{R}$ has a multivariate Student-t distribution with ν degrees of freedom.

5.3.1 Variance Portfolio Estimator

Theorem 32 *When $\mathbf{R}$ has a multivariate Student-t distribution with $\nu \geqslant 5$ degrees of freedom, the asymptotic estimation risk of $\hat{\mathbf{x}}_V$ is*

$$\begin{aligned} \bar{L}_V &= tr(\Sigma \frac{1}{T} L_V Q_V L_V) \\ &= \frac{1}{T} \frac{(\nu-2)}{(\nu-4)} \left(\mathbf{x}_V^t \Sigma \mathbf{x}_V\right) [N-M], \end{aligned}$$

where $\mathbf{x}_V^t \Sigma \mathbf{x}_V$ is the variance of the optimal portfolio $\mathbf{x}_V$.

Proof. Let $\tilde{R} := (\mathbf{R}-\boldsymbol{\mu})^t \mathbf{x}_V$. Then

$$\begin{aligned} & Q_V \\ =\ & Cov((\mathbf{R}-\boldsymbol{\mu})\tilde{R}) \\ =\ & Cov_{\tilde{R}}\left(E\left[(\mathbf{R}-\boldsymbol{\mu})\tilde{R} \mid \tilde{R}\right]\right) + E_{\tilde{R}}\left(Cov\left[(\mathbf{R}-\boldsymbol{\mu})\tilde{R} \mid \tilde{R}\right]\right), \text{ by decomposition of variance,} \\ =\ & Cov_{\tilde{R}}\left(\tilde{R}E\left[(\mathbf{R}-\boldsymbol{\mu}) \mid \tilde{R}\right]\right) + E_{\tilde{R}}\left(\tilde{R}^2 Cov\left[(\mathbf{R}-\boldsymbol{\mu}) \mid \tilde{R}\right]\right). \end{aligned} \tag{5.20}$$

Now $\left[\begin{array}{cc} (\mathbf{R}-\boldsymbol{\mu})^t & \tilde{R} \end{array}\right]$ has a multivariate Student-t distribution with ν degrees of freedom, with mean $\mathbf{0}$ and covariance matrix

$$\left[\begin{array}{cc} \Sigma & \Sigma\mathbf{x}_V \\ \mathbf{x}_V^t\Sigma & \mathbf{x}_V^t\Sigma\mathbf{x}_V \end{array}\right],$$

by definition of $\tilde{R}$. Then from Proposition 45 in the Appendix we have

$$E\left[(\mathbf{R}-\boldsymbol{\mu}) \mid \tilde{R}\right] = \mathbf{0} + \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\tilde{R},$$

and

$$Cov\left[(\mathbf{R}-\boldsymbol{\mu}) \mid \tilde{R}\right] = \left(\frac{\nu+\tilde{R}^2/(\frac{\nu-2}{\nu}\mathbf{x}_V^t\Sigma\mathbf{x}_V)}{\nu-1}\right)\frac{\nu-2}{\nu}\left[\Sigma-\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\right].$$

Using the last two expressions in (5.20) we get

$$\begin{aligned}
& Q_V \\
= \; & Cov_{\tilde{R}}\left(\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\tilde{R}^2\right)+E_{\tilde{R}}\left(\frac{\nu-2}{\nu}\left(\frac{\nu+\tilde{R}^2/(\frac{\nu-2}{\nu}\mathbf{x}_V^t\Sigma\mathbf{x}_V)}{\nu-1}\right)\left[\Sigma-\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\right]\tilde{R}^2\right) \\
= \; & E\left(\left[\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\tilde{R}^2\right]\left[\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\tilde{R}^2\right]^t\right) \\
& -E\left(\left[\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\tilde{R}^2\right]\right)E\left(\left[\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\tilde{R}^2\right]^t\right) \\
& +\frac{\nu-2}{\nu}\left[\Sigma-\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\right]E_{\tilde{R}}\left(\left(\frac{\nu+\tilde{R}^2/(\frac{\nu-2}{\nu}\mathbf{x}_V^t\Sigma\mathbf{x}_V)}{\nu-1}\right)\tilde{R}^2\right) \\
= \; & \frac{1}{\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)^2}\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\left(E(\tilde{R}^4)-[E(\tilde{R}^2)]^2\right) \\
& +\frac{\nu-2}{\nu(\nu-1)}\left[\Sigma-\frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\right]E_{\tilde{R}}\left(\nu\tilde{R}^2+\tilde{R}^4/(\frac{\nu-2}{\nu}\mathbf{x}_V^t\Sigma\mathbf{x}_V)\mathbf{x}\right) \qquad (5.21)
\end{aligned}$$

Since $\tilde{R}$ has a Student-t distribution with variance $\mathbf{x}_V^t\Sigma\mathbf{x}_V$, we have $E(\tilde{R}^2) = \mathbf{x}_V^t\Sigma\mathbf{x}_V$ and $E(\tilde{R}^4) = 3\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)^2(\nu-2)/(\nu-4)$ - see for example Johnson, Kotz, and Balakrishnan, p. 365. Then, plugging these last results in (5.21) yields

$$\begin{aligned}
Q_V \; = \; & \Sigma\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)\frac{\nu-2}{\nu(\nu-1)}\left(\nu+3\frac{\nu}{(\nu-4)}\right) \\
& -\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\left(3\frac{(\nu-2)}{(\nu-4)}-1-\frac{\nu-2}{\nu(\nu-1)}\left(\nu+3\frac{\nu}{(\nu-4)}\right)\right) \\
= \; & \Sigma\left(\mathbf{x}_V^t\Sigma\mathbf{x}_V\right)\frac{(\nu-2)}{(\nu-4)} \\
& -\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\left(3\frac{(\nu-2)}{(\nu-4)}-1-\frac{\nu-2}{\nu(\nu-1)}\left(\nu+3\frac{\nu}{(\nu-4)}\right)\right)
\end{aligned}$$

The result then follows directly from reasoning identical to the proof of Theorem (28). ■

5.3.2 Least Absolute Deviation Portfolio Estimator

Theorem 33 *When* $\mathbf{R}$ *has a multivariate Student-t distribution with* $\nu \geqslant 3$ *degrees of freedom, for* $\alpha = 50\%$*, the asymptotic risk of* $\hat{\mathbf{x}}_\alpha$ *is*

$$\begin{aligned}
\bar{L}_{0.5} &= tr(\Sigma \frac{1}{T} L_{0.5} Q_{0.5} L_{0.5}) \\
&= \left(\frac{\nu-1}{\nu-2}\right)^2 \left(\frac{1}{f_{\mathbf{x}_V^t \mathbf{R}}(q_{0.5})}\right)^2 \frac{1}{4T} \left(\mathbf{x}_V^t \Sigma \mathbf{x}_V\right) [N-M] \\
&= \frac{\pi}{4T} \frac{(\nu-1)^2}{\nu-2} \left(\frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})}\right)^2 \left(\mathbf{x}_V^t \Sigma \mathbf{x}_V\right) [N-M].
\end{aligned}$$

where $f_{\mathbf{x}_V^t \mathbf{R}}(.)$ *is the density of random variable* $\mathbf{x}_V^t \mathbf{R}$*, and where* $\mathbf{x}_V^t \Sigma \mathbf{x}_V$ *is the variance of the optimal portfolio* $\mathbf{x}_V$ *.*

Proof. Let $R_V := \mathbf{x}_V^t \mathbf{R}$. Notice that R_V is univariate Student-t with ν degrees of freedom, and for $\alpha = 50\%$, $q_\alpha = \mathbf{x}_V^t \boldsymbol{\mu}$. Now $\begin{bmatrix} \mathbf{R}^t & R_V \end{bmatrix}$ has a multivariate Student-t distribution with mean $\begin{bmatrix} \boldsymbol{\mu} & \mathbf{x}_V^t \boldsymbol{\mu} \end{bmatrix}$ and covariance matrix

$$\begin{bmatrix} \Sigma & \Sigma \mathbf{x}_V \\ \mathbf{x}_V^t \Sigma & \mathbf{x}_V^t \Sigma \mathbf{x}_V \end{bmatrix},$$

by definition of R_V. Then from Proposition (45) in the Appendix we have

$$\begin{aligned}
\boldsymbol{\tau} &= E\left[\mathbf{R} \mid \mathbf{x}_V^t \mathbf{R} = q_\alpha\right] = E\left[\mathbf{R} \mid R_V = \mathbf{x}_V^t \boldsymbol{\mu}\right] \\
&= \boldsymbol{\mu} + \frac{1}{\mathbf{x}_V^t \Sigma \mathbf{x}_V} \Sigma \mathbf{x}_V (\mathbf{x}_V^t \boldsymbol{\mu} - \mathbf{x}_V^t \boldsymbol{\mu}) \\
&= \boldsymbol{\mu}.
\end{aligned}$$

and

$$
\begin{aligned}
& Cov\left[\mathbf{R} \mid \tilde{R} = \mathbf{x}_V^t \boldsymbol{\mu}\right] \\
= & \left(\frac{\nu + (\mathbf{x}_V^t\boldsymbol{\mu} - \mathbf{x}_V^t\boldsymbol{\mu})^2/(\frac{\nu-2}{\nu}\mathbf{x}_V^t\Sigma\mathbf{x}_V)}{\nu - 1}\right)\frac{\nu-2}{\nu}\left(\Sigma - \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\mathbf{x}_V\mathbf{x}_V^t\Sigma\right) \\
= & \left(\frac{\nu-2}{\nu-1}\right)\left(\Sigma - \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\Sigma\Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\Sigma\right), \text{ from (5.6)}, \\
= & \left(\frac{\nu-2}{\nu-1}\right)\left(\Sigma - \frac{1}{\mathbf{x}_V^t\Sigma\mathbf{x}_V}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{b}\mathbf{b}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\right)
\end{aligned}
$$

Then, the same reasoning as in Theorem (31) leads to the following simple expression for the asymptotic estimation risk $\bar{L}_{0.5}$ of $\hat{\mathbf{x}}_{0.5}$:

$$
\bar{L}_{0.5} = \left(\frac{\nu-1}{\nu-2}\right)^2\left(\frac{1}{f_{\mathbf{x}_V^t\mathbf{R}}(q_{0.5})}\right)^2\frac{1}{4T}(N-M)
$$

Now,

$$
\begin{aligned}
\left(\frac{1}{f_{\mathbf{x}_V^t\mathbf{R}}(q_{0.5})}\right)^2 & = \left(\frac{1}{\sqrt{\pi}}\frac{\nu^{\nu/2}\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{(\frac{\nu-2}{\nu}\mathbf{x}_V^t\Sigma\mathbf{x}_V)^{-1/2}}{(\nu + (q_\alpha - \mathbf{x}_V^t\boldsymbol{\mu})^2/(\frac{\nu-2}{\nu}\mathbf{x}_V^t\Sigma\mathbf{x}_V))^{(\nu+1)/2}}\right)^{-2} \\
& = \left(\frac{1}{\sqrt{\pi(\nu-2)}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}(\mathbf{x}_V^t\Sigma\mathbf{x}_V)^{-1/2}\right)^{-2} \\
& = \pi(\nu-2)\left(\frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})}\right)^2\mathbf{x}_V^t\Sigma\mathbf{x}_V
\end{aligned}
$$

since $q_{0.5} = \mathbf{x}_V^t\boldsymbol{\mu}$, and because $\mathbf{x}_V^t\mathbf{R}$ has a Student-t distribution with ν degrees of freedom. The result follows. ■

5.3.3 Comparing the Asymptotic Estimation Risk of the Least Absolute Deviation and Variance Estimators

Here we use the theoretical results from the last two Theorems, and simulation results, in order to compare the asymptotic estimation risk of the least absolute deviation and variance portfolio estimators under the assumption that returns have a multivariate Student-t distribution.

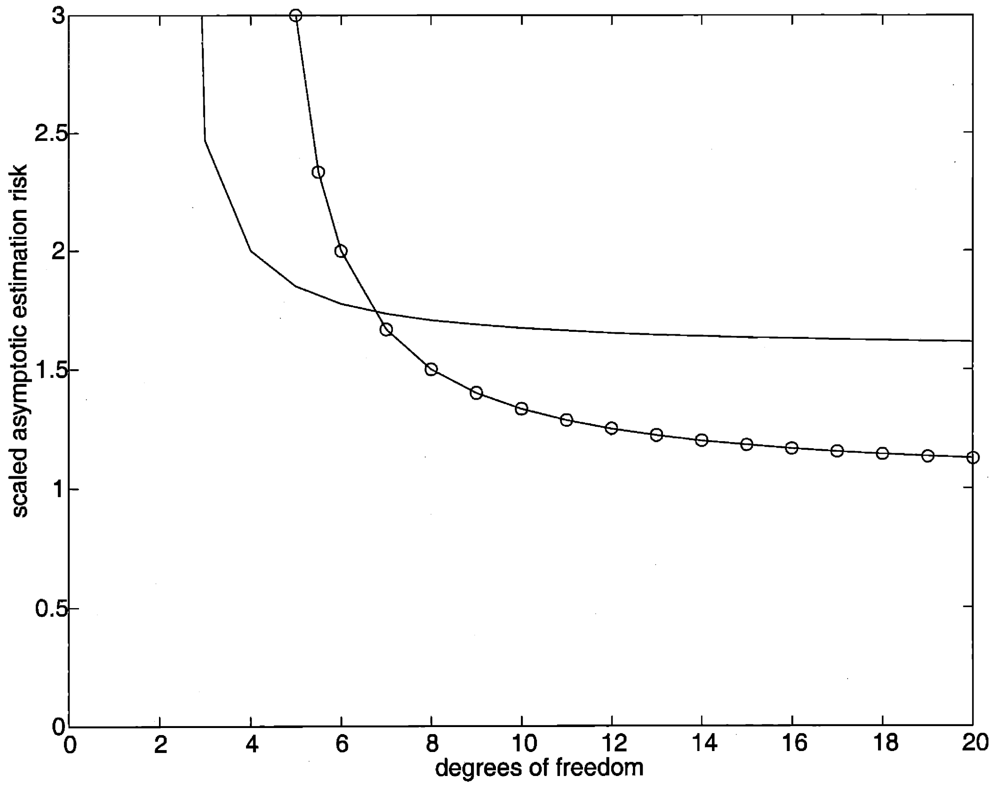

Figure 5-1: Scaled estimation risk as a function of degrees of freedom of multivariate Student-t: -o = variance, – = LAD.

Theoretical Performance of Least Absolute Deviation and Variance Portfolio Estimators

Let us define the scaled asymptotic estimation risk of estimator $\hat{\mathbf{x}}$ to be

$$\bar{L}_s = \frac{\text{asymptotic estimation risk of } \hat{\mathbf{x}}}{\frac{1}{T}\mathbf{x}_V^t \Sigma \mathbf{x}_V (N-M)}.$$

Note that the scaling factor in the denominator is just the asymptotic estimation risk of the variance portfolio estimator, under the assumption that returns are multivariate Gaussian. Using the results from the last two theorems, Figure 5-1 plots $\bar{L}_s$ for the LAD and variance estimators, for the multivariate Student-t,.as a function of the degrees of freedom of the distribution.

Figure 5-1 shows that as the degrees of freedom increases, and the multivariate Student-t approaches the multivariate Gaussian, $\bar{S}_V$ approaches one, and $\bar{S}_{0.5}$ approaches $\pi/2 \simeq 1.5$, which is their scaled asymptotic estimation risk under the Gaussian. Figure 5-1 also suggests that, asymptotically at least, we should expect the 50%-shortfall portfolio estimator to outperform the variance portfolio estimator as soon as the degrees of freedom of the distribution is, or goes below, six.

Simulated Performance Least Absolute Deviation and Variance Portfolio Estimators

Here we use simulation to validate the theoretical results presented previously. For simplicity we will focus on the estimation of

$$\mathbf{x}_V = \arg \min_{\mathbf{e}^t x=1} \mathbf{x}^t \Sigma \mathbf{x},$$

in the case where $N = 10$, and where Σ is equicorrelated (the correlation matrix has equal off-diagonal elements $\rho = 0.5$) with equal variances equal to one - i.e.

$$\Sigma = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & \ddots & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}.$$

Now define the scaled loss of portfolio estimator $\hat{\mathbf{x}}$ to be equal to

$$\bar{l}_s(\hat{\mathbf{x}}) = \frac{\hat{\mathbf{x}}^t \Sigma \hat{\mathbf{x}} - \mathbf{x}_V^t \Sigma \mathbf{x}_V}{\frac{1}{T} \mathbf{x}_V^t \Sigma \mathbf{x}_V (N - M)}. \tag{5.22}$$

Table 5.1 shows the average scaled loss over 500 Monte Carlo replications and for $T = 500$ observations, and the scaled asymptotic estimation risk for $\hat{\mathbf{x}}_V$ and $\hat{\mathbf{x}}_{0.5}$, for the Student-t with degrees of freedom equal to 3, 5, 7, and ∞ (i.e. the Gaussian). The simulated results agree with the theoretical results for all distributions, except for $\hat{\mathbf{x}}_V$ and the Student-t with 5 degrees of freedom[1]. Finally, Figure 5-2 shows the average scaled loss over 500 Monte Carlo replications, as a function of T, for the estimators $\hat{\mathbf{x}}_V$ and $\hat{\mathbf{x}}_{0.5}$, and for the four distributions we just considered.

[1] We ran another Monte Carlo with 500 replications of the following experiment: $N = 2$, $T = 10000$, and we obtained, for $\hat{\mathbf{x}}_V$, $\bar{l}_s = 3.0085$ with standard error (0.3075), which agrees with the theoretical result $L_s = 3$.

Distribution		$\hat{\mathbf{x}}_V$	$\hat{\mathbf{x}}_{0.5}$
Student-t (d.f. = 3)	$\bar{l}_s$	5.2503 (0.1834)	2.4844 (0.0535)
	$\bar{L}_s$	NA	2.4674
Student-t (d.f. = 5)	$\bar{l}_s$	2.1483 (0.0543)	1.8522 (0.0383)
	$\bar{L}_s$	3	1.8506
Student-t (d.f. = 7)	$\bar{l}_s$	1.6399 (0.0368)	1.7628 (0.0367)
	$\bar{L}_s$	1.6667	1.7349
Gaussian	$\bar{l}_s$	1.0342 (0.0216)	1.6279 (0.0353)
	$\bar{L}_s$	1	1.5708

Table 5.1: Average scaled loss, and scaled asymptotic statistical risk, over 500 Monte Carlo replications

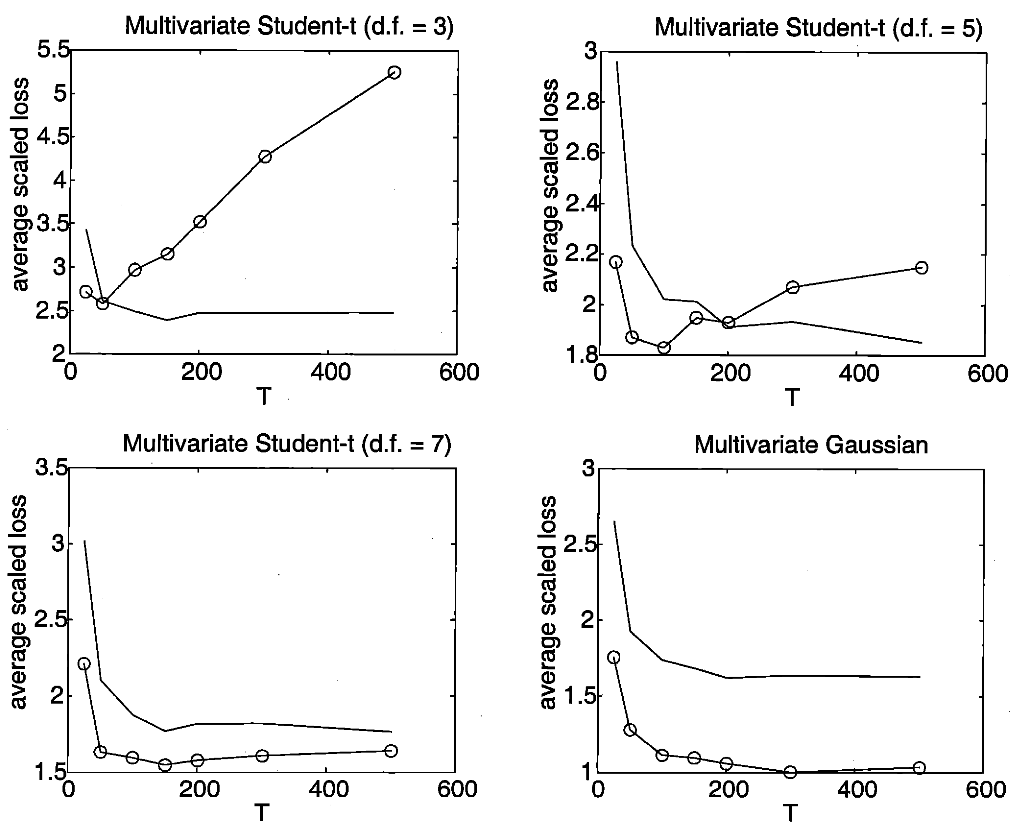

Figure 5-2: Average scaled loss over 500 MC: -o = $\hat{\mathbf{x}}_V$, $--$ = $\hat{\mathbf{x}}_{0.5}$

5.4 Other Robust Portfolio Estimators

In this Section we introduce three additional robust portfolio estimators: the Huber portfolio estimator, the trimean portfolio estimator, and the trimmed mean portfolio estimator.

5.4.1 The Huber Portfolio Estimator

The Huber contrast function - with parameter γ - can be algebraically expressed as

$$\eta_\gamma(z) = \begin{cases} z^2 & \text{if } |z| \leqslant \gamma \\ 2\gamma|z| - \gamma^2 & \text{if } z > \gamma \end{cases}, \tag{5.23}$$

and it is graphed in Figure 5-3. We will define the sample Huber portfolio estimator as the solution to

$$\begin{array}{ll} \text{minimize} & \frac{1}{T}\sum_i^T \eta_\gamma(\mathbf{x}^t\mathbf{R}_i - q) \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array} \tag{5.24}$$

In particular, compare (5.24) to (5.4) and (5.3). Intuitively, the Huber portfolio estimator should be

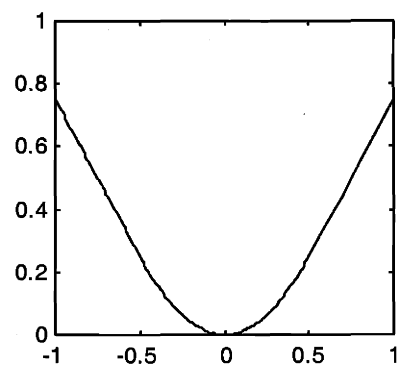

Figure 5-3: Huber contrast function, $\gamma = 0.5$.

less sensitive to outliers than the variance portfolio estimator, since deviations from q are only penalized linearly instead of quadratically. When returns have heavy tails and tail-dependence this may be a good thing, as we saw in the case of the LAD estimator.

Quadratic Programming Formulation of the Sample Huber Portfolio Optimization Problem

The sample Huber portfolio optimization problem has a simple QP formulation. We will use the following Lemma from Mangasarian and Musicant (2000).

Lemma 34 *The Huber contrast function $\eta_\gamma(z)$ of (5.23) is given by*

$$\eta_\gamma(z) = \min_{t \in \mathbb{R}} t^2 + 2\gamma|z-t|, \; z \in \mathbb{R}.$$

This last Lemma allows us to write Problem (5.24) as

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, q\in\mathbb{R}, t\in\mathbb{R}^T} & \frac{1}{T}\sum_{i=1}^T t_i^2 + \frac{1}{T}\sum_{i=1}^T 2\gamma|\mathbf{x}^t\mathbf{R}_i - q - t_i| \\ \text{subject to} & \mathbf{Ax} = \mathbf{b}, \end{array}$$

which can be rewritten as the following QP:

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^N, q\in\mathbb{R}, t\in\mathbb{R}^T, z\in\mathbb{R}^T} & \frac{1}{T}\sum_i^T t_i^2 + \frac{1}{T}\sum_i^T 2\gamma(z_i^+ + z_i^-) \\ \text{subject to} & \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x}^t\mathbf{R}_i - q - t_i = z_i^+ - z_i^-, \; i = 1, \ldots, T, \\ & \mathbf{z}^+, \mathbf{z}^- \geqslant 0, \end{array} \tag{5.25}$$

which has $N + 1 + 3T$ variables and $M + 3T$ constraints, with $N + 1$ dense columns, and a sparse quadratic objective function.

Choice of the γ Parameter

In the Huber portfolio estimation procedure, deviations beyond +/- γ of the location parameter q are penalized linearly instead of quadratically. Intuitively, this will reduce the influence of outliers on the estimation procedure. The choice of the γ parameter in Problem (5.24) can be motivated by minimax arguments, as in Huber (1981). We will choose γ to be proportional to the mean absolute deviation of portfolio $\hat{\mathbf{x}}_{0.5}$. Let R be a random variable. The mean absolute deviation is a robust measure of deviation defined as

$$MAD = E[|R - q_{0.5}|],$$

where $q_{0.5}$ is the median of R's distribution. Under normality, the mean absolute deviation is proportional to the standard deviation, and is equal to

$$MAD = \sigma\sqrt{\frac{2}{\pi}}$$

- see Bouchaud and Potters (2000). Our procedure for choosing γ is outlined below:

1. calculate $(\hat{\mathbf{x}}_{0.5}, \hat{q}_{0.5})$, as described in Chapter 2;

2. calculate the sample MAD of portfolio $\hat{\mathbf{x}}_{0.5}$, defined as

$$MAD(\hat{\mathbf{x}}_{0.5}) = \frac{1}{T}\sum_{i=1}^{T}|\hat{\mathbf{x}}_{0.5}^{t}\mathbf{R}_i - \hat{q}_{0.5}|;$$

3. set

$$\gamma = z_{1-\alpha}\sqrt{\frac{\pi}{2}}MAD(\hat{\mathbf{x}}_{0.5}),$$

   where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal.

As defined above, when returns are multivariate Gaussian, observations that are in the $\alpha\%$ upper and lower tails are de-emphasized (i.e. beyond +/- $z_{1-\alpha}\sigma$ of the mean), and penalized linearly in the Huber portfolio optimization procedure.

5.4.2 Trimean Portfolio Estimator

The trimean portfolio estimator is a linear combination of three shortfall portfolios. For an arbitrary $\alpha \in (0, 0.5)$, we will define the trimean estimator to be

$$\hat{\mathbf{x}}_t = \alpha\hat{\mathbf{x}}_\alpha + (1-2\alpha)\hat{\mathbf{x}}_{0.5} + \alpha\hat{\mathbf{x}}_{1-\alpha}.$$

For example, for $\alpha = 25\%$,

$$\hat{\mathbf{x}}_t = \frac{1}{4}\hat{\mathbf{x}}_{0.25} + \frac{1}{2}\hat{\mathbf{x}}_{0.5} + \frac{1}{4}\hat{\mathbf{x}}_{0.75}.$$

5.4.3 Trimmed Mean Portfolio Estimator

The trimmed mean portfolio estimation procedure begins with the estimation of the shortfall portfolios $(\hat{\mathbf{x}}_\alpha, \hat{q}_\alpha)$ and $(\hat{\mathbf{x}}_{1-\alpha}, \hat{q}_{1-\alpha})$, for $\alpha \in (0, 0.5)$. Then, observations $\mathbf{R}_i$ which are beyond the region defined by the hyperplanes $(\hat{\mathbf{x}}_\alpha, \hat{q}_\alpha)$ and $(\hat{\mathbf{x}}_{1-\alpha}, \hat{q}_{1-\alpha})$ are put aside, and the variance portfolio estimator is used on the rest of the data. We expect the trimmed mean portfolio estimator to be less sensitive to outliers than the variance portfolio estimator applied to the entire set of observations, and this may help the variance estimator when return distributions exhibit heavy-tailedness and tail-dependence.

5.5 Computational Experiments

In this section we apply our portfolio estimators to artificial and real datasets.

5.5.1 Simulated Data

Here we use the same simulation framework as in Section 3. The portfolio estimators that we consider are:

- V: variance;
- LAD: least absolute deviation;
- H1: Huber, with $\gamma = 1.96\sqrt{\frac{\pi}{2}}MAD(\hat{\mathbf{x}}_{0.5})$;
- H2: Huber, with $\gamma = 1.64\sqrt{\frac{\pi}{2}}MAD(\hat{\mathbf{x}}_{0.5})$;
- T1: trimean, $\hat{\mathbf{x}}_t = \frac{1}{4}\hat{\mathbf{x}}_{0.25} + \frac{1}{2}\hat{\mathbf{x}}_{0.5} + \frac{1}{4}\hat{\mathbf{x}}_{0.75}$;
- T2: trimean, $\hat{\mathbf{x}}_t = 0.3\hat{\mathbf{x}}_{0.3} + 0.4\hat{\mathbf{x}}_{0.5} + 0.3\hat{\mathbf{x}}_{0.7}$;
- TM1: trimmed mean, $\alpha = 5\%$;
- TM2: trimmed mean, $\alpha = 10\%$.

The following figures report the average scaled loss (5.22) over 500 Monte Carlo replications, for different values of T. Figure 5-4 compares V and LAD to H1 and H2; Figure 5-5 compares V and LAD to T1 and T2; and finally Figure 5-6 compares V and LAD to TM1 and TM2. The simulation reveals that:

1. For large values of T, i.e. asymptotically, all the robust estimators listed above, except LAD, outperform estimator V as soon as the degrees for freedom of the Student-t are less than or equal to 7. LAD outperforms V when the degrees of freedom is less than or equal to 5.

2. The difference in performance between different versions of the estimators, H1 and H2, T1 and T2, and TM1 and TM2, is negligible for all values of T.

3. For small values of T, the situation is mitigated, with estimator V having notably better performance than T1, T2, TM1, and TM2 for the $T = 25$, the smallest T under consideration. Estimators H1 and H2, however, seem never to perform worse than V, for any T, even for the Gaussian!

4. The Huber estimators seem to be the safest alternative to estimator V, since its performance is never worse than V, and yet is significantly better than V for low degrees for freedom of the Student-t and large values of T.

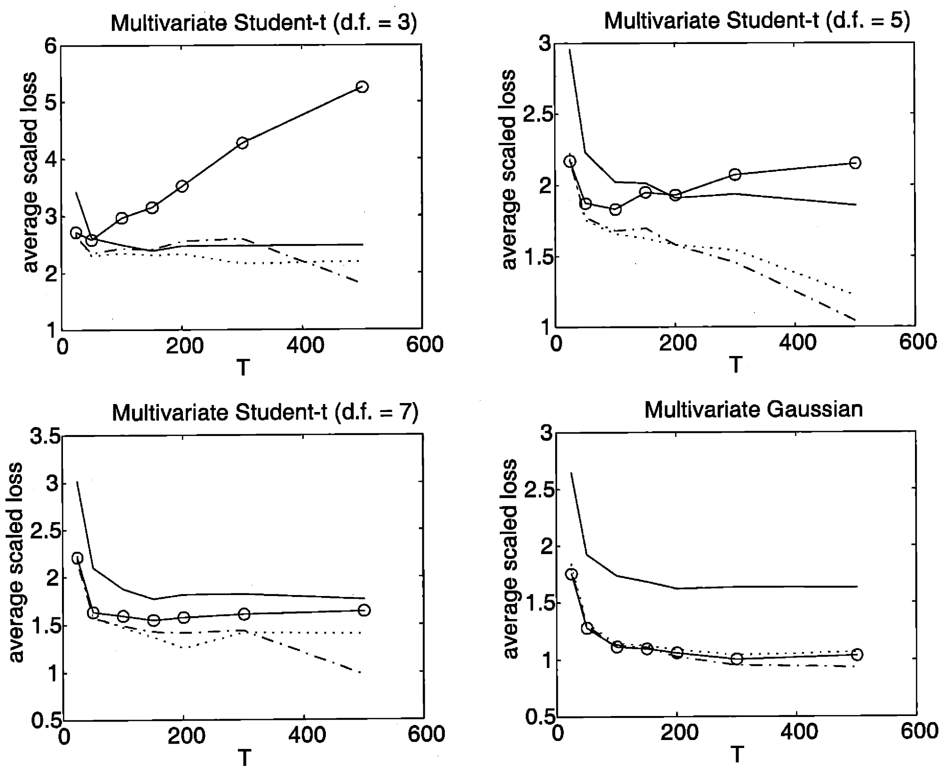

Figure 5-4: Average scaled loss over 500 Monte Carlo replications. -o = variance; – = LAD; - . = H1; $\cdots$ = H2.

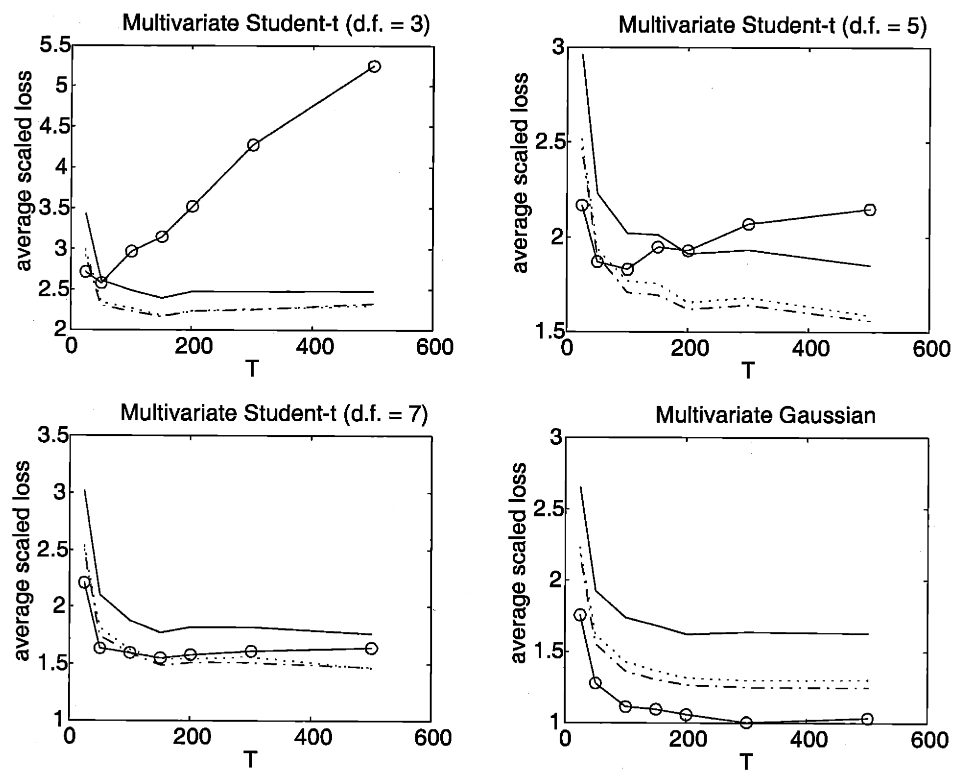

Figure 5-5: Average scaled loss over 500 Monte Carlo replications. -o = variance; – = LAD; - . = T1; $\cdots$ = T2.

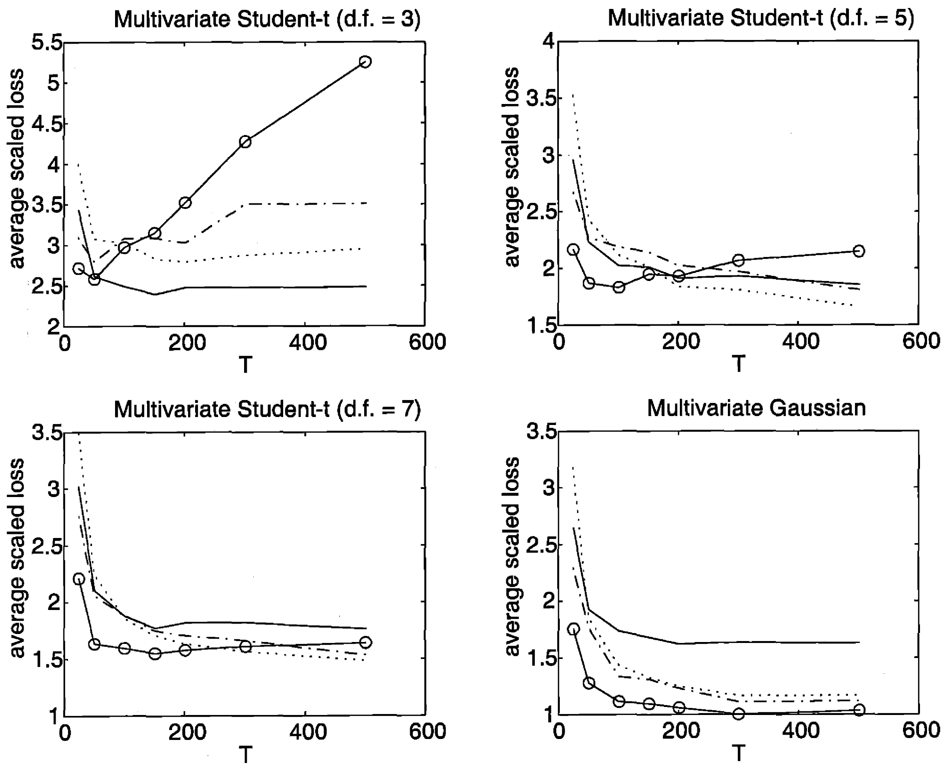

Figure 5-6: Average scaled loss over 500 Monte Carlo replications. -o = variance; – – = LAD; - . = TM1; ··· = TM2.

mean	standard deviation	skewness	kurtosis
0.2227	4.8955	0.5431	8.7186

Table 5.2: BTK stock return sample statistics, averaged over 16 stocks.

5.5.2 Historical Data

Here we work with a dataset consisting of the daily returns of stocks in the BTK biotechnology index. The data span the time period 3/20/1997 to 3/22/2001. The index has 17 companies, but we removed CRA because it did not have return data before 4/28/1999. We also removed three dates: 11/1/1999, 1/28/1998, and 5/8/1997, because some companies had no data on those days. The final dataset we work with has 1008 observations on each of 16 stocks. Table 5.2 summarizes some sample statistics. Notice in particular that the average kurtosis is 8.7, whereas if data were normally distributed, one would expect this number to be close to 3. This average kurtosis figure indicates the presence of heavier tails than the Gaussian.

In this experiment, we build portfolios using the estimators

- V: variance;
- LAD: least absolute deviation;
- H: Huber, with $\gamma = 1.96 MAD(\hat{\mathbf{x}}_{0.5})$;
- T: trimean, $\hat{\mathbf{x}}_t$=$0.3\hat{\mathbf{x}}_{0.3} + 0.4\hat{\mathbf{x}}_{0.5} + 0.3\hat{\mathbf{x}}_{0.7}$;
- TM: trimmed mean, $\alpha = 10\%$.

The only constraint that we impose on the portfolios is that the weights sum to one. For each estimator, starting on 9/1/1997, and at the beginning of each month after that until 3/1/2001, we use the data of the previous 6 months for estimation. We then collect the ex-post returns of the portfolio estimate over the next month. For each estimator, we collect all the ex-post returns into one vector, spanning 9/1/1997 to 3/22/2001. We also calculate the returns of the equally-weighted portfolio EQ, rebalanced every month, over the same period. Table 5.3 reports some sample statistics of these ex-post returns. The statistics that we consider are:

- **mean:** the sample mean of daily ex-post returns;

- **STD:** the sample standard deviation of daily ex-post returns;

- **information ratio:** we have defined the information ratio here as a signal to noise ratio equal to $\frac{\sqrt{260}(\text{mean})}{\text{STD}}$, where the standardization by $\sqrt{260}$ makes the information ratio an annual estimate, assuming 260 trading days per year;

- **α-VaR** for $\alpha = 5\%$ and 1%: the sample α-quantile of the daily ex-post return distribution;

- **α-CVaR** for $\alpha = 5\%$ and 1%: the sample conditional mean of the daily ex-post return distribution, given they are below the α-quantile;

- **MaxDD:** the maximum drawdown, defined as the largest percent decrease in the value of the portfolio over the period under consideration.

- **CRet**: cumulative return.

The numbers in bold indicate the ex-post performance of the corresponding estimator was **better** than the performance of estimator V, according to the corresponding statistic. The following insights emerge from examining the table:

1. estimator H outperforms V according to every measure of performance which we consider. As our experiments on artificial data showed, the Huber estimator seems like the safest robust estimator.

2. estimator TM vastly outperforms V according to cumulative return. Extreme returns also seem to be controlled better in TM, as both the 1%-Var and 1%-CVar are the lowest of all portfolio estimators, and MaxDD is also the lowest among each estimator.

3. estimator EQ under-performs estimator V according to every measure of performance, except the mean ex-post daily return. This shows that portfolio optimization may add value compared to a simple equally-weighted rebalancing strategy.

We also repeat the previous experiment, except that we add an expected return constraint to the optimization problem. That is, at every month t's beginning, we add the constraint

$$\mathbf{x}^t \boldsymbol{\mu}_t = r_p,$$

	V	LAD	H	T	TM	EQ
mean	0.2334	**0.2419**	**0.2375**	**0.2374**	**0.2667**	**0.2345**
STD	2.6726	2.7348	**2.6665**	2.7107	2.8196	3.1455
information ratio	1.4080	**1.4261**	**1.4359**	**1.4122**	**1.5254**	1.2020
5%-VaR	4.1230	4.1825	**4.0809**	4.1748	4.1850	5.0497
1%-VaR	6.8058	7.6358	**6.7576**	7.2469	**6.3173**	8.5239
5%-CVaR	5.4477	5.7555	**5.4102**	5.5908	5.4991	7.1423
1%-CVaR	7.9494	8.4141	**7.9449**	8.1057	**7.7089**	10.3449
MaxDD	-32.68	-36.73	**-32.42**	-33.90	**-30.92**	-47.44
CRet	486.5393	**523.1102**	**509.2406**	**502.4341**	**662.9031**	423.2045

Table 5.3: Portfolio Estimator Performance on BTK Data, convextity constraint only. All statistics are expressed as percentages, except the information ratio.

for rp in the range 160%/260 (which corresponds to an annualized target return of 160%) to 0, where μ_t is the vector of mean returns over month t. This is equivalent to having a perfect forecast of average monthly returns. We now proceed as in the previous experiment, collecting the ex-post returns of estimators V, LAD and H. The resulting minimum variance frontiers (the lowest level of standard deviation for every level of target return) is plotted in Figure 5-7. It is clear that estimator H improves upon V on the mean-variance scale, offering lower risk, as measured by standard deviation, for every level of expected return.

Finally, we repeat the initial experiment but with a five month rolling window, adding the constraint

$$\mathbf{x}^t\bar{\mathbf{R}}_{t-1} = r_p,$$

where rp in the range 160%/260 (which corresponds to an annualized target return of 160%) to 0, where $\bar{\mathbf{R}}_{t-1}$ is the vector of mean returns calculated over the last five months. Five months was chosen arbitrarily, and corresponds to 100 trading days. Figure 5-8 shows the resulting minimum variance frontiers of the ex-post returns for V, LAD, and H. Table 5.4 reports the results for $r_{p1} = 100\%/260, r_{p2} = 80\%/260, r_{p3} = 60\%/260, r_{p4} = 40\%/260$, and Figure 5-9 shows the cumulative return for these four values of target return r_p, for V and LAD (the cumulative returns for H were similar to V). In Table 5.4, the numbers in bold indicate the ex-post performance of the corresponding estimator was **better** than the performance of estimator V, according to the corresponding statistic. The cumulative returns and information ratio of LAD are impressive compared to V, with the information ratio of LAD being about 20% higher than that of V, and cumulative returns being about 40% higher. The other statistics

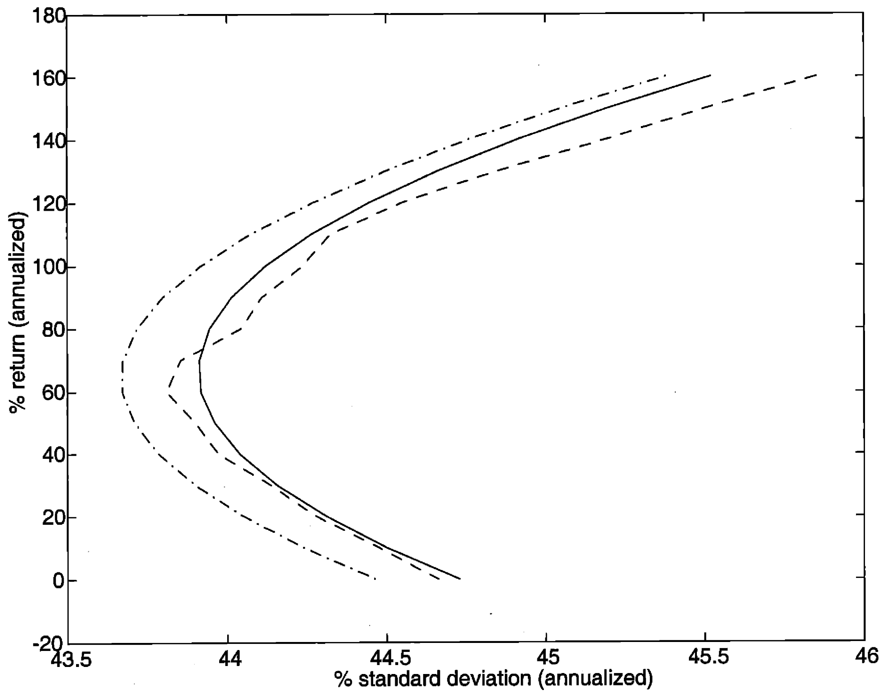

Figure 5-7: Minimum variance (standard deviation) frontier, with perfect forecast of returns: – = V; -- = LAD; -. = H.

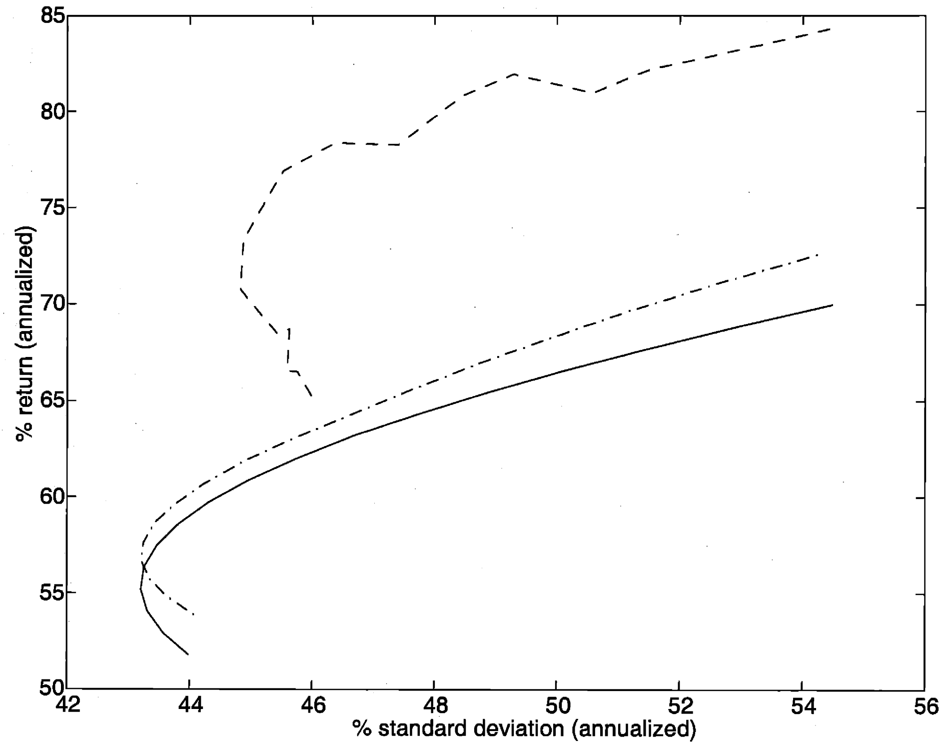

Figure 5-8: Minimum variance (standard deviation) frontier, estimating returns from past: – = V; - - = LAD; -. = H.

of the LAD are more mitigated, though the 1%-VaR is consistently lower for V. H always has lower STD and CVaR, at both 1% and 5%, and higher information ratio and cumulative returns.

Figure 5-9 shows the cumulative returns from estimators V and LAD.

	V (r_{p1})	LAD (r_{p1})	H (r_{p1})	V (r_{p2})	LAD (r_{p2})	H (r_{p2})
mean	0.2429	**0.3011**	**0.2468**	0.2342	**0.2959**	**0.2379**
STD	2.8946	2.9396	**2.8868**	2.7890	2.8233	**2.7831**
information ratio	1.3532	**1.6518**	**1.3786**	1.3539	**1.6901**	**1.3782**
5%-VaR	4.3786	4.6094	**4.3324**	4.2980	4.5060	**4.2956**
1%-VaR	8.3108	**8.0370**	8.3232	8.1151	**7.9135**	**8.0487**
5%-CVaR	6.2348	6.3200	**6.1858**	5.9455	5.9619	**5.9141**
1%-CVaR	9.2525	9.4662	**9.1182**	8.7808	**8.6382**	**8.6617**
MaxDD	-45.21	**-44.65**	**-43.70**	-40.89	**-38.28**	**-39.47**
CRet	504.0782	**903.7305**	**526.8074**	474.0644	**887.7804**	**494.2505**

	V (r_{p3})	LAD (r_{p3})	H (r_{p3})	V (r_{p4})	LAD (r_{p4})	H (r_{p4})
mean	0.2254	**0.2799**	**0.2294**	0.2167	**0.2635**	**0.2216**
STD	2.7175	2.7821	**2.7141**	2.6828	2.8166	**2.6823**
information ratio	1.3377	**1.6224**	**1.3626**	1.3025	**1.5084**	**1.3322**
5%-VaR	4.1586	4.2642	4.1920	4.0724	4.3380	4.1322
1%-VaR	7.4643	**6.6035**	**7.2389**	6.8855	**6.6551**	6.9184
5%-CVaR	5.6923	5.7784	**5.6661**	5.4915	5.6309	**5.4800**
1%-CVaR	8.3091	8.3235	**8.2099**	7.9230	8.1587	**7.9029**
MaxDD	-36.35	**-34.25**	**-35.11**	-32.05	**-30.26**	-0.32.09
CRet	440.5369	**765.5921**	**460.2708 5**	404.3046	**641.6139**	**426.934**

Table 5.4: V, LAD and H Performance on BTK Data: 5 month window, expected return constraint. All statistics are expressed as percentages, except for the information ratio.

5.6 Conclusion

In this chapter we considered the problem of estimating portfolios from return data, when the return data are generated from an elliptically symmetric distribution. We showed that when the distribution of returns has heavier tails and more tail dependence than them multivariate Gaussian, there are alternatives to the variance portfolio estimator which may have superior performance. We started by introducing the LAD portfolio estimator. We showed that when returns have a multivariate Student-t distribution with less than 6 degrees of freedom, the LAD estimator asymptotically outperforms the variance estimator. This motivated our consideration of other alternative robust portfolio estimators. Among those alternatives, the Huber portfolio estimator is the most robust, behaving almost like the variance estimator when data are Gaussian, but vastly outperforming it under departures from normality. Our experiments on simulated and historical data suggest that robust portfolio estimation may have a role to play in actual implementations of portfolio selection in the financial markets.

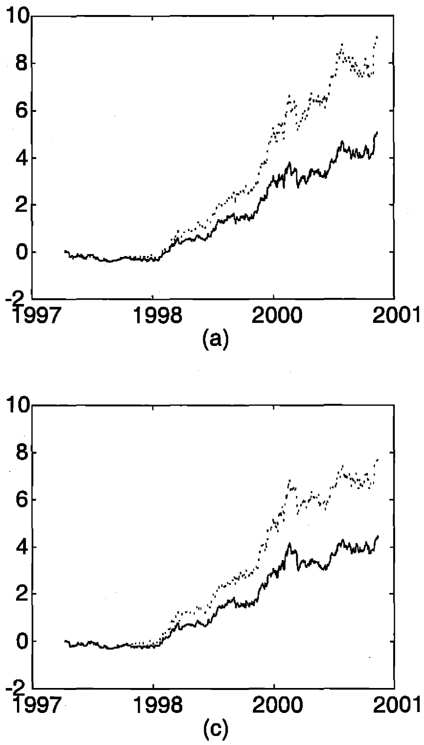

Figure 5-9: Cumulative returns for V (–) and LAD($\cdots$). (a) $r_{p1} = 100\%/260$; (b) $r_{p2} = 80\%/260$; (c) $r_{p3} = 60\%/260$; (d) $r_{p4} = 40\%/260$.

Chapter 6

Regularization in Portfolio Estimation

Let $\mathbf{R}$ be a random return vector in $\mathbb{R}^N$ with mean $\boldsymbol{\mu}$, positive definite covariance matrix Σ, and a continuous density, and suppose that $\mathbf{R}$ has an elliptically symmetric distribution. As in Chapter 5, suppose that we do not know the exact distribution of $\mathbf{R}$, but that we are given independent and identically distributed realizations $\mathbf{R}_1, \ldots, \mathbf{R}_T$ of the random return vector $\mathbf{R}$. Suppose we want to estimate the optimal variance portfolio defined as

$$\mathbf{x}_V = \arg \min_{\mathbf{Ax}=\mathbf{b}} \mathbf{x}^t \Sigma \mathbf{x}, \tag{6.1}$$

where $\mathbf{A}$ is an $(M \times N)$ deterministic matrix with linearly independent rows, $\mathbf{b}$ is an M-dimensional vector, and $\mathbf{b} \neq \mathbf{0}$ so the problem is nontrivial. In Chapter 5, we saw that we could estimate $\mathbf{x}_V$ with

$$(\hat{\mathbf{x}},\hat{q}) = \arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q \in \mathbb{R}}} \left(\frac{1}{T} \sum_{i=1}^{T} \eta(\mathbf{x}^t \mathbf{R}_i - q) \right), \tag{6.2}$$

where q is a measure of location, and where $\eta(.)$ is a cost function, which is $(.)$ for variance portfolio estimation, $|.|$ for LAD portfolio estimation, and

$$\eta(z) = \begin{cases} z^2 & \text{if } |z| \leqslant \gamma \\ 2\gamma|z| - \gamma^2 & \text{if } z > \gamma \end{cases}, \tag{6.3}$$

for some $\gamma > 0$ in Huber portfolio estimation.

Now suppose that the number of observations T is of the same order of magnitude, or less than

N. Then, asymptotic results such as those from the last two chapters are logically useless, and the estimates of the portfolio weights will be very imprecise. Sample portfolio estimation using (6.2) runs the risk of overfitting the data, because there are not enough observations relative to the number of parameters (portfolio weights) that we are estimating. In what follows, we show how to improve the precision of the weight estimates via regularization. Specifically, we consider the family of regularized portfolio estimators of the form

$$(\hat{\mathbf{x}}(\lambda),\hat{q}(\lambda)) = \arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q\in\mathbb{R}}} \left(\frac{1}{T} \sum_{i=1}^{T} \eta(\mathbf{x}^t\mathbf{R}_i - q) + \lambda ||\mathbf{x} - \mathbf{x}_m||_p^p \right), \tag{6.4}$$

where:

- $\lambda \geqslant 0$ is the regularization parameter;
- $||.||_p$ is the L_p-norm in $\mathbb{R}^N$ for $\mathbf{x} \in \mathbb{R}^N$, $||\mathbf{x}||_p = \sum_{j=1}^N |x_i|^p$, where $x_i, i = 1, \ldots, N$, is the ith coordinate of $\mathbf{x}$;
- $\mathbf{x}_m \in \mathbb{R}^N$ is a prior portfolio, and is deterministic.

We will only consider the values of $p = 1$ and 2, which correspond respectively to L1 and L2 regularization[1]. Regularization will be motivated below from a Bayesian perspective. But we note that the inclusion of the penalty term $\lambda ||\mathbf{x} - \mathbf{x}_m||_p^p$ can be explained at an intuitive level as follows. The λ parameter in (6.4) penalizes the objective function in such a way that extreme deviations from the prior are unlikely. The term $\lambda ||\mathbf{x} - \mathbf{x}_m||_p^p$ is a penalty that reflects the investor's a priori confidence in the portfolio $\mathbf{x}_m$. $\mathbf{x}_m$ could be the investor's current position, or a benchmark. A large λ reflects strong confidence in $\mathbf{x}_m$, and conversely, a small λ reflects weak confidence in $\mathbf{x}_m$. In the limit, if $\lambda = 0$, then the estimate $\hat{\mathbf{x}}(0)$ is just, respectively, the - unregularized - sample portfolio estimator. An appropriate choice of λ will reduce the variability of the portfolio estimate, without biasing the estimate too much. We below provide evidence that this is a good thing.

The rest of this chapter is organized as follows. In Section 1, we motivate the use of regularization in portfolio estimation from a Bayesian perspective. In Section 2, we motivate regularization with

[1]Note that in least-squares linear regression L2 regularization corresponds to ridge regression - see Hoerl and Kennard (1970a, 1970b) and Gruber (1998) - and L1 regularization corresponds to the "lasso" - see Tibshirani (1994).

a bias vs. variance argument. In Section 3 we suggest an algorithm to choose the optimal value of λ. In Section 4 we carry out computational experiments on simulated and historical data that show that regularized portfolio estimators may have better performance than their unregularized ($\lambda = 0$) counterparts, according to several measures of performance including the information ratio, standard deviation, VaR, and CVaR.

6.1 Regularization from a Bayesian Perspective

6.1.1 Literature on Bayesian Portfolio Selection

The classical Bayesian approach to portfolio selection - see Bawa, Brown and Klein (1979) - is to assume prior distributions on both the mean of the returns, and on their covariance matrix. Assume there exists an informative prior distribution on the parameters $\boldsymbol{\mu}$ and Σ. Call the density of this prior distribution $f_0(\boldsymbol{\mu}, \Sigma)$. Then, given the sample $\mathbf{R}_1, \ldots, \mathbf{R}_T$, the distribution of the parameters can be updated using the formula

$$f(\boldsymbol{\mu}, \Sigma | \mathbf{R}_1, \ldots, \mathbf{R}_T) \sim f(\mathbf{R}_1, \ldots, \mathbf{R}_T | \boldsymbol{\mu}, \Sigma) f_0(\boldsymbol{\mu}, \Sigma),$$

where the term on the left is called the posterior distribution, and where $\sim$ stands for "is proportional to". In this Bayesian framework, the posterior distribution $f(\boldsymbol{\mu}, \Sigma | \mathbf{R}_1, \ldots, \mathbf{R}_T)$ is then used as an input to the mean-variance portfolio selection problem. Using informative priors typically "shrinks" the sample estimate of the parameters towards a prior mean, where the degree of shrinkage is related to the degree of confidence in the prior. Jorion (1986) shrinks the sample mean towards a constant times a vector of ones, where the optimal shrinkage parameter is obtained by empirical Bayes methods. Pastor (2000) uses an informative prior to estimate the mean, where the prior reflects a priori belief in an asset pricing model, such as the CAPM or the APT with a specified set of factors or benchmark portfolios. Ledoit (1995) uses a prior on the covariance matrix to shrink the sample covariance matrix towards a structured covariance matrix such as the identity matrix. His choice of shrinkage parameter depends on the estimation of a matrix norm which serves as a distance between covariance matrices. Frost and Savarino (1986) use an empirical Bayes approach to shrink the sample estimates of both the mean and covariance matrix towards respectively, a constant times a vector of ones, and a matrix with equal diagonal terms and equal off diagonal terms.

Our approach is different in that instead of having a prior on the covariance matrix Σ, we have a prior on the weights $\mathbf{x}$. This may be more natural in situations where a prior portfolio choice $\mathbf{x}_m$ exists, as was discussed in the introduction. In addition, our approach also works for cases in which we are interested in estimating optimal shortfall portfolios $\mathbf{x}_\alpha$ as opposed to optimal variance portfolios $\mathbf{x}_V$. In such cases, the estimation of the covariance Σ is not necessarily useful. We bypass the estimation of the underlying distribution, and in particular Σ, and directly focus on the estimation of the optimal weights that minimize sample risk plus a penalty that depends on the prior distribution for the weights $\mathbf{x}$.

Note that we do not consider the problem of estimating the mean $\boldsymbol{\mu}$. That is, if $\boldsymbol{\mu}$ is part of the constraint matrix $\mathbf{A}$, $\boldsymbol{\mu}$ will be assumed to be given.

6.1.2 Bayesian View of Regularization in Portfolio Selection

For $p = 2$ and $p = 1$, we can think of our estimation problem as fitting into the following Bayesian framework. Assume that given the weight vector $\mathbf{x}$ and location parameter q, deviations $\nu = \mathbf{x}^t\mathbf{R} - q$, conditional on $\mathbf{x}$ and q, have a density

$$f(\nu|\mathbf{x},q) \propto e^{-\eta(\mathbf{x}^t\mathbf{R}-q)}.$$

Notice that in the absence of a prior on $\mathbf{x}$, the maximum likelihood estimate of the parameter $\mathbf{x}$ (along with an estimate of the location parameter q) given $\mathbf{R}_1, \dots, \mathbf{R}_T$, is just

$$(\hat{\mathbf{x}},\hat{q}) = \arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q\in\mathbb{R}}} \frac{1}{T}\sum_{i=1}^{T} \eta(\mathbf{x}^t\mathbf{R}_i - q),$$

which is the sample "risk minimizing" portfolio estimator.

Priors on x and q

We will consider the following two forms for the prior on $\mathbf{x}$.

- The multivariate normal density

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{(N-M)/2}(2\lambda)^{-(N-M)/2}} \exp\left[-\lambda||(\mathbf{x}-\mathbf{x}_m)||_2^2\right].$$

- The multivariate Laplacian density

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2)^{(N-M)}(\lambda)^{-(N-M)}} \exp\left[-\lambda||(\mathbf{x}-\mathbf{x}_m)||_1^1\right].$$

Notice that in both cases, the density $f_{\mathbf{x}}(\mathbf{x})$ integrates to one over the $N-M$ dimensional manifold defined by the constraint set $\mathbf{Ax} = \mathbf{b}$. Notice also that higher values of λ correspond to lower values for the variance of the elements of $\mathbf{x}$. This matches our interpretation of λ as the degree of confidence in $\mathbf{x}_m$.

We will also assume a diffuse prior on the location parameter q,

$$f_q(q) \propto \text{ constant.}$$

Posterior Distributions of x

Now the posterior distribution of $\mathbf{x}$ and q, given the values $\nu_1, \ldots, \nu_T$ - where $\nu_i = \mathbf{x}^t\mathbf{R}_i - q$ - is

$$\begin{aligned} f(\mathbf{x}, q \mid \nu_1, \ldots, \nu_T) &\propto f(\nu_1, \ldots, \nu_T \mid \mathbf{x}, q) f_{\mathbf{x}}(\mathbf{x}) f_q(q) \\ &\propto f(\nu_1, \ldots, \nu_T \mid \mathbf{x}, q) f_{\mathbf{x}}(\mathbf{x}). \end{aligned}$$

It is easy to see that the posterior log-likelihood

$$\log f(\mathbf{x}, q \mid \nu_1, \ldots, \nu_T)$$

is maximized over $\mathbf{x} \in \mathcal{X}$ and $q \in \mathbb{R}$ by solving the regularized problem

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{T}\textstyle\sum_{i=1}^{T} \eta(\mathbf{x}^t\mathbf{R}_i - q) + \lambda||\mathbf{x}-\mathbf{x}_m||_p^p \\ \text{subject to} \quad & \mathbf{Ax} = \mathbf{b} \\ & q \in \mathbb{R}, \end{aligned}$$

where

- $p = 2$ corresponds to a multivariate normal prior on $\mathbf{x}$,
- $p = 1$ corresponds to a multivariate Laplacian prior on $\mathbf{x}$.

We have not yet answered the question of how to find λ, the degree of confidence in the prior. We address that issue in Section 3.

6.2 Approximation Error vs. Estimation Error

As in Chapter 4, let us define the estimation risk of portfolio estimator $\hat{\mathbf{x}}$, satisfying $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$, as

$$L_{\hat{\mathbf{x}}} = E(\hat{\mathbf{x}}^t \Sigma \hat{\mathbf{x}}) - \mathbf{x}_V^t \Sigma \mathbf{x}_V, \tag{6.5}$$

where the expectation $E(.)$ is taken with respect to the sample $\mathbf{R}_1, \cdots, \mathbf{R}_T$, of which $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{R}_1, \cdots, \mathbf{R}_T)$ is a function. We argued in Chapter 5 that estimation risk is the extra risk, over the optimal portfolio variance $\mathbf{x}_V^t \Sigma \mathbf{x}_V$, that is incured by not knowing the distribution of $\mathbf{R}$, and having to estimate the portfolio $\hat{\mathbf{x}}$.

Now consider the family of portfolio estimators

$$(\hat{\mathbf{x}}_\lambda, \hat{q}_\lambda) = \arg \min_{\substack{\mathbf{A}\mathbf{x}=\mathbf{b} \\ q \in \mathbb{R}}} \left(\frac{1}{T} \sum_{i=1}^{T} \eta(\mathbf{x}^t \mathbf{R}_i - q) + \lambda \|\mathbf{x} - \mathbf{x}_m\|_p^p \right),$$

indexed by λ, where $\eta(.)$,$\mathbf{x}_m$, p are given. Define

$$(\mathbf{x}_\lambda, q_\lambda) = \arg \min_{\substack{\mathbf{A}\mathbf{x}=\mathbf{b} \\ q \in \mathbb{R}}} \left(E[\eta(\mathbf{x}^t R - q)] + \lambda \|\mathbf{x} - \mathbf{x}_m\|_p^p \right),$$

Now notice that the estimation risk of $\hat{\mathbf{x}}_\lambda$ can be written as

$$\begin{aligned} L_{\hat{\mathbf{x}}_\lambda} &= [L_{\hat{\mathbf{x}}_\lambda} - L_{\mathbf{x}_\lambda}] + [L_{\mathbf{x}_\lambda}] \\ &= \underbrace{[E(\hat{\mathbf{x}}_\lambda^t \Sigma \hat{\mathbf{x}}_\lambda) - \mathbf{x}_\lambda^t \Sigma \mathbf{x}_\lambda]}_{(a)} + \underbrace{[\mathbf{x}_\lambda^t \Sigma \mathbf{x}_\lambda - \mathbf{x}_V^t \Sigma \mathbf{x}_V]}_{(b)}, \end{aligned}$$

where we call (a) the estimation error, and (b) the approximation error. Notice that when λ is decreased, estimation error is increased but approximation error is decreased. Inversely, when λ is increased, estimation error is decreased, but approximation error is increased. In particular, if $\lambda = 0$, then the approximation error is 0, since by elliptical symmetry of $\mathbf{R}$,

$$\arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q\in\mathbb{R}}} E[\eta(\mathbf{x}^t R - q)] = (\mathbf{x}_V, q_V)$$

On the other hand, if $\lambda = \infty$, then the estimation error is 0. Let $\lambda_0 \geqslant 0$ be the value of λ that minimizes estimation risk, i.e.

$$\lambda_0 = \arg \min_{\lambda > 0} L_{\hat{\mathbf{x}}_\lambda},$$

where $L_{\hat{\mathbf{x}}_\lambda}$. Then λ_0 is between 0 and ∞, and optimally balances estimation error and approximation error.

Of course, in practice, we do not know λ_0, and we have to estimate a λ from the data, which is what we explain how to do in the next section. Let us also note that the degree of approximation error will be intimately linked to the choice of the prior portfolio $\mathbf{x}_m$.

6.3 Choosing the λ Parameter

As argued in the previous section, the optimal value of λ minimizes estimation risk, defined as

$$L_{\hat{\mathbf{x}}} = E(\hat{\mathbf{x}}^t \Sigma \hat{\mathbf{x}}) - \mathbf{x}_V^t \Sigma \mathbf{x}_V.$$

One way to pick λ is to choose a λ value that minimizes an estimate of the term $E(\hat{\mathbf{x}}^t \Sigma \hat{\mathbf{x}})$, the variance of the estimator. We propose to evaluate the variance of the estimator by p-fold cross validation.

We next describe p-fold cross-validation. Let λ be given.

Step 1: remember that our data consists in T observations $\mathbf{R}_1, \ldots, \mathbf{R}_T$. Suppose for simplicity that T is a multiple of p. Divide the T observations into p subsets of T/p observations. Call these subsets $T(j)$, for $j = 1, \cdots, p$.

Step 2: for every $j = 1, \cdots, p$ calculate

$$(\hat{\mathbf{x}}_\lambda(j), \hat{q}_\lambda(j)) = \arg \min_{\substack{\mathbf{Ax}=\mathbf{b} \\ q \in \mathbb{R}}} \left(\frac{1}{card(T \backslash T(j))} \sum_{i \in T \backslash T(j)} [\eta(\mathbf{x}^t \mathbf{R}_i - q)] + \lambda ||\mathbf{x} - \mathbf{x}_m||_p^p \right),$$

where $card(T \backslash T(j))$ is the cardinality of set $T \backslash T(j)$.

Step 3: for every $j = 1, \cdots, p$ the sum of squared errors

$$PE_\lambda(j) = \sum_{i \in T(j)} [(\hat{\mathbf{x}}_\lambda^t(j)\mathbf{R}_i - \hat{q}_\lambda(j))^2].$$

Step 4: calculate the total sum of squared errors

$$PE_\lambda = \sum_{j=1}^{p} PE_\lambda(j).$$

Now given a list of candidate values of λ, one strategy is to chooses the value of λ with minimum total sum of squared errors, which is an estimate of the variance of the portfolio estimator. The extreme case with $p = T$ corresponds to "leave-one-out" cross validation.

6.4 Computational Experiments

6.4.1 Simulated Data

This experiment uses data generated by a multivariate normal distribution with $N = 100$, a correlation matrix with off-diagonal elements equal to 0.5, and with standard deviations generated by a lognormal distribution with dispersion parameter 1/2. Call the resulting covariance matrix Σ, and call $\mathbf{x}_V$ the solution to

$$\begin{aligned} &\text{minimize} \quad \mathbf{x}^t \Sigma \mathbf{x} \\ &\text{subject to} \quad \mathbf{x}^t \mathbf{e} = 1. \end{aligned}$$

We will consider the estimators (where the only constraint is the convexity constraint):

- V: variance;

- V2: L2-regularized variance;
- LAD;
- LAD1: L1-regularized LAD;
- LAD2: L2-regularized LAD;
- H: Huber;
- H2: L2-regularized Huber.

For $T = 60, 120, 200, 400$, we repeat 100 times the following steps:

Step 1: generate the standard deviations according to the description above, and calculate $\mathbf{x}_V$.

Step 2: generate a dataset with T observations, and calculate the estimators. We use five-fold cross-validation to choose the λ parameter for regularization, and we use an equally weighted portfolio as the prior $\mathbf{x}_m$. Note that for $T = 60$, the estimators V and H are not well-defined.

Step 3: calculate the scaled loss for each estimator, where the scaled loss of estimator is defined as

$$l_s(\hat{\mathbf{x}}) = \frac{\hat{\mathbf{x}}^t \Sigma \hat{\mathbf{x}} - \mathbf{x}_V^t \Sigma \mathbf{x}_V}{\mathbf{x}_V^t \Sigma \mathbf{x}_V}.$$

Notice that l_s will always be positive.

Table 6.1 reports the average scaled loss of each estimator, over the 100 replications of the simulation, as a function of T. Standard errors are in parentheses. EQ corresponds to the equally-weighted prior $\mathbf{x}_m$. The conclusion is that regularization helps significantly when the number of assets is of the same order of magnitude as the number of observations. For example, with $T = 120$, V is almost 500% more risky than the optimal portfolio $\mathbf{x}_V$, but V2 is only 115% more risky than $\mathbf{x}_V$. Even at $T = 200$, the improvement of V2 over V is substantial. Though LAD underperforms V and H, LAD1 performs the best out of all estimators. We suspect this is the effect of L1 regularization rather than the effect of the piecewise linear objective in the LAD. In fact, LAD2 performs worse than V2.

Finally, 6.2 shows the results for the experiment repeated with data generated from a multivariate Student-t with three degrees of freedom, and with the same mean and covariance as before. Though

$T =$	**60**	**120**	**200**	**400**
EQ	47.7781	49.9621	49.6050	46.1965
	(1.3507)	(1.3513)	(1.7711)	(1.2289)
V		4.9940	0.9834	0.3307
		(0.0189)	(0.0020)	(0.0006)
V2	1.9723	1.1582	0.7330	0.3061
	(0.0649)	(0.0032)	(0.0018)	(0.0007)
LAD	207.1015	6.8723	1.4738	0.5042
	(53.7248)	(0.2056)	(0.0211)	(0.0051)
LAD1	1.4714	0.8361	0.5522	0.2861
	(0.0467)	(0.0419)	(0.0268)	(0.0057)
LAD2	2.8534	1.4806	0.9178	0.4573
	(0.0865)	(0.0463)	(0.0239)	(0.0099)
H		5.4425	1.0629	0.3377
		(0.2509)	(0.0224	(0.0052)
H2	3.6554	1.3148	0.7697	0.3257
	(0.5703)	(0.0462)	(0.0286)	(0.0068)

Table 6.1: Average loss over 100 Monte Carlo replications (standard error in parenthesis), Gaussian data.

LAD looses to V and H, LAD1 performs the best out of all estimators. V and H perform similarly, as do V2 and H2.

6.4.2 Historical Data

This experiment uses daily return data on 40 indexes, from 5/14/1992 to 5/14/2001, which amounts to 2243 days of data. The indexes cover various combinations of industries and countries. For example, they include the Dow Jones index and the BTK biotech index, as well as international indexes. The complete list of indexes, and the data, are the property of FleetBoston Financial.

We study the performance of the following estimators:

- V: variance;
- V2: L2-regularized variance;
- LAD: least-absolute deviations;
- LAD2: L2-regularized least-absolute deviations (L2-regularized estimation worked better here, so we do not present the results for L1-regularization);

$T =$	**60**	**120**	**200**	**400**
EQ	47.7791	47.1906	46.4628	48.1079
	(1.7598)	(1.2255)	(1.2218)	(1.1320)
V		5.6606	1.4194	0.5802
		(0.2443)	(0.0311)	(0.0094)
V2	2.1940	1.4452	1.0320	0.5531
	0.0702	(0.0462)	(0.0360)	(0.0160)
LAD	202.2074	8.3680	1.8635	0.6544
	47.8392	(0.4122)	(0.0435)	(0.0118)
LAD1	1.7469	0.9589	0.6888	0.4011
	0.0742	(0.0283)	(0.0264)	(0.0105)
LAD2	2.5914	1.6243	1.1243	0.5924
	(0.0765)	(0.0508)	(0.0390)	(0.0114)
H		6.6413	1.4526	0.5602
		(0.3174)	(0.0325)	(0.0090)
H2	4.4259	1.4830	1.0297	0.5361
	(1.0115)	(0.0468)	(0.0385)	(0.0141)

Table 6.2: Average loss over 100 Monte Carlo replications (standard error in parenthesis), Student-t (3 d.f.) data.

- H: Huber estimator, as defined in Chapter 5;
- H2: L2-regularized Huber.

For every estimator, we use the following simple investment strategy: use a 100 day rolling window of past daily returns, and rebalance the portfolio every 5 days. We do not model transaction costs. The portfolio optimization problems that we solve include the convexity constraint, and a target return constraint, i.e.

$$\begin{aligned} \mathbf{x}^t \mathbf{e} &= 1 \\ \mathbf{x}^t \bar{\mathbf{R}} &= r_p, \end{aligned}$$

for r_p in the range 80%/260 to 180%/260, and where $\bar{\mathbf{R}}$ is the sample average return over the 100 day window of past returns. The unregularized estimators V, LAD, and H, also have the box constraints that every estimated weight must be in the interval [-5,5]. Notice that this actually **improves** the performance of V, LAD, and H: without these box constraints, the information ratios that we obtained are all close to 1 (as opposed to 2.5, as we will see below). The resulting stream of ex-post portfolio returns is collected for each estimator/target return combination. We record the following statistics of the ex-post returns of each estimator/target return combination.

- **mean:** the sample mean of daily ex-post returns;
- **STD:** the sample standard deviation of daily ex-post returns;
- **information ratio:** we have defined the information ratio here as a signal to noise ratio equal to $\frac{\sqrt{260}(\text{mean})}{\text{STD}}$, where the standardization by $\sqrt{260}$ makes the information ratio an annual estimate, assuming 260 trading days per year;
- **α-VaR** for $\alpha = 5\%$ and 1%: the sample α-quantile of the daily ex-post return distribution;
- **α-CVaR** for $\alpha = 5\%$ and 1%: the sample conditional mean of the daily ex-post return distribution, given they are below the α-quantile;
- **MaxDD:** the maximum drawdown, defined as the largest percent decrease in the value of the portfolio over the period under consideration;
- **CRet**: cumulative return.

Figure 6-2 shows the resulting minimum-variance frontiers of the ex-post returns for V2, LAD2, and H2. The V2 estimator has a better mean-variance profile than the other two estimators. Table 6.4 reports the results for $r_{p1} = 180\%/260, r_{p2} = 160\%/260, r_{p3} = 140\%/260, r_{p4} = 120\%/260$. Again, in Table 6.4, the numbers in bold indicate the ex-post performance of the corresponding estimator was **better** than the performance of estimator V2, according to the corresponding statistic. The striking result is the dramatic improvement in the performance of the regularized portfolio estimators, as opposed to their unregularized counterparts, according to every measure of performance that we are considering. Sharpe ratios, for example, are improved by 20%. For example, V with target return r_{p1} has an information ratio of 2.47, but V2 has an information ratio of 3.02. We attribute this improvement in performance to the fact that regularization, with the λ parameter chosen by cross-validation, optimally balances the approximation and estimation error of the estimation procedure. Given that we have 40 indexes and only 100 returns, achieving this balance results in significant improvement in performance. Figures 6-3, 6-4, 6-5 show, in mean-standard deviation space, estimators V, LAD, and H respectively, and their L2-regularized counterpart.

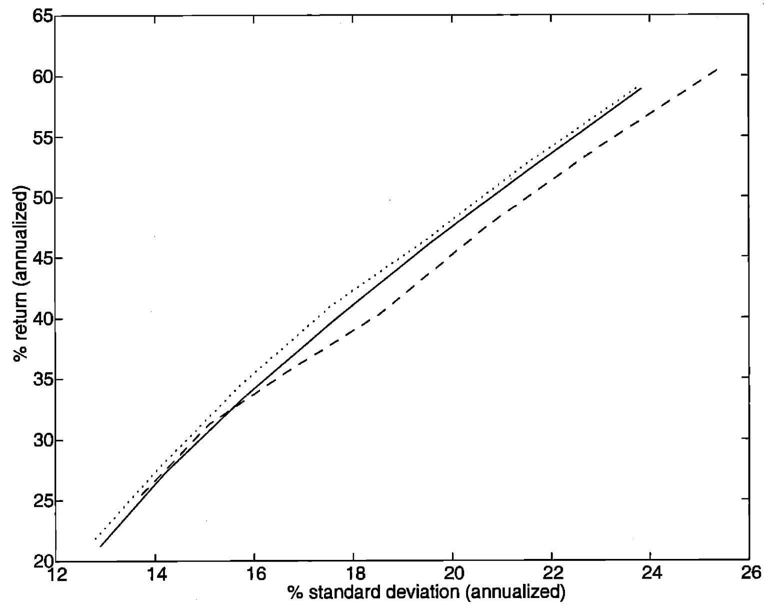

Figure 6-1: Minimum variance (standard deviation) frontier, estimating returns from past: – = V; - - = LAD; -. = H.

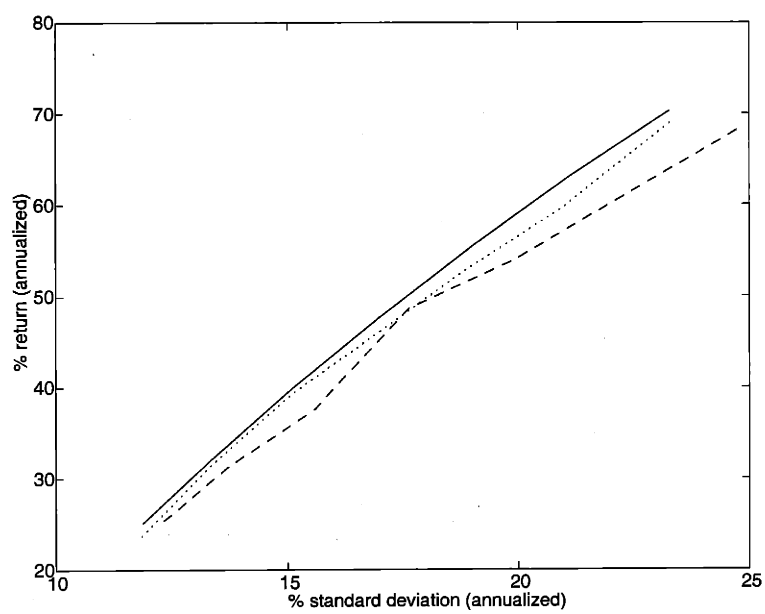

Figure 6-2: Minimum variance (standard deviation) frontier, estimating returns from past: – = V2; - - = LAD2; -. = H2.

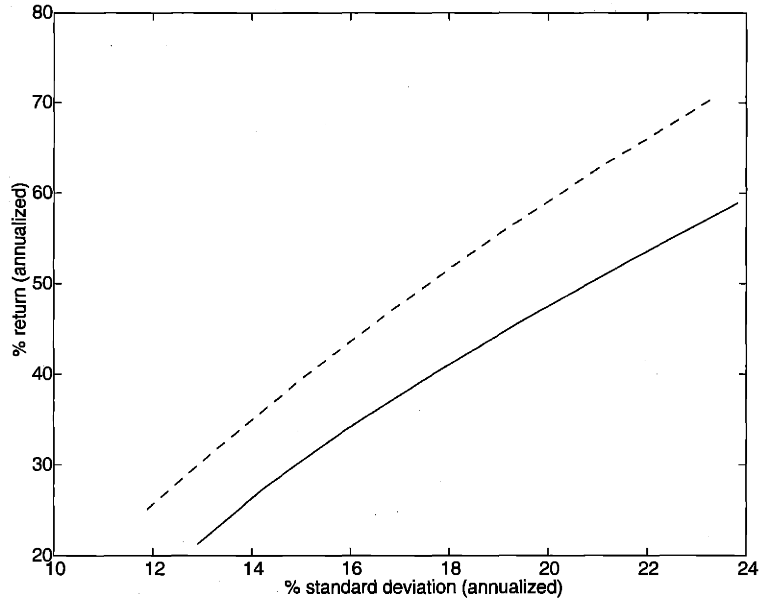

Figure 6-3: Minimum variance (standard deviation) frontier, estimating returns from past: – = V; - - = V2.

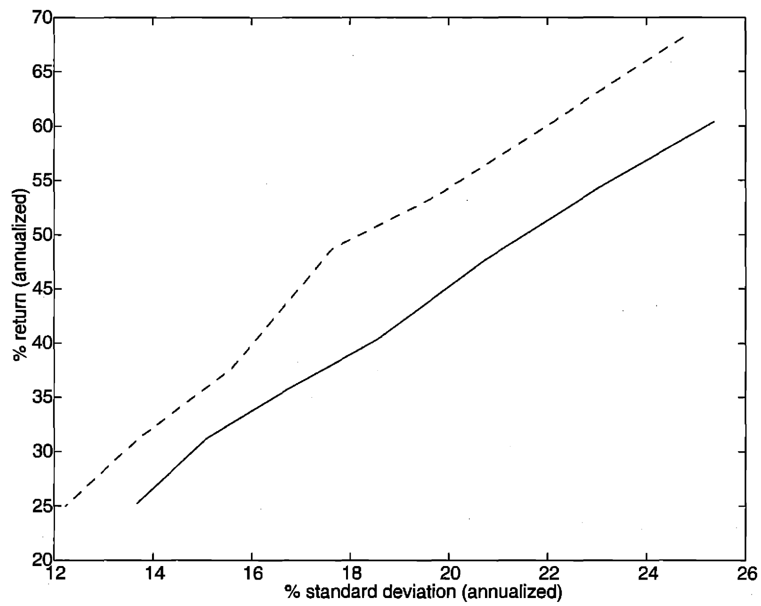

Figure 6-4: Minimum variance (standard deviation) frontier, estimating returns from past: – = LAD; - - = LAD2.

	V (r_{p1})	LAD (r_{p1})	H (r_{p1})	V (r_{p2})	LAD (r_{p2})	H (r_{p2})
mean	0.2265	**0.2323**	**0.2267**	0.2019	**0.2082**	**0.2031**
STD	1.4776	1.5725	**1.4714**	1.3421	1.4252	**1.3352**
information ratio	2.4720	2.3818	**2.4839**	2.4253	2.3559	**2.4525**
5%-VaR	2.3150	2.4514	**2.2790**	2.1203	2.2537	**2.0865**
1%-VaR	3.5635	3.7846	3.5746	3.3625	3.5228	3.3771
5%-CVaR	3.1012	3.3481	3.1157	2.8469	3.0614	2.8581
1%-CVaR	4.2158	4.6147	4.2218	3.8755	4.2014	3.8853
MaxDD	-41.33	**-31.75**	-41.32	-39.99	**-29.47**	-40.24
CRet	2297.6	**2454.5**	**23057**	1615.3	**1752.1**	**1648.6**

	V (r_{p3})	LAD (r_{p3})	H (r_{p3})	V (r_{p4})	LAD (r_{p4})	H (r_{p4})
mean	0.1774	**0.1830**	**0.1784**	0.1528	**0.1551**	**0.1571**
STD	1.2125	1.2850	**1.2069**	1.0909	1.1502	**1.0852**
information ratio	2.3592	2.2958	**2.3836**	2.2587	2.1749	**2.3346**
5%-VaR	1.9643	2.0663	**1.9329**	1.7822	1.8432	**1.7382**
1%-VaR	3.1134	3.3295	**3.0777**	2.8849	3.0695	**2.7556**
5%-CVaR	2.6036	2.8185	2.6141	2.3811	2.5734	2.3866
1%-CVaR	3.5725	3.8268	**3.5681**	3.3227	3.5458	**3.2832**
MaxDD	-38.97	**-31.34**	-39.17	-38.04	**-30.64**	**-37.77**
CRet	1125.8	**1212.7**	**1145.2**	771.2	**792.8**	**829.2**

Table 6.3: V, LAD and H Performance on index data. 100 day rolling window. All statistics are expressed as percentages, except for the Sharpe ratio.

6.5 Conclusion

In this chapter we have shown that in situations where the number of assets is of the same order of magnitude as the number of observations, L1 and L2 regularization can improve the performance of portfolio estimators. We justified regularization from a Bayesian perspective by arguing that intuitively, deviations from a prior portfolio should be penalized. From a frequentist perspective, we argued that there existed values of the regularization parameter such that approximation error and estimation error were optimally balanced. We suggested choosing the regularization parameter by using p-fold cross-validation. Experiments with artificial and real data showed that regularization, in the context of portfolio selection, works in practice.

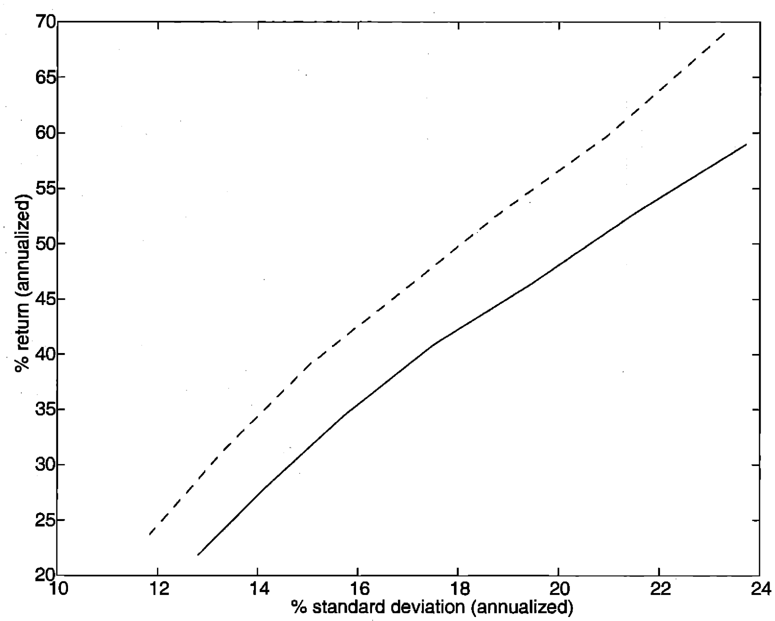

Figure 6-5: Minimum variance (standard deviation) frontier, estimating returns from past: – = H; - - = H2.

	V2 (r_{p1})	**LAD2 (r_{p1})**	**H2 (r_{p1})**	**V2 (r_{p2})**	**LAD2 (r_{p2})**	**H2 (r_{p2})**
mean	0.2702	0.2621	0.2652	0.2422	0.2329	0.2297
STD	1.4428	1.5334	1.4437	1.3079	1.3746	**1.3006**
information ratio	3.0198	2.7562	2.9618	2.9859	2.7326	2.8482
5%-VaR	2.1340	2.2578	**2.1158**	1.9549	2.0579	1.9662
1%-VaR	3.3162	3.5945	**3.3085**	3.1178	3.2700	**3.1129**
5%-CVaR	2.9348	3.1960	2.9676	2.6816	2.9005	2.7176
1%-CVaR	4.3542	4.8266	**4.1545**	4.0087	4.4207	**3.8792**
MaxDD	-22.69	-23.60	-25.39	-22.72	-24.84	-24.69
CRet	4494.7	3898.8	4166.4	3026.2	2592.8	2505.9

	V2 (r_{p3})	**LAD2 (r_{p3})**	**H2 (r_{p3})**	**V2 (r_{p4})**	**LAD2 (r_{p4})**	**H2 (r_{p4})**
mean	0.2131	0.2070	0.2035	0.1829	**0.1873**	0.1760
STD	1.1774	1.2310	**1.1696**	1.0518	1.0937	**1.0476**
information ratio	2.9190	2.7116	2.8049	2.8038	2.7617	2.7094
5%-VaR	1.7482	1.8569	1.7807	1.5843	1.6540	**1.5827**
1%-VaR	2.8585	3.0822	2.9173	2.6016	2.7324	2.6363
5%-CVaR	2.4450	2.6198	2.4748	2.2236	2.3597	2.2533
1%-CVaR	3.6707	4.0600	**3.5955**	3.3555	3.7134	**3.3204**
MaxDD	-22.56	-21.91	-23.84	-22.30	-22.89	-23.14
CRet	1987.6	1789.3	1712.7	1264.9	**1347.2**	1134.6

Table 6.4: V2, LAD2 and H2 Performance on index data. 100 day rolling window. All statistics are expressed as percentages, except for the information ratio.

Chapter 7

Conclusion

Let $\mathbf{R}$ be a random return vector in $\mathbb{R}^N$ with mean $\boldsymbol{\mu}$, and assume that a sample of return observations $\mathbf{R}_1, \ldots, \mathbf{R}_T$ is available. In the introduction to this thesis, we noted that we could rewrite the sample variance minimization problem

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^t \hat{\Sigma} \mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1 \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p, \end{aligned}$$

where r_p is an arbitrary target return, and where $\hat{\Sigma}$ is the sample covariance matrix, as

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \min_q \tfrac{1}{T} \textstyle\sum_{i=1}^{T} (\mathbf{x}^t \mathbf{R}_i - q)^2 \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1 \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p. \end{aligned} \tag{7.1}$$

The rest of this thesis examined the properties of portfolio selection where the objective in Problem (7.1) is replaced by a piecewise linear function. Specifically, we considered the alternative portfolio selection problem

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \min_q \tfrac{1}{T} \textstyle\sum_{i=1}^{T} [\rho_\alpha(\mathbf{x}^t \mathbf{R}_i - q)] \\ \text{subject to} \quad & \mathbf{x}^t \mathbf{e} = 1 \\ & \mathbf{x}^t \boldsymbol{\mu} = r_p, \end{aligned} \tag{7.2}$$

where

$$\rho_\alpha(z) = z - \frac{1}{\alpha} z 1_{\{z<0\}}, \tag{7.3}$$

for $\alpha \in (0,1)$. We called the term

$$\hat{s}_\alpha(\mathbf{x}) = \min_q \frac{1}{T} \sum_{i=1}^{T} [\rho_\alpha(\mathbf{x}^t \mathbf{R}_i - q)], \tag{7.4}$$

the (sample) α-shortfall, or shortfall, of portfolio $\mathbf{x} \in \mathbb{R}^N$.

In the presence of return distribution asymmetry, we showed that the shortfall has an intuitive advantage over variance. As a measure of risk in portfolio selection, the shortfall can capture downside, tail risk for $\alpha \leq 1/2$. The shortfall is related to VaR and CVaR, which are quantile-based measures of risk which are currently receiving recognition in the field of risk management. And if returns are elliptically symmetric, shortfall is proportional to the standard deviation. Moreover, classical mean-variance portfolio analysis results such as two-fund separation in the presence of riskless asset, and the concept of an asset's beta, generalize to shortfall.

We showed that the sample shortfall portfolio optimization problem can be formulated as a linear program with a number of constraints proportional to the number of observations. We derived a central limit theorem for the shortfall portfolio estimator. In the presence of departures from normality within the class of elliptically symmetric distributions, and in particular in the presence of heavy tails and tail dependence, we showed that the shortfall portfolio estimator can outperform the variance portfolio estimator. We also introduced other "robust" portfolio estimators, such as the Huber portfolio estimator, that outperform the variance portfolio estimator under departures from normality within the class of elliptically symmetric distributions.

Finally, we introduced the concept of regularization in portfolio estimation, which consists in adding a penalty to the portfolio optimization objective which is proportional to the norm of the difference between the portfolio estimate and a prior portfolio. We considered L1 and L2 regularization, corresponding to penalties that are proportional respectively to the L1 and L2 norm of the portfolio. We showed that when the number of return observations was of the same order of magnitude as, or less than, the dimension N of the portfolio, regularization could dramatically improve portfolio estimator performance.

7.1 Future Research

Using definition (7.4) and (7.3), we can rewrite the α-shortfall of portfolio $\mathbf{x}$ as

$$\begin{aligned}
\hat{s}_\alpha(\mathbf{x}) &= \min_q \frac{1}{T}\sum_{i=1}^{T}\left[\rho_\alpha(\mathbf{x}^t\mathbf{R}_i - q)\right] \\
&= \min_q \frac{1}{T}\sum_{i=1}^{T}\left[|\mathbf{x}^t\mathbf{R}_i - q|1_{\{\mathbf{x}^t\mathbf{R}_i > q\}} + \left(\frac{1}{\alpha}-1\right)|\mathbf{x}^t\mathbf{R}_i - q|1_{\{\mathbf{x}^t\mathbf{R}_i \leqslant q\}}\right].
\end{aligned}$$

For $\alpha = 50\%$, we saw in this thesis that shortfall is the mean absolute deviation

$$\hat{s}_\alpha(\mathbf{x}) = \min_q \frac{1}{T}\sum_{i=1}^{T}|\mathbf{x}^t\mathbf{R}_i - q|.$$

A natural extension of the shortfall is a measure of portfolio risk which we call α-variance, and which we define as

$$\begin{aligned}
\hat{\sigma}^2_\alpha(\mathbf{x}) &= \min_q \frac{1}{T}\sum_{i=1}^{T}\left[\rho^2_\alpha(\mathbf{x}^t\mathbf{R}_i - q)\right] \\
&= \min_q \frac{1}{T}\sum_{i=1}^{T}\left[(\mathbf{x}^t\mathbf{R}_i - q)^2 1_{\{\mathbf{x}^t\mathbf{R}_i > q\}} + \left(\frac{1}{\alpha}-1\right)(\mathbf{x}^t\mathbf{R}_i - q)^2 1_{\{\mathbf{x}^t\mathbf{R}_i \leqslant q\}}\right].
\end{aligned}$$

For $\alpha = 50\%$, the α-variance $\sigma^2_\alpha(\mathbf{x})$ is the variance

$$\hat{\sigma}^2_\alpha(\mathbf{x}) = \min_q \frac{1}{T}\sum_{i=1}^{T}(\mathbf{x}^t\mathbf{R}_i - q)^2$$

of portfolio $\mathbf{x}$. In Figure 7-1, we show $\rho_\alpha(.)$ and $\rho^2_\alpha(.)$ for $\alpha = 10\%$ and 50%.

We conjecture that the α-variance is related to third order stochastic dominance. α-variance portfolio optimization would fit naturally into the framework developed in Chapter 3. Like shortfall, the α parameter can be chosen to penalize downside (below the location parameter) returns more severely than upside returns.

The central limit theorems that we proved assumed that the constraint set $\mathbf{A}\mathbf{x} = \mathbf{b}$ was determin-

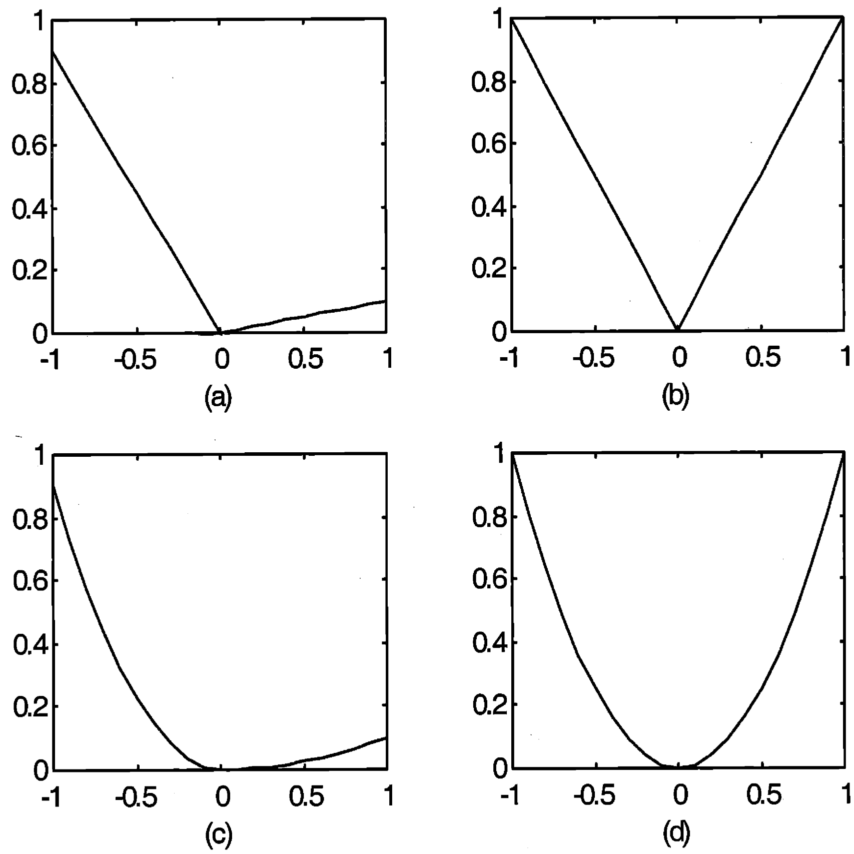

Figure 7-1: (a): $\alpha\rho_\alpha(.)$, $\alpha = 10\%$; (b): $\rho_\alpha(.)$, $\alpha = 50\%$; (c): $\alpha\rho^2_\alpha(.)$, $\alpha = 10\%$; (d): $\rho^2_\alpha(.)$, $\alpha = 50\%$

istic. An extension of this framework would be to let $\mathbf{Ax} = \mathbf{b}$ depend on the sample $\mathbf{R}_1, \cdots, \mathbf{R}_T$. In particular, the case where $\mathbf{Ax} = \mathbf{b}$ contains the constraint

$$\mathbf{x}^t\bar{\mathbf{R}} = r_p,$$

where rp is the target return, and $\bar{\mathbf{R}}$ is the sample mean, might be considered. Another extension would be to consider higher term asymptotics, which might be a useful tool to cope with the issue of small samples that is going to plague any portfolio estimation problem with a large number of assets.

On the robustness front, other portfolio estimators might be experimented with, in addition to the ones we considered. Any robust procedure used in the location and regression problems could conceivably be adapted to the portfolio selection problem. Of these, let us mention high breakdown estimation procedures such as the least median of squares, the least trimmed sum of squared, and the least trimmed sum of absolute deviations (see Hawkins and Olive, 1999). Their optimization, however, will be more difficult than the optimization of the convex portfolio selection problems that we considered here.

We feel that the topic of regularization in portfolio estimation deserves further attention. Though we have showed that it works extremely well in practice, we feel that our mathematical motivation could be significantly improved upon. And maybe more importantly, better ways of choosing the regularization parameter could be devised, that might be less computationally intensive than the p-fold cross-validation.

Finally, the extension of the themes that animated this thesis - namely, portfolio selection under departures from normality, and the application of linear programming to portfolio optimization - to the multi-period framework would undoubtedly be of practical interest.

Appendix A

Review of Stochastic Dominance for Classes of Utility Functions

The expected utility maximization paradigm[1] states that investors make investment decisions in order to maximize their expected utility. That is, faced with two investment choices, whose random returns can be described by X and Y, an investor with utility $u(.)$ chooses X over Y if and only if $E[u(X)] > E[u(Y)]$. Instead of focusing on specific utility functions, which may be hard to articulate for any particular investor, the financial economics literature has also analyzed the properties of classes of utility functions. The following classes have been considered.

$$\begin{aligned}
\mathbb{U}_1 \; &: \; = \{u(x) \mid u(x) \text{ is finite for every finite x, } u'(x) > 0 \; \forall x \in \mathbb{R}\}; \\
\mathbb{U}_2 \; &: \; = \{u(x) \mid u(x) \in \mathbb{U}_1, \; -\infty < u''(x) < 0 \; \forall x \in \mathbb{R}\}; \\
\mathbb{U}_3 \; &: \; = \{u(x) \mid u(x) \in \mathbb{U}_2, \; \infty > u'''(x) > 0 \; \forall x \in \mathbb{R}\};
\end{aligned}$$

Note that $\mathbb{U}_1 \supset \mathbb{U}_2 \supset \mathbb{U}_3$. $\mathbb{U}_1$ corresponds to the class of (nonsatiable) wealth seeking investors. $\mathbb{U}_2$ corresponds to the class of wealth-seeking, risk-averse investors.(investors in this class will refuse to participate in bets were the expected return is 0, which follows directly from Jensen's inequality). $\mathbb{U}_3$ corresponds to the class of wealth-seeking, risk-averse investors with a preference for skewness preference.

Stochastic dominance theorems can be used to determine if, between two investment alternatives X

[1] The material in this appendix can be found in Huang and Litzenberger (1988), Bawa (1975), and Levy (1992).

and Y, one is preferred by all investors belonging to a certain class of utility functions. A distribution F stochastically dominates a distribution G for all utility functions in class $\mathbb{U}_i$ if F is always preferred to G by any investor with utility function $u \in \mathbb{U}_i$. This last condition can be expressed mathematically as

$$E_F(u) > E_G(u) \text{ for all } u \in \mathbb{U}_i,$$

where $E_F(.)$ stands for the expectation with respect to distribution F. Stochastic dominance for all utility functions in $\mathbb{U}_1$ is called First Order Stochastic Dominance (FSD). Similarly, stochastic dominance for utility functions in $\mathbb{U}_2$ is called Second Order Stochastic Dominance (SSD), and for all utility functions in $\mathbb{U}_3$ is called Third Order Stochastic Dominance (FSD). Necessary and sufficient conditions for each type of stochastic dominance are given in the theorem below, adapted from Bawa (1975).

Theorem 35 *F FSD G if and only if*

$$F(x) \leqslant G(x) \ \forall x \in \mathbb{R}, \textit{ and } < \textit{ for some } x.$$

F SSD G if and only if

$$\int_a^x F(t)dt \leqslant \int_a^x G(t)dt \ \forall x \in \mathbb{R}, \textit{ and } < \textit{ for some } x.$$

F TSD G if and only if

$$\begin{aligned} \mu_F &\geqslant \mu_G \textit{ and} \\ \int_a^x \int_a^t F(y)dydt &\leqslant \int_a^x \int_a^t G(y)dydt \ \forall x \in \mathbb{R}, \textit{ and } < \textit{ for some } x. \end{aligned}$$

A.1 Stochastic Dominance and Lower Partial Moment (LPM) Conditions on Distributions

Define the i^{th} order lower partial moment of distribution F, calculated at point c as

$$LPM_i(c;F) := \int_{-\infty}^{x} (x-t)^i dF(t).$$

Notice that $LPM_0(c;F) = F(c)$ is the probability that the return is below c, which is Roy's (1952) Safety-First measure of risk. The following theorem is due to Bawa (1975).

Theorem 36 (Bawa, 1975) *F FSD G if and only if*

$$LPM_0(c;F) \leqslant LPM_0(c;G) \ \forall c \in \mathbb{R}, \text{ and } < \text{ for some } x.$$

F SSD G if and only if

$$LPM_1(c;F) \leqslant LPM_1(c;G) \ \forall x \in \mathbb{R}, \text{ and } < \text{ for some } x.$$

F TSD G if and only if

$$\begin{aligned} \mu_F &\geqslant \mu_G \text{ and} \\ LPM_2(c;F) &\leqslant LPM_2(c;G) \ \forall x \in \mathbb{R}, \text{ and } < \text{ for some } x. \end{aligned}$$

Proof. The proof for FSD is obvious. For SSD, notice that

$$\begin{aligned} \int_{-\infty}^{c} F(t)dt &= \left[F(t)t \mid_a^c - \int_a^c tdF(t) \right] \\ &= \left[F(c)c - F(a)a - \int_a^c tdF(t) \right] \\ &= \int_{-\infty}^{c} (c-t)dF(t) \\ &= LPM_1(c;F). \end{aligned}$$

For TSD, notice that

$$\begin{aligned} LPM_2(c;F) &= \int_{-\infty}^{c} (c-t)^2 dF(t) \\ &= \left[(c-t)^2 F(t)\right]_{-\infty}^{x} + 2\int_{-\infty}^{c} (c-t)F(t)dt \\ &= 2\left\{[(c-t)\int_{-\infty}^{t} F(y)dy]_{-\infty}^{c} + \int_{-\infty}^{c}\int_{-\infty}^{t} F(y)dydt\right\} \\ &= 2\int_{-\infty}^{c}\int_{-\infty}^{t} F(y)dydt, \end{aligned}$$

which finishes the proof. ∎

A.2 Stochastic Dominance and Quantile-Based Conditions on Distributions

First and second-order conditions for stochastic dominance, can be restated in terms of quantile-based risk measures, namely VaR and CVaR. We define the α-quantile of distribution F as

$$q_\alpha(F) := \inf\{x|F(x) \geqslant \alpha\}.$$

Then, the α-level VaR of distribution F is defined as

$$VaR_\alpha(F) = -q_\alpha(F)$$

and the α-level CVaR of distribution F is defined as

$$CVaR_\alpha(F) = -E[X \mid X \leqslant q_\alpha(F)].$$

The following theorem is due to Levy and Kroll (1978). They provide a geometric justification for their result. We offer an analytical derivation of the result, assuming that the densities of the variables under consideration are well-defined.

Theorem 37 (Levy and Kroll, 1978) *Let X and Y be two random variables with distributions F*

and G respectively, and assume that their densities are well-defined. Then F FSD G if and only if

$$q_\alpha(F) \geqslant q_\alpha(G) \ \forall \alpha \in (0,1), \textit{ and } > \textit{ for some } \alpha,$$

or, alternatively,

$$VaR_\alpha(F) \leqslant VaR_\alpha(G) \ \forall \alpha \in (0,1), \textit{ and } > \textit{ for some } \alpha.$$

F SSD G if and only if

$$E(X|X \leqslant q_\alpha(F)) \geqslant E(Y|Y \leqslant q_\alpha(G)) \ \forall \alpha \in (0,1), \textit{ and } > \textit{ for some } \alpha,$$

or, alternatively,

$$CVaR_\alpha(F) \leqslant CVaR_\alpha(G) \ \forall \alpha \in (0,1), \textit{ and } > \textit{ for some } \alpha,$$

Proof. The proof of the FSD condition is obvious when one remembers the definitions of $F(x)$ and $q_\alpha(F)$, and is omitted. To prove the SSD condition, we need to show that

$$\begin{aligned} \int_{-\infty}^{x} F(t)dt \quad &\leqslant \quad \int_{-\infty}^{x} G(t)dt \ \forall x \in \mathbb{R}, \text{ and } < \text{ for some } x \\ \iff \int_{-\infty}^{q_\alpha(F)} tdF(t) \ &\geqslant \int_{-\infty}^{q_\alpha(G)} tG(t) \ \forall \alpha \in (0,1), \text{ and } > \text{ for some } \alpha. \end{aligned}$$

To show $(\Rightarrow)$ notice that for any $\alpha \in (0,1)$ and any $x \in \mathbb{R}$, using integration by parts and the fact that $F(q_\alpha(F)) = \alpha$ for any distribution F, we have

$$\begin{aligned} & \int_{-\infty}^{q_\alpha(F)} tdF(t) - \int_{-\infty}^{q_\alpha(G)} tdG(t) \\ = \ & [tF(t)]_a^{q_\alpha(F)} - \int_{-\infty}^{q_\alpha(F)} F(t)dt - [tG(t)]_a^{q_\alpha(G)} + \int_{-\infty}^{q_\alpha(G)} G(t)dt \\ = \ & q_\alpha(F)\alpha - \int_{-\infty}^{q_\alpha(F)} F(t)dt - q_\alpha(G)\alpha + \int_{-\infty}^{q_\alpha(G)} G(t)dt \\ = \ & \int_{-\infty}^{x} (G(t) - F(t))dt + (q_\alpha(F) - x)\alpha - \int_{x}^{q_\alpha(F)} F(t)dt + (x - q_\alpha(G))\alpha - \int_{q_\alpha(G)}^{x} G(t)dt. \quad \text{(A.1)} \end{aligned}$$

Now assume that $\int_{-\infty}^{x} F(t)dt \leqslant \int_{-\infty}^{x} G(t)dt \ \forall x \in \mathbb{R}$, and let $\alpha \in (0,1)$ be given. Choose $x = q_\alpha(G)$. Then

$$(x - q_\alpha(G))\alpha - \int_{q_\alpha(G)}^{x} G(t)dt = 0,$$

and if $q_\alpha(F) \geqslant x$,

$$(q_\alpha(F) - x)\alpha - \int_{x}^{q_\alpha(F)} F(t)dt \geqslant 0,$$

since for $t \in [x, q_\alpha(F)]$, $F(t) \leqslant \alpha$, and Equation (A.1) is greater than or equal to 0; and if $q_\alpha(F) \leqslant x$,

$$\begin{aligned}(q_\alpha(F) - x)\alpha - \int_{x}^{q_\alpha(F)} F(t)dt &= -(x - q_\alpha(F))\alpha + \int_{q_\alpha(F)}^{x} F(t)dt \\ &\geqslant 0,\end{aligned}$$

since for $t \in [q_\alpha(F), x]$, $F(t) \geqslant \alpha$,and Equation (A.1) is greater than or equal to 0. It follows that

$$\int_{-\infty}^{x} F(t)dt \leqslant \int_{-\infty}^{x} G(t)dt \forall x \in \mathbb{R} \Rightarrow \int_{-\infty}^{q_\alpha(F)} tdF(t) - \int_{-\infty}^{q_\alpha(G)} tdG(t) \geqslant 0 \ \forall \alpha \in (0,1).$$

Also, assume that $\int_{-\infty}^{x} F(t)dt < \int_{-\infty}^{x} G(t)dt$ for a given $x \in \mathbb{R}$. Let $\alpha := G(x)$. We then have $q_\alpha(G) \leqslant x$. Then

$$(x - q_\alpha(G))\alpha - \int_{q_\alpha(G)}^{x} G(t)dt = 0,$$

since $G(t) = \alpha$ for $t \in [q_\alpha(G), x]$, and we also have

$$(q_\alpha(F) - x)\alpha - \int_{x}^{q_\alpha(F)} F(t)dt \geqslant 0,$$

as shown above. It follows that Equation (A.1) is greater than 0,

$$\int_{-\infty}^{x} F(t)dt < \int_{-\infty}^{x} G(t)dt \Rightarrow \int_{-\infty}^{q_\alpha(F)} tdF(t) > \int_{-\infty}^{q_\alpha(G)} tdG(t) \text{ for } \alpha = G(x).$$

We are done proving ($\Rightarrow$).

To show ($\Leftarrow$) notice that for any $x \in \mathbb{R}$ and any $\alpha \in (0,1)$, we can rewrite Equation (A.1) as

$$\begin{aligned} &\int_{-\infty}^{x} (F(t) - G(t))dt \\ = &\int_{-\infty}^{q_\alpha(G)} tdG(t) - \int_{-\infty}^{q_\alpha(F)} tdF(t) + (q_\alpha(F) - x)\alpha \\ &- \int_{x}^{q_\alpha(F)} F(t)dt + (x - q_\alpha(G))\alpha - \int_{q_\alpha(G)}^{x} G(t)dt. \end{aligned} \tag{A.2}$$

Now assume that $\int_{-\infty}^{q_\alpha(F)} tdF(t) \geqslant \int_{-\infty}^{q_\alpha(G)} tdG(t)$ for all $\alpha \in (0,1)$, and let x be given. Choose $\alpha = F(x)$. Then $q_\alpha(F) \leq x$ and

$$(q_\alpha(F) - x)\alpha - \int_{x}^{q_\alpha(F)} F(t)dt = 0$$

since $F(t) = \alpha$ for $t \in [q_\alpha(F), x]$, and if $q_\alpha(G) \leqslant x$,

$$(x - q_\alpha(G))\alpha - \int_{q_\alpha(G)}^{x} G(t)dt \leqslant 0,$$

since for $t \in [q_\alpha(G), x]$, $G(t) \geqslant \alpha$, and Equation (A.2) is less than or equal to 0; and if $q_\alpha(G) \geqslant x$,

$$(x - q_\alpha(G))\alpha - \int_{q_\alpha(G)}^{x} G(t)dt \leqslant 0,$$

since for $t \in [x, q_\alpha(G)]$, $G(t) \leqslant \alpha$,and Equation (A.2) is less than or equal to 0. It follows that

$$\int_{-\infty}^{q_\alpha(F)} tdF(t) - \int_{-\infty}^{q_\alpha(G)} tdG(t) \geqslant 0 \ \forall \alpha \in (0,1) \Rightarrow \int_{-\infty}^{x} F(t)dt \leqslant \int_{-\infty}^{x} G(t)dt \forall x \in \mathbb{R}.$$

Also assume that $\int_{-\infty}^{q_\alpha(F)} tdF(t) > \int_{-\infty}^{q_\alpha(G)} tdG(t)$ for a given $\alpha \in (0,1)$. Choose $x = q_\alpha(F)$. Then

$$(q_\alpha(F) - x)\alpha - \int_{x}^{q_\alpha(F)} F(t)dt = 0,$$

and if $q_\alpha(G) \leqslant x$, and as shown above,

$$(x - q_\alpha(G))\alpha - \int_{q_\alpha(G)}^{x} G(t)dt \leqslant 0.$$

It follows that Equation (A.2) is less than 0,

$$\int_{-\infty}^{q_\alpha(F)} tdF(t) > \int_{-\infty}^{q_\alpha(G)} tdG(t) \Rightarrow \int_{-\infty}^{x} F(t)dt \leqslant \int_{-\infty}^{x} G(t)dt \text{ for } x = q_\alpha(F).$$

We are done proving ($\Leftarrow$). ∎

Appendix B

Multivariate Elliptically Symmetric Distributions

This appendix is dedicated to reviewing some facts about multivariate elliptically symmetric distributions, and to presenting some results related to those distributions. The results that we present are used in Chapter 5.

B.1 Definition

Let $\mathbf{R}$ be a random vector in $\mathbb{R}^N$. In the following definition, a singular density refers to a density that is defined over a hyperplane that has dimension less than N.

Definition 38 $\mathbf{R}$ *has a multivariate elliptical distribution if its (possibly singular) density can be expressed as*

$$f(\mathbf{r}) = |\prod_{i|\lambda_i \neq 0} \lambda_i|^{-1/2}\; g[(\mathbf{r}-\boldsymbol{\mu})^t \Omega^-(\mathbf{r}-\boldsymbol{\mu}); K], \tag{B.1}$$

where Ω is the nonnegative definite dispersion matrix with eigenvalues $\lambda_i, i = 1, \ldots, N$, K is the rank of Ω, Ω^- is its generalized inverse and $\boldsymbol{\mu}$ is the vector of medians.

The generalized inverse of Ω is defined as the matrix Ω^- that satisfies

$$\Omega\Omega^-\Omega = \Omega.$$

Then, when Ω is invertible, $\Omega^- = \Omega^{-1}$. If the mean of $\mathbf{R}$ is well-defined, then $\boldsymbol{\mu}$ is equal to the mean of $\mathbf{R}$. If $\mathbf{R}$ has a well-defined covariance matrix Σ, then Ω is proportional to the covariance matrix of $\mathbf{R}$, i.e.$\Omega = \tau\Sigma$ for some $\tau > 0$.

If $K < N$, then $f(\mathbf{r})$ integrates to one over a hyperplane of dimension K, containing $\boldsymbol{\mu}$. For example, consider in $\mathbb{R}^2$ the vector $\mathbf{R} =(R, 2R+1)^t$, where $R \sim N(1,1)$. $\mathbf{R}$ has mean $\boldsymbol{\mu} =(1,3)^t$ and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \sqrt{3}\mathbf{v}\mathbf{v}^t.$$

with $\mathbf{v} = (1/\sqrt{3}, 2/\sqrt{3})^t$. Then

$$\Sigma^- = \frac{1}{\sqrt{3}}\mathbf{v}\mathbf{v}^t$$

As defined, $\mathbf{R}$ has a singular multivariate Gaussian distribution

$$f(\mathbf{r}) = (\sqrt{3})^{-1/2}\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\mathbf{r} - \boldsymbol{\mu})^t\Sigma^-(\mathbf{r} - \boldsymbol{\mu})\right].$$

Let $M = \{\mathbf{r} \in \mathbb{R}^2 | \mathbf{r} = \boldsymbol{\mu}+k\mathbf{v},\ k \in \mathbb{R}\}$. M is a hyperplane of dimension one. It is easy to verify that

$$\int_M f(\mathbf{r})d\mathbf{r} = 1$$

- but note that $\int_{\mathbb{R}^2} f(\mathbf{r})d\mathbf{r}$ is not well-defined.

B.2 Characteristic Function and Stability Under Affine Transformations

Suppose that $\mathbf{R}$ has a multivariate elliptically symmetric distribution with density given by (B.1), and let $E(.)$ denote the expectation with respect to $\mathbf{R}$. Then its characteristic function is, by definition, a function $\phi_{\mathbf{R}} : \mathbb{R}^N \to \mathbb{R}$, such that

$$\begin{aligned} \phi_{\mathbf{R}}(\mathbf{z}) &= E(e^{i\mathbf{z}^t\mathbf{R}}) \\ &= e^{i\mathbf{z}^t\boldsymbol{\mu}}\psi(\mathbf{z}^t\Omega^-\mathbf{z}) \text{ for some function } \psi(.) \text{ - see for example Muirhead, 1982.} \end{aligned} \tag{B.2}$$

Proposition 39 *Suppose that* $\mathbf{R}$ *has a multivariate elliptically symmetric distribution. Any affine transformation of* $\mathbf{R}$ *also has a multivariate elliptically symmetric distribution..*

Proof. Let $\mathbf{U} = H\mathbf{R} + \mathbf{v}$, where H is an $(M \times N)$ matrix and $\mathbf{v} \in \mathbb{R}^M$. The characteristic function of $\mathbf{U}$ is

$$\begin{aligned} \phi_{\mathbf{U}}(\mathbf{z}) &= E(e^{i\mathbf{z}^t(H\mathbf{R}+\mathbf{v})}) \\ &= e^{i\mathbf{z}^t(H\boldsymbol{\mu}+\mathbf{v})}\psi(\mathbf{z}^tH\Omega^-H^t\mathbf{z}), \end{aligned} \tag{B.3}$$

so $\mathbf{U}$ has a multivariate elliptically symmetric distribution with mean $H\boldsymbol{\mu} + \mathbf{v}$ and dispersion matrix $H\Omega^-H^t$. ■

B.3 Conditional Expectation and Conditional Covariance

Let $\mathbf{R}$ have a multivariate elliptically symmetric distribution with density given by (B.1), characteristic function given by (B.2), and partition $\mathbf{R}$, $\boldsymbol{\mu}$, and Ω as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \tag{B.4}$$

where Ω_{11} is $N_1 \times N_1$ and Ω_{22} is $(N - N_1) \times (N - N_1)$. To get the conditional mean and conditional covariance of $\mathbf{R}_1$ given $\mathbf{R}_2$, we will need the following Lemma.

Lemma 40 *Let Ω, a positive nonnegative definite matrix, be partitioned as in (B.4) Then*

$$N(\Omega_{22}) \subset N(\Omega_{12})$$

and

$$C(\Omega_{21}) \subset C(\Omega_{22}),$$

where $N(.)$ stands for the null-space of the argument, and $C(.)$ stands for the column space of the argument.

Proof. See Muirhead (1982), Lemma 1.2.10. ∎

The following Proposition is based on Problem 1.28 in Muirhead (1982).

Proposition 41 *Let $\mathbf{R}$ have a multivariate elliptically symmetric distribution, and let $\mathbf{R}$ be partitioned as in (B.4). Then, conditional on $\mathbf{R}_2$, $\mathbf{R}_1$ has mean*

$$E(\mathbf{R}_1|\mathbf{R}_2) = \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-}(\mathbf{R}_2 - \boldsymbol{\mu}_2)$$

covariance matrix proportional to $\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-}\Omega_{21}$, that is,

$$Cov(\mathbf{R}_1|\mathbf{R}_2) = h(\mathbf{R}_2)\Omega_{11.2}$$

for some function $h(.)$.

Proof. By Lemma 40 we have $C(\Omega_{21}) \subset C(\Omega_{22})$. Therefore, there exists a $N_1 \times (N - N_1)$ matrix B such that $\Omega_{21} = \Omega_{22}B^t$, or $\Omega_{21}^t = \Omega_{12} = B\Omega_{22}$. We can therefore write

$$\Omega_{12}\Omega_{22}^{-}\Omega_{22} = B\Omega_{22}\Omega_{22}^{-}\Omega_{22} = B\Omega_{22} = \Omega_{12}, \tag{B.5}$$

where the second equality follows by definition of the generalized inverse. Let

$$F = \begin{bmatrix} \mathbf{I}_{N_1} & -\Omega_{12}\Omega_{22}^{-} \\ \mathbf{0} & \mathbf{I}_{N-N_1} \end{bmatrix}. \tag{B.6}$$

Now consider the random vector

$$
\begin{aligned}
F\mathbf{R} &= \begin{bmatrix} \mathbf{I}_{N_1} & -\Omega_{12}\Omega_{22}^{-} \\ \mathbf{0} & \mathbf{I}_{N-N_1} \end{bmatrix} \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{R}_1 - \Omega_{12}\Omega_{22}^{-}\mathbf{R}_2 \\ \mathbf{R}_2 \end{pmatrix}.
\end{aligned}
$$

The characteristic function of $F\mathbf{R}$ is

$$
\begin{aligned}
\phi_{F\mathbf{R}}(\mathbf{z}) &= E(e^{i\mathbf{z}^t(F\mathbf{R})}) \\
&= e^{i\mathbf{z}^t(F\boldsymbol{\mu})}\psi(\mathbf{z}^t F\Omega F^t \mathbf{z}),
\end{aligned}
$$

for some function $\psi(.)$. $F\mathbf{R}$ therefore has an elliptically symmetric distribution with mean

$$
F\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 - \Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{pmatrix},
$$

and dispersion matrix (letting $\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-}\Omega_{21}$)

$$
\begin{aligned}
F\Omega F^t &= \begin{bmatrix} \mathbf{I}_{N_1} & -\Omega_{12}\Omega_{22}^{-} \\ \mathbf{0} & \mathbf{I}_{N-N_1} \end{bmatrix} \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} \\ -\Omega_{22}^{-}\Omega_{21} & \mathbf{I}_{N-N_1} \end{bmatrix} \\
&= \begin{bmatrix} \Omega_{11.2} & \mathbf{0} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} \\ -\Omega_{22}^{-}\Omega_{21} & \mathbf{I}_{N-N_1} \end{bmatrix} \quad \text{using (B.5)} \\
&= \begin{bmatrix} \Omega_{11.2} & \mathbf{0} \\ \mathbf{0} & \Omega_{22} \end{bmatrix}. \qquad \text{(B.7)}
\end{aligned}
$$

The rank of $F\Omega F^t$ is equal to then rank of Ω, since F is invertible. Since $F\mathbf{R} = (\mathbf{U}, \mathbf{R}_2)$ has an elliptically symmetric distribution, its density can be written as

$$
\begin{aligned}
& f(\mathbf{u}, \mathbf{r}_2) \\
= \; & |\prod_{i|\lambda_i \neq 0} \lambda_{Fi}|^{-1/2} \\
& g[(\mathbf{u} - (\boldsymbol{\mu}_1 - \Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))^t \Omega_{11.2}^{-}(\mathbf{u} - (\boldsymbol{\mu}_1 - \Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2) + (\mathbf{r}_2 - \boldsymbol{\mu}_2)^t \Omega_{22}^{-}(\mathbf{r}_2 - \boldsymbol{\mu}_2)); K], \quad \text{(B.8)}
\end{aligned}
$$

for some $g(.)$, where λ_{Fi} are the eigenvalues $F\Omega F^t$. Therefore, conditional on $\mathbf{R}_2$, $\mathbf{R}_1$ has mean

$$E(\mathbf{R}_1|\mathbf{R}_2) = \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-}(\mathbf{R}_2 - \boldsymbol{\mu}_2)$$

covariance matrix proportional to $\Omega_{11.2}$, that is,

$$\begin{aligned} Cov(\mathbf{R}_1|\mathbf{R}_2) &= h(\mathbf{R}_2)\Omega_{11.2} \\ &= h(\mathbf{R}_2)(\Omega_{11} - \Omega_{12}\Omega_{22}^{-}\Omega_{21}). \end{aligned}$$

■

B.4 Examples of Multivariate Elliptical Distributions

B.4.1 Multivariate Gaussian Distribution

Definition 42 *Random vector $\mathbf{R} \in \mathbb{R}^N$ is said to have a multivariate Gaussian distribution if its density has the form*

$$f(\mathbf{r}) = \frac{1}{(2\pi)^{K/2}}|\prod_{i|\lambda_i \neq 0} \lambda_i|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^t\Omega^{-}(\mathbf{r}-\boldsymbol{\mu})\right].$$

where K is the rank of Ω, and $\lambda_i, i = 1, \ldots, N$ are its eigenvalues.

If $\mathbf{R}$ has the density given above, then

$$E(\mathbf{R}) = \boldsymbol{\mu}$$

and

$$Cov(\mathbf{R}) = \Omega.$$

The following proposition

Proposition 43 *Let $\mathbf{R}$ have a multivariate Gaussian distribution, and let $\mathbf{R}$ be partitioned as in (B.4).*

Then, conditional on $\mathbf{R}_2$, $\mathbf{R}_1$ *has mean*

$$E(\mathbf{R}_1|\mathbf{R}_2) = \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-}(\mathbf{R}_2 - \boldsymbol{\mu}_2)$$

covariance matrix equal to

$$Cov(\mathbf{R}_1|\mathbf{R}_2) = \Omega_{11} - \Omega_{12}\Omega_{22}^{-}\Omega_{21}.$$

Proof. Partitioning $\mathbf{R}$, $\boldsymbol{\mu}$, and Ω as in (B.4), we see that the marginal density of $\mathbf{R}_1$ is

$$f(\mathbf{r}_1) = \frac{1}{(2\pi)^{K_1/2}}|\prod_{i|\theta_i\neq 0}\theta_i|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{r}_1-\boldsymbol{\mu}_1)^t\Omega_{11}^{-}(\mathbf{r}_1-\boldsymbol{\mu}_1)\right]$$

where K_1 is the rank of Ω_{11}, and $\theta_i, i = 1, \ldots, N_1$ are its eigenvalues. Let F be defined as in (B.6), so that

$$F\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 - \Omega_{12}\Omega_{22}^{-}\mathbf{R}_2 \\ \mathbf{R}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{U} \\ \mathbf{R}_2 \end{pmatrix}.$$

Using (B.8), we see that the joint density of $\mathbf{U}$ and $\mathbf{R}_2$ is

$$\begin{aligned} & f(\mathbf{u}, \mathbf{r}_2) \\ = \; & \frac{1}{(2\pi)^{K/2}}|\prod_{i|\lambda_i\neq 0}\lambda_{Fi}|^{-1/2} \\ & \exp\left[-\frac{1}{2}\left[(\mathbf{u} - (\boldsymbol{\mu}_1 - \Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))^t\Omega_{11.2}^{-}(\mathbf{u} - (\boldsymbol{\mu}_1 - \Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2)) + (\mathbf{r}_2-\boldsymbol{\mu}_2)^t\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)\right]\right] \end{aligned}$$

where $\lambda_{Fi}, i = 1, \ldots, N$ are the eigenvalues of $F\Omega F^t$, and K is equal to the rank of $F\Omega F^t$ (which is equal to the rank of Ω by invertibility of F). Order the eigenvalues of $F\Omega F^t$ in such a way that $\lambda_{Fi}, i \leqslant N_1$ are the eigenvalues of $\Omega_{11.2}^{-}$, and $\lambda_{Fi}, N_1 < i \leqslant N$ are the eigenvalues of $\Omega_{11.2}^{-}$. Call $K_1$the rank of $\Omega_{11.2}$, and K_2 the rank of Ω_{22}. Notice that $K = K_1 + K_2$. The conditional distribution of $\mathbf{U}$

given $\mathbf{R}_2$ is by definition equal to

$$\begin{aligned}
& f(\mathbf{u}|\mathbf{r}_2) \\
= \; & f(\mathbf{u},\mathbf{r}_2)/f(\mathbf{r}_2) \\
= \; & \frac{1}{(2\pi)^{K/2}} |\prod_{i|\lambda_i\neq 0} \lambda_{Fi}|^{-1/2} \\
& \exp\left[-\frac{1}{2}\left[(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))^t\Omega_{11.2}^{-}(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))+(\mathbf{r}_2-\boldsymbol{\mu}_2)^t\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)\right]\right] \\
& / \left(\frac{1}{(2\pi)^{(K_2)/2}} |\prod_{N_1<i|\lambda_{Fi}\neq 0} \lambda_{Fi}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{r}_2-\boldsymbol{\mu}_2)^t\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)\right]\right) \\
= \; & \frac{1}{(2\pi)^{K_1/2}} \\
& |\prod_{i\leqslant N_1|\lambda_{Fi}\neq 0} \lambda_{Fi}|^{-1/2} \exp\left[-\frac{1}{2}\left[(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))^t\Omega_{11.2}^{-}(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))\right]\right].
\end{aligned}$$

Therefore, using the definition of $\mathbf{U}$, it can be seen that the conditional distribution of $\mathbf{R}_1$ given $\mathbf{R}_2$ is

$$\begin{aligned}
& f(\mathbf{r}_1|\mathbf{r}_2) \\
= \; & \frac{1}{(2\pi)^{K_{11}/2}} |\prod_{i\leqslant N_1|\lambda_{Fi}\neq 0} \lambda_{Fi}|^{-1/2} \\
& \exp\left[-\frac{1}{2}\left[(\mathbf{r}_1-(\boldsymbol{\mu}_1+\Omega_{12}\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)))^t\Omega_{11.2}^{-}(\mathbf{r}_1-(\boldsymbol{\mu}_1+\Omega_{12}\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)))\right]\right].
\end{aligned}$$

It follows that the conditional mean of $\mathbf{R}_1$ given $\mathbf{R}_2$ is

$$E(\mathbf{R}_1|\mathbf{R}_2) = \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2),$$

and the conditional covariance of $\mathbf{R}_1$ given $\mathbf{R}_2$ is

$$Cov(\mathbf{R}_1|\mathbf{R}_2) = \Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-}\Omega_{21},$$

where the second equality follows by definition of $\Omega_{11.2}$. ■

B.4.2 Multivariate Student-t Distribution

Definition 44 *Random vector* $\mathbf{R} \in \mathbb{R}^N$ *is said to have a multivariate Student t-distribution with* ν *degrees of freedom* ($\nu \in \mathbb{N}^+$, *the set of all positive integers) if its density has the form*

$$f(\mathbf{r}) = C_K(\nu)\frac{|\prod_{i|\lambda_i\neq 0}\lambda_i|^{-1/2}}{[\nu + (\mathbf{r}-\boldsymbol{\mu})^t\Omega^-(\mathbf{r}-\boldsymbol{\mu})]^{(\nu+K)/2}},$$

where K *is the rank of* Ω, *and* $\lambda_i, i = 1, \ldots, N$ *are its eigenvalues, and where*

$$C_K(\nu) = \frac{\nu^{\nu/2}\Gamma(\frac{\nu+K}{2})}{\pi^{K/2}\Gamma(\frac{\nu}{2})}.$$

$\Gamma(.)$ *is the gamma function, defined as*

$$\Gamma(x) = \int_0^\infty x^{u-1}e^{-u}du, \quad x > 0.$$

It can be shown - see Press (1972), Ch. 6, Section 2 - that if $\mathbf{R}$ has the density given above, and $\nu > 1$, then the mean of $\mathbf{R}$ is well defined, and

$$E(\mathbf{R}) = \boldsymbol{\mu}.$$

If ν is equal to 1, then $\mathbf{R}$ is said to have a multivariate Cauchy distribution, and its mean does not exist. Also, if $\nu > 2$

$$Cov(\mathbf{R}) = \frac{\nu}{\nu-2}\Omega.$$

Proposition 45 *Let* $\mathbf{R}$ *have a multivariate Student-t distribution with* ν *degrees of freedom, and let* $\mathbf{R}$ *be partitioned as in (B.4). Then, conditional on* $\mathbf{R}_2$, $\mathbf{R}_1$ *has mean*

$$E(\mathbf{R}_1|\mathbf{R}_2) = \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^-(\mathbf{R}_2 - \boldsymbol{\mu}_2)$$

covariance matrix equal to

$$Cov(\mathbf{R}_1|\mathbf{R}_2) = \frac{(\nu + (\mathbf{r}_2 - \boldsymbol{\mu}_2)^t\Omega_{22}^-(\mathbf{r}_2 - \boldsymbol{\mu}_2))}{(\nu + K_2 - 2)} \left(\Omega_{11} - \Omega_{12}\Omega_{22}^-\Omega_{21}\right).$$

Proof. Partitioning $\mathbf{R}$, $\boldsymbol{\mu}$, and Ω as in (B.4), we see that the marginal density of $\mathbf{R}_1$ is

$$f(\mathbf{r}_1) = C_{K_1}(\nu)\frac{|\prod_{i|\theta_i\neq 0}\theta_i|^{-1/2}}{[\nu + (\mathbf{r} - \boldsymbol{\mu})^t\Omega_{11}^-(\mathbf{r} - \boldsymbol{\mu})]^{(\nu+K_1)/2}}$$

where K_1 is the rank of Ω_{11}, and $\theta_i, i = 1, \ldots, N_1$ are its eigenvalues. Let

$$F\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 - \Omega_{12}\Omega_{22}^-\mathbf{R}_2 \\ \mathbf{R}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{U} \\ \mathbf{R}_2 \end{pmatrix}.$$

Using (B.8), we see that the joint density of $\mathbf{U}$ and $\mathbf{R}_2$ is

$$\begin{aligned} & f(\mathbf{u}, \mathbf{r}_2) \\ = \; & C_K(\nu)\frac{|\prod_{i|\lambda_i\neq 0}\lambda_{Fi}|^{-1/2}}{[\nu + (\mathbf{u} - (\boldsymbol{\mu}_1 - \Omega_{12}\Omega_{22}^-\boldsymbol{\mu}_2))^t\Omega_{11.2}^-(\mathbf{u} - (\boldsymbol{\mu}_1 - \Omega_{12}\Omega_{22}^-\boldsymbol{\mu}_2)) + (\mathbf{r}_2 - \boldsymbol{\mu}_2)^t\Omega_{22}^-(\mathbf{r}_2 - \boldsymbol{\mu}_2)]^{(\nu+K)/2}} \end{aligned}$$

where $\lambda_{Fi}, i = 1, \ldots, N$ are the eigenvalues of $F\Omega F^t$, and K is equal to the rank of $F\Omega F^t$ (which is equal to the rank of Ω by invertibility of F). Order the eigenvalues of $F\Omega F^t$ in such a way that $\lambda_{Fi}, i \leqslant N_1$ are the eigenvalues of $\Omega_{11.2}^-$, and $\lambda_{Fi}, N_1 < i \leqslant N$ are the eigenvalues of $\Omega_{11.2}^-$. Call $K_1$the rank of $\Omega_{11.2}$, and K_2 the rank of Ω_{22}. Notice that $K = K_1 + K_2$. Now the conditional distribution of

$\mathbf{U}$ given $\mathbf{R}_2$ is

$$
\begin{aligned}
& f(\mathbf{u}|\mathbf{r}_2) \\
= \; & f(\mathbf{u},\mathbf{r}_2)/f(\mathbf{r}_2) \\
= \; & C_K(\nu)|\prod_{i|\lambda_i\neq 0}\lambda_{Fi}|^{-1/2} \\
& [\nu+(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))^t\Omega_{11.2}^{-}(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2)+(\mathbf{r}_2-\boldsymbol{\mu}_2)^t\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)]^{-(\nu+K)/2} \\
& /\left(C_{K_2}(\nu)\frac{|\prod_{N_1<i|\lambda_{Fi}\neq 0}\lambda_{Fi}|^{-1/2}}{[\nu+(\mathbf{r}_2-\boldsymbol{\mu}_2)^t\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)]^{(\nu+K_2)/2}}\right) \\
= \; & \frac{C_K(\nu)}{C_{K_2}(\nu)}(\nu+a_{22})^{(\nu+K_2)/2}|\prod_{i\leqslant N_1|\lambda_{Fi}\neq 0}\lambda_{Fi}|^{-1/2} \\
& [\nu+(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))^t\Omega_{11.2}^{-}(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2)+a_{22}]^{-(\nu+K)/2} \\
= \; & \frac{C_K(\nu)}{C_{K_2}(\nu)}(\nu+a_{22})^{(\nu+K_2)/2}|\prod_{i\leqslant N_1|\lambda_{Fi}\neq 0}\lambda_{Fi}|^{-1/2} \\
& \left[(\nu+K_2)+(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))^t\left[\Omega_{11.2}^{-}\frac{(\nu+K_2)}{(\nu+a_{22})}\right](\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2)\right]^{-((\nu+K_2)+K_1)/2} \\
& \frac{(\nu+K_2)^{(\nu+K)/2}}{(\nu+a_{22})^{(\nu+K)/2}} \\
= \; & \frac{C_K(\nu)}{C_{K_2}(\nu)}(\nu+K_2)^{(\nu+K_2)/2}(\nu+a_{22})^{-K_1/2}(\nu+K_2)^{K_1/2}|\prod_{i\leqslant N_1|\lambda_{Fi}\neq 0}\lambda_{Fi}|^{-1/2} \\
& \left[(\nu+K_2)+(\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2))^t\left[\Omega_{11.2}^{-}\frac{(\nu+K_2)}{(\nu+a_{22})}\right](\mathbf{u}-(\boldsymbol{\mu}_1-\Omega_{12}\Omega_{22}^{-}\boldsymbol{\mu}_2)\right]^{-((\nu+K_2)+K_1)/2}
\end{aligned}
\tag{B.9}
$$

where $a_{22}=(\mathbf{r}_2-\boldsymbol{\mu}_2)^t\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)$. Notice that

$$
\begin{aligned}
(\nu+a_{22})^{-K_1/2}(\nu+K_2)^{K_1/2}|\prod_{i\leqslant N_1|\lambda_{Fi}\neq 0}\lambda_{Fi}|^{-1/2} &= \left[\prod_{i\leqslant N_1|\lambda_{Fi}\neq 0}\left(\frac{(\nu+a_{22})}{(\nu+K_2)}\lambda_{Fi}\right)\right]^{-1/2} \\
&= \left[\prod_{i|\zeta_i\neq 0}\zeta_i\right]^{-1/2},
\end{aligned}
$$

where $\zeta_i, i = 1, , N_1$ are the eigenvalues of $\left[\Omega_{11.2}\frac{(\nu+a_{22})}{(\nu+K_2)}\right]$, and

$$\begin{aligned}\frac{C_K(\nu)}{C_{K_2}(\nu)}(\nu+K_2)^{(\nu+K_2)/2} &= \left(\frac{\nu^{\nu/2}\Gamma(\frac{\nu+K}{2})}{\pi^{K/2}\Gamma(\frac{\nu}{2})}\right) / \left(\frac{\nu^{\nu/2}\Gamma(\frac{\nu+K_2}{2})}{\pi^{K_2/2}\Gamma(\frac{\nu}{2})}\right)\\ &= \frac{(\nu+K_2)^{(\nu+K_2)/2}\Gamma(\frac{\nu+K_2+K_1}{2})}{\pi^{K_1/2}\Gamma(\frac{\nu+K_2}{2})}\\ &= C_{K_1}(\nu+K_2).\end{aligned}$$

It then follows from (B.9) and the definition of the multivariate Student t-distribution, that conditionally on $\mathbf{R}_2$, $\mathbf{U}$ must have a multivariate Student t-distribution with $(\nu + K_2)$ degrees of freedom, and dispersion matrix $\left[\Omega_{11.2}\frac{(\nu+a_{22})}{(\nu+K_2)}\right]$. Therefore, using the definition of $\mathbf{U}$, it can be seen that the conditional density of $\mathbf{R}_1$ given $\mathbf{R}_2$ is

$$\begin{aligned}& f(\mathbf{r}_1|\mathbf{r}_2)\\ =\; & C_{K_1}(\nu+K_2)\left[\prod_{i|\zeta_i\neq 0}\zeta_i\right]^{-1/2}\\ & [(\nu+K_2)\\ & +(\mathbf{r}_1-(\boldsymbol{\mu}_1+\Omega_{12}\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)))^t\left[\Omega_{11.2}\frac{(\nu+a_{22})}{(\nu+K_2)}\right]^{-}(\mathbf{r}_1-(\boldsymbol{\mu}_1+\Omega_{12}\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2)))]^{-(\nu+K_2))+K_1}\end{aligned}$$

It follows that the conditional mean of $\mathbf{R}_1$ given $\mathbf{R}_2$ is

$$E(\mathbf{R}_1|\mathbf{R}_2) = \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2),$$

and the conditional covariance of $\mathbf{R}_1$ given $\mathbf{R}_2$ is

$$\begin{aligned}Cov(\mathbf{R}_1|\mathbf{R}_2) &= \frac{(\nu+K_2)}{(\nu+K_2-2)}\left[\Omega_{11.2}\frac{(\nu+a_{22})}{(\nu+K_2)}\right]\\ &= \frac{(\nu+a_{22})}{(\nu+K_2-2)}\left(\Omega_{11}-\Omega_{12}\Omega_{22}^{-}\Omega_{21}\right)\\ &= \frac{(\nu+(\mathbf{r}_2-\boldsymbol{\mu}_2)^t\Omega_{22}^{-}(\mathbf{r}_2-\boldsymbol{\mu}_2))}{(\nu+K_2-2)}\left(\Omega_{11}-\Omega_{12}\Omega_{22}^{-}\Omega_{21}\right),\end{aligned}$$

where the second equality by definition of $\Omega_{11.2}$, and the third by definition of a_{22}. ■

Appendix C

An Alternative Polynomial Time Algorithm to Solve the Sample α-Shortfall Portfolio Optimization Problem

Assume for simplicity that $K := T\alpha$ is integer. Then, even though formulation (3.7) has an exponential number of constraints, it can be used to develop an algorithm which will solve the sample α-shortfall portfolio optimization problem in polynomial time.

C.1 Complexity According to Formulation (3.7)

The linear program corresponding to formulation (3.7) can be solved in polynomial time. To show this, let us first prove that we can solve the Separation Problem for the polyhedron $P = \{\mathbf{x} \in \mathbb{R}^n, z \in \mathbb{R} \mid \mathbf{Ax} = \mathbf{b}, \frac{1}{K}\sum_{i \in S}(\mathbf{x}^t\mathbf{R}_i) \geqslant z\}$ and where S ranges over all K-element subsets of $\{1, \cdots, T\}$ in polynomial time. The Separation Problem is to:

(a) either decide that $(\mathbf{x}, z) \in P$, or

(b) find a constraint (indexed by S), such that $\frac{1}{K}\sum_{i \in S}(\mathbf{x}^t\mathbf{R}_i) \geqslant z$ is violated.

Let a pair $(\mathbf{x}, z)$ be given. Checking whether $\mathbf{Ax} = \mathbf{b}$ holds can be done in $O(N)$ steps. To find out if an inequality $\frac{1}{K}\sum_{i\in S}(\mathbf{x}^t\mathbf{R}_i) \geqslant z$ is violated for some S which is a K-element subsets of $\{1, \cdots, T\}$, do the following. First calculate

$$\begin{aligned} r_1(\mathbf{x}) \ &: \ = \mathbf{x}^t\mathbf{R}_1, \\ &\vdots \\ r_T(\mathbf{x}) \ &: \ = \mathbf{x}^t\mathbf{R}_T, \end{aligned}$$

which will take $O(NT)$ steps. Then order the sequence $[r_1(\mathbf{x}), \cdots, r_T(\mathbf{x})]$ from smallest to largest, and call the resulting sequence $\left[r_{(1)}(\mathbf{x}), \cdots, r_{(T)}(\mathbf{x})\right]$, which will take $O(T)$ steps If $\frac{1}{K}\sum_{i=1}^{K} r_{(i)}(\mathbf{x}) < z$, we have found a violated constraint. Else if $\frac{1}{K}\sum_{i=1}^{K} r_{(i)}(\mathbf{x}) \geqslant z$, then $(\mathbf{x}, z) \in P$, because $\frac{1}{K}\sum_{i=1}^{K} r_{(i)}(\mathbf{x})$ is a lower bound for $\frac{1}{K}\sum_{i\in S} r_{(i)}(\mathbf{x})$ for all S, by virtue of our ordering. So the Separation Problem can be solved in $O(NT)$ steps. Using Theorem 8.5 in Bertsimas and Tsitsiklis (1997), it follows that the shortfall optimization problem can be solved in time polynomial in $NT + 1$ and $\log U$, where U is a bound on the size of the problem parameter values.

C.2 A Practical Algorithm to Solve the α-Sample Shortfall Portfolio Optimization Problem According to Formulation (3.7)

The algorithm used in Theorem 8.5 by Bertsimas and Tsitsiklis (1997) (see previous Section) is the Ellipsoid Method for Optimization. This algorithm can be used to prove theoretical results, but typically performs poorly in practice. We propose the following practical algorithm to solve the sample

α-shortfall portfolio optimization problem .

Step 1: Start by solving the problem with the equality constraint, and a small subset of the inequality constraints.

Step 2: Find a violated constraint in the entire set of inequalities (that is equivalent to (b) in the Separation Problem above). If no constraints are violated, then the solution is optimal and the algorithm terminates.

Step 3: Add the violated constraint to the constraint set, and solve using the dual-simplex method. This is efficient because the last dual solution stays feasible after a constraint is added, yielding a (typically good) starting point. Go to Step 2.

Appendix D

Proofs from Chapter 1

In this appendix we find optimal (the tightest) upper and lower bounds on the shortfall of a distribution under two different set of assumptions: 1.) we are given the mean μ, variance σ^2, and α-quantile q_α of the distribution; and 2.) we are given the mean μ and variance σ^2 of the distribution. We also provide optimal bounds for the α-quantile q_α of a distribution, given its mean μ and variance σ^2, as not every triplet $(\mu, \sigma^2, q_\alpha)$ in fact corresponds to a distribution. The results are summarized in Table (2.1).

D.1 Bounds on q_α Given μ and σ^2

We find a lower bound on the α-quantile q_α of a distribution, given its mean μ, variance σ^2. We first consider an intermediate problem, that of finding an upper bound on the probability that a random variable with mean μ and variance σ^2 is below a given value q. This problem can be expressed as

$$\begin{array}{ll} \max_f & \int_{-\infty}^{q} f(z)dz \\ \text{s.t.} & \int_{-\infty}^{\infty} f(z)dz = 1 \\ & \int_{-\infty}^{\infty} zf(z)dz = \mu \\ & \int_{-\infty}^{\infty} z^2 f(z)dz = \mu^2 + \sigma^2 \\ & f(z) \geqslant 0. \end{array}$$

Its dual formulation, for which strong duality holds (see Issi, 1960), is

$$\begin{aligned} \min_{u_1,u_2,u_3} \quad & u_1 + u_2\mu + u_3(\mu^2+\sigma^2) \\ \text{s.t.} \quad & u_1 + u_2 z + u_3 z^2 \geqslant 1 \forall z \in (-\infty, q) \\ & u_1 + u_2 z + u_3 z^2 \geqslant 0 \ \forall z \in (q, \infty). \end{aligned} \tag{D.1}$$

Theorem 46 *The optimal upper bound on the probability that a random variable $X \sim (\mu, \sigma^2)$ is below a given scalar q is*

$$\max_{X\sim(\mu,\sigma^2)} \Pr(X \leqslant q) = \begin{cases} \sigma^2/[\sigma^2 + (q-\mu)^2] & \textit{if } q - \mu \leqslant 0, \\ 1 & \textit{if } q - \mu > 0. \end{cases}$$

Proof. The optimal upper bound on the probability that a random variable $X \sim (\mu, \sigma^2)$ is below a given scalar q is the solution to Problem (D.1), which can be rewritten as

$$\begin{aligned} \min_{g(z)} \quad & E_{Z\sim(\mu,\sigma^2)}[g(Z)] \\ \text{s.t.} \quad & g(z) \geqslant 1 \ \forall z \in (-\infty, q) \\ & g(z) \geqslant 0 \ \forall z \in (q, \infty) \\ & g(z) := u_1 + u_2 z + u_3 z^2. \end{aligned}$$

The optimal solution to the preceding problem is also the optimal solution to

$$\begin{aligned} \min_{a,b,c} \quad & E_{Z\sim(\mu,\sigma^2)}[a(Z-b)^2 + c] \\ \text{s.t.} \quad & a(z-b)^2 + c \geqslant 1 \ \forall z \in (-\infty, q) \\ & a(z-b)^2 + c \geqslant 0 \ \forall z \in (q, \infty). \end{aligned}$$

Note that $a \geqslant 0$ is a necessary condition for the function $g(z) := a(z-b)^2 + c$ to be feasible.

Now for any feasible $g(z)$ with $b \leqslant q$, we have $g(z) \geqslant g_1(z) := 1$, and $g_1(z)$ is feasible, $E_{Z\sim(\mu,\sigma^2)}[g_1(Z)] = 1$. For any feasible $g(z)$ with $b \geqslant q$, we have $g(z) \geqslant g_2(z) := a(z-b)^2$ with $g_2(q) = 1$. Therefore

$$a(q-b)^2 = 1 \Leftrightarrow a = \frac{1}{(q-b)^2},$$

so that $g_2(z) = (z-q)^2/(q-b)^2$, and $E_{Z\sim(\mu,\sigma^2)}[g_2(Z)] = [\sigma^2 + (b-\mu)^2]/(b-q)^2 := m(b)$. Notice that

$$m(b) = 1 + \frac{[\sigma^2 + (q-\mu)^2]}{(b-q)^2} + \frac{2(q-\mu)}{(b-q)}$$

and

$$\begin{aligned}\frac{\partial m(b)}{\partial b} &= -\frac{2}{(b-q)^3}[\sigma^2 + (q-\mu)^2 + (q-\mu)(b-q)] \\ &= -\frac{2}{(b-q)^3}[\sigma^2 + (q-\mu)^2 + (q-\mu)(b-q)],\end{aligned}$$

so that

$$\frac{\partial m(b)}{\partial b} = 0 \text{ iff } b - q = \frac{[\sigma^2 + (q-\mu)^2]}{\mu - q}.$$

$m(b)$ is $+\infty$ at $b = q$, $\lim_{b\to+/-\infty} m(b) = 1$, and $m(b)$ has a unique local minimum at $b = [\sigma^2 + (q-\mu)^2]/(\mu - q) + q$. It follows that

$$\arg\min_{b\geqslant q} m(b) = \begin{cases} [\sigma^2 + (q-\mu)^2]/(\mu-q) + q & \text{if } q - \mu \leqslant 0, \\ +\infty & \text{if } q - \mu > 0, \end{cases}$$

and we are done. ■

Now to find a lower bound on the α-quantile $q_\alpha := \inf\{q | \Pr(X \leqslant q) \geqslant \alpha\}$, all we need to do is set the upper bound in the Theorem above to α, which yields

$$-\sigma\sqrt{\frac{(1-\alpha)}{\alpha}} + \mu \leqslant q_\alpha.$$

To find an upper bound on the α-quantile q_α of a distribution with mean μ variance σ^2, we consider the intermediate problem of finding an lower bound on the probability that a random variable with mean μ and variance σ^2 is below a given value q. This problem can be expressed as

$$\begin{array}{ll} \min_f & \int_{-\infty}^{q} f(z)dz \\ \text{s.t.} & \int_{-\infty}^{\infty} f(z)dz = 1 \\ & \int_{-\infty}^{\infty} zf(z)dz = \mu \\ & \int_{-\infty}^{\infty} z^2 f(z)dz = \mu^2 + \sigma^2 \\ & f(z) \geqslant 0. \end{array}$$

Its dual formulation is

$$\begin{array}{ll} \max_{u_1,u_2,u_3} & u_1 + u_2\mu + u_3(\mu^2 + \sigma^2) \\ \text{s.t.} & u_1 + u_2 z + u_3 z^2 \leqslant 1 \forall z \in (-\infty, q) \\ & u_1 + u_2 z + u_3 z^2 \leqslant 0 \ \forall z \in (q, \infty). \end{array} \tag{D.2}$$

Theorem 47 *The optimal lower bound on the probability that a random variable $X \sim (\mu, \sigma^2)$ is below a given scalar q is*

$$\min_{X\sim(\mu,\sigma^2)} \Pr(X \leqslant q) = \begin{cases} 0 & \text{if } q - \mu \leqslant 0, \\ 1 - \sigma^2/[\sigma^2 + (q-\mu)^2] & \text{if } q - \mu > 0. \end{cases}$$

Proof. The optimal lower bound on the probability that a random variable $X \sim (\mu, \sigma^2)$ is below a given scalar q is the solution to Problem (D.2), which can be rewritten as

$$\begin{array}{ll} \max_{g(z)} & E_{Z\sim(\mu,\sigma^2)}[g(Z)] \\ \text{s.t.} & g(z) \leqslant 1 \ \forall z \in (-\infty, q) \\ & g(z) \leqslant 0 \ \forall z \in (q, \infty) \\ & g(z) := u_1 + u_2 z + u_3 z^2. \end{array}$$

The optimal solution - $g(z)$ - to the preceding problem is also the optimal solution to

$$\begin{array}{ll} \max_{a,b,c} & E_{Z\sim(\mu,\sigma^2)}[a(Z-b)^2 + c] \\ \text{s.t.} & a(z-b)^2 + c \leqslant 1 \ \forall z \in (-\infty, q) \\ & a(z-b)^2 + c \leqslant 0 \ \forall z \in (q, \infty). \end{array}$$

Note that $a \leqslant 0$ is a necessary condition for the function $g(z) := a(z-b)^2 + c$ to be feasible.

Now for any feasible $g(z)$ with $b \geqslant q$, $g(z) \leqslant g_1(z) := 0$, and $g_1(z)$ is feasible, $E_{Z\sim(\mu,\sigma^2)}[g_1(Z)] = 0$. For any feasible $g(z)$ with $b \leqslant q$, $g(z) \leqslant g_2(z) := a(z-b)^2 + 1$ with $g_2(q) = 0$. Therefore

$$a(q-b)^2 + 1 = 0 \Leftrightarrow a = -\frac{1}{(q-b)^2},$$

so that $g_2(z) = -(z-q)^2/(q-b)^2 + 1$, and $E_{Z\sim(\mu,\sigma^2)}[g_2(Z)] = 1 - [\sigma^2 + (b-\mu)^2]/(b-q)^2 := m(b)$. Notice that

$$m(b) = -\frac{[\sigma^2 + (q-\mu)^2]}{(b-q)^2} - \frac{2(q-\mu)}{(b-q)}$$

and

$$\begin{aligned}
\frac{\partial m(b)}{\partial b} &= \frac{2}{(b-q)^3}[\sigma^2 + (q-\mu)^2 + (q-\mu)(b-q)] \\
&= \frac{2}{(b-q)^3}[\sigma^2 + (q-\mu)^2 + (q-\mu)(b-q)],
\end{aligned}$$

so that

$$\frac{\partial m(b)}{\partial b} = 0 \text{ iff } b - q = \frac{[\sigma^2 + (q-\mu)^2]}{\mu - q}.$$

$m(b)$ is $-\infty$ at $b = q$, $\lim_{b\to+/-\infty} m(b) = 0$, and $m(b)$ has a unique local maximum at $b = [\sigma^2 + (q-\mu)^2]/(\mu-q) + q$. It follows that

$$\arg\min_{b\leqslant q} m(b) = \begin{cases} -\infty & \text{if } q-\mu \leqslant 0, \\ [\sigma^2 + (q-\mu)^2]/(\mu-q) + q & \text{if } q-\mu > 0, \end{cases}$$

and we are done. ■

Now to find an upper bound all we need to do is set the lower bound in the Theorem above to α, which yields

$$q_\alpha \leqslant \sigma\sqrt{\frac{\alpha}{(1-\alpha)}} + \mu.$$

D.2 Bounds on Shortfall Given μ, $\sigma^2$and q_α

In this Section we find lower and upper bounds on the shortfall s_α of a distribution, given its mean μ, variance σ^2 and α-quantile q_α. Let us first consider the search for an upper bound on the shortfall. The problem we are trying to solve can be expressed as

$$\begin{array}{ll} \max_f & \mu - \frac{1}{\alpha}\int_{-\infty}^{q_\alpha} z f(z)dz \\ \text{s.t.} & \int_{-\infty}^{\infty} f(z)dz = 1 \\ & \int_{-\infty}^{\infty} z f(z)dz = \mu \\ & \int_{-\infty}^{\infty} z^2 f(z)dz = \mu^2 + \sigma^2 \\ & \int_{-\infty}^{q_\alpha} f(z)dz = \alpha \\ & f(z) \geqslant 0. \end{array}$$

Its dual formulation is

$$\begin{array}{lll} \mu + & \min_{u_1,u_2,u_3,u_4} & \frac{1}{\alpha}[u_1 + u_2\mu + u_3(\mu^2+\sigma^2) + u_4\alpha] \\ & \text{s.t.} & u_1 + u_2 z + u_3 z^2 + u_4 \geqslant -z \ \forall z \in (-\infty, q_\alpha) \\ & & u_1 + u_2 z + u_3 z^2 \geqslant 0 \ \forall z \in (q_\alpha, \infty). \end{array} \tag{D.3}$$

The dual can be solved in closed form. The proof relies on the geometry of the underlying problem.

Theorem 48 ***(Optimal upper bound on the shortfall of a distribution)*** *The optimal upper bound on the shortfall s_α of a distribution, given its mean μ, variance σ^2 and α-quantile q_α is*

$$\max_{X\sim(\mu,\sigma^2,q_\alpha)} s_\alpha(X) = \sigma\sqrt{(1-\alpha)/\alpha}.$$

Proof. The optimal upper bound on the shortfall of a distribution, given its mean μ, variance σ^2 and α-quantile q_α, is the solution to Problem (D.3), which can be rewritten as

$$\begin{array}{lll} \mu + & \min_{g(z)} & \frac{1}{\alpha} E_{Z\sim(\mu,\sigma^2,q_\alpha)}[g(Z)] \\ & \text{s.t.} & g(z) \geqslant -z \ \forall z \in (-\infty, q_\alpha) \\ & & g(z) \geqslant 0 \ \forall z \in (q_\alpha, \infty) \\ & & g(z) := u_1 + u_2 z + u_3 z^2 + u_4 1_{\{z \leqslant q_\alpha\}}. \end{array}$$

where $1_{\{z \leqslant q_\alpha\}} = 1$ if $z \leqslant q_\alpha$ and 0 otherwise. The optimal solution - $g(z)$ - to the preceding problem is also the optimal solution to

$$\begin{aligned}
\min_{a,b,c,u_4} \quad & E_{Z\sim(\mu,\sigma^2,q_\alpha)}[a(Z-b)^2 + c + u_4 1_{\{z\leqslant q_\alpha\}}] \\
\text{s.t.} \quad & a(z-b)^2 + c + u_4 \geqslant -z \; \forall z \in (-\infty, q_\alpha) \\
& a(z-b)^2 + c \geqslant 0 \; \forall z \in (q_\alpha, \infty).
\end{aligned} \tag{D.4}$$

Note that $a \geqslant 0$ is a necessary condition for the function $g(z) := a(z-b)^2 + c + u_4 1_{\{z\leqslant q_\alpha\}}$ to be feasible.

Suppose $g(z)$ is optimal. Then it is necessary that $a(z-b)^2 + c + u_4 = -z$ for some $z \in (-\infty, q_\alpha)$; otherwise one could choose $u_4' < u_4$ so that $g'(z) := a(z-b)^2 + c + u_4' 1_{\{z\leqslant q_\alpha\}} < g(z) \; \forall z \in \mathbb{R}$ and $g'(z)$ is feasible.

Now consider the following two cases:

(a) If $g(q_\alpha) = -q_\alpha$ then $b > q_\alpha$ and it follows that $c = 0$. Furthermore, a must be such that $g(z)$ is tangent to the line $-z$ at q_α; otherwise, one could choose $a' < a$, and u_4' such that $g'(z) := a'(z-b)^2 + c + u_4' 1_{\{z\leqslant q_\alpha\}} < g(z) \; \forall z \in \mathbb{R}$ and $g'(z)$ is feasible.

If $g(z^*) = -z^*$ for some $z^* < q_\alpha$, then $g(z)$ must be tangent to the line $-z$ at z^*. Furthermore, one must have $b \geqslant q_\alpha$; otherwise one could choose $a' < a$ and u_4' such that $g'(z) := a'(z-q_\alpha)^2 + u_4' 1_{\{z\leqslant q_\alpha\}} < g(z) \; \forall z \in \mathbb{R}$ and $g'(z)$ is feasible.

We have just shown that Problem (D.4) has the same solution as

$$\begin{aligned}
\min \quad & E[a(z-b)^2 + u_4 1_{\{z\leqslant q_\alpha\}}] \\
\text{s.t.} \quad & a(z-b)^2 + u_4 \geqslant -z \; \forall z \in (-\infty, q_\alpha) \\
& a \geqslant 0, \; b \geqslant q_\alpha, \text{ and } a(z^*-b)^2 + u_4 = -z^* \;\text{ for some } z \in (-\infty, q_\alpha).
\end{aligned} \tag{D.5}$$

Now suppose that $g(z) := a(z-b)^2 + u_4 1_{\{z\leqslant q_\alpha\}}$ a feasible solution of Problem (D.5). We know that

$$\frac{\partial g}{\partial a}(z^*) = -1 \Leftrightarrow 2a(z^*-b) = -1,$$

so that

$$g(z^*) = a(z^*-b)^2 + u_4 = \frac{1}{4a} + u_4 = -z^* = \frac{1}{2a} - b,$$

and

$$u_4 = \frac{1}{4a} - b,$$

with $z^* = b - 1/(2a) \leqslant q_\alpha$. We therefore have

$$\begin{aligned} E[g(z)] &= a[\sigma^2 + (\mu - b)^2] + u_4\alpha \\ &= a[\sigma^2 + (\mu - b)^2] + [\frac{1}{4a} - b]\alpha := m(a,b), \end{aligned}$$

with the constraints $a \geqslant 0$, $b \geqslant q_\alpha$, and $1/(2a) - b \leqslant q_\alpha$. Problem (D.5) is therefore equivalent to

$$\begin{aligned} \min_{a,b} \quad & m(a,b) = a[\sigma^2 + (\mu - b)^2] + [\tfrac{1}{4a} - b]\alpha \\ \text{s.t.} \quad & b - 1/(2a) \leqslant q_\alpha \\ & b \geqslant q_\alpha, \ a \geqslant 0. \end{aligned} \tag{D.6}$$

The Lagrangean for the above problem can be written as

$$\begin{aligned} L \ &: = L(a, b, \lambda_1, \lambda_2, \lambda_3) \\ &= a[\sigma^2 + (\mu - b)^2] + [\frac{1}{4a} - b]\alpha \\ &\quad + \lambda_1(b - 1/(2a) - q_\alpha) + \lambda_2(q_\alpha - b) + \lambda_3(-a), \end{aligned}$$

with $\lambda_1, \lambda_2, \lambda_3 \geqslant 0$. The Kuhn-Tucker necessary conditions are

$$\begin{aligned} \frac{\partial L}{\partial a} &= [\sigma^2 + (\mu - b)^2] - \frac{1}{4a^2}\alpha + \lambda_1 \frac{1}{2a^2} - \lambda_3 = 0 \\ \frac{\partial L}{\partial b} &= -2a(\mu - b) - \alpha + \lambda_1 - \lambda_2 = 0 \\ \lambda_1(b - 1/(2a) - q_\alpha) &= 0 \\ \lambda_2(q_\alpha - b) &= 0 \\ \lambda_3(-a) &= 0. \end{aligned}$$

Since $a = 0$ is not possible, we must have $\lambda_3 = 0$. We are left with the following four cases:

Case 1: $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 = 0$. Then $b = q_\alpha$, and $1/(2a) = 0$, so this scenario is impossible.

Case 2: $\lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0$. Then the optimal parameters are $a^* = \sigma_\alpha/(2\sigma)$ and $b^* = \mu + \alpha\sigma/\sigma_\alpha$, where $\sigma_\alpha := \sqrt{\alpha(1-\alpha)}$. Plugging in these parameters into the function $m(a, b)$ yields the optimal value for Problem (D.6), $m(a^*, b^*) = \sigma\sigma_\alpha - \mu\alpha$. This is the optimal solution when

$$\begin{aligned} b^* - \frac{1}{2a^*} &\leqslant q_\alpha \\ b^* &\geqslant q_\alpha, \end{aligned}$$

or

$$\begin{aligned} \mu + \frac{\sigma}{\sigma_\alpha}(\alpha - 1) &\leqslant q_\alpha \\ \mu + \frac{\alpha\sigma}{\sigma_\alpha} &\geqslant q_\alpha, \end{aligned}$$

which is equivalent to

$$-\sigma\sqrt{\frac{(1-\alpha)}{\alpha}} \leqslant q_\alpha - \mu \leqslant \sigma\sqrt{\frac{\alpha}{(1-\alpha)}}.$$

Case 3: $\lambda_1 = 0, \lambda_2 > 0, \lambda_3 = 0$. Then the optimal parameters are $a^* = \sqrt{\alpha}/\left[2\sqrt{\sigma^2 + (\mu - q_\alpha)^2}\right]$ and $b^* = q_\alpha$. In this case the optimal value for Problem (D.6) is $m(a^*, b^*) = \sqrt{\alpha}\sqrt{\sigma^2 + (\mu - q_\alpha)^2} - q_\alpha\alpha$. For this to be the optimal solution, there must exist $\lambda_2^* > 0$ such that

$$\frac{\partial L}{\partial b} = -2a^*(\mu - q_\alpha) - \alpha - \lambda_2^* = 0.$$

That is, we must have

$$-2a^*(\mu - q_\alpha) - \alpha > 0,$$

or

$$q_\alpha - \mu > \sigma\sqrt{\frac{\alpha}{(1-\alpha)}},$$

where $\sigma_\alpha := \sqrt{\alpha(1-\alpha)}$.

Case 4: $\lambda_1 > 0, \lambda_2 = 0, \lambda_3 = 0$. Then the optimal parameters are $a^* = 1/[2(\sqrt{[\sigma^2 + (\mu - q_\alpha)^2]/(1-\alpha)})]$ and $b^* = q_\alpha + \sqrt{[\sigma^2 + (\mu - q_\alpha)^2]/(1-\alpha)}$. It follows that the optimal value for Problem (D.6) is $m(a^*, b^*) = \sqrt{1-\alpha}\sqrt{\sigma^2 + (\mu - q_\alpha)^2} - q_\alpha \alpha - (\mu - q_\alpha)$. For this to be the optimal solution, there must exist $\lambda_1^* > 0$ such that

$$\frac{\partial L}{\partial b} = -2a^*(\mu - b^*) - \alpha + \lambda_1^* = 0.$$

That is, we must have

$$\alpha + 2a^*(\mu - b^*) > 0,$$

or

$$q_\alpha - \mu < -\sigma\sqrt{\frac{(1-\alpha)}{\alpha}},$$

where $\sigma_\alpha := \sqrt{\alpha(1-\alpha)}$.

We know that given μ and σ^2, possible q_α must satisfy $-\sigma\sqrt{(1-\alpha)/\alpha} \leqslant q_\alpha - \mu \leqslant \sigma\sqrt{\alpha/(1-\alpha)}$, so that only Case 2 above is feasible. The result follows. ■

Next we consider the search for a lower bound on the shortfall s_α of a distribution, given its mean μ, variance σ^2 and α-quantile q_α. This problem can be expressed as

$$\begin{array}{ll} \min_f & \mu - \frac{1}{\alpha}\int_{-\infty}^{q_\alpha} z f(z)dz \\ \text{s.t.} & \int_{-\infty}^{\infty} f(z)dz = 1 \\ & \int_{-\infty}^{\infty} z f(z)dz = \mu \\ & \int_{-\infty}^{\infty} z^2 f(z)dz = \mu^2 + \sigma^2 \\ & \int_{-\infty}^{q_\alpha} f(z)dz = \alpha \\ & f(z) \geqslant 0. \end{array}$$

Its dual formulation is

$$\begin{array}{lll} \mu + & \max_{u_1,u_2,u_3,u_4} & \frac{1}{\alpha}[u_1 + u_2\mu + u_3(\mu^2+\sigma^2) + u_4\alpha] \\ & \text{s.t.} & u_1 + u_2 z + u_3 z^2 + u_4 \leqslant -z \ \forall z \in (-\infty, q_\alpha) \\ & & u_1 + u_2 z + u_3 z^2 \leqslant 0 \ \forall z \in (q_\alpha, \infty). \end{array} \tag{D.7}$$

Theorem 49 ***(Optimal lower bound on the shortfall of a distribution)*** *The optimal lower bound on the shortfall s_α of a distribution, given its mean μ, variance σ^2 and α-quantile q_α is*

$$\min_{X\sim(\mu,\sigma^2,q_\alpha)} s_\alpha(X) = \begin{cases} -(q_\alpha - \mu) & \textit{if } q_\alpha - \mu \leqslant 0, \\ (q_\alpha - \mu)(1-\alpha)/\alpha & \textit{if } q_\alpha - \mu > 0. \end{cases}$$

Proof. The optimal upper bound on the shortfall of a distribution, given its mean μ, variance σ^2 and α-quantile q_α, is the solution to Problem (D.7). Let $g(z) := u_1 + u_2 z + u_3 z^2 + u_4 1_{\{z \leqslant q_\alpha\}}$ be feasible. Let $g_1(z) := -z + q_\alpha(1 - 1_{\{z\leqslant q_\alpha\}})$ and $g_2(z) := -q_\alpha 1_{\{z\leqslant q_\alpha\}}$, and notice that both g_1 and g_2 are feasible. Then if $\min_z\{u_1 + u_2 z + u_3 z^2\} \leqslant q_\alpha$, $g(z) \leqslant g_1(z)$, and if $\min_z\{u_1 + u_2 z + u_3 z^2\} > q_\alpha$, then $g(z) \leqslant g_2(z)$. It follows that the optimal solution to Problem (D.7) is

$$\begin{aligned} & \mu + \frac{1}{\alpha}\max\{E_{Z\sim(\mu,\sigma^2,q_\alpha)}[g_1(z)], E_{Z\sim(\mu,\sigma^2,q_\alpha)}[g_2(z)]\} \\ = \ & \mu + \frac{1}{\alpha}\max\{-\mu + q_\alpha(1-\alpha), -q_\alpha\alpha\}, \end{aligned}$$

and the result follows. ■

D.3 Bounds on Shortfall Given μ and σ^2

The results above allow us to bound shortfall given μ and σ^2 :

Proposition 50 *The optimal lower bound on the shortfall s_α of a distribution, given its mean μ and variance σ^2 is 0. The optimal upper bound on the shortfall s_α of a distribution, given its mean μ, variance σ^2 is $\sigma\sqrt{(1-\alpha)/\alpha}$.*

Appendix E

Proofs from Chapter 2

E.1 Proof of Lemma 15

Proof. The gradient:

Notice that

$$E[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)] = (\mathbf{x}^t\boldsymbol{\mu} - q) - \frac{1}{\alpha}E[(\mathbf{x}^t\mathbf{R}-q)1_{\{\mathbf{x}^t\mathbf{R}-q\leqslant 0\}}]$$

We will first prove (4.14) for the k-th component of $\nabla_{\mathbf{x}}E[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)]$. We have

$$\frac{\partial \rho_\alpha(\mathbf{x}^t\mathbf{R}-q)}{\partial x_k} = \mu_k - \frac{1}{\alpha}\frac{\partial}{\partial x_k}E[(\mathbf{x}^t\mathbf{R} - q)1_{\{\mathbf{x}^t\mathbf{R}-q\leqslant 0\}}].$$

Writing the expectations above as a bivariate integrals in the variables $U = \sum_{i\neq k} x_i R_i$, and $V = R_k$ and differentiating with respect to x_k we obtain

$$\begin{aligned}
\frac{\partial \rho_\alpha(\mathbf{x}^t\mathbf{R}-q)}{\partial x_k} &= \mu_k - \frac{1}{\alpha}\frac{\partial}{\partial x_k}\iint_{\mathbb{R}^2}(u + x_k v - q)1_{\{u+x_k v-q\leqslant 0\}}f_{U,V}(u,v)dudv \\
&= \mu_k - \frac{1}{\alpha}\frac{\partial}{\partial x_k}\int_{-\infty}^{\infty}\int_{-\infty}^{q-x_k v}(u + x_k v - q)f_{U,V}(u,v)dudv \\
&= \mu_k - \frac{1}{\alpha}\int_{-\infty}^{\infty}\int_{-\infty}^{q-x_k v} v f_{U,V}(u,v)dudv \\
&\quad -\frac{1}{\alpha}\int_{-\infty}^{\infty}(-v)0 f_{U,V}(q - x_k v, v)dv \\
&= \mu_k - \frac{1}{\alpha}E[R_k 1_{\{\mathbf{x}^t\mathbf{R}-q\leqslant 0\}}].
\end{aligned}$$

The gradient of $E[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)]$ with respect to q is found using the same arguments.

The Hessian:

Let j and k, $j \neq k$, be fixed indices in $\{1, \cdots, N\}$. We will prove (4.15) for the jk-th component of $\nabla^2_{\mathbf{x}} E[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)]$. We have

$$\frac{\partial E[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)]}{\partial x_k} = \mu_k - \frac{1}{\alpha} E(R_k 1_{\{\mathbf{x}^t\mathbf{R}-q \leqslant 0\}}).$$

Let $U = \sum_{i \neq k,j} x_i R_i$, and $V = R_j$, and $W = R_k$, which have a joint density $f_{U,V,W}(.,.,.)$ since $\mathbf{R}$ is assumed to have a density. Differentiating the last equation with respect to x_j we obtain

$$\begin{aligned}
\frac{\partial E[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)]}{\partial x_j \partial x_k} &= -\frac{1}{\alpha}\frac{\partial}{\partial x_j} \iint_{\mathbb{R}^3} w 1_{\{u + x_j v + x_k w \leqslant q\}} f_{U,V,W}(u,v,w) du dv dw \\
&= -\frac{1}{\alpha}\frac{\partial}{\partial x_j} \int_{-\infty}^{\infty} w \int_{-\infty}^{\infty} \int_{-\infty}^{q - x_j v - x_k w} f_{U,V,W}(u,v,w) du dv dw \\
&= -\frac{1}{\alpha} \int_{-\infty}^{\infty} w \int_{-\infty}^{\infty} (-v) f_{U,V,W}(q - x_j v - x_k w, v, w) dv dw \\
&= -\frac{1}{\alpha} \iint_{\mathbb{R}^N} r_k(-r_j) f_{\mathbf{R},\mathbf{x}^t\mathbf{R}}(\mathbf{r}, q) d\mathbf{r} \\
&= -f_{\mathbf{x}^t\mathbf{R}}(q) \frac{1}{\alpha} \iint_{\mathbb{R}^N} r_k(-r_j) f_{\mathbf{R}|\mathbf{x}^t\mathbf{R}}(\mathbf{r}|q) d\mathbf{r} \\
&= f_{\mathbf{x}^t\mathbf{R}}(q) \frac{1}{\alpha} E[R_k R_j \mid \mathbf{x}^t\mathbf{R} - q = 0].
\end{aligned}$$

The case where $j = k$ is handled the same way.

Now, let k in $\{1, \cdots, N\}$ be fixed, and let $U = \sum_{i \neq k} x_i R_i$ and $V = R_k$. We then have

$$\begin{aligned}
\frac{\partial E[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)]}{\partial q \partial x_k} &= -\frac{1}{\alpha}\frac{\partial}{\partial q} \iint_{\mathbb{R}^2} v 1_{\{u + x_k v \leqslant q\}} f_{U,V}(u,v) du dv \\
&= -\frac{1}{\alpha}\frac{\partial}{\partial q} \int_{-\infty}^{\infty} v \int_{-\infty}^{q - x_k v} f_{U,V}(u,v) du dv \\
&= -\frac{1}{\alpha} \int_{-\infty}^{\infty} v f_{U,V}(q - x_k v, v) du dv \\
&= \frac{1}{\alpha} \iint_{\mathbb{R}^N} r_k f_{\mathbf{R},\mathbf{x}^t\mathbf{R}}(\mathbf{r}, q) d\mathbf{r} \\
&= -f_{\mathbf{x}^t\mathbf{R}}(q) \frac{1}{\alpha} \iint_{\mathbb{R}^N} r_k f_{\mathbf{R}|\mathbf{x}^t\mathbf{R}}(\mathbf{r}|q) d\mathbf{r} \\
&= -f_{\mathbf{x}^t\mathbf{R}}(q) \frac{1}{\alpha} E[R_k \mid \mathbf{x}^t\mathbf{R} - q = 0].
\end{aligned}$$

The expression for $\nabla_q^2 E[\rho_\alpha(\mathbf{x}^t\mathbf{R}-q)]$ follows in similar fashion. ■

E.2 Proof of Lemma 16

Proof. (a) Note that the gradient and Hessian of $E[\rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W}+\mathbf{z}^t\mathbf{W})]$ exist by Assumption (B) and Lemma 15, and we can expand $E[\rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W}+\mathbf{z}^t\mathbf{W})-\rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W})]$ around $\mathbf{y}_\alpha$ to obtain

$$\begin{aligned}
& E[\rho_\alpha((\mathbf{y}_\alpha+\mathbf{z})^t\mathbf{W})-\rho_\alpha(\mathbf{y}_\alpha^t\mathbf{W})] \\
= \;& \mathbf{z}^t\nabla_{\mathbf{y}}E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}+\frac{1}{2}\mathbf{z}^t[\nabla_{\mathbf{y}}^2E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}]\mathbf{z}+o(||\mathbf{z}||^2). \qquad \text{(E.1)}
\end{aligned}$$

Now because $\mathbf{y}_\alpha$ solves

$$\begin{array}{ll}
\text{minimize} & E[\rho_\alpha(\mathbf{y}^t\mathbf{W})] \\
\text{subject to} & \mathbf{A}_0\mathbf{y}=\mathbf{b},
\end{array}$$

we must have (writing the Lagrangean of this last problem and taking its derivative with respect to $\mathbf{y}$)

$$\nabla_{\mathbf{y}}E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}+\mathbf{A}_0^t\boldsymbol{\lambda}_\alpha=\mathbf{0},$$

for some $\boldsymbol{\lambda}_\alpha\in\mathbb{R}^M$. This means that

$$\nabla_{\mathbf{y}}E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}=-\mathbf{A}_0^t\boldsymbol{\lambda}_\alpha,$$

and for all $\mathbf{z}\in\mathcal{Z}$

$$\begin{aligned}
\mathbf{z}^t\nabla_{\mathbf{y}}E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha} &= -\mathbf{z}^t\mathbf{A}_0^t\boldsymbol{\lambda}_\alpha \\
&= \mathbf{0}, \qquad \text{(E.2)}
\end{aligned}$$

by definition of $\mathcal{Z}$. Therefore, the linear part in (E.1) vanishes for all $\mathbf{z}\in\mathcal{Z}$, proving (a).

(b) Consider the quadratic part in (E.1). Use (4.15) to write,

$$\begin{aligned}\nabla^2_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha} &= f_{\mathbf{x}^t\mathbf{R}}(q_\alpha)\begin{bmatrix} Cov[\mathbf{R} \mid \mathbf{x}^t_\alpha\mathbf{R}{-}q_\alpha = 0] & \mathbf{0} \\ \mathbf{0}^t & 0 \end{bmatrix} \\ &\quad + f_{\mathbf{x}^t\mathbf{R}}(q_\alpha)\begin{bmatrix} -E[\mathbf{R}|\ \mathbf{x}^t_\alpha\mathbf{R}{-}q_\alpha = 0] \\ 1 \end{bmatrix}\begin{bmatrix} -E[\mathbf{R}|\ \mathbf{x}^t_\alpha\mathbf{R}{-}q_\alpha = 0] \\ 1 \end{bmatrix}^t\end{aligned}$$

We see that the rank of $\nabla^2_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}$ is equal to the rank of $Cov[\mathbf{R} \mid \mathbf{x}^t_\alpha\mathbf{R}{-}q_\alpha = 0]$ plus one. From Assumption (B), we know that the rank of $Cov[\mathbf{R} \mid \mathbf{x}^t_\alpha\mathbf{R}{-}q_\alpha = 0]$ is $N-1$, so the rank of $\nabla^2_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}$ is N. Let us write

$$\nabla^2_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha} = \sum_{i=1}^{N+1} \lambda_i \mathbf{v}_i \mathbf{v}_i^t,$$

where $\lambda_1 = 0 < \lambda_2 < \cdots < \lambda_{N+1}$, and $\mathbf{v}_i^t\mathbf{v}_j = 0$ for all $i \neq j$, $||\mathbf{v}_i|| = 1$ for all i - i.e the $\mathbf{v}_i$ form an orthonormal basis in $\mathbb{R}^{N+1}$. Therefore for all $\mathbf{z} \in \mathcal{Z}$,

$$\begin{aligned}\mathbf{z}^t[\nabla^2_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}]\mathbf{z} &= \sum_{i=1}^{N+1} \lambda_i(\mathbf{z}^t\mathbf{v}_i)^2 \geqslant \lambda_2 \sum_{i=2}^{N+1}(\mathbf{z}^t\mathbf{v}_i)^2 \\ &= \lambda_2||\mathbf{z}||^2 - \lambda_2(\mathbf{z}^t\mathbf{v}_1)^2 = \lambda_2||\mathbf{z}||^2 - \lambda_2||\mathbf{z}||^2||\mathbf{v}_1||^2\cos(\mathbf{z},\mathbf{v}_1)^2 \\ &\geqslant ||\mathbf{z}||^2\bar{\lambda}, \end{aligned} \tag{E.3}$$

with $\bar{\lambda} := \lambda_2[1 - (\sup_{\mathbf{z}\in\mathcal{Z}}\cos(\mathbf{z},\mathbf{v}_1))^2]$. Using the fact that $\nabla^2_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}$ has rank N it is easy to verify that

$$\mathbf{v}^t[\nabla^2_{\mathbf{y}} E[\rho_\alpha(\mathbf{y}^t\mathbf{W})]|_{\mathbf{y}_\alpha}]\mathbf{v} = 0 \text{ if and only if } \mathbf{v} \text{ is a multiple of } \mathbf{y}_\alpha = (\mathbf{x}^t_\alpha, q_\alpha)^t.$$

Therefore $\mathbf{v}_1$ must be a multiple of $\mathbf{y}_\alpha$, so $\mathbf{A}_0\mathbf{v}_1 \neq \mathbf{0}$, and $\mathbf{v}_1 \notin \mathcal{Z}$ by definition of $\mathcal{Z}$. Therefore $\sup_{\mathbf{z}\in\mathcal{Z}}\cos(\mathbf{z},\mathbf{v}_1) < 1$ and $\bar{\lambda} > 0$, proving (b).

(c) Using (E.2) and (E.3) in (E.1) we see that for all $\mathbf{z} \in \mathcal{Z}$

$$\begin{aligned} & E[\rho_\alpha((\mathbf{y}_\alpha + \mathbf{z})^t \mathbf{W}) - \rho_\alpha(\mathbf{y}_\alpha^t \mathbf{W})] \\ \geqslant\ & \frac{1}{2}||\mathbf{z}||^2 \bar{\lambda} + o(||\mathbf{z}||^2), \end{aligned}$$

which shows that there exists $\epsilon_\alpha > 0$ such that

$$E[\rho_\alpha((\mathbf{y}_\alpha + \mathbf{z})^t \mathbf{W}) - \rho_\alpha(\mathbf{y}_\alpha^t \mathbf{W})] \geqslant \frac{1}{4}\bar{\lambda}||\mathbf{z}||^2 \text{ for all } \mathbf{z} \in \mathcal{Z} \text{ and } ||\mathbf{z}|| \leqslant \epsilon_\alpha.$$

The results follow with $\gamma := \frac{1}{4}\bar{\lambda}$, and by noticing that by convexity of $E[\rho_\alpha((\mathbf{y}_\alpha + \mathbf{z})^t \mathbf{W}) - \rho_\alpha(\mathbf{y}_\alpha^t \mathbf{W})]$ any local minimum is also a global minimum. ■

Appendix F

An Inverse of a Partitioned Matrix

Let $\mathbf{R} \in \mathbb{R}^N$ have mean $\boldsymbol{\mu}$ and covariance matrix Σ, possibly nonnegative definite. Define the matrix

$$M = \begin{bmatrix} \Gamma & -\boldsymbol{\mu} & \mathbf{A}^t \\ -\boldsymbol{\mu}^t & 1 & \mathbf{0} \\ \mathbf{A} & \mathbf{0} & \mathbf{0}_M \end{bmatrix} \tag{F.1}$$

where $\Gamma = E[\mathbf{R}\mathbf{R}^t]$, and $\mathbf{A}$ is an $M \times N$ dimensional matrix of rank $M < N$ with linearly independent rows.

Proposition 51 *Assume that M as given in (F.1) is invertible, and assume that $\Gamma = E[\mathbf{R}\mathbf{R}^t]$ is positive definite. If Σ is nonsingular, then the inverse of M can be written as*

$$M^{-1} = \begin{bmatrix} \Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1} & \cdot & \cdot \\ \boldsymbol{\mu}^t\left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right] & \frac{1}{\eta} - \boldsymbol{\mu}^t\Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\boldsymbol{\mu} & \cdot \\ \left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Sigma^{-1} & \left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Sigma^{-1}\boldsymbol{\mu} & -\left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1} \end{bmatrix}.$$

where we have only written the lower triangular part of M^{-1}, and where

$$\eta = 1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}.$$

If Σ is singular, then the inverse of M can be written as

$$M^{-1} = \begin{bmatrix} \Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega & \cdot & \cdot \\ \boldsymbol{\mu}^t\left(\Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) & \frac{1}{\delta} & \cdot \\ \Phi\mathbf{A}\Gamma^{-1}\left(I + \boldsymbol{\mu}\boldsymbol{\mu}^t\frac{1}{\delta}\Omega\right) & \frac{1}{\delta}\Phi\mathbf{A}\Gamma^{-1}\boldsymbol{\mu} & \Phi\left(\frac{1}{\delta}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi - \mathbf{I}\right) \end{bmatrix},$$

where we have only written the lower triangular part of M^{-1}, and where

$$\Phi = \left(\mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)^{-1},$$

$$\Omega = \left[\Gamma^{-1} - \Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\right],$$

and

$$\delta = \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}.$$

Proof. Case 1: Σ nonsingular.

Let $\mathbf{a} \in \mathbb{R}^N$, $b \in \mathbb{R}$, $\mathbf{c} \in \mathbb{R}^M$. We will solve for $\mathbf{x} \in \mathbb{R}^N$, $q \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^M$ the system of equations

$$M_V \begin{pmatrix} \mathbf{x} \\ q \\ \boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ b \\ \mathbf{c} \end{pmatrix}.$$

Using the definition (F.1) of M, the system of equations can be rewritten as

$$\begin{cases} \Gamma\mathbf{x} & -\boldsymbol{\mu}q & +\mathbf{A}^t\boldsymbol{\gamma} & = \mathbf{a} \\ -\boldsymbol{\mu}^t\mathbf{x} & +q & & = b \\ \mathbf{A}\mathbf{x} & & & = \mathbf{c}. \end{cases}$$

$$\Leftrightarrow \begin{cases} & \mathbf{x} & = \Gamma^{-1}(\mathbf{a} + \boldsymbol{\mu}q - \mathbf{A}^t\boldsymbol{\gamma}) \\ -\boldsymbol{\mu}^t[\Gamma^{-1}(\mathbf{a} + \boldsymbol{\mu}q - \mathbf{A}^t\boldsymbol{\gamma})] & +q & = b \\ \mathbf{A}[\Gamma^{-1}(\mathbf{a} + \boldsymbol{\mu}q - \mathbf{A}^t\boldsymbol{\gamma})] & & = \mathbf{c}. \end{cases} \tag{F.2}$$

Remember first that if $\Sigma = \Gamma - \boldsymbol{\mu}\boldsymbol{\mu}^t$ is nonsingular, its inverse can be written as

$$\Sigma^{-1} = \Gamma^{-1} + \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}, \tag{F.3}$$

and we must have

$$1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu} \neq 0. \tag{F.4}$$

Now, using (F.4), we can solve the second equation in (F.2) for q, yielding

$$q = \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}[b + \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{a} - \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\gamma] \tag{F.5}$$

which can be plugged into the third equation in (F.2) to yield

$$\begin{aligned}
c &= \mathbf{A}\left[\Gamma^{-1}\left(\mathbf{a} + \boldsymbol{\mu}\frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\left[b + \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{a} - \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\gamma\right] - \mathbf{A}^t\gamma\right)\right] \\
&= \left(\mathbf{A}\Gamma^{-1} + \mathbf{A}\frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\right)\mathbf{a} + \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}b \\
&\quad - \left(\frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t + \mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)\gamma \\
&= \mathbf{A}\left(\Gamma^{-1} + \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\right)\mathbf{a} + \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}b \\
&\quad - \left[\mathbf{A}\left(\frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1} + \Gamma^{-1}\right)\mathbf{A}^t\right]\gamma,
\end{aligned}$$

so

$$\gamma = [\mathbf{A}\Sigma^{-1}\mathbf{A}^t]^{-1}\left(\mathbf{A}\Sigma^{-1}\mathbf{a} + \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}b - \mathbf{c}\right) \tag{F.6}$$

where we have used (F.3). Putting (F.6) into (F.5) yields the expression

$$q = \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\left[b + \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{a} - \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t[\mathbf{A}\Sigma^{-1}\mathbf{A}^t]^{-1}\left(\mathbf{A}\Sigma^{-1}\mathbf{a} + \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}b - \mathbf{c}\right)\right] \tag{F.7}$$

Finally, using (F.7) and (F.6) in the $\mathbf{x}$ equation of (F.2) yields

$$\begin{aligned}
\mathbf{x} &= \Gamma^{-1}(\mathbf{a}+\mu q-\mathbf{A}^t\gamma) \\
&= \Gamma^{-1}\mathbf{a} \\
&\quad +\Gamma^{-1}\mu\frac{1}{1-\mu^t\Gamma^{-1}\mu}\left\{b+\mu^t\Gamma^{-1}\mathbf{a}-\mu^t\Gamma^{-1}\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\left[\mathbf{A}\Sigma^{-1}\mathbf{a}+\frac{1}{1-\mu^t\Gamma^{-1}\mu}\mathbf{A}\Gamma^{-1}\mu b-\mathbf{c}\right]\right\} \\
&\quad -\Gamma^{-1}\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\left[\mathbf{A}\Sigma^{-1}\mathbf{a}+\frac{1}{1-\mu^t\Gamma^{-1}\mu}\mathbf{A}\Gamma^{-1}\mu b-\mathbf{c}\right] \\
&= \left(\Gamma^{-1}+\frac{1}{1-\mu^t\Gamma^{-1}\mu}\Gamma^{-1}\mu\mu^t\Gamma^{-1}\right)\mathbf{a}+\Gamma^{-1}\mu\frac{1}{1-\mu^t\Gamma^{-1}\mu}b \\
&\quad -\left(\Gamma^{-1}+\frac{1}{1-\mu^t\Gamma^{-1}\mu}\Gamma^{-1}\mu\mu^t\Gamma^{-1}\right)\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\left[\mathbf{A}\Sigma^{-1}\mathbf{a}+\frac{1}{1-\mu^t\Gamma^{-1}\mu}\mathbf{A}\Gamma^{-1}\mu b-\mathbf{c}\right] \\
&= \Sigma^{-1}\mathbf{a}+\Gamma^{-1}\mu\frac{1}{1-\mu^t\Gamma^{-1}\mu}b \\
&\quad -\Sigma^{-1}\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\mathbf{A}\Sigma^{-1}\mathbf{a} \\
&\quad -\frac{1}{1-\mu^t\Gamma^{-1}\mu}\Sigma^{-1}\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\mathbf{A}\Gamma^{-1}\mu b+\Sigma^{-1}\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\mathbf{c} \\
&= \left(\Sigma^{-1}-\Sigma^{-1}\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\mathbf{A}\Sigma^{-1}\right)\mathbf{a} \\
&\quad +\frac{1}{1-\mu^t\Gamma^{-1}\mu}\left(\mathbf{I}-\Sigma^{-1}\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\mathbf{A}\right)\Gamma^{-1}\mu b \\
&\quad +\Sigma^{-1}\mathbf{A}^t\left[\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right]^{-1}\mathbf{c}. \qquad \text{(F.8)}
\end{aligned}$$

Now write

$$M^{-1}=\begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}.$$

We can get the blocks in M^{-1} by writing

$$\begin{pmatrix} \mathbf{x} \\ q \\ \gamma \end{pmatrix} = M^{-1} \begin{pmatrix} \mathbf{a} \\ b \\ \mathbf{c} \end{pmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{pmatrix} \mathbf{a} \\ b \\ \mathbf{c} \end{pmatrix},$$

and then matching the terms $\mathbf{a}$, b and $\mathbf{c}$ terms in (F.6), (F.7) and (F.8) with M_{11} through M_{33}. This yields the following form for the inverse of M:

$$M^{-1} = \begin{bmatrix} \Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1} & \cdot & \cdot \\ \frac{1}{\eta}\boldsymbol{\mu}^t\Gamma^{-1}\left[\mathbf{I} - \mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right] & \frac{1}{\eta^2}\left[\eta - \boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\right] & \cdot \\ \left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Sigma^{-1} & \frac{1}{\eta}\left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu} & -\left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1} \end{bmatrix},$$

where we have only written the lower triangular part of M^{-1}, and where

$$\eta = 1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}^t. \tag{F.9}$$

The last expression for M^{-1} can be simplified somewhat by noticing that

$$\begin{aligned} & \frac{1}{\eta}\Gamma^{-1}\boldsymbol{\mu} \\ = & \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\Gamma^{-1}\boldsymbol{\mu}, \text{ by (F.9)}, \\ = & \left(\Gamma^{-1} + \frac{1}{1 - \boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\right)\boldsymbol{\mu} \\ = & \Sigma^{-1}\boldsymbol{\mu}, \end{aligned}$$

yielding

$$M^{-1} = \begin{bmatrix} \Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1} & . & . \\ \boldsymbol{\mu}^t\left[\Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\right] & \frac{1}{\eta} - \boldsymbol{\mu}^t\Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}\boldsymbol{\mu} & . \\ \left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Sigma^{-1} & \left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1}\mathbf{A}\Sigma^{-1}\boldsymbol{\mu} & -\left(\mathbf{A}\Sigma^{-1}\mathbf{A}^t\right)^{-1} \end{bmatrix}.$$

Case 2: Σ singular.

Let $\mathbf{a} \in \mathbb{R}^N$, $b \in \mathbb{R}$, $\mathbf{c} \in \mathbb{R}^M$. We will solve for $\mathbf{x} \in \mathbb{R}^N$, $q \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^M$ the system of equations

$$M \begin{pmatrix} \mathbf{x} \\ q \\ \boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ b \\ \mathbf{c} \end{pmatrix}.$$

Using the definition (F.1) of M, the system of equations can be rewritten as

$$\left\{ \begin{array}{llll} \Gamma\mathbf{x} & -\boldsymbol{\mu}q & +\mathbf{A}^t\boldsymbol{\gamma} & = \mathbf{a} \\ -\boldsymbol{\mu}^t\mathbf{x} & +q & & = b \\ \mathbf{A}\mathbf{x} & & & = \mathbf{c}. \end{array} \right.$$

$$\Leftrightarrow \left\{ \begin{array}{lll} & \mathbf{x} & = \Gamma^{-1}(\mathbf{a} + \boldsymbol{\mu}q - \mathbf{A}^t\boldsymbol{\gamma}) \\ -\boldsymbol{\mu}^t[\Gamma^{-1}(\mathbf{a} + \boldsymbol{\mu}q - \mathbf{A}^t\boldsymbol{\gamma})] & +q & = b \\ \mathbf{A}[\Gamma^{-1}(\mathbf{a} + \boldsymbol{\mu}q - \mathbf{A}^t\boldsymbol{\gamma})] & & = \mathbf{c}. \end{array} \right. \tag{F.10}$$

Solving the third equation in (F.10) for $\boldsymbol{\gamma}$ we get

$$\begin{aligned} \boldsymbol{\gamma} &= (\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\left[\mathbf{A}\Gamma^{-1}(\mathbf{a} + \boldsymbol{\mu}q) - \mathbf{c}\right] \\ &= \Phi\left[\mathbf{A}\Gamma^{-1}(\mathbf{a} + \boldsymbol{\mu}q) - \mathbf{c}\right], \end{aligned} \tag{F.11}$$

with

$$\Phi = (\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}.$$

Note that Φ exists since $\mathbf{A}$ has linearly independent rows. Plugging (F.11) into the second equation in (F.10) becomes

$$\begin{aligned} b &= -\boldsymbol{\mu}^t\Gamma^{-1}\left(\mathbf{a}+\boldsymbol{\mu}q-\mathbf{A}^t\Phi\left[\mathbf{A}\Gamma^{-1}(\mathbf{a}+\boldsymbol{\mu}q)-\mathbf{c}\right]\right)+q \\ &= -\boldsymbol{\mu}^t[\Gamma^{-1}-\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}]\mathbf{a}+\left(1-\boldsymbol{\mu}^t\left[\Gamma^{-1}-\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\right]\boldsymbol{\mu}\right)q-\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{c} \\ &= -\boldsymbol{\mu}^t\Omega\mathbf{a}+\left(1-\boldsymbol{\mu}^t\Omega\boldsymbol{\mu}\right)q-\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi^{-1}\mathbf{c}, \end{aligned}$$

with

$$\Omega=\left[\Gamma^{-1}-\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\right].$$

Therefore,

$$q=\frac{1}{\delta}\boldsymbol{\mu}^t\Omega\mathbf{a}+\frac{1}{\delta}b+\frac{1}{\delta}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{c}, \qquad \text{(F.12)}$$

with

$$\delta=1-\boldsymbol{\mu}^t\Omega\boldsymbol{\mu},$$

with $\delta \neq 0$ from the assumption that M is invertible. Therefore, using (F.12) in (F.11) we get

$$\begin{aligned} \boldsymbol{\gamma} &= (\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\left[\mathbf{A}\Gamma^{-1}(\mathbf{a}+\boldsymbol{\mu}q)-\mathbf{c}\right] \\ &= (\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\left[\mathbf{A}\Gamma^{-1}\left(\mathbf{a}+\boldsymbol{\mu}\frac{1}{\delta}\left(\boldsymbol{\mu}^t\Omega\mathbf{a}+b+\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t(\mathbf{A}\Gamma^{-1}\mathbf{A}^t)^{-1}\mathbf{c}\right)\right)-\mathbf{c}\right] \\ &= \Phi\mathbf{A}\Gamma^{-1}\left(\mathbf{I}+\frac{1}{\delta}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\mathbf{a} \\ &\quad +\frac{1}{\delta}\Phi\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}b \\ &\quad +\frac{1}{\delta}\Phi\left(\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi-\mathbf{I}\right)\mathbf{c}. \end{aligned} \qquad \text{(F.13)}$$

Also, using (F.12) and (F.13) in the first equation of (F.10) yields

$$\begin{aligned}
\mathbf{x} &= \Gamma^{-1}(\mathbf{a}+\boldsymbol{\mu}q-\mathbf{A}^t\gamma) \\
&= \Gamma^{-1}\mathbf{a}+\Gamma^{-1}\boldsymbol{\mu}q-\Gamma^{-1}\mathbf{A}^t\gamma \\
&= \Gamma^{-1}\mathbf{a} \\
&\quad +\Gamma^{-1}\boldsymbol{\mu}\left(\frac{1}{\delta}\boldsymbol{\mu}^t\Omega\mathbf{a}+\frac{1}{\delta}b+\frac{1}{\delta}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{c}\right) \\
&\quad -\Gamma^{-1}\mathbf{A}^t\left[\Phi\mathbf{A}\Gamma^{-1}\left(\mathbf{I}+\frac{1}{\delta}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\mathbf{a}+\frac{1}{\delta}\Phi\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}b+\frac{1}{\delta}\Phi\left(\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi-\mathbf{I}\right)\mathbf{c}\right] \\
&= \left[\Gamma^{-1}+\frac{1}{\delta}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega-\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\left(\mathbf{I}+\frac{1}{\delta}\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\right]\mathbf{a} \\
&\quad +\left(\frac{1}{\delta}\Gamma^{-1}\boldsymbol{\mu}-\frac{1}{\delta}\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\right)b \\
&\quad +\left[\frac{1}{\delta}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi-\Gamma^{-1}\mathbf{A}^t\Phi\left(\frac{1}{\delta}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi-\mathbf{I}\right)\right]\mathbf{c} \\
&= \left(\mathbf{I}+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\right)\Omega\mathbf{a} \qquad \text{(F.14)} \\
&\quad +\frac{1}{\delta}\Omega\boldsymbol{\mu}b \\
&\quad \left(\mathbf{I}+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\right)\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{c},
\end{aligned}$$

where the last equation follows by definition of Ω.

Now write

$$M^{-1}=\begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}.$$

We can get the blocks in M^{-1} by writing

$$\begin{pmatrix} \mathbf{x} \\ q \\ \gamma \end{pmatrix} = M^{-1} \begin{pmatrix} \mathbf{a} \\ b \\ \mathbf{c} \end{pmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{pmatrix} \mathbf{a} \\ b \\ \mathbf{c} \end{pmatrix},$$

and then matching the terms $\mathbf{a}$, b and $\mathbf{c}$ terms in (F.13), (F.12) and (F.14) with M_{11} through M_{33}. This yields the following general form for the inverse of M:

$$M^{-1} = \begin{bmatrix} \Omega + \frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega & \cdot & \cdot \\ \frac{1}{\delta}\boldsymbol{\mu}^t\Omega & \frac{1}{\delta} & \cdot \\ \Phi\mathbf{A}\Gamma^{-1}\left(I + \boldsymbol{\mu}\boldsymbol{\mu}^t\frac{1}{\delta}\Omega\right) & \frac{1}{\delta}\Phi\mathbf{A}\Gamma^{-1}\boldsymbol{\mu} & \Phi\left(\frac{1}{\delta}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi - \mathbf{I}\right) \end{bmatrix},$$

where we have only written the lower triangular part of M^{-1}, and where

$$\Phi = \left(\mathbf{A}\Gamma^{-1}\mathbf{A}^t\right)^{-1},$$

$$\Omega = \left[\Gamma^{-1} - \Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\right],$$

and

$$\delta = 1 - \boldsymbol{\mu}^t\Omega\boldsymbol{\mu}.$$

When Σ is singular, notice that

$$\boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu} = 1 \quad \text{(F.15)}$$

(otherwise

$$\begin{aligned}
& \Sigma\left(\Gamma^{-1}+\frac{1}{1-\boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\right) \\
= & \left(\Gamma-\boldsymbol{\mu}\boldsymbol{\mu}^t\right)\left(\Gamma^{-1}+\frac{1}{1-\boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\right), \text{ by definition of the covariance,} \\
= & \mathbf{I},
\end{aligned}$$

and we have a contradiction). Therefore, δ simplifies to

$$\delta=\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}.$$

Also, notice that

$$\begin{aligned}
\left(\Omega+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right)\boldsymbol{\mu} &= \frac{1}{\delta}\left(\delta\Omega\boldsymbol{\mu}+\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\left[\Gamma^{-1}-\Gamma^{-1}\mathbf{A}^t\Phi\mathbf{A}\Gamma^{-1}\right]\boldsymbol{\mu}\right), \text{ by definition of } \Omega, \\
&= \frac{1}{\delta}\left(\delta\Omega\boldsymbol{\mu}-\delta\Omega\boldsymbol{\mu}+\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}\right), \text{ by definition of } \delta, \\
&= \frac{\boldsymbol{\mu}^t\Gamma^{-1}\boldsymbol{\mu}}{\delta}\Omega\boldsymbol{\mu} \\
&= \frac{1}{\delta}\Omega\boldsymbol{\mu}, \text{ by (F.15),}
\end{aligned}$$

so we can rewrite M^{-1} as

$$\begin{aligned}
& M^{-1} \\
= & \begin{bmatrix}
\Omega+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega & \cdot & \cdot \\
\boldsymbol{\mu}^t\left(\Omega+\frac{1}{\delta}\Omega\boldsymbol{\mu}\boldsymbol{\mu}^t\Omega\right) & \frac{1}{\delta} & \cdot \\
\Phi\mathbf{A}\Gamma^{-1}\left(I+\boldsymbol{\mu}\boldsymbol{\mu}^t\frac{1}{\delta}\Omega\right) & \frac{1}{\delta}\Phi\mathbf{A}\Gamma^{-1}\boldsymbol{\mu} & \Phi\left(\frac{1}{\delta}\mathbf{A}\Gamma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^t\Gamma^{-1}\mathbf{A}^t\Phi-\mathbf{I}\right)
\end{bmatrix}.
\end{aligned}$$

■

Appendix G

Alternative Proof of Asymptotic Normality of the Variance Portfolio Estimator

We present an alternative, more direct proof of the asymptotic normality of the variance portfolio estimator, a result we obtained in Chapter 4. Let $\mathbf{R}$ be a random return vector in $\mathbb{R}^N$ with mean $\boldsymbol{\mu}$ and covariance matrix Σ, which we assume is positive definite. Also, assume that $\mathbf{R}$ has a continuous density. Let $\mathbf{R}_1, \cdots, \mathbf{R}_T$ be realizations of variable $\mathbf{R}$. Suppose that we want to estimate the parameter

$$\mathbf{x}_V = \arg \min_{\mathbf{Ax}=\mathbf{b}} \mathbf{x}^t \Sigma \mathbf{x},$$

where $\mathbf{A}$ is an $(M \times N)$ matrix with linearly independent rows, and $\mathbf{b}$ is an M-dimensional vector. Assume that $\mathbf{b} \neq \mathbf{0}$ so the problem is non-trivial. Notice that $\mathbf{x}_V$ is unique by strict convexity of $\mathbf{x}^t \Sigma \mathbf{x}$ and convexity of the set $\mathbf{Ax} = \mathbf{b}$. The variance portfolio estimator is

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{Ax}=\mathbf{b}} \mathbf{x}^t \hat{\Sigma} \mathbf{x}, \tag{G.1}$$

where $\hat{\Sigma} = \sum_{i=1}^T (\mathbf{R}_i - \bar{\mathbf{R}})(\mathbf{R}_i - \bar{\mathbf{R}})^t / T$ is the sample covariance matrix, and $\bar{\mathbf{R}} = \sum_{i=1}^T \mathbf{R}_i / T$ is the sample mean.

Proposition 52 *Suppose Assumption (A) from Chapter 4 holds. Let* $\hat{\mathbf{x}}$ *be defined as in (G.1). Then*

$$\sqrt{T}(\hat{\mathbf{x}} - \mathbf{x}_V) \rightsquigarrow N(\mathbf{0}, VQV)$$

where

$$Q = Cov[(\mathbf{R} - \boldsymbol{\mu})[(\mathbf{R} - \boldsymbol{\mu})^t \mathbf{x}_V],$$

and where $V = \Sigma^{-1} - \Sigma^{-1}\mathbf{A}^t(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}\mathbf{A}\Sigma^{-1}$ *is the upper-left* $(N \times N)$ *corner of* M^{-1} *where*

$$M = \begin{bmatrix} \Sigma & \mathbf{A}^t \\ \mathbf{A} & \mathbf{0} \end{bmatrix}.$$

Proof. Notice that, along with a unique vector of Langrange multipliers $\boldsymbol{\lambda}_V \in \mathbb{R}^M$, $\mathbf{x}_V$ uniquely solves the system of linear equations

$$\begin{cases} \Sigma\mathbf{x}_V & +\mathbf{A}^t\boldsymbol{\lambda}_V & = \mathbf{0} \\ \mathbf{A}\mathbf{x}_V & +\mathbf{0}\boldsymbol{\lambda}_V & = \mathrm{b}. \end{cases} \tag{G.2}$$

Let

$$M = \begin{bmatrix} \Sigma & \mathbf{A}^t \\ \mathbf{A} & \mathbf{0} \end{bmatrix}.$$

M is invertible by uniqueness of $\mathbf{x}_V$ and $\boldsymbol{\lambda}_V$ in (G.2). $\hat{\Sigma}$ is positive definite with probability one from the assumption that $\mathbf{R}$ has a continuous density. Notice then that, along with a unique set of Langrange multipliers $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^M$, with probability one $\hat{\mathbf{x}}$ uniquely solves

$$\begin{cases} \hat{\Sigma}\hat{\mathbf{x}} & +\mathbf{A}^t\hat{\boldsymbol{\lambda}} & = \mathbf{0} \\ \mathbf{A}\hat{\mathbf{x}} & +\mathbf{0}\hat{\boldsymbol{\lambda}} & = \mathrm{b}, \end{cases} \tag{G.3}$$

and let

$$\hat{M} = \begin{bmatrix} \hat{\Sigma} & \mathbf{A}^t \\ \mathbf{A} & \mathbf{0}_M \end{bmatrix}.$$

With probability one, $\hat{M}$ is invertible by uniqueness of $\hat{\mathbf{x}}$ and $\hat{\boldsymbol{\lambda}}$ in (G.3).

Using (G.3), write

$$\begin{aligned} \begin{bmatrix} \mathbf{0}_N \\ \mathbf{b} \end{bmatrix} &= \hat{M} \begin{bmatrix} \mathbf{x}_V \\ \boldsymbol{\lambda}_V \end{bmatrix} + \hat{M} \begin{bmatrix} \hat{\mathbf{x}} - \mathbf{x}_V \\ \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_V \end{bmatrix} \qquad \text{(G.4)} \\ &= \begin{bmatrix} \hat{\Sigma}\mathbf{x}_V + \mathbf{A}^t \boldsymbol{\lambda}_V \\ \mathbf{b} \end{bmatrix} + \hat{M} \begin{bmatrix} \hat{\mathbf{x}} - \mathbf{x}_V \\ \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_V \end{bmatrix}. \end{aligned}$$

Since $\hat{M}$ is invertible with probability one, we can rewrite (G.4) (after multiplying both sides by $\sqrt{T}$) as

$$\begin{aligned} \sqrt{T} \begin{bmatrix} \hat{\mathbf{x}} - \mathbf{x}_V \\ \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_V \end{bmatrix} &= -\sqrt{T}\hat{M}^{-1} \left\{ \begin{bmatrix} \hat{\Sigma}\mathbf{x}_V + \mathbf{A}^t \boldsymbol{\lambda}_V \\ \mathbf{b} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix} \right\} \\ &= -(M + o_P(1))^{-1} \begin{bmatrix} \sqrt{T}(\hat{\Sigma}\mathbf{x}_\mathbf{V} + \mathbf{A}^t \boldsymbol{\lambda}_V) \\ \mathbf{0}_M \end{bmatrix}. \qquad \text{(G.5)} \end{aligned}$$

Notice that

$$\begin{aligned} \sqrt{T}(\hat{\Sigma}\mathbf{x}_V + \mathbf{A}^t \boldsymbol{\lambda}_V) &= \sqrt{T}[(\hat{\Sigma} - \Sigma)\mathbf{x}_V + \Sigma\mathbf{x}_V + \mathbf{A}^t \boldsymbol{\lambda}_V) \\ &= \sqrt{T}(\hat{\Sigma} - \Sigma)\mathbf{x}_V \\ &= \sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T}[(\mathbf{R}_i - \bar{\mathbf{R}})(\mathbf{R}_i - \bar{\mathbf{R}})^t - E(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t]\right)\mathbf{x}_V \\ &= \sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T}[(\mathbf{R}_i - \boldsymbol{\mu})(\mathbf{R}_i - \boldsymbol{\mu})^t\mathbf{x}_V - E(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t\mathbf{x}_V]\right) \qquad \text{(G.6)} \\ &\quad + \sqrt{T}[(\boldsymbol{\mu} - \bar{\mathbf{R}})(\boldsymbol{\mu} - \bar{\mathbf{R}})^t]\mathbf{x}_V. \end{aligned}$$

Notice that by the classical central limit theorem, the first term in (G.6) converges in distribution to a normal vector with mean $\mathbf{0}$ and covariance

$$Q = E\{(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^t[(\mathbf{R}-\boldsymbol{\mu})^t\mathbf{x}_V]^2\} - E[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^t\mathbf{x}_V]E[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^t\mathbf{x}_V]^t.$$

The second term in (G.6) converges to zero in probability. So by Slutsky's Lemma (see for example van der Vaart, Lemma 2.8) $\sqrt{T}(\hat{\Sigma}\mathbf{x}_V + \mathbf{A}^t\boldsymbol{\lambda}_V)$ converges in distribution to a normal vector with mean $\mathbf{0}$ and covariance $\mathbf{Q}$, and from (G.5)

$$\sqrt{T}\begin{bmatrix} \hat{\mathbf{x}} - \mathbf{x}_V \\ \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_V \end{bmatrix} \rightsquigarrow \mathbf{U},$$

where $\rightsquigarrow$ stands for "converges in distribution", and where $\mathbf{U}$ has mean $\mathbf{0}$ and covariance

$$\mathbf{Q}_\mathbf{U} = M^{-1}\begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} M^{-1}.$$

The inverse of M is

$$M^{-1} = \begin{bmatrix} \Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1}) & -\Sigma^{-1}\mathbf{A}^t\mathbf{F} \\ -\mathbf{F}\mathbf{A}\Sigma^{-1} & \mathbf{F} \end{bmatrix},$$

where $\mathbf{F} = -(\mathbf{A}\Sigma^{-1}\mathbf{A}^t)^{-1}$. Then

$$
\begin{aligned}
&\mathbf{Q_U} \\
= &\begin{bmatrix} \Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1}) & -\Sigma^{-1}\mathbf{A}^t\mathbf{F} \\ -\mathbf{F}\mathbf{A}\Sigma^{-1} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\
&\begin{bmatrix} \Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1}) & -\Sigma^{-1}\mathbf{A}^t\mathbf{F} \\ -\mathbf{F}\mathbf{A}\Sigma^{-1} & \mathbf{F} \end{bmatrix} \\
= &\begin{bmatrix} \Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1})\mathbf{Q} & \mathbf{0} \\ -\mathbf{F}\mathbf{A}\Sigma^{-1}\mathbf{Q} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1}) & -\Sigma^{-1}\mathbf{A}^t\mathbf{F} \\ -\mathbf{F}\mathbf{A}\Sigma^{-1} & \mathbf{F} \end{bmatrix} \\
= &\begin{bmatrix} \Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1})\mathbf{Q}\Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1}) & -\Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1})\mathbf{Q}\Sigma^{-1}\mathbf{A}^t\mathbf{F} \\ -\mathbf{F}\mathbf{A}\Sigma^{-1}\mathbf{Q}\Sigma^{-1}(\mathbf{I}_N + \mathbf{A}^t\mathbf{F}\mathbf{A}\Sigma^{-1}) & \mathbf{F}\mathbf{A}\Sigma^{-1}\mathbf{Q}\Sigma^{-1}\mathbf{A}^t\mathbf{F} \end{bmatrix}.
\end{aligned}
$$

The result then follows. ■

Bibliography

[1] Artzner, P., Delbaen, F., Eber, J.M., and D. Heath. Coherent Measures of Risk. Mathematical Finance **9** (1999), 203-228.

[2] Bawa, V. Safety-First, Stochastic Dominance, and Optimal Portfolio Choice. Journal of Financial and Quantitative Analysis **33** (1978), 255-271.

[3] Bawa, V. Optimal Rules for Ordering Uncertain Prospects. Journal of Financial Economics **2** (1975), 95-121.

[4] Bawa, Vijay S., Brown, Stephen J., and Roger W. Klein. *Estimation Risk and Optimal Portfolio Choice.* North-Holland Publishing Company, 1979.

[5] Bawa, V., and E. B. Lindenberg Capital Market Equilibrium in a Mean-Lower Partial Moment Framework. Journal of Financial Economics **5** (1977): 189-200.

[6] Bertsimas, Dimitris, and Ioana Popescu. Optimal Inequalities in Probability: a Convex Programming Approach. Working paper, Operations Research Center, MIT, 1999.

[7] Bookstaber, R., and R. Clarke. Option Portfolio Strategies: Measurement and Evaluation. Journal of Business **57** (1984), 469-492.

[8] Bouchaud, Jean-Phillipe, and Marc Potters. *Theory of Financial Risks.* Cambridge University Press, 2000.

[9] Campbell, John, Lo, Andrew, and Craig MacKinlay. *The Econometrics of Financial Markets.* Princeton University Press, 1997.

[10] Chamberlain, G. A Characterization of the Distributions that Imply Mean-Variance Utility Functions. Journal of Economic Theory **29** (1983), 185-201.

[11] Dodge, Yadolah, and Jana Jureckova. *Adaptive Regression.* Springer, 2000.

[12] Dowd, K. *Beyond Value at Risk.* John Wiley and Sons, 1998.

[13] Duffie, D., and J. Pan. An Overview of Value at Risk. The Journal of Derivatives **4** (1997), 7-49.

[14] Embrechts, Paul, McNeil, Alexander, and Daniel Strauman. Correlation and Dependence in Risk Management: Properties and Pitfalls, ETH RiskLab Report, 1999.

[15] Fishburn, P. C. Mean-Risk Analysis with Risk Associated with Below-Target Returns. The American Economic Review **67** (1977), 116-126.

[16] Frost, Peter A., and James E. Savarino. An Empirical Bayes Approach to Efficient Portfolio Selection. Journal of Financial and Quantitative Analysis **21** (1986), 293-305.

[17] Garman, M. Taking VaR to Pieces. Risk **10** (1997), 70-71.

[18] Gibbons, Michael, Ross, Stephen A., and Shanken. A Test of the Efficiency of a Given Portfolio. Econometrica **57** (1989), 1121-1152.

[19] Grootveld, H., and W. Hallerback. Variance vs. Downside Risk: Is there Really that much Difference? European Journal of Operations Research **114** (1999), 304-319.

[20] Gruber, Marvin H. J. *Improving efficiency by Shrinkage.* New York: Marcel Dekker, 1998.

[21] Hamao, Y., Masulis, R., and V. Ng. Correlations in Price Changes and Volatility Across International Stock Markets. The Review of Financial Studies **3** (1990), 282-307.

[22] Harlow, W.V., Rao, R. Asset Pricing in a Generalized Mean-Lower Partial Moment Framework: Theory and Evidence. The Journal of Financial and Quantitative Analysis **24** (1989) 285-311.

[23] Hawkings, Douglas M., and David Olive. Applications and Algorithms for Least Trimmed Sum of Absolute Deviations Regressions. Computational Statistics and Data Analysis, **32** (1998) 119-134.

[24] Hoerl, A.E., and Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics **12** (1970a), 55-67.

[25] Hoerl, A.E., and Kennard, R.W. Ridge Regression: Applications to Nonorthogonal Problems. Technometrics **12** (1970b), 69-82.

[26] Hoerl, R.W., Schuenemeyer, J.H. and Hoerl, A.E. A Simulation of Biased Estimation and Subset Selection Regression Techniques. Technometrics **28** (1986), 369-380.

[27] Huang, C.f., and R. H. Litzenberger. Foundations for Financial Economics. Prentice Hall, 1988.

[28] Huber, P. J. *Robust Statistics.* New York: Wiley, 1981.

[29] Ingersoll, J. E. *Theory of Financial Decision Making.* Rowman & Littlefield, 1987.

[30] Issi, K. On Sharpness of Tchebycheff-Type Inequalities. Annals of the Institute of Statistical Mathematics **14** (1963), 185-197.

[31] Jobson, J.D., and B. Korkie. Estimation for Markowitz Efficient Portfolios. Journal of the American Statistical Association **75** (1980), 544-554.

[32] Jorion, Philippe. Bayes-Stein Estimation for Portfolio Analysis. Journal of Financial and Quantitative Analysis **21** (1986), 279-292.

[33] Jorion, P. *Value at Risk.* McGraw-Hill, 1997.

[34] King, M., and S. Wadhani. Transmission of Volatility Between Markets. The Review of Financial Studies **3** (1990), 5-35.

[35] Koenker, Roger, and Gilbert Basset, Jr. Regression Quantiles. Econometrica **46** (1978).

[36] Koenker, R. W., and V. D'Orey. Computing Regression Quantiles. Royal Statistical Society **36** (1987): 383-393.

[37] Ledoit, Olivier. *Essays on Risk and Return in the Stock Market.* Ph.D. Thesis, Sloan School of Management, MIT (1995).

[38] Lemus, G. *Portfolio Optimization via Quantile-Based Risk Measures.* Ph.D. Thesis, Sloan School of Management, MIT (1999).

[39] Lemus, G., Samarov, A., and R. Welsch. Portfolio Analysis Based on Value-at-Risk. Proceedings of the 52nd Session of ISI **3** (19999), 221-222.

[40] Levy, H. Stochastic Dominance and Expected Utility: Survey and Analysis. Management Science **38** (1992), 555-593.

[41] Levy, H., and Y. Kroll. Ordering Uncertain Options with Borrowing and Lending. The Journal of Finance **31** (1978), 553-574.

[42] Lindskog, Filip. *Modelling Dependence with Copulas and Applications to Risk Management.* ETH Master's Thesis, 2000.

[43] Markowitz, H. M. *Portfolio Selection.* Blackwell, 1959.

[44] Mangasarian, Olvi, and David Musicant. Robust Linear and Support Vector Regression. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000), 1-6.

[45] Muirhead, R. *Aspects of Multivariate Statistical Theory.* New York: Wiley, 1982.

[46] Neelakandan, H. *Volatility-Correlation Dynamics in Financial Markets.* Master's Thesis, Sloan School of Management, MIT, 1994.

[47] Pastor, Lubos. Portfolio Selection and Asset Pricing Models. Journal of Finance **55** (2000), 179-223.

[48] Portnoy, S., and R. Koenker. The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators. Statistical Sciences **12** (1997), 279-300.

[49] Pflug, G. Ch. *Optimization of Stochastic Models.* Kluwer Academic Publishers, 1996.

[50] Pollard, D. *Convergence of Stochastic Processes.* New York: Springer-Verlag, 1984.

[51] Pollard, D. Asymptotics for Least Absolute Deviation Regression Estimators. Econometric Theory **7** (1991), 186-199.

[52] Scaillet, O. Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall. Unpublished, 2000.

[53] Press, S. J. *Applied Multivariate Analysis.* Holt, Rinehart and Wilson, 1972.

[54] Roy, A.D. Safety First and the Holding of Assets. Econometrica **20** (1952), 431-449.

[55] Smith, James E. Generalized Chebyshev Inequalities: Theory and Applications in Decision Analysis. Operations Research **43** (1995), 807-825.

[56] Tasche, Dirk. Risk Contributions and Performance Measurement. Working Paper, TU Munich, 2000.

[57] Tasche, Dirk. Modern Portfolio Theory with Homogeneous Risk Measures. Slides from talk given in Rotterdam, February 2001.

[58] Tibshirani, R. Regression Selection and Shrinkage via the Lasso. Technical Report (June 1994).

[59] Uryasev, S., and R. T. Rockafellar. Optimization of Conditional Value-at-Risk. Unpublished Research Report, 1999.

[60] Van de Geer, S. Asymptotic Normality of minimum L1-norm estimators in linear regression, Report MS-R8806, Stichting Mathematisch Centrum, Amsterdam, June 1990.

[61] Van der Vart, A.W. *Asymptotic Statistics.* Cambridge University Press, 1998.

[62] Wahba, Grace. *Spline Models for Observational Data.* CMB-NSF Regional Conference Series in Applied Mathematics **59**. Society for Industrial and Applied Mathematics, Philadelphia, 1990.