A Thesis for the Degree of Ph.D. in Engineering

Estimation of Optimal Portfolio Weights Using Shrinkage Technique

March 2010

Graduate School of Science and Technology
Keio University

Takuya Kinkawa

Contents

1 Preliminaries **1**

1.1 Introduction 1

1.2 Summary of the present thesis 5

1.3 Review of the Stein-type estimator for the mean vector 6

1.4 Estimation problem of the mean-variance optimal portfolio weights 8

1.4.1 Definitions and assumptions 8

1.4.2 A class of classical estimators 11

2 Shrinkage toward the Origin or a Fixed Point **13**

2.1 Introduction 13

2.2 Case in which there are no constraints on portfolio weights 15

2.3 Case in which there are linear constraints on portfolio weights 22

2.4 Proofs of Theorems 2.1, 2.2, and 2.3 23

2.4.1 Lemmas 23

2.4.2 Proof of Theorem 2.1 27

2.4.3 Proof of Theorem 2.2 30

2.4.4 Proof of Theorem 2.3 31

2.5 Comparison of estimators 32

2.5.1 Risk comparison by Monte Carlo simulation 32

2.5.2 Comparison based on actual asset returns data 38

3 Shrinkage toward the Grand Mean or a Linear Subspace 43

3.1 Introduction 43

3.2 Case in which there are no constraints on portfolio weights 45

3.3 Case in which there are linear constraints on portfolio weights 50

3.4 Proofs of Theorems 3.1, 3.2, and 3.3 52

3.4.1 Proofs of Theorems 3.1 and 3.2 52

3.4.2 Proof of Theorem 3.3 58

4 Concluding Remarks 61

Bibliography 64

Acknowledgements 70

List of Tables

2.1 Risk values of estimators for the case $p = 10$ and $N = 60$ 34

2.2 Risk values of estimators for some pairs of p and N 37

2.3 Summary statistics of asset's excess returns 40

2.4 Comparison of estimators based on actual asset returns data 42

Chapter 1

Preliminaries

1.1. Introduction

Investors, individuals and institutions, hold various assets, such as cash (deposits), stocks, bonds, commodities, and real estate. They must determine the best allocation of these assets. Investing risky assets sometimes yields profits, but may also incur losses. In general, assets with higher expected returns involve higher risks, that is, investing is a trade-off between risk and return. *Portfolio Theory* in finance tells us how to determine the optimal combination (portfolio) of assets (cf., for example, Ingersoll 1987 and Elton, Gruber, Brown, and Goetzmann 2007). The determination of optimal portfolio requires taking account of many factors, such as investment purpose, investment horizon, amount of funds, and investors' risk tolerance. A numerous number of studies have discussed widely the optimal portfolio selection problem from various perspectives. Mean-variance optimization, which was introduced by Markowitz (1952), is one of the standard frameworks employed to determine optimal portfolio weights. In the present thesis, we investigate the estimation problem of single period mean-variance optimal portfolio weights for the case in which there are no constraints or linear equality constraints on portfolio weights.

The mean-variance optimization requires estimators for the mean vector and the covariance matrix of excess returns on the risky assets over the risk-free rate. A classical approach adopts the usual sample estimates for the mean vector and the covariance matrix. However, the optimal portfolio weights obtained by the classical approach are instable and unreliable. For example, Best and Grauer (1991) have shown that the estimation error of the mean vector yields portfolios that are very different from the true optimal portfolio. Chopra and Ziemba (1993) have suggested that the effect of the estimation error of the mean vector on the estimated optimal portfolio is greater than that of the covariance matrix.

Klein and Bawa (1976, 1977), Brown (1979), Chen and Brown (1983), Alexander and Resnick (1985), and Michaud (1989) have also pointed out this problem. Brandt (2009) has surveyed not only these problems but also more general portfolio choice problems. In finance, these problems are referred to as *estimation risk* problems (cf., Brown 1978 and Bawa, Brown, and Klein 1979).

In an attempt to reduce the estimation error for the optimal portfolio weights, application of Stein-type shrinkage estimators of the mean vector has been proposed. James and Stein (1961) first exhibited an estimator with uniformly smaller risk than that of the minimum variance unbiased estimator when the number of variables is larger than 2. This estimator is referred to as the *James-Stein* estimator in the literature. Since the breakthrough by James and Stein (1961), the shrinkage estimators have been studied extensively in the field of mathematical statistics. The minimum variance unbiased estimator is usually used to estimate the mean vector. However if the restriction of unbiasedness is removed, the James-Stein estimator can improve upon the minimum variance unbiased estimator for any set of true parameter values, and its improvement is large near the origin. Although the James-Stein estimator has bias, its gives an uniform improvement in terms of sum of mean squared errors.

Jobson, Korkie, and Ratti (1979) and Jobson and Korkie (1980, 1981) first applied the James-Stein estimators in the mean-variance optimal portfolio selection problem. Jorion (1985,

1986, 1991) proposed the adoption of an estimator that shrinks the sample mean toward the grand mean based on the evidence of mean reversion in financial markets. Frost and Savarino (1986) applied empirical Bayes type estimators. Michaud (1998) surveyed a number of estimation methods for the optimal portfolio weights, including shrinkage techniques and re-sampling methods. Grauer and Hakansson (1995, 2001) investigated the effectiveness of the Stein-type estimators in the optimal portfolio selection. Recently, Kan and Zhou (2007) discussed the problem of investing in two funds, namely a risk free asset and the tangency portfolio, and proposed the combination of a sample tangency portfolio and a sample global minimum variance portfolio. Okhrin and Schmid (2007) compared several types of Stein-type estimators for portfolio weights and showed that, for moderate sample sizes, Stein-type estimators improve upon the classical estimator, which we obtain by plugging in the sample estimates.

In the estimation problem of the mean vector, the expectation of the quadratic loss is usually used to evaluate the goodness of an estimator. In the field of mathematical statistics, a numerous number of previous studies have presented dominance results of various shrinkage estimators under the quadratic loss. However, the loss function used in the estimation problem of the optimal portfolio weights is different from the quadratic loss function. One of the desirable loss functions in this problem is defined as the difference between the utility of the true optimal portfolio and that of an estimated portfolio (cf., for example, Brown 1976, 1978, Bawa, Brown, and Klein 1979, Jorion 1986, Kashima 2001, 2005, and Mori 2004).

Most of the related studies have demonstrated the effectiveness of applying Stein-type shrinkage estimators for the optimal portfolio weights by numerical simulations or empirical studies. Since numerical simulations are usually performed only for some selected sets of parameter values, these results do not guarantee the improvement of the Stein-type estimators upon the classical estimator for any set of parameter values. Since actual data sets contain various distinctive characteristics, a successful result in one empirical study does not necessarily guarantee

that in another empirical study. The problem of determining whether or not the Stein-type estimators dominate the classical estimator analytically is the fundamental one to be addressed.

There are a few studies that have addressed the problem of showing analytically the dominance of the Stein-type estimators for the mean-variance optimal portfolio weights. Kashima (2001, 2005) and Mori (2004) have shown some interesting analytical results under the loss function used in the estimation problem of the mean-variance optimal portfolio weights. Kashima (2001, 2005) has shown that the estimation problem of the mean-variance optimal portfolio weights reduces to the estimation problem of the mean vector under quadratic loss when the covariance matrix is known. Therefore, when the covariance matrix is known, we can apply the abundant existing results on Stein-type estimators provided in the mathematical statistics literature (cf., for example, Lehmann and Casella 1998) to the mean-variance optimal portfolio choice problem. However, the estimation problem of the mean-variance optimal portfolio weights does not reduce to the estimation problem of the mean vector under quadratic loss when the covariance matrix is unknown.

Mori (2004) presented a dominance result of a Stein-type estimator in the estimation problem of the mean-variance optimal portfolio weights when the covariance matrix is unknown and is estimated. Although another interesting class of Stein-type estimators has been proposed by Baranchik (1970) and a similar class was adopted by Kashima (2001, 2005), the dominance for the corresponding class has not been addressed when the covariance matrix is unknown. The estimator given by Mori (2004) shrinks the sample mean toward the origin. However, in some cases the other shrinkage targets are more appropriate, including a fixed point and a linear subspace. One important property of the Stein-type estimators is that their improvements are larger when the mean vector is close to the shrinkage target. This property implies that the Stein-type estimators are effective, especially when we choose a shrinkage target pertinently, taking account of a prior information concerning the mean vector. For example, Jorion (1985,

1986, 1991) selected the grand mean as a shrinkage target based on the evidence of mean reversion in financial markets. Therefore, it is desirable for us to present the dominance results for a broader class of Stein-type estimators in the mean-variance optimal portfolio selection problem, which shrink not only toward the origin but also toward the other shrinkage targets.

1.2. Summary of the present thesis

In the present thesis, we describe the analytically obtained dominance results for a class of Stein-type estimators for the mean-variance optimal portfolio weights and clarify the conditions for some previously proposed estimators in finance to have smaller risks than the classical estimator, which we obtain by plugging in the sample estimates.

Section 1.3 presents a brief review of the Stein-type shrinkage estimators for the mean vector. In Section 1.4, we present the definition of the loss function to evaluate goodness of an estimator for the mean-variance optimal portfolio weights and fundamental results for the classical estimator.

Chapter 2 presents dominance results for the Stein-type estimators that shrink toward the origin or a fixed point, which are based on Kinkawa and Shinozaki (2010). The class of estimators considered in this chapter extends the estimator given by Mori (2004) in the following two ways. First, we introduce a general class of estimators given by Baranchik (1970). Second, we consider estimators that shrink not only toward the origin but also toward an arbitrary fixed point. The obtained results enable us to clarify the conditions for Kan and Zhou's (2007) two-fund rule estimator and Garlappi, Uppal, and Wang's (2007) estimator to have smaller risks than the classical estimator. Furthermore, we propose an estimator using a prior information concerning the Sharpe ratio, which also has smaller risk than the classical estimator. In this chapter, we also present risk behaviors of the previous estimators by Monte Carlo simulation and the results

of empirical studies.

Chapter 3 presents dominance results for the Stein-type estimators that shrink toward the grand mean or a linear subspace, which are based on Kinkawa and Shinozaki (2009). Jorion's (1985, 1986, 1991) Bayes-Stein estimator, which is very popular in finance, and Kan and Zhou's (2007) three-fund rule estimator belong to this class of estimators. Thus, we clarify the conditions for these estimators to have smaller risks than the classical estimator. Black and Litterman's (1992) estimator, which is also very popular in finance, also belongs to this class.

In Chapters 2 and 3, we also present dominance results for the case with linear constraints on portfolio weights, which is similar to that considered by Mori (2004). Concluding remarks are presented in Chapter 4.

1.3. Review of the Stein-type estimator for the mean vector

In the framework of statistical estimation theory, an estimator is considered to be desirable if it has smaller expected loss than the other estimators. The quadratic loss function is usually used in the estimation problem of the mean vector. Let $\boldsymbol{x}$ be a $p \times 1$ random vector, which is distributed as the multivariate normal distribution $N_p(\boldsymbol{\mu}, I)$. The quadratic loss function for an estimator $\boldsymbol{\delta}(\boldsymbol{x})$ of $\boldsymbol{\mu}$ is defined as

$$L(\boldsymbol{\mu}; \boldsymbol{\delta}(\boldsymbol{x})) = (\boldsymbol{\delta}(\boldsymbol{x}) - \boldsymbol{\mu})'(\boldsymbol{\delta}(\boldsymbol{x}) - \boldsymbol{\mu}).$$

The goodness of $\boldsymbol{\delta}(\boldsymbol{x})$ is evaluated in terms of the expected loss:

$$R(\boldsymbol{\mu}; \boldsymbol{\delta}(\boldsymbol{x})) = E\left[L(\boldsymbol{\mu}; \boldsymbol{\delta}(\boldsymbol{x}))\right],$$

which is generally a function of $\boldsymbol{\mu}$ and is referred to as the risk function of $\boldsymbol{\delta}(\boldsymbol{x})$. Although a natural estimator of $\boldsymbol{\mu}$ is $\boldsymbol{x}$ itself, James and Stein (1961) demonstrated that the following

James-Stein estimator has smaller risk than $\boldsymbol{x}$ under the quadratic loss when p is larger than 2:

$$\boldsymbol{\delta}_{JS}(\boldsymbol{x}) = \left(1 - \frac{p-2}{\boldsymbol{x}'\boldsymbol{x}}\right)\boldsymbol{x},$$

that is, $R(\boldsymbol{\mu}; \boldsymbol{\delta}_{JS}(\boldsymbol{x})) < R(\boldsymbol{\mu}; \boldsymbol{x})$ for all $\boldsymbol{\mu}$. Note that an estimator $\boldsymbol{\delta}_1(\boldsymbol{x})$ is said to *dominate* an estimator $\boldsymbol{\delta}_2(\boldsymbol{x})$ if for all $\boldsymbol{\mu}$, $R(\boldsymbol{\mu}; \boldsymbol{\delta}_1(\boldsymbol{x})) \leq R(\boldsymbol{\mu}; \boldsymbol{\delta}_2(\boldsymbol{x}))$ with strict inequality for some $\boldsymbol{\mu}$.

Although the James-Stein estimator is biased, it dominates the unbiased estimator $\boldsymbol{x}$ in terms of expected loss. One important property of the James-Stein estimator is that it has much smaller risk than $\boldsymbol{x}$ when the true parameter value is close to the origin, which is the shrinkage target. The risk of the James-Stein estimator is less than that of $\boldsymbol{x}$, even when the true parameter value is far from the shrinkage target.

The mathematical statistics literature contains a considerable number of studies on Stein-type estimators. The James-Stein estimator shrinks the unbiased estimator toward the origin, but sometimes results in over-shrinkage and has the opposite sign of the unbiased estimator. The positive-part James-Stein estimator overcomes this disadvantage by prohibiting the sign change and dominates the crude James-Stein estimator, as shown in Baranchik (1964). Strawderman (1971) and Berger (1976) have derived estimators that improve upon the James-Stein estimator and are admissible. Estimators that shrink toward the grand mean, or more generally toward a linear subspace, have also been reported (cf., for example, Lindley 1962 and Casella and Hwang 1987). Furthermore, a considerable number of papers have been dedicated to Stein-type estimation for various types of covariance matrices, loss functions, and distributions other than the normal distribution (cf., for example, Saleh 2006).

In the present thesis, we investigate the estimation problem of the optimal portfolio weights. The number of variables p corresponds to the number of risky assets included in an investor's portfolio. When we estimate the mean vector of risky asset's returns using the James-Stein estimator, the mean of each risky asset is estimated using not only the historical return data of

its asset but also the data for the other assets. Although the James-Stein estimator may not provide an improvement when each mean is estimated, this estimator provides an improvement when we estimate all of the means simultaneously. This property leads to the idea of applying the shrinkage estimators of the mean vector to determine simultaneously portfolio weights for risky assets. Indeed, in Chapters 2 and 3, we demonstrate that the Stein-type estimators for the mean-variance optimal portfolio weights dominate the classical estimator when $p > 2$.

1.4. Estimation problem of the mean-variance optimal portfolio weights

In Section 1.3, we have made a review of the Stein-type estimator for the mean-vector. However, the loss function used to evaluate the goodness of estimators for the mean-variance optimal portfolio weights is different from the quadratic one. In this section, we give the definition of the loss function in the estimation problem of the mean-variance optimal portfolio and discuss some basic properties of the classical estimator.

1.4.1. Definitions and assumptions

We assume that an investor chooses portfolio weights $\boldsymbol{w}$ so as to maximize the mean-variance objective function

$$u(\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{\mu} - \frac{\tau}{2}\boldsymbol{w}'\Sigma\boldsymbol{w}, \tag{1.1}$$

where $\boldsymbol{\mu}$ and Σ are respectively the $p \times 1$ mean vector and the $p \times p$ covariance matrix of excess returns on the p risky assets over the risk free rate, and τ is the degree of risk aversion. Since $\boldsymbol{\mu}$ and Σ are unknown, we need to estimate them.

Let $\boldsymbol{w}(\boldsymbol{\mu}, \Sigma)$ be a true (but unknown) optimal portfolio weights based on $\boldsymbol{\mu}$ and Σ. Letting $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ denote the estimators of $\boldsymbol{\mu}$ and Σ respectively, let $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ be an estimator for the

mean-variance optimal portfolio weights based on the estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$. We define the loss function of the estimator $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ as

$$L\left(\boldsymbol{\mu}, \Sigma; \hat{\boldsymbol{w}}\left(\hat{\boldsymbol{\mu}}, \hat{\Sigma}\right)\right) = u\left(\boldsymbol{w}\left(\boldsymbol{\mu}, \Sigma\right)\right) - u\left(\hat{\boldsymbol{w}}\left(\hat{\boldsymbol{\mu}}, \hat{\Sigma}\right)\right), \tag{1.2}$$

and define the risk function as $R(\boldsymbol{\mu}, \Sigma; \hat{\boldsymbol{w}}(\hat{\boldsymbol{\mu}}, \hat{\Sigma})) = E[L(\boldsymbol{\mu}, \Sigma; \hat{\boldsymbol{w}}(\hat{\boldsymbol{\mu}}, \hat{\Sigma}))]$. Kashima (2001, 2005) and Mori (2004) have defined the same loss. Brown (1976, 1978), Jorion (1986), Kan and Zhou (2007), Okhrin and Schmid (2007), and Golosnoy and Okhrin (2007) have also adopted the same or similar loss. In this context an estimator $\hat{\boldsymbol{w}}_1(\hat{\boldsymbol{\mu}}_1, \hat{\Sigma}_1)$ is said to *dominate* an estimator $\hat{\boldsymbol{w}}_2(\hat{\boldsymbol{\mu}}_2, \hat{\Sigma}_2)$ if for all $(\boldsymbol{\mu}, \Sigma)$, $R(\boldsymbol{\mu}, \Sigma; \hat{\boldsymbol{w}}_1(\hat{\boldsymbol{\mu}}_1, \hat{\Sigma}_1)) \leq R(\boldsymbol{\mu}, \Sigma; \hat{\boldsymbol{w}}_2(\hat{\boldsymbol{\mu}}_2, \hat{\Sigma}_2))$ with strict inequality for some $(\boldsymbol{\mu}, \Sigma)$ (cf., Definition 3 of Kashima 2001).

When we have no constraints on portfolio weights, the solution for the maximization problem of Equation (1.1) is given as

$$\boldsymbol{w}(\boldsymbol{\mu}, \Sigma) = \frac{1}{\tau}\Sigma^{-1}\boldsymbol{\mu}. \tag{1.3}$$

Thus, the loss (1.2) of the estimator $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\mu}}, \hat{\Sigma}) = \tau^{-1}\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$ is written as

$$L(\boldsymbol{\mu}, \Sigma; \hat{\boldsymbol{w}}(\hat{\boldsymbol{\mu}}, \hat{\Sigma})) = (2\tau)^{-1}(\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \Sigma^{-1}\boldsymbol{\mu})'\Sigma(\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \Sigma^{-1}\boldsymbol{\mu}). \tag{1.4}$$

Furthermore, when Σ is known, the loss is written as $(2\tau)^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\Sigma^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$. Therefore, as shown in Kashima (2001, 2005), the estimation problem of the mean-variance optimal portfolio weights reduces to the estimation problem of the mean vector under quadratic loss. However, when the covariance matrix is unknown, the estimation problem of the mean-variance optimal portfolio weights is not reduced in this way.

We also consider the case in which linear constraints $A'\boldsymbol{w} = \boldsymbol{b}$ are imposed, where A is a $p \times q$ matrix of rank $A = q$ and $\boldsymbol{b}$ is a $q \times 1$ vector. In this case, the solution for the maximization problem of Equation (1.1) is

$$\boldsymbol{w}_A(\boldsymbol{\mu}, \Sigma) = \frac{1}{\tau}F_1(A, \Sigma)\boldsymbol{\mu} + F_2(A, \Sigma)\boldsymbol{b}, \tag{1.5}$$

where, as in Mori (2004), we define

$$F_1(A,\Sigma) = \Sigma^{-1} - \Sigma^{-1}A(A'\Sigma^{-1}A)^{-1}A'\Sigma^{-1} \text{ and } F_2(A,\Sigma) = \Sigma^{-1}A(A'\Sigma^{-1}A)^{-1}.$$

The loss function of $\hat{\boldsymbol{w}}_A(\hat{\boldsymbol{\mu}},\hat{\Sigma})$ based on the estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ is defined as

$$\begin{aligned} L(\boldsymbol{\mu},\Sigma;\hat{\boldsymbol{w}}_A(\hat{\boldsymbol{\mu}},\hat{\Sigma})) =& u(\boldsymbol{w}_A(\boldsymbol{\mu},\Sigma)) - u(\hat{\boldsymbol{w}}_A(\hat{\boldsymbol{\mu}},\hat{\Sigma})) \\ =& u(\boldsymbol{w}_A(\boldsymbol{\mu},\Sigma)) \\ & - \boldsymbol{b}'F_2(A,\hat{\Sigma})\boldsymbol{\mu} + (\tau/2)\boldsymbol{b}'[F_2(A,\hat{\Sigma})]'\Sigma F_2(A,\hat{\Sigma})\boldsymbol{b} - (1/\tau)\hat{\boldsymbol{\mu}}'F_1(A,\hat{\Sigma})\boldsymbol{\mu} \\ & + 1/(2\tau)\hat{\boldsymbol{\mu}}'F_1(A,\hat{\Sigma})\Sigma F_1(A,\hat{\Sigma})\hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\mu}}'F_1(A,\hat{\Sigma})\Sigma F_2(A,\hat{\Sigma})\boldsymbol{b}. \end{aligned}$$

We notice that this loss function differs from not only the quadratic loss function but also the loss function of Equation (1.4) which is derived when we have no constraints on portfolio weights. Therefore, the improved estimators when we have no constraints do not necessarily improve upon the classical estimator if we have some constraints on portfolio weights.

Hereafter, we assume that we have N observations on excess returns on p risky assets: $\boldsymbol{x}_i$, $i = 1,\ldots,N$, and that they are independently and identically distributed as the multivariate normal distribution $N_p(\boldsymbol{\mu},\Sigma)$. It is well recognized that actual asset returns deviate from normality and are heterogeneous, especially for high frequency data. However, the dominance for a wide class of Stein-type estimators is not investigated even under the normality and independence assumption, so we believe that our results will provide useful insights for the general case.

The sample estimates of $\boldsymbol{\mu}$ and Σ are given as follows.

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i \sim N_p(\boldsymbol{\mu},\Sigma/N), \quad S = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})' \sim W_p(N-1,\Sigma/N),$$

where $W_p(N-1,\Sigma/N)$ denotes the Wishart distribution with $N-1$ degrees of freedom and covariance matrix Σ/N.

1.4.2. A class of classical estimators

When we have no constraints on portfolio weights, substituting $\bar{\boldsymbol{x}}$ and S for $\boldsymbol{\mu}$ and Σ in Equation (1.3), we have an estimator of the mean-variance optimal portfolio weights $\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$. Slightly more generally, we consider its scalar multiple and define the classical estimator as

$$\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S) = \frac{c}{\tau}S^{-1}\bar{\boldsymbol{x}}. \tag{1.6}$$

Letting θ^2 denote $\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$ and assuming $N > p+4$, as is the case for Equation (41) of Kan and Zhou (2007), an expression for the risk of $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$ can be expressed as

$$\begin{aligned} &2\tau R\left(\boldsymbol{\mu},\Sigma;\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)\right) \\ &= \left(1-\frac{2cN}{N-p-2}\right)\theta^2 + \frac{c^2N^2(N-2)}{(N-p-1)(N-p-2)(N-p-4)}\left(\theta^2+\frac{p}{N}\right). \end{aligned} \tag{1.7}$$

Note that the risk depends on $\boldsymbol{\mu}$ and Σ only through θ^2, that is, the risk has the same value for any $\boldsymbol{\mu}$ and Σ as long as θ^2 has the same value. The unbiased estimator of $\boldsymbol{w}$ is given by $\hat{\boldsymbol{w}}_{ub} = \tau^{-1}(N-p-2)N^{-1}S^{-1}\bar{\boldsymbol{x}}$, because $E\left[N^{-1}S^{-1}\right] = (N-p-2)^{-1}\Sigma^{-1}$ when $N > p+2$ from Muirhead (1982, p.97). From the risk expression in Equation (1.7), we can easily see that the unbiased estimator dominates $\hat{\boldsymbol{w}}_C(1;\bar{\boldsymbol{x}},S) = \tau^{-1}S^{-1}\bar{\boldsymbol{x}}$. However, the unbiased estimator is not admissible even if we restrict to the class of estimators $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$. The value of c that yields minimum risk is

$$c^* = \frac{(N-p-1)(N-p-4)}{N(N-2)}\frac{\theta^2}{\theta^2+p/N}. \tag{1.8}$$

Since $\boldsymbol{\mu}$ and Σ are unknown parameters, the classical estimator given by Equation (1.6) with $c = c^*$ is not available in practice. A reasonable choice of c that does not depend on θ^2 is given by an upper bound of c^*. c^* is bounded from above by

$$c^{**} = \frac{(N-p-1)(N-p-4)}{N(N-2)}, \tag{1.9}$$

and $\hat{\boldsymbol{w}}_C(c^{**};\bar{\boldsymbol{x}},S) = c^{**}\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$ dominates $\hat{\boldsymbol{w}}_{ub}$.

When we have linear constraints on portfolio weights $A'\boldsymbol{w} = \boldsymbol{b}$, similarly to Mori (2004), we define the classical estimator as

$$\hat{\boldsymbol{w}}_{C,A}(c; \bar{\boldsymbol{x}}, S) = \frac{c}{\tau} F_1(A, S)\bar{\boldsymbol{x}} + F_2(A, S)\boldsymbol{b}.$$

Letting φ^2 denote $\boldsymbol{\mu}' F_1(\Sigma, A)\boldsymbol{\mu}$ and assuming that $N > \max(p+1, p-q+4)$, from the proof of Theorem 2.1 in Mori (2004), we can easily obtain the risk expression of $\hat{\boldsymbol{w}}_{C,A}(c; \bar{\boldsymbol{x}}, S)$ as

$$\begin{aligned}
&2\tau R\left(\boldsymbol{\mu}, \Sigma; \hat{\boldsymbol{w}}_{C,A}(c; \bar{\boldsymbol{x}}, S)\right) \\
&= \left(1 - \frac{2cN}{N-p+q-2}\right)\varphi^2 + \frac{c^2 N^2 (N-2)}{(N-p+q-1)(N-p+q-2)(N-p+q-4)}\left(\varphi^2 + \frac{p-q}{N}\right) \\
&\quad + \text{terms which do not contain } c. \qquad (1.10)
\end{aligned}$$

This is minimized when c is equal to

$$c^{\dagger} = \frac{(N-p+q-1)(N-p+q-4)}{N(N-2)} \frac{\varphi^2}{\varphi^2 + (p-q)/N}.$$

Although $c^{\dagger}$ depends on the unknown parameters $\boldsymbol{\mu}$ and Σ, it is bounded from above by

$$c^{\dagger\dagger} = \frac{(N-p+q-1)(N-p+q-4)}{N(N-2)}.$$

$\hat{\boldsymbol{w}}_{C,A}(c^{\dagger\dagger}; \bar{\boldsymbol{x}}, S) = c^{\dagger\dagger}\tau^{-1} F_1(S, A)\bar{\boldsymbol{x}} + F_2(S, A)\boldsymbol{b}$ dominates the unbiased estimator that corresponds to the choice $c = (N-p+q-2)N^{-1}$.

Chapter 2

Shrinkage toward the Origin or a Fixed Point

2.1. Introduction

Jobson, Korkie, and Ratti (1979) have suggested the effectiveness of the James-Stein estimator that shrinks toward the origin in the problem of the mean-variance optimal portfolio selection. Recently, Kan and Zhou (2007) proposed an estimator for the mean-variance optimal portfolio weights, which also shrinks toward the origin. However, they have not investigated the effectiveness of these estimators analytically. Mori (2004) has shown analytically a dominance result of a Stein-type estimator for the mean-variance optimal portfolio weights, which shrinks toward the origin, when the covariance matrix is estimated by the sample estimator. However, the dominance for another interesting class of Stein-type estimators proposed by Baranchik (1970) has not been addressed. Furthermore, Mori (2004) has not considered estimators that shrink toward an arbitrary fixed point. In this chapter, we extend the estimator described by Mori (2004) and present dominance results for a broader class of estimators. Furthermore, we propose an estimator using a prior information concerning the Sharpe ratio, which also has a

smaller risk than the classical estimator.

For the case in which Σ is known, Baranchik (1970) has first introduced the following class of estimators of $\boldsymbol{\mu}$

$$\hat{\boldsymbol{\mu}}_{Stein} = \left(1 - \frac{r((\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\Sigma^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0))}{(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\Sigma^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)}\right)(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) + \boldsymbol{\mu}_0, \tag{2.1}$$

where $r(\cdot)$ is a nondecreasing function and $\boldsymbol{\mu}_0$ is an arbitrary constant vector. Kashima (2001) has also proposed an estimator of the same form under the assumption that returns on risky assets are generated by a linear regression model and has introduced the function $r(\cdot)$ provided by Berger (1976). Here we first consider the Stein-type estimators of the form $(1 - r(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})$ $/(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}))\bar{\boldsymbol{x}}$ as an estimator of $\boldsymbol{\mu}$, which we obtain by replacing Σ in Equation (2.1) by the sample estimate S and setting $\boldsymbol{\mu}_0 = \mathbf{0}$. Lin and Tsai (1973) have shown the dominance results for this class under the quadratic loss function when the covariance matrix is unknown. One of main differences between the class of estimators given in this chapter and the one given by Mori (2004) is that we introduce a non-decreasing function $r(\cdot)$ instead of a constant. By introducing the function $r(\cdot)$, we can evaluate analytically some estimators provided in previous studies. Another is that we consider the estimators that shrink not only toward the origin but also toward an arbitrary fixed point, which will be discussed later.

The remainder of this chapter is organized as follows. Section 2.2 gives the dominance results of a class of Stein-type estimator for the mean-variance optimal portfolio weights when we have no constraints on portfolio weights. In this section, we also show that some estimators provided in previous studies belong to our class. Section 2.3 gives the dominance results when we have linear constraints on portfolio weights. Section 2.4 gives the proofs of the theorems stated in Sections 2.2 and 2.3. Section 2.5 illustrates the risk behaviors of the classical estimator and various Stein-type estimators that belong to the class given in the theorems in Section 2.2.

2.2. Case in which there are no constraints on portfolio weights

Denoting $a^+ = \max(0, a)$, we have the following dominance result under the loss (1.2).

Theorem 2.1. *Let $p > 2$ and $N > p + 4$. If $c_c \geq c_s \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$, $r(\cdot)$ is nondecreasing and $0 \leq r(\cdot) \leq 2(p-2)(N-p-2)^{-1}$, then $\hat{\boldsymbol{w}}_C(c_c; \bar{\boldsymbol{x}}, S) = c_c \tau^{-1} S^{-1} \bar{\boldsymbol{x}}$ is dominated by the following Stein-type estimator*

$$\hat{\boldsymbol{w}}_S(c_s; \bar{\boldsymbol{x}}, S) = \frac{c_s}{\tau} S^{-1} \left(1 - \frac{r(\bar{\boldsymbol{x}}' S^{-1} \bar{\boldsymbol{x}})}{\bar{\boldsymbol{x}}' S^{-1} \bar{\boldsymbol{x}}} \right) \bar{\boldsymbol{x}}. \tag{2.2}$$

Furthermore, the estimator $\hat{\boldsymbol{w}}_S(c_s; \bar{\boldsymbol{x}}, S)$ is dominated by its positive-part Stein-type estimator $\hat{\boldsymbol{w}}_S^+(c_s; \bar{\boldsymbol{x}}, S) = (c_s/\tau) S^{-1} (1 - r(\bar{\boldsymbol{x}}' S^{-1} \bar{\boldsymbol{x}})/(\bar{\boldsymbol{x}}' S^{-1} \bar{\boldsymbol{x}}))^+ \bar{\boldsymbol{x}}$ if they are not identical.

The proof is given in Section 2.4.2.

Now, we present some remarks on Theorem 2.1. Similar remarks apply to the other theorems in the present thesis.

1. *The conditions $p > 2$ and $N > p + 2$.*

The condition $p > 2$ implies that the Stein-type estimator $\hat{\boldsymbol{w}}_S(c_s; \bar{\boldsymbol{x}}, S)$ improves upon the classical estimator $\hat{\boldsymbol{w}}_C(c_c; \bar{\boldsymbol{x}}, S)$ in terms of the expected loss, when the number of risky assets included in portfolio is greater than 2. The condition $N > p + 4$ is required to guarantee the existence of the expectation of S^{-1}. However, since the number of observations is usually large enough, this condition is not restrictive.

2. *The condition $c_c \geq c_s \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$.*

When c_c satisfies the condition $c_c \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$, we can choose c_s satisfying $c_c \geq c_s \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$. However, note that the condition $c_c \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$ is not satisfied for $c_c = c^{**}$ given in Equation (1.9). Actually, $\hat{\boldsymbol{w}}_S(c_s; \bar{\boldsymbol{x}}, S)$ does not improve upon $\hat{\boldsymbol{w}}_C(c_s; \bar{\boldsymbol{x}}, S)$ when $c_s < (N-p-1)(N-p-$

$2)N^{-1}(N-2)^{-1}$, as we will show in Section 2.4.2. In this case, $\hat{\boldsymbol{w}}_C(c_s;\bar{\boldsymbol{x}},S)$ also does not improve upon $\hat{\boldsymbol{w}}_S(c_s;\bar{\boldsymbol{x}},S)$. That is, it depends on the value of θ^2, the estimator of which has a smaller risk than the other when $c_s < (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$.

3. *Shrinkage factor as a function of* $\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}$.

$\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}$ measures the difference between the zero vector, that is the shrinkage target, and the sample mean. When the sample mean is far from the shrinkage target, the value of the shrinkage factor $(1-r(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})/(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}))^+$ is close to 1, and thus the Stein-type estimator is also close to the classical estimator. This partially explains one important property of the Stein-type estimators that even if we set the shrinkage target incorrectly, the Stein-type estimators do not have larger risks than the classical estimator.

4. *The condition* $0 \leq r(\cdot) \leq 2(p-2)(N-p-2)^{-1}$.

Since the upper bound $2(p-2)(N-p-2)^{-1}$ is small when $N \gg p$, the value of the shrinkage factor tends to be close to 1 and the Stein-type estimator is close to the classical estimator. This is quite natural because the estimation error of the classical estimator is small when there are large number of observations. Application of the Stein-type estimators is more effective when a sufficiently large number of observations cannot be obtained.

We also notice that, as given in Theorem 2.1 of Lin and Tsai (1973), the estimator $(1-r(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})/(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}))\bar{\boldsymbol{x}}$ of $\boldsymbol{\mu}$ improves upon the sample estimator $\bar{\boldsymbol{x}}$ when $0 < r(\cdot) < 2(p-2)(N-p+2)^{-1}$ under quadratic loss, which is different from the condition in Theorem 2.1.

5. *The role of the function* $r(\cdot)$.

By introducing the non-decreasing function $r(\cdot)$, we have succeeded in presenting general dominance results for a class of shrinkage estimators for the mean-variance optimal portfolio weights. As will be seen in the following examples, from the general results, we are able to clarify the conditions for the previously proposed estimators to dominate the classical estimator. Further,

general dominance results may enable us to develop new shrinkage estimators pertinently. We first note that the risks of $\hat{\boldsymbol{w}}_C(c_c; \bar{\boldsymbol{x}}, S)$ and $\hat{\boldsymbol{w}}_S(c_s; \bar{\boldsymbol{x}}, S)$ depend on $\boldsymbol{\mu}$ and Σ only through $\theta^2 = \boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$ from the expression of the risk given by Equation (2.11) in Section 2.4.2. Since $\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}$ is an estimator of the squared Sharpe ratio $\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$, we may choose the function $r(\cdot)$ using a prior information concerning the Sharpe ratio. We will present one example in this direction in the following Example 2.4.

We, here, present some examples of the choice of the function $r(\cdot)$. Although some choices have been given by Lin and Tsai (1973), which lead to generalized Bayes minimax estimators under quadratic loss, we present others here. Although we give the results for the case with no constraints on the portfolio weights, we have similar results for the case with linear constraints, which will be given in Section 2.3.

Example 2.1. A simple choice of $r(\cdot)$ is given by $r_1(v) = d_1 v/(v + d_2)$, $v \geq 0$, where d_1 and d_2 are nonnegative constants. We see from Theorem 2.1 that the corresponding Stein-type estimator improves upon $\hat{\boldsymbol{w}}_C(c_c; \bar{\boldsymbol{x}}, S)$ if $0 < d_1 \leq 2(p-2)(N-p-2)^{-1}$ and $d_2 \geq 0$.

Example 2.2. *Kan and Zhou's two-fund rule estimator.* One of the estimators provided by Kan and Zhou (2007) is

$$\hat{\boldsymbol{w}}_{KZ2} = \frac{1}{\tau}\frac{(N-p-1)(N-p-4)}{N(N-2)}S^{-1}\left(1 - \frac{p/N}{\tilde{\theta}^2(\hat{\theta}^2) + p/N}\right)\bar{\boldsymbol{x}}, \tag{2.3}$$

where

$$\tilde{\theta}^2(\hat{\theta}^2) = \frac{(N-p-2)\hat{\theta}^2 - p}{N} + \frac{2(\hat{\theta}^2)^{p/2}(1+\hat{\theta}^2)^{-(N-2)/2}}{NB_{\hat{\theta}^2/(1+\hat{\theta}^2)}(p/2, (N-p)/2)}, \quad \hat{\theta}^2 = \bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}},$$

and $B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\mathrm{d}y$.

We see that $\hat{\theta}^2 \sim p(N-p)^{-1}F_{p,N-p}(N\theta^2)$, where $F_{p,N-p}(N\theta^2)$ denotes the noncentral F distribution with p and $N-p$ degrees of freedom and noncentrality parameter $N\theta^2$. Thus the unbiased estimator of $N\theta^2$ is $(N-p-2)\hat{\theta}^2 - p$. Kubokawa, Robert, and Saleh (1993) have,

however, provided an estimator $N\tilde{\theta}^2(\hat{\theta}^2)$ that improves upon the unbiased estimator under squared error loss. The improvement is significant when θ^2 is small. Kan and Zhou (2007) obtained the two-fund rule estimator by substituting $\tilde{\theta}^2(\hat{\theta}^2)$ for θ^2 in Equation (1.8).

If we set $r(\cdot)$ equal to $r_{KZ2}(\hat{\theta}^2) = d_1\hat{\theta}^2/(\tilde{\theta}^2(\hat{\theta}^2) + d_2/N)$, then $\hat{\boldsymbol{w}}_S$ given in Theorem 2.1 is of the same form as the two-fund rule estimator given in Equation (2.3). Therefore, the two-fund rule estimator can be considered as a Stein-type estimator shrinking $\bar{\boldsymbol{x}}$ toward the origin. Now, we show that $r_{KZ2}(\cdot)$ is nonnegative and nondecreasing if d_1 and d_2 satisfy $0 < d_1 < 2(p-2)N(N-p-2)^{-1}$ and $d_2 \geq p$. Let $v = \bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}$. We first show that $r_{KZ2}(\cdot)$ is nondecreasing. Since

$$\frac{\mathrm{d}}{\mathrm{d}v}\left(\frac{d_1 v}{\tilde{\theta}^2(v) + d_2/N}\right) = d_1\frac{\tilde{\theta}^2(v) + d_2/N - v(\mathrm{d}\tilde{\theta}^2(v)/\mathrm{d}v)}{(\tilde{\theta}^2(v) + d_2/N)^2},$$

we only need to prove that $v(\mathrm{d}\tilde{\theta}^2(v)/\mathrm{d}v) < \tilde{\theta}^2(v) + d_2/N$ if $d_2 \geq p$.

Let $f(v) = 2v^{p/2}(1+v)^{-(m+p-2)/2}/B_{v/(1+v)}(p/2, m/2)$. Since $f(v) = 2v^{p/2}(1+v)^{-(m+p-2)/2}$ $\times[\int_0^v t^{p/2-1}(1+t)^{-(m+p)/2}\mathrm{d}t]^{-1}$ as shown in the Appendix of Kan and Zhou (2007), we have $\mathrm{d}\tilde{\theta}^2(v)/\mathrm{d}v = (N-p-2)N^{-1} - (1/2)v^{-1}(1+v)^{-1}f(v)\tilde{\theta}^2(v) \leq (N-p-2)N^{-1}$ by direct differentiation. Since $N\tilde{\theta}^2(v) = (N-p-2)v - p + f(v)$ and $f(v) > 0$, we see that if $d_2 \geq p$,

$$v(\mathrm{d}\tilde{\theta}^2(v)/\mathrm{d}v) \leq v(N-p-2)N^{-1} = \tilde{\theta}^2(v) + (p - f(v))N^{-1} \leq \tilde{\theta}^2(v) + d_2N^{-1}.$$

Next, we prove that $r_{KZ2}(v)$ is nonnegative. As shown in the Appendix of Kan and Zhou (2007), $p - f(v) = (N-p-2)\int_0^v t^{p/2}(1+t)^{-N/2}\mathrm{d}t/(\int_0^v t^{p/2-1}(1+t)^{-N/2}\mathrm{d}t)$. From Equation (2.3) of Kubokawa et al. (1993), we see that $0 \leq p - f(v) \leq p(N-p-2)(p+2)^{-1}v \leq (N-p-2)v$, and thus $N\tilde{\theta}^2(v) = (N-p-2)v - p + f(v) \geq 0$.

For the case of the two-fund rule estimator, $d_1 = p/N$ and $d_2 = p$, and the corresponding $r_{KZ2}(\cdot)$ satisfies the condition given in Theorem 2.1, but the choice $c_s = (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$ does not. Therefore, from the argument given in Section 2.4.2, we see that

$\hat{\boldsymbol{w}}_{KZ2}$ with $c_s = (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$ does not uniformly improve upon the classical estimator.

Example 2.3. *Garlappi, Uppal, and Wang's estimator.* Garlappi, Uppal, and Wang (2007) have proposed an estimator that shrinks toward the origin and has different shrinkage factors for each subset of assets (Proposition 3). Setting the number of subsets equal to 1, the estimator reduces to the following:

$$\hat{\boldsymbol{w}}_{GUW} = \frac{1}{\tau} S^{-1} \left[1 - \min \left(\frac{\sqrt{\varepsilon}}{\sqrt{\hat{\theta}^2}}, 1 \right) \right] \bar{\boldsymbol{x}},$$

where $\varepsilon = ep(N-p)^{-1}$ and e is a constant such that $P[(N-p)p^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})' S^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) \leq e] = 1-\alpha$, that is, $e = F^{-1}_{p,N-p}(1-\alpha)$, where $F^{-1}_{p,N-p}(\cdot)$ is the inverse function of the central F distribution function with p and $N-p$ degrees of freedom. We note that $\hat{\theta}^2 \sim (N-p)N^{-1}F_{p,N-p}$ under the null hypothesis that $\theta^2 = 0$. This estimator is based on the idea that an investor will not invest in risky assets when the Sharpe ratio (cf., Sharpe 1966, 1994) of the estimated optimal portfolio is not significantly different from 0, but he/she will do, according to his/her aversion to uncertainty, which is reflected in α, when the Sharpe ratio is significantly different from 0. Although their estimator is not designed to dominate the classical estimator, it belongs to our class if we set $r(\cdot)$ equals to $r_{GUW}(\hat{\theta}^2) = \min(\sqrt{\varepsilon\hat{\theta}^2}, \hat{\theta}^2)$. $r_{GUW}(\cdot)$ is an increasing function, but it will be larger than the upper bound $2(p-2)(N-p-2)^{-1}$ given in Theorem 2.1. Thus, we may modify $r_{GUW}(\cdot)$ as

$$r^*_{GUW}(\hat{\theta}^2) = \min \left(\sqrt{\varepsilon\hat{\theta}^2}, \hat{\theta}^2, \frac{2(p-2)}{N-p-2} \right).$$

Note that $r^*_{GUW}(\cdot)$ reduces to

$$r^*_{GUW}(\hat{\theta}^2) = \min \left(\hat{\theta}^2, \frac{2(p-2)}{N-p-2} \right), \tag{2.4}$$

when $\varepsilon \geq 2(p-2)(N-p-2)^{-1}$, as is the case if $p = 10$, $N = 60$, and $\alpha = 0.05$.

Example 2.4. *Using prior information concerning the Sharpe ratio.* As discussed in Section 1.4.2, the value of c that results in minimum risk for the classical type estimator is given by Equation (1.8), which is a function of the squared Sharpe ratio $\theta^2 = \boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$ (cf., Sharpe 1966, 1994) of the tangency portfolio $\Sigma^{-1}\boldsymbol{\mu}(\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu})^{-1}$. Since θ^2 is unknown, we consider constructing an estimator by choosing a suitable function $r(\cdot)$ which reflects prior information about θ^2 when it is available. Wang (2005) and Garlappi et al. (2007) have also proposed utilizing information concerning the Sharpe ratio to estimate the optimal portfolio. However, our estimator will also have uniformly smaller risk than the classical estimator.

By replacing c in Equation (1.6) with c^* given by Equation (1.8), we obtain

$$\hat{\boldsymbol{w}}_C = \frac{c^{**}}{\tau}\frac{\theta^2}{\theta^2 + p/N}S^{-1}\boldsymbol{x} = \frac{c^{**}}{\tau}\left(1 - \frac{(p/N)\hat{\theta}^2}{\theta^2 + p/N}\frac{1}{\hat{\theta}^2}\right)S^{-1}\boldsymbol{x}.$$

Thus, we may choose $r(\hat{\theta}^2) = (p/N)\hat{\theta}^2(\theta_0^2 + p/N)^{-1}$, where θ_0^2 is determined reflecting the prior information of the squared Sharpe ratio. To take account of the condition for the dominance given in Theorem 2.1, we modify it as

$$r_{sr}(\hat{\theta}^2) = \min\left(\frac{(p/N)\hat{\theta}^2}{\theta_0^2 + p/N}, \frac{2(p-2)}{N-p-2}\right). \tag{2.5}$$

We impose the upper bound $2(p-2)(N-p-2)^{-1}$ on $r_{sr}(\cdot)$ so as not to allow the estimator to have a larger risk than the classical estimator for some θ^2, at least when we set $c_s = c_c = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$. When the upper bound is not imposed, the estimator $\hat{\boldsymbol{w}}_{Sharpe}$ could have a larger risk than the classical estimator for some large value of θ^2. When $\theta_0^2 \leq (p/2)(N-p-2)(p-2)^{-1}N^{-1}\hat{\theta}^2 - pN^{-1}$, that is, when the estimator $\hat{\theta}^2 = \bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}$ of θ^2 has a much larger value than the prior θ_0^2, then $r_{sr} = 2(p-2)(N-p-2)^{-1}$, and thus the estimator $\hat{\boldsymbol{w}}_{Sharpe}$ does not reflect the prior value θ_0^2.

We expect that this estimator will significantly improve upon the classical estimator if the prior information $\theta^2 = \theta_0^2$ is approximately correct. We also note that, when we set $\theta_0^2 = 0$, Equation (2.5) reduces to Equation (2.4). Finally we note that the estimator using the prior

information concerning the Sharpe ratio belongs to the class of estimators given in Theorems 2.1 and 2.3.

In the context of estimating the mean vector, Stein-type estimator shrinking toward an arbitrary constant vector improves upon the sample mean vector. Furthermore, since some investors have certain prior information about excess returns on the assets, it is reasonable to consider estimators of the mean-variance optimal portfolio weights which reflect that prior information. However, we have some technical difficulties in showing the general dominance of the Stein-type estimators that shrink toward an arbitrary constant vector under the loss of Equation (1.2). Nevertheless we can show the following.

Theorem 2.2. *Let $p > 2$ and $N > p+4$. If $c_c \geq c_s = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$ and d is a constant which satisfies $0 < d < 2(p-2)(N-p-2)^{-1}$, then $\hat{\boldsymbol{w}}_C(c_c; \bar{\boldsymbol{x}}, S) = c_c\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$ is dominated by the following Stein-type estimator*

$$\frac{c_s}{\tau}S^{-1}\left[\left(1 - \frac{d}{(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'S^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)}\right)(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) + \boldsymbol{\mu}_0\right], \tag{2.6}$$

where $\boldsymbol{\mu}_0$ is a $p \times 1$ arbitrary constant vector.

The proof is given in Section 2.4.3.

We notice that the condition $c_c \geq c_s = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$ on c_c and c_s differs from the one given in Theorem 2.1. When $c_c \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$, the Stein-type estimator with $c_s = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$ improves upon the classical estimator.

2.3. Case in which there are linear constraints on portfolio weights

Now we turn to the case in which we have linear constraints on portfolio weights as discussed in Mori (2004). By replacing $\boldsymbol{\mu}$ and Σ in Equation (1.5) by $c(1 - d/(\bar{\boldsymbol{x}}' F_1(S, A)\bar{\boldsymbol{x}}))\bar{\boldsymbol{x}}$ and S respectively, where d is a positive constant, he has provided the following Stein-type estimator for the mean-variance optimal portfolio weights:

$$\hat{\boldsymbol{w}}_{Mori} = \frac{c}{\tau} F_1(A, S) \left(1 - \frac{d}{\bar{\boldsymbol{x}}' F_1(A, S)\bar{\boldsymbol{x}}}\right) \bar{\boldsymbol{x}} + F_2(A, S)\boldsymbol{b}. \tag{2.7}$$

By replacing the constant d in Equation (2.7) with a non-decreasing function $r(\cdot)$, we give a broader class of Stein-type estimators of $\boldsymbol{w}_A(\boldsymbol{\mu}, \Sigma)$ in the following.

Theorem 2.3. *Let $p > q+2$ and $N > \max(p+1, p-q+4)$. If $c_c \geq c_s \geq (N-p+q-1)(N-p+q-2)N^{-1}(N-2)^{-1}$, $r_A(\cdot)$ is nondecreasing and $0 \leq r_A(\cdot) \leq 2(p-q-2)(N-p+q-2)^{-1}$, then $\hat{\boldsymbol{w}}_{C,A}(c_c; \bar{\boldsymbol{x}}, S) = c_c\tau^{-1}F_1(A, S)\bar{\boldsymbol{x}} + F_2(A, S)\boldsymbol{b}$ is dominated by the following Stein-type estimator*

$$\hat{\boldsymbol{w}}_{S,A}(c_s; \bar{\boldsymbol{x}}, S) = \frac{c_s}{\tau} F_1(A, S) \left(1 - \frac{r_A(\bar{\boldsymbol{x}}' F_1(A, S)\bar{\boldsymbol{x}})}{\bar{\boldsymbol{x}}' F_1(A, S)\bar{\boldsymbol{x}}}\right) \bar{\boldsymbol{x}} + F_2(A, S)\boldsymbol{b}.$$

Furthermore, the estimator $\hat{\boldsymbol{w}}_{S,A}(c_s; \bar{\boldsymbol{x}}, S)$ is dominated by its positive-part Stein-type estimator $\hat{\boldsymbol{w}}^+_{S,A}(c_s; \bar{\boldsymbol{x}}, S) = c_s\tau^{-1}F_1(A, S)(1 - r_A(\bar{\boldsymbol{x}}' F_1(A, S)\bar{\boldsymbol{x}})/(\bar{\boldsymbol{x}}' F_1(A, S)\bar{\boldsymbol{x}}))^+\bar{\boldsymbol{x}} + F_2(A, S)\boldsymbol{b}$ if they are not identical.

The proof is given in Section 2.4.4.

When the function $r_A(\cdot)$ is a constant and $c_s = c_c$, Theorem 2.3 reduces to Theorem 2.3 of Mori (2004) except for the upper bound of $r_A(\cdot)$. The upper bound given by Mori (2004) depends on c and is smaller than ours. Furthermore, he has not given the result when $c_s \neq c_c$.

We notice that $p-q$ in Theorem 2.3 plays the same role with p in Theorem 2.1. This means that when there are linear equality constraints on the portfolio weights, the effective range of

the number of risky assets p and the function $r(\cdot)$ is narrower than that for the case with no constraints. We need q more risky assets to expect that the Stein-type estimator behaves in the same way with the case with no constraints.

Next, similarly to Theorem 2.2, we give the following.

Theorem 2.4. *Let $p > q+2$ and $N > \max(p+1, p-q+4)$. If $c_c \geq c_s = (N-p+q-1)(N-p+q-2)N^{-1}(N-2)^{-1}$ and d_A is a constant which satisfies $0 \leq d_A \leq 2(p-q-2)(N-p+q-2)^{-1}$, then $\hat{\boldsymbol{w}}_{C,A}(c_c; \bar{\boldsymbol{x}}, S) = c_c\tau^{-1}F_1(A,S)\bar{\boldsymbol{x}} + F_2(A,S)\boldsymbol{b}$ is dominated by the following Stein-type estimator*

$$\frac{c_s}{\tau}F_1(A,S)\left[\left(1 - \frac{d_A}{(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'F_1(A,S)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)}\right)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0) + \boldsymbol{\mu}_0\right] + F_2(A,S)\boldsymbol{b},$$

where $\boldsymbol{\mu}_0$ is a $p \times 1$ arbitrary constant vector.

The proof is omitted, because we can show it by a similar argument to the proofs of Theorems 2.2 and 2.3.

In practice, a possible choice of $\boldsymbol{\mu}_0$ is given by $\boldsymbol{\mu}_0 = \alpha\mathbf{1}$ where α is a constant. However, when there is a constraint that $\mathbf{1}'\boldsymbol{w}$ is equal to a constant value, the estimator given in Theorem 2.4 reduces to the estimator given in Theorem 2.3 because $F(\mathbf{1}, S)\mathbf{1} = 0$. Thus, in this case, the estimator given in Theorem 2.4 is meaningful only for the case in which we have a prior information that returns of risky assets do not have a common value.

2.4. Proofs of Theorems 2.1, 2.2, and 2.3

2.4.1. Lemmas

First, we give some lemmas used to prove Theorems 2.1, 2.2, and 2.3 in the following subsection. Throughout this section we suppose that $\boldsymbol{y} \sim N_p(\boldsymbol{\eta}, I/N)$ and $W \sim W_p(N-1, I/N)$ and that $\boldsymbol{y}$ and W are independent.

Lemma 2.1. *Put* $\omega_{11} = \boldsymbol{y}'W^{-1}\boldsymbol{y}(\boldsymbol{y}'\boldsymbol{y})^{-1}$, *then* ω_{11} *is independent of* $\boldsymbol{y}$ *and* $N\omega_{11}^{-1} \sim \chi^2_{N-p}$. *If the expectation of each term of the following equations exists, then*

(i) $E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\dfrac{\boldsymbol{y}'W^{-2}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\right] = \dfrac{N-2}{N-p-1}E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right]$,

(ii) $E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\dfrac{\boldsymbol{a}'W^{-1}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\right] = E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\dfrac{\boldsymbol{a}'\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}\right]$,

(iii) $E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})^2\dfrac{\boldsymbol{y}'W^{-2}\boldsymbol{y}}{(\boldsymbol{y}'W^{-1}\boldsymbol{y})^2}\right] = \dfrac{N-2}{N-p-1}E\left[\dfrac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})^2}{\boldsymbol{y}'\boldsymbol{y}}\right]$,

(iv) $E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\dfrac{\boldsymbol{a}'W^{-2}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\right] = \dfrac{N-2}{N-p-1}E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\dfrac{\boldsymbol{a}'\boldsymbol{y}\omega_{11}}{\boldsymbol{y}'\boldsymbol{y}}\right]$,

where $\boldsymbol{a}$ *is a non-random* $p \times 1$ *vector.*

Proof. Let Q be an orthogonal matrix satisfying $Q\boldsymbol{y} = (\boldsymbol{y}'\boldsymbol{y})^{1/2}\boldsymbol{e}_1$, where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$. Denote $U = QWQ'$ and let it be partitioned as

$$U = \begin{pmatrix} u_{11} & \boldsymbol{u}_{21}{}' \\ \boldsymbol{u}_{21} & U_{22} \end{pmatrix},$$

where u_{11} is a scalar, $\boldsymbol{u}_{21}$ is a $(p-1) \times 1$ vector and U_{22} is a $(p-1) \times (p-1)$ matrix. Since $Q'\boldsymbol{e}_1 = (\boldsymbol{y}'\boldsymbol{y})^{-1/2}\boldsymbol{y}$, we see that $\boldsymbol{e}_1{}'U^{-1}\boldsymbol{e}_1 = \boldsymbol{y}'W^{-1}\boldsymbol{y}(\boldsymbol{y}'\boldsymbol{y})^{-1}$ and $\omega_{11} = \boldsymbol{e}_1{}'U^{-1}\boldsymbol{e}_1 = (u_{11} - \boldsymbol{u}_{21}{}'U_{22}^{-1}\boldsymbol{u}_{21})^{-1}$. $N\omega_{11}^{-1}$ has χ^2_{N-p} distribution from Theorem 3.2.10 of Muirhead (1982) because $U \sim W_p(N-1, I/N)$. From this and Theorem 3.2.12 of Muirhead (1982), we see that ω_{11} and $\boldsymbol{y}$ are independent.

From Theorem 3.2.10 of Muirhead (1982), we also easily see that $U_{22}^{-1/2}\boldsymbol{u}_{21} \sim N_{p-1}(\boldsymbol{0}, I/N)$ and $U_{22} \sim W_{p-1}(N-1, I/N)$ and that $U_{22}^{-1/2}\boldsymbol{u}_{21}$ and U_{22} are independent. Thus, we have $E[\boldsymbol{u}_{21}{}'U_{22}^{-2}\boldsymbol{u}_{21}] = (p-1)(N-p-1)^{-1}$ for $N > p+1$. From this and the fact that ω_{11}^{-1} and $U_{22}^{-1}\boldsymbol{u}_{21}$ are independent from Theorem 3.2.12 of Muirhead (1982), we have

$$\begin{aligned} E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\frac{\boldsymbol{y}'W^{-2}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\right] &= E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{\boldsymbol{y}'\boldsymbol{y}\omega_{11}^2(1+\boldsymbol{u}_{21}'U_{22}^{-2}\boldsymbol{u}_{21})}{\boldsymbol{y}'\boldsymbol{y}\omega_{11}}\right] \\ &= \frac{N-2}{N-p-1}E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right]. \end{aligned}$$

Since $E[U_{22}^{-1}\boldsymbol{u}_{21}] = \mathbf{0}$, we have

$$E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\frac{\boldsymbol{a}'W^{-1}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\right] = E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{(\boldsymbol{y}'\boldsymbol{y})^{1/2}\boldsymbol{a}'Q'}{\boldsymbol{y}'\boldsymbol{y}\omega_{11}}\begin{pmatrix}1\\ -U_{22}^{-1}\boldsymbol{u}_{21}\end{pmatrix}\omega_{11}\right]$$
$$= E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{\boldsymbol{a}'\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}\right].$$

Similarly, we have

$$E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})^2\frac{\boldsymbol{y}'W^{-2}\boldsymbol{y}}{(\boldsymbol{y}'W^{-1}\boldsymbol{y})^2}\right] = E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})^2\frac{\boldsymbol{y}'\boldsymbol{y}\omega_{11}^2(1+\boldsymbol{u}_{21}'U_{22}^{-2}\boldsymbol{u}_{21})}{(\boldsymbol{y}'\boldsymbol{y}\omega_{11})^2}\right]$$
$$= \frac{N-2}{N-p-1}E\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})^2}{\boldsymbol{y}'\boldsymbol{y}}\right],$$

and

$$E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\frac{\boldsymbol{a}'W^{-2}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\right]$$
$$= E\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})}{\boldsymbol{y}'\boldsymbol{y}\omega_{11}}(\boldsymbol{y}'\boldsymbol{y})^{1/2}\boldsymbol{a}'Q'\left\{\begin{pmatrix}1\\ -U_{22}^{-1}\boldsymbol{u}_{21}\end{pmatrix}(1+\boldsymbol{u}_{21}'U_{22}^{-2}\boldsymbol{u}_{21})\omega_{11}^2 + \begin{pmatrix}0\\ -U_{22}^{-1}\boldsymbol{u}_{21}\end{pmatrix}\omega_{11}\right\}\right]$$
$$= \frac{N-2}{N-p-1}E\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})}{\boldsymbol{y}'\boldsymbol{y}\omega_{11}}(\boldsymbol{y}'\boldsymbol{y})^{1/2}\boldsymbol{a}'Q'\boldsymbol{e}_1\omega_{11}^2\right] = \frac{N-2}{N-p-1}E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{\boldsymbol{a}'\boldsymbol{y}\omega_{11}}{\boldsymbol{y}'\boldsymbol{y}}\right].$$

□

Lemma 2.2. *Let $r(\cdot)$ be differentiable, and denote its derivative by $r'(\cdot)$. If the expectation of each term of the following equations exists, then*

(i) $(N-p-2)E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right] = NE\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right] + 2E\left[r'(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\boldsymbol{y}'\boldsymbol{y}\omega_{11}^2\right]$,

(ii) $E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{\boldsymbol{\eta}'\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}\right] = E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right] - \frac{2}{N}E\left[r'(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right] - \frac{p-2}{N}E\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})}{\boldsymbol{y}'\boldsymbol{y}}\right]$,

where $\omega_{11} = \boldsymbol{y}'W^{-1}\boldsymbol{y}(\boldsymbol{y}'\boldsymbol{y})^{-1}$.

Proof. Put $\omega_0 = N\omega_{11}^{-1}$ and $h(\omega_0) = Nr(\boldsymbol{y}'\boldsymbol{y}N\omega_0^{-1})\omega_0^{-1}$. Using integration by parts, we have the identity $E[h(\omega_0)\omega_0] = (N-p)E[h(\omega_0)] + 2E[h'(\omega_0)\omega_0]$ (cf., for example, Equation (2.15)

of Efron and Morris 1976). Since $h'(\omega_0) = -N^2\boldsymbol{y}'\boldsymbol{y}r'(\boldsymbol{y}'\boldsymbol{y}N\omega_0^{-1})\omega_0^{-3} - Nr(\boldsymbol{y}'\boldsymbol{y}N\omega_0^{-1})\omega_0^{-2}$, we have (i).

Put $\boldsymbol{z} = (z_1, \ldots, z_p)' = \sqrt{N}\boldsymbol{y}$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)' = \sqrt{N}\boldsymbol{\eta}$, and $g(z_k) = r(N^{-1}\sum_{j=1}^p {z_j}^2\omega_{11})$ $z_k(\sum_{j=1}^p {z_j}^2)^{-1}$. Since $\boldsymbol{z} \sim N(\boldsymbol{\gamma}, I)$, from Stein's lemma (cf., for example, Lemma 3.5.1 in Anderson 2003), we have $E[g(z_k)(z_k - \gamma_k)] = E[\partial g(z_k)/\partial z_k]$ and

$$\begin{aligned}\frac{\partial g(z_k)}{\partial z_k} =& 2N^{-1}\omega_{11}r'\left(N^{-1}\sum_{j=1}^p {z_j}^2\omega_{11}\right)\frac{{z_k}^2}{\sum_{j=1}^p {z_j}^2} \\ &+ r\left(N^{-1}\sum_{j=1}^p {z_j}^2\omega_{11}\right)\left(\frac{1}{\sum_{j=1}^p {z_j}^2} - \frac{2{z_k}^2}{(\sum_{j=1}^p {z_j}^2)^2}\right).\end{aligned}$$

Since

$$E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{(\boldsymbol{y}-\boldsymbol{\eta})'\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}\right] = \sum_{j=1}^p E\left[g(z_j)(z_j - \gamma_j)\right] = \sum_{j=1}^p E\left[\frac{\partial g(z_j)}{\partial z_j}\right],$$

we have (ii).

□

Before providing Lemma 2.3, we let $\tilde{A} = \Sigma^{-1/2}A$ and define P as the orthogonal matrix such that

$$P\tilde{A}\tilde{A}'P' = \Lambda, \quad \Lambda = \begin{pmatrix} \Lambda_1 & O \\ O & O \end{pmatrix}, \quad P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix},$$

where Λ is a diagonal matrix with eigenvalues of $\tilde{A}\tilde{A}'$ on its diagonal, Λ_1 is a $q \times q$ diagonal matrix with positive diagonal elements, P_1 is a $q \times p$ matrix and P_2 is a $(p-q) \times p$ matrix.

Denoting $W = \Sigma^{-1/2}S\Sigma^{-1/2}$, we define a matrix V and partition it as

$$V = PWP' = \begin{pmatrix} P_1WP_1' & P_1WP_2' \\ P_2WP_1' & P_2WP_2' \end{pmatrix} = \begin{pmatrix} V_{11} & V_{21}' \\ V_{21} & V_{22} \end{pmatrix}.$$

Then the inverse of V is partitioned and is expressed as

$$V^{-1} = \begin{pmatrix} V^{(11)} & V^{(21)'} \\ V^{(21)} & V^{(22)} \end{pmatrix} = \begin{pmatrix} V_{11\cdot 2}^{-1} & -V_{11\cdot 2}^{-1}V_{21}'V_{22}^{-1} \\ -V_{22}^{-1}V_{21}V_{11\cdot 2}^{-1} & V_{22}^{-1} + V_{22}^{-1}V_{21}V_{11\cdot 2}^{-1}V_{21}'V_{22}^{-1} \end{pmatrix},$$

where $V_{11\cdot 2} = V_{11} - V_{21}'V_{22}^{-1}V_{21}$.

Lemma 2.3. *Let $L = P_1\tilde{A}$, then*

(i) $F_1(\tilde{A}, W) = P_2'V_{22}^{-1}P_2$,

(ii) $F_1(\tilde{A}, W)^2 = P_2'V_{22}^{-2}P_2$,

(iii) $F_2(\tilde{A}, W) = P_1'(L')^{-1} - P_2'V_{22}^{-1}V_{21}(L')^{-1}$,

(iv) $F_1(\tilde{A}, W)F_2(\tilde{A}, W) = -P_2'V_{22}^{-2}V_{21}(L')^{-1}$.

Proof. (i) and (iii) are given by Lemma 3.1 in Mori (2004). (ii) and (iv) are easily obtained from (i) and (iii). □

2.4.2. Proof of Theorem 2.1

First, assuming that $c_c = c_s = c$, we provide an expression of the risk difference between the classical type estimator $\hat{\boldsymbol{w}}_C(c; \bar{\boldsymbol{x}}, S)$ and the Stein-type estimator $\hat{\boldsymbol{w}}_S(c; \bar{\boldsymbol{x}}, S)$. Next, we derive the conditions for the Stein-type estimator $\hat{\boldsymbol{w}}_S(c_s; \bar{\boldsymbol{x}}, S)$ to improve upon the classical estimator $\hat{\boldsymbol{w}}_C(c_c; \bar{\boldsymbol{x}}, S)$. Finally, we prove that for any $c > 0$ the positive-part Stein-type estimator $\hat{\boldsymbol{w}}_S^+(c; \bar{\boldsymbol{x}}, S)$ improves upon the Stein-type estimator $\hat{\boldsymbol{w}}_S(c; \bar{\boldsymbol{x}}, S)$ if they are not identical.

For simplicity, we assume that the function $r(\cdot)$ is differentiable in the proofs. Even if $r(\cdot)$ is not differentiable, the proofs go through by applying Riemann integration and replacing the terms $r'(x)\mathrm{d}x$ by $\mathrm{d}r(x)$.

Risk difference

Putting $\boldsymbol{y} = \Sigma^{-1/2}\bar{\boldsymbol{x}}$, $\boldsymbol{\eta} = \Sigma^{-1/2}\boldsymbol{\mu}$ and $W = \Sigma^{-1/2}S\Sigma^{-1/2}$, we see that $\boldsymbol{y} \sim N_p(\boldsymbol{\eta}, I/N)$ and $W \sim W_p(N-1, I/N)$, and $\boldsymbol{y}$ and W are independent. The loss difference between $\hat{\boldsymbol{w}}_C(c; \bar{\boldsymbol{x}}, S)$

and $\hat{\boldsymbol{w}}_S(c;\bar{\boldsymbol{x}},S)$ under the loss function given by Equation (1.2) is given as

$$
\begin{aligned}
&2\tau\Delta L \\
&= 2c^2 r(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})\frac{\bar{\boldsymbol{x}}'S^{-1}\Sigma S^{-1}\bar{\boldsymbol{x}}}{\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}} - 2cr(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})\frac{\boldsymbol{\mu}'S^{-1}\bar{\boldsymbol{x}}}{\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}} - c^2 r(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})^2\frac{\bar{\boldsymbol{x}}'S^{-1}\Sigma S^{-1}\bar{\boldsymbol{x}}}{(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})^2} \qquad (2.8)\\
&= 2c^2 r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\frac{\boldsymbol{y}'W^{-2}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}} - 2cr(\boldsymbol{y}'W^{-1}\boldsymbol{y})\frac{\boldsymbol{\eta}'W^{-1}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}} - c^2 r(\boldsymbol{y}'W^{-1}\boldsymbol{y})^2\frac{\boldsymbol{y}'W^{-2}\boldsymbol{y}}{(\boldsymbol{y}'W^{-1}\boldsymbol{y})^2}.
\end{aligned}
$$

Putting $\omega_{11} = (\boldsymbol{y}'W^{-1}\boldsymbol{y})(\boldsymbol{y}'\boldsymbol{y})^{-1}$, we have from Lemmas 2.1 (i) and 2.2 (i) in Section 2.4.1

$$
\begin{aligned}
&E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\frac{\boldsymbol{y}'W^{-2}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\right] \\
&= \frac{N-2}{N-p-1}E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right] \qquad (2.9)\\
&= \frac{N(N-2)}{(N-p-1)(N-p-2)}E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right] + \frac{2(N-2)}{(N-p-1)(N-p-2)}E\left[r'(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\boldsymbol{y}'\boldsymbol{y}\omega_{11}^2\right].
\end{aligned}
$$

From Lemmas 2.1 (ii) and 2.2 (ii), we have

$$
\begin{aligned}
&E\left[r(\boldsymbol{y}'W^{-1}\boldsymbol{y})\frac{\boldsymbol{\eta}'W^{-1}\boldsymbol{y}}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\right] \\
&= E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{\boldsymbol{\eta}'\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}\right] \qquad (2.10)\\
&= E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right] - \frac{2}{N}E\left[r'(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right] - \frac{p-2}{N}E\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})}{\boldsymbol{y}'\boldsymbol{y}}\right].
\end{aligned}
$$

Finally, applying Lemma 2.1 (iii), we have the risk difference as

$$
\begin{aligned}
2\tau\Delta R =& 2c\left(c\frac{N(N-2)}{(N-p-1)(N-p-2)} - 1\right)E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right] \qquad (2.11)\\
&+ 2c\frac{p-2}{N}E\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})}{\boldsymbol{y}'\boldsymbol{y}}\right] - c^2\frac{N-2}{N-p-1}E\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})^2}{\boldsymbol{y}'\boldsymbol{y}}\right] + C,
\end{aligned}
$$

where

$$
C = 4c^2\frac{N-2}{(N-p-1)(N-p-2)}E\left[r'(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\boldsymbol{y}'\boldsymbol{y}\omega_{11}^2\right] + \frac{4c}{N}E\left[r'(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right] \geq 0.
$$

Conditions for improving upon $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$

Assuming that $c \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$ and $0 \leq r(\cdot) \leq 2(p-2)(N-p-2)^{-1}$, we first show that $\hat{\boldsymbol{w}}_S(c;\bar{\boldsymbol{x}},S)$ improves upon $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$.

We note that since both $r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})$ and $-1/(\boldsymbol{y}'\boldsymbol{y})$ are nondecreasing functions of $\boldsymbol{y}'\boldsymbol{y}$, from Lemma 6.6 of Lehmann and Casella (1998), we have $E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})(-(\boldsymbol{y}'\boldsymbol{y})^{-1})\right] \geq E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right]$ $\times E\left[-(\boldsymbol{y}'\boldsymbol{y})^{-1}\right]$. Thus we have

$$
\begin{aligned}
& E\left[\left(2\frac{p-2}{N} - c\frac{N-2}{N-p-1}r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right)\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})}{\boldsymbol{y}'\boldsymbol{y}}\right] \\
& \geq E\left[\left(2\frac{p-2}{N} - c\frac{N-2}{N-p-1}\frac{2(p-2)}{N-p-2}\right)\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})}{\boldsymbol{y}'\boldsymbol{y}}\right] \\
& \geq 2\frac{p-2}{N}\left(c\frac{N(N-2)}{(N-p-1)(N-p-2)} - 1\right)E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right]E\left[-\frac{1}{\boldsymbol{y}'\boldsymbol{y}}\right].
\end{aligned}
$$

Therefore, we have

$$
2\tau\Delta R \geq 2c\left(c\frac{N(N-2)}{(N-p-1)(N-p-2)} - 1\right)\left(1 - \frac{p-2}{N}E\left[\frac{1}{\boldsymbol{y}'\boldsymbol{y}}\right]\right)E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right].
$$

Since $N\boldsymbol{y}'\boldsymbol{y}$ has the noncentral χ^2 distribution with p degrees of freedom and noncentrality parameter $N\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$, we have $E\left[(\boldsymbol{y}'\boldsymbol{y})^{-1}\right] = E\left[N(p-2+2Z)^{-1}\right]$, where Z is a random variable having the Poisson distribution with mean $(N/2)\boldsymbol{\mu}\Sigma^{-1}\boldsymbol{\mu}$. Thus, we see that $2\tau\Delta R \geq 0$. Here, we note that if $N > p+4$, the risk of $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$, which is given by Equation (1.7), is finite, and that this condition is also necessary for the finiteness of the risk of $\hat{\boldsymbol{w}}_S(c;\bar{\boldsymbol{x}},S)$.

Finally, from Equation (1.7), the risk of $\hat{\boldsymbol{w}}_C(c_1;\bar{\boldsymbol{x}},S)$ is larger than that of $\hat{\boldsymbol{w}}_C(c_2;\bar{\boldsymbol{x}},S)$ when $c_1 > c_2 \geq (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$. Therefore, we see that if $c_c \geq c_s \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$, $\hat{\boldsymbol{w}}_S(c_s;\bar{\boldsymbol{x}},S)$ improves upon $\hat{\boldsymbol{w}}_C(c_c;\bar{\boldsymbol{x}},S)$.

Case in which $c < (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$

Here we show that $\hat{\boldsymbol{w}}_S(c;\bar{\boldsymbol{x}},S)$ does not dominate $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$ when $c < (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$. Using Equation (2.9), Equation (2.10), and Lemma 2.1 (iii), we express the risk difference between $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$ and $\hat{\boldsymbol{w}}_S(c;\bar{\boldsymbol{x}},S)$ as

$$
\begin{aligned}
& 2\tau\Delta R \\
& = \frac{2c^2(N-2)}{N-p-1}E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right] - 2cE\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\boldsymbol{\eta}'\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}\right] - \frac{c^2(N-2)}{N-p-1}E\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})^2}{\boldsymbol{y}'\boldsymbol{y}}\right]. \quad (2.12)
\end{aligned}
$$

Let $\lim_{u\to\infty} r(u) = b > 0$. We see that when $\boldsymbol{\eta}'\boldsymbol{\eta} \to \infty$, $E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\boldsymbol{\eta}'\boldsymbol{y}(\boldsymbol{y}'\boldsymbol{y})^{-1}\right] \to b$ and $E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\omega_{11}\right] \to bE\left[\omega_{11}\right] = bN(N-p-2)^{-1}$. Therefore, when $\boldsymbol{\eta}'\boldsymbol{\eta} \to \infty$, the sum of the first and second terms of Equation (2.12) approaches

$$2c^2 b\frac{N(N-2)}{(N-p-1)(N-p-2)} - 2cb = 2cb\left(c\frac{N(N-2)}{(N-p-1)(N-p-2)} - 1\right).$$

Therefore, we see that the risk difference given by Equation (2.12) is negative when $\boldsymbol{\eta}'\boldsymbol{\eta}$ is sufficiently large if $c < (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$, and that $\hat{\boldsymbol{w}}_S(c;\bar{\boldsymbol{x}},S)$ does not improve upon $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$.

Positive-part Stein-type estimator $\hat{\boldsymbol{w}}_S^+(c;\bar{\boldsymbol{x}},S)$

Defining $g(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}}) = 1 - r(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})/(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})$, we express the risk difference between $\hat{\boldsymbol{w}}_S(c;\bar{\boldsymbol{x}},S)$ and $\hat{\boldsymbol{w}}_S^+(c;\bar{\boldsymbol{x}},S)$ as

$$\begin{aligned}2\tau\Delta R =& c^2 E\left[\left\{g(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})\right\}^2 \bar{\boldsymbol{x}}'S^{-1}\Sigma S^{-1}\bar{\boldsymbol{x}} - \left\{g(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})^+\right\}^2 \bar{\boldsymbol{x}}'S^{-1}\Sigma S^{-1}\bar{\boldsymbol{x}}\right] \\ &+ 2cE\left[\left\{g(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})^+ - g(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})\right\}\boldsymbol{\mu}'S^{-1}\bar{\boldsymbol{x}}\right].\end{aligned}$$

The first term of the right-hand side is clearly nonnegative and is equal to zero if and only if $g \equiv g^+$. Therefore, we need only to show that the second term is nonnegative. In a similar way to the proof of Lemma 2.1 (ii), the second term can be expressed as

$$E\left[\left\{g(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})^+ - g(\bar{\boldsymbol{x}}'S^{-1}\bar{\boldsymbol{x}})\right\}\boldsymbol{\mu}'S^{-1}\bar{\boldsymbol{x}}\right] = E\left[E\left[\left\{g(\boldsymbol{y}'\boldsymbol{y}\omega_{11})^+ - g(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\right\}\boldsymbol{\eta}'\boldsymbol{y}\omega_{11}|\omega_{11}\right]\right],$$

which we can show is nonnegative by applying the proof of Lemma 3.5.2 of Anderson (2003).

2.4.3. Proof of Theorem 2.2

First, assuming $c_c = c_s = c$, we evaluate the risk difference between $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$ and the estimator $\hat{\boldsymbol{w}}_{Stein}(c;\bar{\boldsymbol{x}},S) \equiv c\tau^{-1}S^{-1}[(1 - r((\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*)\ ((\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*)^{-1})\ \bar{\boldsymbol{x}}^* + \boldsymbol{\mu}_0]$, where $\bar{\boldsymbol{x}}^* = \bar{\boldsymbol{x}} - \boldsymbol{\mu}_0$ and $r(\cdot)$ is a nondecreasing function. Putting $\boldsymbol{\mu}^* = \boldsymbol{\mu} - \boldsymbol{\mu}_0$, we have the loss difference

between $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$ and $\hat{\boldsymbol{w}}_{Stein}(c;\bar{\boldsymbol{x}},S)$ under the loss function given by Equation (1.2) as

$$\begin{aligned}2\tau L =&\frac{2c^2r((\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*)}{(\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*}\boldsymbol{\mu}_0'S^{-1}\Sigma S^{-1}\bar{\boldsymbol{x}}^* - \frac{2cr((\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*)}{(\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*}\boldsymbol{\mu}_0'S^{-1}\bar{\boldsymbol{x}}^* \\ &+\frac{2c^2r((\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*)}{(\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*}(\bar{\boldsymbol{x}}^*)'S^{-1}\Sigma S^{-1}\bar{\boldsymbol{x}}^* - \frac{2cr((\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*)}{(\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*}\boldsymbol{\mu}^{*\prime}S^{-1}\bar{\boldsymbol{x}}^* \\ &-\frac{c^2r((\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*)^2}{((\bar{\boldsymbol{x}}^*)'S^{-1}\bar{\boldsymbol{x}}^*)^2}(\bar{\boldsymbol{x}}^*)'S^{-1}\Sigma S^{-1}\bar{\boldsymbol{x}}^*.\end{aligned} \tag{2.13}$$

Since $\bar{\boldsymbol{x}}^* \sim N(\boldsymbol{\mu}^*, \Sigma/N)$ and the last three terms of the right-hand side are of the same form as Equation (2.8), the expectation of their sum is nonnegative when the conditions given in Theorem 2.1 are satisfied.

We show that for the first and second terms of Equation (2.13), the expectation of their sum is zero when $r(\cdot)$ is a constant and $c = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$. Putting $\boldsymbol{y} = \Sigma^{-1/2}\bar{\boldsymbol{x}}^*$, $\boldsymbol{\eta} = \Sigma^{-1/2}\boldsymbol{\mu}^*$, $W = \Sigma^{-1/2}S\Sigma^{-1/2}$, and $\boldsymbol{\eta}_0 = \Sigma^{-1/2}\boldsymbol{\mu}_0$, wee see that $\boldsymbol{y} \sim N(\boldsymbol{\eta}, I/N)$ and $W \sim W_p(N-1, I/N)$. From Lemma 2.1 (ii) and (iv), the expectation is written as

$$\begin{aligned}&2cE\left[\frac{r(\boldsymbol{y}'W^{-1}\boldsymbol{y})}{\boldsymbol{y}'W^{-1}\boldsymbol{y}}\left(c\boldsymbol{\eta}_0'W^{-2}\boldsymbol{y} - \boldsymbol{\eta}_0'W^{-1}\boldsymbol{y}\right)\right] \\ &= 2c^2\frac{N-2}{N-p-1}E\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{\boldsymbol{\eta}_0'\boldsymbol{y}\omega_{11}}{\boldsymbol{y}'\boldsymbol{y}}\right] - 2cE\left[\frac{r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})}{\boldsymbol{y}'\boldsymbol{y}}\boldsymbol{\eta}_0'\boldsymbol{y}\right] \\ &= 2cE\left[r(\boldsymbol{y}'\boldsymbol{y}\omega_{11})\frac{\boldsymbol{\eta}_0'\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}\left(c\omega_{11}\frac{N-2}{N-p-1} - 1\right)\right].\end{aligned} \tag{2.14}$$

Now we put $r(\cdot) = const.$ Since $\boldsymbol{y}$ and ω_{11} are independent and $E[\omega_{11}] = N(N-p-2)^{-1}$, we can easily see that Equation (2.14) is zero for $c = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$.

Finally, from Equation (1.7), the risk of $\hat{\boldsymbol{w}}_C(c_1;\bar{\boldsymbol{x}},S)$ is larger than that of $\hat{\boldsymbol{w}}_C(c_2;\bar{\boldsymbol{x}},S)$ when $c_1 > c_2 \geq (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$. Therefore, we see that if $c_c \geq c_s = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$, the estimator $\hat{\boldsymbol{w}}_{Stein}(c_s;\bar{\boldsymbol{x}},S)$ improves upon $\hat{\boldsymbol{w}}_C(c_c;\bar{\boldsymbol{x}},S)$.

2.4.4. Proof of Theorem 2.3

Putting $\tilde{A} = \Sigma^{-1/2}A$ and noting that $F_1(A,S) = \Sigma^{-1/2}F_1(\tilde{A},W)\Sigma^{-1/2}$ and $F_2(A,S) = \Sigma^{-1/2}$ $F_2(\tilde{A},W)$, from Lemma 3.2 (iii) and (iv) of Mori (2004) , we have the loss difference between

$\hat{\boldsymbol{w}}_{C,A}(c;\bar{\boldsymbol{x}},S)$ and $\hat{\boldsymbol{w}}_{S,A}(c;\bar{\boldsymbol{x}},S)$ under the loss function (1.2) as

$$
\begin{aligned}
&2\tau\Delta L \\
&= -2cr_A(\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}})\frac{\bar{\boldsymbol{x}}'F_1(A,S)\boldsymbol{\mu}}{\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}}} - c^2r_A(\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}})^2\frac{\bar{\boldsymbol{x}}'F_1(A,S)\Sigma F_1(A,S)\bar{\boldsymbol{x}}}{(\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}})^2} \\
&\quad + 2c^2r_A(\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}})\frac{\bar{\boldsymbol{x}}'F_1(A,S)\Sigma F_1(A,S)\bar{\boldsymbol{x}}}{\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}}} + 2\tau cr_A(\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}})\frac{\bar{\boldsymbol{x}}'F_1(A,S)\Sigma F_2(A,S)\boldsymbol{b}}{\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}}} \\
&= -2cr_A(\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{y})\frac{\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{\eta}}{\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{y}} - c^2r_A(\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{y})^2\frac{\boldsymbol{y}'F_1(\tilde{A},W)^2\boldsymbol{y}}{(\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{y})^2} \\
&\quad + 2c^2r_A(\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{y})\frac{\boldsymbol{y}'F_1(\tilde{A},W)^2\boldsymbol{y}}{\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{y}} + 2\tau cr_A(\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{y})\frac{\boldsymbol{y}'F_1(\tilde{A},W)F_2(\tilde{A},W)\boldsymbol{b}}{\boldsymbol{y}'F_1(\tilde{A},W)\boldsymbol{y}}.
\end{aligned}
$$

We put $\tilde{\boldsymbol{y}} = P_2\boldsymbol{y}$, $\tilde{\boldsymbol{\eta}} = P_2\boldsymbol{\eta}$, $V_{21} = P_2WP_1'$, and $V_{22} = P_2WP_2'$. Using Lemma 2.3 (i) $\sim$ (iv), we have

$$
\begin{aligned}
2\tau\Delta L = &- 2cr_A(\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{y}})\frac{\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{\eta}}}{\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{y}}} - c^2r_A(\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{y}})^2\frac{\tilde{\boldsymbol{y}}'V_{22}^{-2}\tilde{\boldsymbol{y}}}{(\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{y}})^2} \\
&+ 2c^2r_A(\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{y}})\frac{\tilde{\boldsymbol{y}}'V_{22}^{-2}\tilde{\boldsymbol{y}}}{\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{y}}} - 2\tau cr_A(\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{y}})\frac{\tilde{\boldsymbol{y}}'V_{22}^{-2}V_{21}(L')^{-1}\boldsymbol{b}}{\tilde{\boldsymbol{y}}'V_{22}^{-1}\tilde{\boldsymbol{y}}}.
\end{aligned}
$$

Let $\mathrm{vec}(X)$ denote the operation that stacks the columns of a matrix X into a vector. Since $\mathrm{vec}(V_{22}^{-1/2}V_{21})$ follows $N_{(p-q)\times q}(\boldsymbol{0}, I/N)$ and is independent of V_{22} and $\tilde{\boldsymbol{y}}$, the expectation of the fourth term is zero. Therefore, we have the similar expression of the risk difference to the expectation of Equation (2.8) except for the number of variables, and we have Theorem 2.3 by a similar argument to the proof of Theorem 2.1. We note that if $N > \max(p+1, p-q+4)$ the risk of $\hat{\boldsymbol{w}}_{C,A}(c;\bar{\boldsymbol{x}},S)$, whose expression is given in Equation (1.10), is finite and that this condition is also necessary for the finiteness of the risk of $\hat{\boldsymbol{w}}_{S,A}(c;\bar{\boldsymbol{x}},S)$.

2.5. Comparison of estimators

2.5.1. Risk comparison by Monte Carlo simulation

In this section, we investigate the risk behaviors of six estimators by Monte Carlo simulation for the case with no constraints. The six estimators are (i) the "optimal" estimator $\hat{\boldsymbol{w}}_{opt}(c^*) =$

$c^*\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$, with c^* given by Equation (1.8), (ii) a classical estimator $\hat{\boldsymbol{w}}_C(c) = c\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$, which is a benchmark for the other estimators, (iii) a Stein-type estimator with $r(\cdot) = (p-2)(N-p-2)^{-1}$: $\hat{\boldsymbol{w}}_{SD}(c) = c\tau^{-1}S^{-1}[(1-(p-2)(N-p-2)^{-1}/\hat{\varphi}^2)^+(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)+\boldsymbol{\mu}_0]$, where $\hat{\varphi}^2 = (\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'S^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)$, (iv) Kan and Zhou's two-fund rule estimator: $\hat{\boldsymbol{w}}_{KZ2}(c) = c\tau^{-1}S^{-1}[1-pN^{-1}\hat{\theta}^2/(\tilde{\theta}^2(\hat{\theta}^2)+pN^{-1})]\bar{\boldsymbol{x}}$, (v) estimators that use prior information concerning the Sharpe ratio: $\hat{\boldsymbol{w}}_{Sharpe}(c) = c\tau^{-1}S^{-1}[1-r_{sr}(\hat{\theta}^2)/\hat{\theta}^2]^+\bar{\boldsymbol{x}}$, and (vi) Garlappi, Uppal, and Wang's estimator: $\hat{\boldsymbol{w}}_{GUW}(c) = c\tau^{-1}S^{-1}[1-r_{GUW}(\hat{\theta}^2)/\hat{\theta}^2]^+\bar{\boldsymbol{x}}$, where we set $\alpha = 0.01$ and 0.05. Although the "optimal" estimator is not available in practice, we present its risk values in order to compare them with the risks of the other estimators.

We generated N random vectors from the p-variate normal distribution $N((\theta^2/p)^{1/2}\mathbf{1}, I)$ for some selected values of θ^2. For the estimators, the averaged losses over 100,000 repetitions were used to approximate their risks. As can be seen from Equation (1.7) and the expression of Equation (2.11) in Section 2.4.2, risks of the classical estimators and the estimators of the form of Equation (2.2) in Theorem 2.1 depend on $\boldsymbol{\mu}$ and Σ only through $\theta^2 = \boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$. That is, the risk values are the same for any $\boldsymbol{\mu}$ and Σ if the value of θ^2's is constant. The risks of the estimators that have the form of Equation (2.6) in Theorem 2.2 depend on $\boldsymbol{\mu}$ and Σ only through θ^2 and $\varphi^2 = (\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\Sigma^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)$ when $c_s = (N-p-1)(N-p-2)N^{-1}(N-1)^{-1}$. Although the risks of the estimators given in Theorem 2.2 depend on $\boldsymbol{\mu}$ and Σ when $c_s \neq (N-p-1)(N-p-2)N^{-1}(N-1)^{-1}$, we set $\boldsymbol{\mu}_0 = (a^2/p)^{1/2}\mathbf{1}$ in $\hat{\boldsymbol{w}}_{SD}(c)$ where a is a scalar. In this section we set the degree of risk aversion to $\tau = 3$, following Kan and Zhou (2007). As seen in Section 1.4, τ is just a scaling factor in the risk function, and the risk for other values of τ can be obtained by simply scaling the risk for $\tau = 3$.

Firstly, we report on the results for the case $p = 10$ and $N = 60$. The risk values of $\hat{\boldsymbol{w}}_{opt}(c^*)$ are given at the top of Table 2.1. Panel A of Table 2.1 presents the risk values of the six estimators for various values of $\theta^2 = \boldsymbol{\mu}'\boldsymbol{\mu}$ for the choice $c = c^{**} = (N-p-1)(N-$

Table 2.1: Risk values of estimators for the case $p = 10$ and $N = 60$

	Risk × 10,000							Improvements(%)						
θ^2	0	0.05	0.1	0.5	1	5	10	0	0.05	0.1	0.5	1	5	10
$\hat{\boldsymbol{w}}_{opt}(c^*)$	0	68	116	326	509	1801	3389							
A: $c = c^{**} = (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$														
$\hat{\boldsymbol{w}}_C(c^{**})$	**224**	240	**256**	383	**541**	1809	**3393**							
$\hat{\boldsymbol{w}}_{SD}(c^{**})$														
$a^2 = 0$	**34**	100	**150**	356	**533**	1819	**3405**	**85**	58	**41**	7	**1**	-1	**0**
$a^2 = 0.05$	82	50	77	329	524	1819	3406	64	79	70	14	3	-1	0
$a^2 = 0.1$	**113**	58	**66**	307	**518**	1819	**3406**	**49**	76	**74**	20	**4**	-1	**0**
$a^2 = 0.5$	187	169	162	193	438	1820	3408	17	30	37	50	19	-1	0
$a^2 = 1$	**204**	205	**211**	253	**351**	1819	**3409**	**9**	15	**18**	34	**35**	-1	**0**
$a^2 = 5$	220	233	248	367	516	1619	3413	2	3	3	4	5	11	-1
$a^2 = 10$	**222**	236	**252**	375	**530**	1754	**3203**	**1**	1	**2**	2	**2**	3	**6**
$\hat{\boldsymbol{w}}_{KZ2}(c^{**})$	29	88	138	360	538	1822	3409	87	63	46	6	1	-1	0
$\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$														
$\theta_0^2 = 0$	**4**	83	**152**	404	**567**	1837	**3422**	**98**	65	**41**	-5	**-5**	-2	**-1**
$\theta_0^2 = 0.05$	14	72	128	399	567	1837	3422	94	70	50	-4	-5	-2	-1
$\theta_0^2 = 0.1$	**32**	76	**120**	390	**567**	1837	**3422**	**86**	68	**53**	-2	**-5**	-2	**-1**
$\theta_0^2 = 0.5$	126	146	166	330	541	1837	3422	44	39	35	14	0	-2	-1
$\theta_0^2 = 1$	**165**	182	**199**	337	**511**	1837	**3422**	**27**	24	**22**	12	**6**	-2	**-1**
$\theta_0^2 = 5$	210	226	242	369	528	1802	3417	6	6	6	4	2	0	-1
$\theta_0^2 = 10$	**217**	233	**249**	376	**534**	1803	**3390**	**3**	3	**3**	2	**1**	0	**0**
$\hat{\boldsymbol{w}}_{GUW}(c^{**})$														
$\alpha = 0.01$	0	83	163	620	888	2332	3984	100	66	36	-62	-64	-29	-17
$\alpha = 0.05$	0	82	158	534	778	2198	3841	100	66	38	-40	-44	-22	-13
B: $c = \tilde{c}^{**} = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$														
$\hat{\boldsymbol{w}}_C(\tilde{c}^{**})$	244	260	276	404	564	1842	3439	-9	-8	-8	-6	-4	-2	-1
$\hat{\boldsymbol{w}}_{SD}(\tilde{c}^{**})$														
$a^2 = 0$	37	104	155	363	541	1837	3436	83	57	40	5	0	-2	-1
$a^2 = 0.05$	89	53	79	330	528	1836	3436	60	78	69	14	2	-1	-1
$a^2 = 0.1$	123	64	69	307	520	1835	3436	45	73	73	20	4	-1	-1
$a^2 = 0.5$	203	187	180	197	432	1831	3435	9	22	30	48	20	-1	-1
$a^2 = 1$	222	225	233	276	357	1826	3434	1	6	9	28	34	-1	-1
$a^2 = 5$	239	254	270	394	549	1635	3412	-7	-6	-5	-3	-1	10	-1
$a^2 = 10$	242	257	273	400	559	1816	3232	-8	-7	-7	-5	-3	0	5
$\hat{\boldsymbol{w}}_{KZ2}(\tilde{c}^{**})$	31	91	141	363	542	1837	3437	86	62	45	5	0	-2	-1
$\hat{\boldsymbol{w}}_{Sharpe}(\tilde{c}^{**})$														
$\theta_0^2 = 0$	4	84	152	402	563	1841	3439	98	65	40	-5	-4	-2	-1
$\theta_0^2 = 0.05$	15	72	128	397	563	1841	3439	93	70	50	-4	-4	-2	-1
$\theta_0^2 = 0.1$	35	78	121	388	563	1841	3439	84	68	53	-1	-4	-2	-1
$\theta_0^2 = 0.5$	137	156	175	331	537	1841	3439	39	35	32	14	1	-2	-1
$\theta_0^2 = 1$	179	196	212	345	514	1841	3439	20	18	17	10	5	-2	-1
$\theta_0^2 = 5$	229	244	260	387	546	1815	3434	-2	-2	-2	-1	-1	0	-1
$\theta_0^2 = 10$	236	252	268	395	554	1826	3416	-5	-5	-5	-3	-2	-1	-1
$\hat{\boldsymbol{w}}_{GUW}(\tilde{c}^{**})$														
$\alpha = 0.01$	0	83	163	613	867	2236	3831	100	66	36	-60	-60	-24	-13
$\alpha = 0.05$	1	82	157	526	757	2113	3709	100	66	39	-37	-40	-17	-9

$p-4)N^{-1}(N-2)^{-1}$, which gives the smallest risk in the class of classical estimators as seen in Equation (1.9) in Section 1.4.2. As we have already seen in Theorems 2.1 and 2.2, the estimators given in Equations (2.2) and (2.6) with $c_s = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$ have smaller risks than $\hat{\boldsymbol{w}}_C(c)$ with $c = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$ for any θ^2. However, when we set $c = (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$, the corresponding classical estimator and the Stein-type estimator do not improve upon each other as seen in Section 2.2. Thus, we mainly investigate the risk behavior of the estimators when we choose $c = c^{**}$. The improvement percentages are also given, which are defined as $1-$ (the proportion of the risk of an estimator to the risk of $\hat{\boldsymbol{w}}_C(c^{**})$). The boldfaced figures in Panel A of Table 2.1 will be used for comparison with those for the other settings of p and N later. The first row in Panel A of Table 2.1 gives the risk values of $\hat{\boldsymbol{w}}_C(c^{**})$. We see that the risk of $\hat{\boldsymbol{w}}_C(c^{**})$ is an increasing function of θ^2, which is consistent with Equation (1.7).

The next seven rows give the risk values of $\hat{\boldsymbol{w}}_{SD}(c^{**})$. In general, we may say that the smaller θ^2 and a^2 are, the smaller the risk of $\hat{\boldsymbol{w}}_{SD}(c^{**})$ is. However, we can confirm the effectiveness of the estimators that shrink toward a fixed point $\boldsymbol{\mu}_0$, since their improvements are larger when a^2 is closer to θ^2. The following row returns the risk values of $\hat{\boldsymbol{w}}_{KZ2}(c^{**})$. Kan and Zhou (2007) expected that the two-fund rule estimator would be better than $\hat{\boldsymbol{w}}_{SD}(c^{**})$ when θ^2 is small, which is consistent with our simulation results. Finally, we note that $\hat{\boldsymbol{w}}_{SD}(c^{**})$ and $\hat{\boldsymbol{w}}_{KZ2}(c^{**})$ have a slightly larger risk than $\hat{\boldsymbol{w}}_C(c^{**})$, when $\theta^2 = 5$ and 10.

In the next seven rows, we present the risk values of $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ for several values of θ_0^2. $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ has sometimes slightly larger risk values than $\hat{\boldsymbol{w}}_C(c^{**})$, except for the setting $\theta_0^2 = 10$. The smallest risk value is attained when $\theta_0^2 = \theta^2$ for all values of θ^2. We expect that the estimator is effective when θ_0^2 is closer to θ^2, which is consistent with the simulation results. However, the improvement is smaller when θ^2 is larger, which can also be seen for the other estimators. For some values of θ^2, $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ has larger risk values than $\hat{\boldsymbol{w}}_{SD}(c^{**})$ for

some choices of a. The difference between the risk behavior of $\hat{\boldsymbol{w}}_{Sharpe}(c)$ and that of $\hat{\boldsymbol{w}}_{SD}(c)$ arises from the investor's prior information used in respective estimators. $\hat{\boldsymbol{w}}_{Sharpe}(c)$ requires only information about θ^2. On the other hand, $\hat{\boldsymbol{w}}_{SD}(c)$ needs information about all elements of $\boldsymbol{\mu}$, which is not directly linked to the information concerning the Sharpe ratio. Thus, we may conclude that $\hat{\boldsymbol{w}}_{Sharpe}(c)$ is useful when we have some rough information concerning the Sharpe ratio of the optimal portfolio.

As is stated previously, the choice $c = c^{**}$ does not satisfy the condition given in Theorems 2.1 and 2.2. However, we have found that the estimators $\hat{\boldsymbol{w}}_{SD}(c^{**})$, $\hat{\boldsymbol{w}}_{KZ2}(c^{**})$, and $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ have much smaller risks than $\hat{\boldsymbol{w}}_{C}(c^{**})$ for small θ^2 and have only slightly larger risks than $\hat{\boldsymbol{w}}_{C}(c^{**})$ for some large θ^2.

The last two rows give the risk values of $\hat{\boldsymbol{w}}_{GUW}(c^{**})$. As discussed in Examples 2.3 and 2.4 of Section 2.2, $\hat{\boldsymbol{w}}_{GUW}(c^{**})$ reduces to $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ with $\theta_0^2 = 0$, if the upper bound $2(p-2)(N-p-2)^{-1}$ is imposed. We therefore present only the risk values of the original estimator provided by Garlappi et al. (2007) except for the choice of c. When $\theta^2 = 0$, the risk value of $\hat{\boldsymbol{w}}_{GUW}(c^{**})$ is almost 0 for $\alpha = 0.01$ and 1×10^{-4} for $\alpha = 0.05$, and these are smaller than the others. However, the risk values of $\hat{\boldsymbol{w}}_{GUW}(c^{**})$ are larger than those of $\hat{\boldsymbol{w}}_{C}(c^{**})$ when $\theta^2 \geq 0.5$.

We have also investigated the risk behavior of the estimators for the choice of $c = \tilde{c}^{**} = (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$. $\tilde{c}^{**}$ is the smallest value of c_s which satisfies the condition given in Theorems 2.1 and 2.2. The results for the case $c = \tilde{c}^{**}$ are given in Panel B of Table 2.1. We have found that the risks of $\hat{\boldsymbol{w}}_{SD}(\tilde{c}^{**})$, $\hat{\boldsymbol{w}}_{KZ2}(\tilde{c}^{**})$, and $\hat{\boldsymbol{w}}_{Sharpe}(\tilde{c}^{**})$ are sometimes slightly larger than $\hat{\boldsymbol{w}}_{C}(c^{**})$ when the shrinkage target is not closer to the true parameter θ^2. However, the estimators have much smaller risks than $\hat{\boldsymbol{w}}_{C}(c^{**})$ when the shrinkage target is closer to the true parameter. We can also confirm that the estimators $\hat{\boldsymbol{w}}_{SD}(\tilde{c}^{**})$, $\hat{\boldsymbol{w}}_{KZ2}(\tilde{c}^{**})$, and $\hat{\boldsymbol{w}}_{Sharpe}(\tilde{c}^{**})$ have smaller risks than $\hat{\boldsymbol{w}}_{C}(\tilde{c}^{**})$ for all θ^2.

Next, we investigate the risk behaviors for other values of p and N. Table 2.2 presents the

Table 2.2: Risk values of estimators for some pairs of p and N

			Risk × 10,000				Improvements(%)			
	a^2, θ_0^2	θ^2	0	0.1	1	10	0	0.1	1	10
A: $p = 10$, $N = 30$										
$\hat{\boldsymbol{w}}_C(c^{**})$			319	382	943	6549				
$\hat{\boldsymbol{w}}_{SD}(c^{**})$	$a^2 = 0$		66	204	937	6597	79	47	1	-1
	$a^2 = 0.1$		132	128	903	6601	59	66	4	-1
	$a^2 = 1$		271	280	689	6613	15	27	27	-1
	$a^2 = 10$		314	369	907	6292	2	3	4	4
$\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$	$\theta_0^2 = 0$		14	172	1022	6656	96	55	-8	-2
	$\theta_0^2 = 0.1$		29	155	1010	6656	91	59	-7	-2
	$\theta_0^2 = 1$		182	251	881	6656	43	34	7	-2
	$\theta_0^2 = 10$		299	362	924	6552	6	5	2	0
B: $p = 10$, $N = 120$										
$\hat{\boldsymbol{w}}_C(c^{**})$			121	137	272	1625				
$\hat{\boldsymbol{w}}_{SD}(c^{**})$	$a^2 = 0$		17	98	269	1627	86	28	1	0
	$a^2 = 0.1$		81	32	264	1627	34	77	3	0
	$a^2 = 1$		115	123	167	1628	5	10	39	0
	$a^2 = 10$		121	136	270	1518	0	1	1	7
$\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$	$\theta_0^2 = 0$		1	116	278	1631	99	15	-2	0
	$\theta_0^2 = 0.1$		37	83	278	1631	69	39	-2	0
	$\theta_0^2 = 1$		104	120	263	1631	15	12	4	0
	$\theta_0^2 = 10$		119	135	270	1623	2	1	1	0
C: $p = 10$, $N = 240$										
$\hat{\boldsymbol{w}}_C(c^{**})$			67	75	145	840				
$\hat{\boldsymbol{w}}_{SD}(c^{**})$	$a^2 = 0$		9	61	144	841	87	18	1	0
	$a^2 = 0.1$		53	17	142	841	21	78	2	0
	$a^2 = 1$		65	71	86	841	3	5	40	0
	$a^2 = 10$		66	74	144	780	0	0	0	7
$\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$	$\theta_0^2 = 0$		1	74	146	841	99	0	-1	0
	$\theta_0^2 = 0.1$		33	55	146	841	50	26	-1	0
	$\theta_0^2 = 1$		61	70	142	841	8	7	2	0
	$\theta_0^2 = 10$		66	74	144	839	1	1	0	0
D: $p = 5$, $N = 60$										
$\hat{\boldsymbol{w}}_C(c^{**})$			125	144	301	1862				
$\hat{\boldsymbol{w}}_{SD}(c^{**})$	$a^2 = 0$		38	114	302	1867	69	20	0	0
	$a^2 = 0.1$		94	56	300	1868	25	61	0	0
	$a^2 = 1$		122	134	212	1869	3	7	29	0
	$a^2 = 10$		125	143	298	1771	0	1	1	5
$\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$	$\theta_0^2 = 0$		12	118	311	1873	91	18	-3	-1
	$\theta_0^2 = 0.1$		39	95	311	1873	69	34	-3	-1
	$\theta_0^2 = 1$		107	126	295	1873	15	12	2	-1
	$\theta_0^2 = 10$		123	142	299	1863	2	1	1	0
E: $p = 25$, $N = 60$										
$\hat{\boldsymbol{w}}_C(c^{**})$			383	458	1133	7884				
$\hat{\boldsymbol{w}}_{SD}(c^{**})$	$a^2 = 0$		34	186	1069	7911	91	59	6	0
	$a^2 = 0.1$		105	109	1012	7913	73	76	11	0
	$a^2 = 1$		290	295	784	7914	24	36	31	0
	$a^2 = 10$		371	439	1090	7535	3	4	4	4
$\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$	$\theta_0^2 = 0$		2	167	1186	7963	100	64	-5	-1
	$\theta_0^2 = 0.1$		15	151	1172	7963	96	67	-3	-1
	$\theta_0^2 = 1$		191	274	1021	7963	50	40	10	-1
	$\theta_0^2 = 10$		353	428	1104	7868	8	7	3	0

risk values of the three estimators for some paired values of p and N. In order to save space, we present only the risks of $\hat{\boldsymbol{w}}_C(c^{**})$, $\hat{\boldsymbol{w}}_{SD}(c^{**})$, and $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$. Panels A, B, C, D, and E in Table 2.2 contain results for $(p, N) = (10, 30)$, $(10, 120)$, $(10, 240)$, $(5, 60)$, and $(25, 60)$, respectively. The corresponding results for $(p, N) = (10, 60)$ given in Panel A of Table 2.1 are boldfaced for comparison.

From Panel A in Table 2.1 and Panels A, B, and C in Table 2.2, we see that the risks of almost all estimators decrease when N increases. However, when $a^2 = \theta^2$, the improvements of $\hat{\boldsymbol{w}}_{SD}(c^{**})$ increase as N increases. On the other hand, the improvements decrease generally when a^2 gets far away from θ^2. Although we set $c = c^{**}$, the risk values of $\hat{\boldsymbol{w}}_{SD}(c^{**})$ are smaller than those of $\hat{\boldsymbol{w}}_C(c^{**})$ for $N = 120$ and 240, even when $\theta^2 = 10$.

From the results given in Panel A of Table 2.1 and Panels D and E of Table 2.2, we see that the improvements of $\hat{\boldsymbol{w}}_{SD}(c^{**})$ and $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ increase when p increases. In the case of $p = 25$, $\hat{\boldsymbol{w}}_{SD}(c^{**})$ has a smaller risk than $\hat{\boldsymbol{w}}_C(c^{**})$ even though we set $c = c^{**}$. Thus, we find that $\hat{\boldsymbol{w}}_{SD}(c^{**})$ is more effective when p is large and when a^2 is close to θ^2. Similarly to the results for $(p, N) = (10, 60)$, we find that $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ has a smaller risk than $\hat{\boldsymbol{w}}_{SD}(c^{**})$ with $\boldsymbol{\mu}_0 = \mathbf{0}$ when $\theta_0^2 = \theta^2$. The improvements of $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ increase when the number p increases, similarly to $\hat{\boldsymbol{w}}_{SD}(c^{**})$. Thus, we also confirm the effectiveness of the estimators $\hat{\boldsymbol{w}}_{Sharpe}$.

2.5.2. Comparison based on actual asset returns data

In Section 2.5.1, we examined the risk behaviors of Stein-type estimators under the normality and independence assumptions. However, actual asset returns deviate from normality and are heterogeneous to some extent. Here, we apply the Stein-type estimators to two actual asset returns data sets and investigate their out-of-sample performance. One data set used in this study consists of monthly returns from July 1985 to February 2008 of 33 stock price indices by industry, which are constructed from stocks listed in the first section of the Tokyo

Stock Exchange. The data is contained in the Nikkei NEEDS database. The other consists of countries' stock market monthly value-weighted dollar returns in French's Data Library, which is available on the web at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french. We use the Country Portfolios included in International Research Returns data from January 1975 through December 2007. We adopt uncollateralized overnight call rate as a risk free rate for the 33 stock price indices, and 1-month Eurodollar deposit rate for the countries' stock returns, which are contained in the Nikkei NEEDS database. Table 2.3 presents summary statistics for the two data sets.

We compared the following three estimators: $\hat{\boldsymbol{w}}_C(c^{**})$, $\hat{\boldsymbol{w}}_{SD}(c^{**})$, and $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$. The procedure to evaluate the effectiveness of the estimators was as follows. Firstly, we estimated the mean-variance optimal portfolio weights $\hat{\boldsymbol{w}}_{t-1}$ using excess returns $\boldsymbol{x}_i$ from $i = t - N$ to $t - 1$ periods for each $t = 1, \ldots, T$. Next, we calculated *ex post* excess returns $y_t = \hat{\boldsymbol{w}}_{t-1}' \boldsymbol{x}_t$ for $t = 1, \ldots, T$. Finally, we computed out-of-sample *ex post* averages $\bar{y} = (1/T) \sum_{t=1}^{T} y_t$, the standard deviation $[\hat{V}(y)]^{1/2} = [1/(T-1) \sum_{t=1}^{T} (y_t - \bar{y})^2]^{1/2}$, and the utility $\hat{u} = \bar{y} - (\tau/2)\hat{V}(y)$, where we set $\tau = 3$. We used $\hat{u}$ to measure the effectiveness of the estimators. We set $N = 60$, 120, and 240.

Here, we present the results not only for the case with no constraints but also for the case with the linear constraint $\mathbf{1}'\boldsymbol{w} = 1$ on portfolio weights. For the latter case, setting $\hat{\zeta}^2 = \bar{\boldsymbol{x}}' F_1(S, \mathbf{1}) \bar{\boldsymbol{x}}$, we compared the following three estimators: $\hat{\boldsymbol{w}}_{C,\mathbf{1}}(c) = c\tau^{-1} F_1(\mathbf{1}, S) \bar{\boldsymbol{x}} + F_2(\mathbf{1}, S)$, $\hat{\boldsymbol{w}}_{SD,\mathbf{1}}(c) = c\tau^{-1} F_1(\mathbf{1}, S)[1 - (p-3)(N-p-1)^{-1}/\hat{\zeta}^2]^{+} \bar{\boldsymbol{x}} + F_2(\mathbf{1}, S)$, and $\hat{\boldsymbol{w}}_{Sharpe,\mathbf{1}}(c) = c\tau^{-1} F_1(\mathbf{1}, S)[1 - r_{sr}(\hat{\zeta}^2)/\hat{\zeta}^2]^{+} \bar{\boldsymbol{x}} + F_2(\mathbf{1}, S)$. When there is a linear constraint $\mathbf{1}'\boldsymbol{w} = 1$, as described in Section 2.3, Stein-type estimators that shrink toward a common value reduce to the estimator that shrinks toward the origin. Therefore, we present the results only for $\hat{\boldsymbol{w}}_{SD,\mathbf{1}}(c)$, which shrinks toward the origin. For the case with the linear constraint, we set $c = c_1^{\dagger\dagger} \equiv (N-p)(N-p-3) N^{-1} (N-2)^{-1}$. For comparison, we also present the results for the

Table 2.3: Summary statistics of asset's excess returns

	avg.(%)	SD(%)	avg./SD	$(\text{avg./SD})^2$
A. 33 Industry Indices				
Fishery, Agriculture & Forestry	−0.16	6.07	−0.027	0.0007
Mining	0.03	7.42	0.004	0.0000
Construction	−0.08	6.10	−0.013	0.0002
Foods	0.12	4.46	0.027	0.0007
Textiles & Apparels	−0.17	5.22	−0.032	0.0011
Pulp & Paper	−0.15	5.95	−0.025	0.0006
Chemicals	0.19	6.65	0.028	0.0008
Pharmaceutical	0.18	4.32	0.042	0.0018
Oil & Coal Products	0.04	6.69	0.005	0.0000
Rubber Products	0.36	6.14	0.059	0.0035
Glass & Ceramics Products	0.10	5.81	0.018	0.0003
Iron & Steel	0.38	7.02	0.054	0.0029
Nonferrous Metals	0.04	6.58	0.006	0.0000
Metal Products	0.09	5.55	0.017	0.0003
Machinery	0.20	5.56	0.037	0.0014
Electric Appliances	0.12	5.63	0.021	0.0004
Transportation Equipments	0.41	4.95	0.082	0.0068
Precision Instruments	0.30	5.46	0.056	0.0031
Other Products	0.32	5.00	0.063	0.0040
Electric Power & Gas	0.14	5.38	0.027	0.0007
Land Transportation	0.23	5.47	0.042	0.0018
Marine Transportation	0.55	7.51	0.074	0.0055
Air Transportation	−0.18	6.74	−0.027	0.0007
Warehousing & Harbor Transportation Services	0.16	6.25	0.026	0.0007
Information & Communication	−0.19	7.32	−0.026	0.0007
Wholesale Trade	0.34	7.05	0.048	0.0023
Retail Trade	0.10	5.82	0.018	0.0003
Banks	−0.18	6.99	−0.026	0.0007
Securities & Commodity Futures	0.20	9.42	0.021	0.0005
Insurance	0.22	6.24	0.035	0.0012
Other Financing Business	−0.01	6.59	−0.002	0.0000
Real Estate	0.36	7.42	0.048	0.0023
Services	0.09	6.23	0.015	0.0002
B. French's Country data				
Australia	0.76	6.44	0.117	0.0138
Belgium	0.83	5.37	0.156	0.0242
Canada	0.57	5.38	0.106	0.0113
France	0.82	6.30	0.131	0.0171
Germany	0.63	5.87	0.108	0.0116
Hong Kong	1.06	8.65	0.122	0.0150
Italy	0.74	7.24	0.102	0.0105
Japan	0.38	6.40	0.060	0.0036
Netherlands	0.81	5.05	0.160	0.0257
Norway	0.86	7.29	0.119	0.0141
Singapore	0.70	7.26	0.096	0.0092
Spain	0.73	6.43	0.114	0.0130
Sweden	0.91	6.71	0.136	0.0184
Switzerland	0.60	4.97	0.121	0.0146
United Kingdom	0.78	5.23	0.150	0.0224

estimator $\hat{\boldsymbol{w}}_{C,\mathbf{1}}(c)$ with $c = \tilde{c}_1^{\dagger\dagger} \equiv (N-p)(N-p-1)N^{-1}(N-2)^{-1}$.

Table 2.4 presents average and standard deviation of *ex post* returns, $\hat{u}$, and $\Delta\hat{u}$, which is the difference between the value of $\hat{u}$ of each estimator and that of $\hat{\boldsymbol{w}}_C(c^{**})$ for the case with no constraints or that of $\hat{\boldsymbol{w}}_{C,\mathbf{1}}(c_1^{\dagger\dagger})$ for the case with the linear constraint $\mathbf{1}'\boldsymbol{w} = 1$. Panels A and B of Table 2.4 present the results for 33 Industry indices and French's Country data respectively. We see that in most cases Stein-type estimators have smaller risks than $\hat{\boldsymbol{w}}_C(c^{**})$ or $\hat{\boldsymbol{w}}_{C,\mathbf{1}}(c_1^{\dagger\dagger})$. However, for some choices of a, $\hat{\boldsymbol{w}}_{SD}(c^{**})$ has a slightly larger risk than $\hat{\boldsymbol{w}}_C(c^{**})$. From Table 2.3, it is reasonable for us to judge that the value of true θ^2 for true optimal portfolio weights is less than 0.01 for the 33 Industry Indices portfolio, and is between 0.01 and 0.1 for French's Country portfolio. We see that the improvements of $\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$ and $\hat{\boldsymbol{w}}_{Sharpe,\mathbf{1}}(c_1^{\dagger\dagger})$ with $\theta_0^2 = 0$, 0.01, 0.1 are large. Thus, we have found that the estimator using a prior information concerning the Sharpe ratio is effective for these data sets. However, we should mention that such a conclusion does not necessarily apply to the other data sets. We need more thorough investigation based on various actual data sets to reach a definite conclusion.

Table 2.4: Comparison of estimators based on actual asset returns data

	$N=60$				$N=120$				$N=240$			
	avg. (%)	SD (%)	$\hat{u}$	$\Delta\hat{u}$	avg. (%)	SD (%)	$\hat{u}$	$\Delta\hat{u}$	avg. (%)	SD (%)	$\hat{u}$	$\Delta\hat{u}$
A. 33 Industry Indices												
A.1. no constraints												
$\hat{\boldsymbol{w}}_C(c^{**})$	−0.4	22.9	−820		0.4	17.5	−417		1.3	15.1	−207	
$\hat{\boldsymbol{w}}_C(\tilde{c}^{**})$	−0.3	24.9	−958	−138	0.4	17.9	−438	−21	1.4	15.3	−213	−5
$\hat{\boldsymbol{w}}_{SD}(c^{**})$												
$a^2=0$	0.8	10.8	−92	728	−0.8	4.2	−107	310	−0.3	6.4	−96	112
$a^2=0.01$	0.4	17.4	−415	405	0.6	13.3	−203	215	2.5	11.4	52	260
$a^2=0.1$	−0.7	22.4	−818	2	0.5	16.3	−349	68	1.9	14.4	−124	83
$a^2=1$	−0.5	22.9	−836	−15	0.4	17.2	−398	19	1.5	14.9	−183	25
$a^2=10$	−0.4	22.9	−826	−6	0.4	17.4	−412	5	1.4	15.1	−200	8
$\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$												
$\theta_0^2=0$	0.3	5.6	−21	799	0.0	0.0	0	417	0.0	0.0	0	207
$\theta_0^2=0.01$	0.2	5.6	−23	797	0.0	0.6	1	418	0.1	1.1	8	215
$\theta_0^2=0.1$	0.1	6.3	−52	768	0.1	4.8	−23	394	0.6	6.5	−5	202
$\theta_0^2=1$	−0.2	14.9	−357	463	0.3	13.8	−253	164	1.2	13.3	−147	60
$\theta_0^2=10$	−0.3	21.7	−741	79	0.4	17.0	−395	22	1.3	14.9	−200	7
A.2. $\mathbf{1}'\boldsymbol{w}=1$												
$\hat{\boldsymbol{w}}_{C,\mathbf{1}}(c_1^{\dagger\dagger})$	−0.7	22.4	−829		0.2	17.2	−419		2.2	14.6	−103	
$\hat{\boldsymbol{w}}_{C,\mathbf{1}}(\tilde{c}_1^{\dagger\dagger})$	−0.6	24.3	−952	−123	0.2	17.6	−439	−20	2.2	14.8	−108	−5
$\hat{\boldsymbol{w}}_{SD,\mathbf{1}}(c_1^{\dagger\dagger})$	0.4	12.2	−183	646	−0.7	6.3	−131	288	0.6	5.7	9	112
$\hat{\boldsymbol{w}}_{Sharpe,\mathbf{1}}(c_1^{\dagger\dagger})$												
$\theta_0^2=0$	−0.1	7.0	−81	748	0.1	4.1	−16	403	0.8	3.2	69	172
$\theta_0^2=0.01$	−0.1	7.0	−83	746	0.1	4.1	−15	404	0.9	3.3	78	181
$\theta_0^2=0.1$	−0.2	7.4	−107	722	0.1	6.0	−40	379	1.4	6.7	75	178
$\theta_0^2=1$	−0.6	15.0	−396	433	0.2	13.7	−263	157	2.0	13.0	−49	54
$\theta_0^2=10$	−0.7	21.3	−756	73	0.2	16.7	−398	21	2.2	14.4	−96	6
B. French's Country data												
B.1. no constraints												
$\hat{\boldsymbol{w}}_C(c^{**})$	1.0	16.9	−326		0.5	11.0	−134		0.3	7.6	−61	
$\hat{\boldsymbol{w}}_C(\tilde{c}^{**})$	1.1	17.7	−363	−38	0.5	11.2	−140	−6	0.3	7.7	−62	−1
$\hat{\boldsymbol{w}}_{SD}(c^{**})$												
$a^2=0$	0.3	8.0	−67	259	−0.3	3.5	−45	89	−0.1	1.2	−14	47
$a^2=0.01$	1.8	16.6	−231	94	1.1	12.0	−111	23	0.8	9.0	−47	14
$a^2=0.1$	1.4	17.4	−314	12	0.7	11.6	−133	1	0.4	8.1	−59	2
$a^2=1$	1.2	17.1	−323	2	0.5	11.2	−134	0	0.3	7.8	−61	0
$a^2=10$	1.1	17.0	−325	1	0.5	11.1	−134	0	0.3	7.7	−61	0
$\hat{\boldsymbol{w}}_{Sharpe}(c^{**})$												
$\theta_0^2=0$	0.1	2.9	1	326	0.0	0.6	−5	129	0.0	0.0	0	61
$\theta_0^2=0.01$	0.2	3.0	2	327	0.0	1.0	1	135	0.0	1.1	2	63
$\theta_0^2=0.1$	0.3	5.3	−11	314	0.2	5.1	−17	117	0.2	4.8	−18	43
$\theta_0^2=1$	0.8	13.7	−198	127	0.4	9.8	−103	31	0.2	7.2	−53	8
$\theta_0^2=10$	1.0	16.5	−309	17	0.5	10.9	−130	4	0.3	7.6	−60	1
B.2. $\mathbf{1}'\boldsymbol{w}=1$												
$\hat{\boldsymbol{w}}_{C,\mathbf{1}}(c_1^{\dagger\dagger})$	1.5	16.4	−255		0.7	11.1	−118		0.4	7.5	−49	
$\hat{\boldsymbol{w}}_{C,\mathbf{1}}(\tilde{c}_1^{\dagger\dagger})$	1.5	17.1	−288	−32	0.7	11.3	−124	−6	0.4	7.6	−50	−1
$\hat{\boldsymbol{w}}_{SD,\mathbf{1}}(c_1^{\dagger\dagger})$	0.8	8.7	−35	221	0.3	5.6	−13	104	0.5	4.2	24	73
$\hat{\boldsymbol{w}}_{Sharpe,\mathbf{1}}(c_1^{\dagger\dagger})$												
$\theta_0^2=0$	0.7	4.8	33	288	0.6	4.3	30	148	0.6	4.2	37	85
$\theta_0^2=0.01$	0.7	4.9	35	290	0.6	4.4	30	148	0.6	4.4	31	80
$\theta_0^2=0.1$	0.9	6.6	24	279	0.6	6.5	−1	117	0.5	5.8	−5	44
$\theta_0^2=1$	1.3	13.6	−144	112	0.7	10.2	−88	29	0.4	7.3	−41	7
$\theta_0^2=10$	1.5	16.1	−241	15	0.7	11.0	−114	3	0.4	7.5	−48	1

Chapter 3

Shrinkage toward the Grand Mean or a Linear Subspace

3.1. Introduction

Jorion (1985, 1986, 1991) proposed the adoption of an estimator of the mean vector that shrinks the sample mean toward the grand mean based on the evidence of mean reversion in financial markets. His estimator is referred to as the *Bayes-Stein* estimator in a number of previous studies in finance (cf., for example, Grauer and Hakansson 1995, 2001, Michaud 1998, Kashima 2001, 2005, Ledoit, O. and Wolf, M. 2003, Okhrin and Schmid 2007, Garlappi et al. 2007, Kan and Zhou 2007, and Brandt 2009). Recently, Kan and Zhou (2007) developed a new estimator by combining a sample tangency portfolio with a sample global minimum variance portfolio. Their estimator also applies the shrinkage toward the grand mean. However, the effectiveness of these estimators has not been investigated analytically. Shrinkage toward the grand mean is a special case of shrinkage toward a linear subspace. In this chapter, we present dominance results for the estimators of the mean-variance optimal portfolio weights, which apply the shrinkage toward a linear subspace.

Lindley (1962) first suggested the idea of modifying the James-Stein estimator and shrinking $\bar{\boldsymbol{x}}$ toward the grand mean instead of a fixed point. More generally, we consider the James-Stein type estimator, which shrinks toward a linear subspace (cf., Lehmann and Casella 1998, Example 6.2). Suppose that a prior information suggests that $\boldsymbol{\mu}$ is close to the subspace $\mathcal{N}(Z) = \{\boldsymbol{\mu} : Z\boldsymbol{\mu} = \mathbf{0}\}$, where Z is an $\ell \times p$ ($\ell \leq p$) matrix of rank $Z = \ell$. Since the maximum likelihood estimator of $\boldsymbol{\mu} \in \mathcal{N}(Z)$ is given by $[I - \Sigma Z'(Z\Sigma Z')^{-1}Z]\bar{\boldsymbol{x}}$ when Σ is known, setting $Y^* = \Sigma Z'(Z\Sigma Z')^{-1}Z$, the Stein-type estimators of $\boldsymbol{\mu}$, which shrink the sample mean toward $(I - Y^*)\bar{\boldsymbol{x}}$, are given as $\hat{\boldsymbol{\mu}} = [(I - Y^*) + (1 - d/\zeta^2)^+ Y^*]\bar{\boldsymbol{x}}$, where $\zeta^2 = \bar{\boldsymbol{x}}'(Y^*)'\Sigma^{-1}Y^*\bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}'Z'(Z\Sigma Z')^{-1}Z\bar{\boldsymbol{x}}$, d is a positive constant and $a^+ = \max(0, a)$ (cf., Casella and Hwang 1987, Example 2). More generally, we consider a class of estimators by replacing d with a function $r(\cdot)$ of $\bar{\boldsymbol{x}}'(Y^*)'\Sigma^{-1}Y^*\bar{\boldsymbol{x}}$ (cf., Baranchik 1970). Since Σ is unknown, we replace Σ by its sample estimator S and write $Y = SZ'(ZSZ')^{-1}Z$ and $\hat{\zeta}^2 = \bar{\boldsymbol{x}}'Y'S^{-1}Y\bar{\boldsymbol{x}}$. We thus obtain an estimator of $\boldsymbol{\mu}$ as

$$\hat{\boldsymbol{\mu}} = \left[(I - Y) + \left(1 - \frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\right)^+ Y\right]\bar{\boldsymbol{x}}.$$

Using this $\hat{\boldsymbol{\mu}}$ and $c^{-1}S$ as estimators of $\boldsymbol{\mu}$ and Σ, respectively, we obtain an estimator for the mean-variance optimal portfolio weights $\tau^{-1}\Sigma^{-1}\boldsymbol{\mu}$:

$$\frac{c_s}{\tau}S^{-1}\left[(I - Y) + \left(1 - \frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\right)Y\right]\bar{\boldsymbol{x}}. \tag{3.1}$$

In this chapter, first, we give a dominance result for the estimator given by Equation (3.1).

Next, we assume that a prior information suggests that $\boldsymbol{\mu} = B\boldsymbol{\beta}$ for some $\boldsymbol{\beta}$, that is, $\boldsymbol{\mu} \in \mathcal{R}(B) = \{\boldsymbol{\mu}; \boldsymbol{\mu} = B\boldsymbol{\beta}\}$, where B is a $p \times k$ non-random matrix of rank $B = k$ and $\boldsymbol{\beta}$ is a $k \times 1$ vector. The generalized least squares estimator of $B\boldsymbol{\beta}$ is given as $B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\bar{\boldsymbol{x}}$ when Σ is known. By replacing Σ by its sample estimator S, we obtain $R\bar{\boldsymbol{x}}$, where $R = B(B'S^{-1}B)^{-1}B'S^{-1}$. We may construct an estimator of $\boldsymbol{\mu}$ that shrinks the sample mean toward

$R\bar{\boldsymbol{x}}$:

$$\frac{c_s}{\tau}S^{-1}\left[R+\left(1-\frac{r(\hat{\xi}^2)}{\hat{\xi}^2}\right)(I-R)\right]\bar{\boldsymbol{x}}, \tag{3.2}$$

where $\hat{\xi}^2=\bar{\boldsymbol{x}}'(I-R)'S^{-1}(I-R)\bar{\boldsymbol{x}}$. In this chapter, we also give a dominance result for the estimator given by Equation (3.2). When we set $B=\mathbf{1}$, the estimator reduces to the one that shrinks toward the grand mean, and it is related to some estimators previously proposed in finance.

The remainder of this chapter is organized as follows. Section 3.2 gives the dominance results of a class of Stein-type estimators for the mean-variance optimal portfolio weights that shrinks toward a linear subspace, when we have no constraints on portfolio weights. In this section, we also show that some estimators provided in previous studies belong to our class. Section 3.3 gives the dominance results when we have linear constraints on portfolio weights. Section 3.4 gives the proofs of the theorems stated in Sections 3.2 and 3.3.

3.2. Case in which there are no constraints on portfolio weights

We have obtained the following theorem under the loss function given in Equation (1.2).

Theorem 3.1. *Let Z be a non-random $\ell\times p$ matrix of rank $Z=\ell$ and $Y=SZ'(ZSZ')^{-1}Z$. Let $2<\ell\leq p$ and $N>p+4$. If $c_c\geq c_s\geq(N-\ell-1)(N-p-2)N^{-1}(N-2)^{-1}$, $r(\cdot)$ is non-decreasing, and $0\leq r(\cdot)\leq 2(\ell-2)(N-p-2)^{-1}$, then $\hat{\boldsymbol{w}}_C(c_c;\bar{\boldsymbol{x}},S)=c_c\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$ is dominated by the following Stein-type estimator for the mean-variance optimal portfolio weights:*

$$\hat{\boldsymbol{w}}_{SZ1}(c_s;\bar{\boldsymbol{x}},S)=\frac{c_s}{\tau}S^{-1}\left[(I-Y)+\left(1-\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\right)Y\right]\bar{\boldsymbol{x}},$$

where $\hat{\zeta}^2=\bar{\boldsymbol{x}}'Y'S^{-1}Y\bar{\boldsymbol{x}}=\bar{\boldsymbol{x}}'Z'(ZSZ')^{-1}Z\bar{\boldsymbol{x}}$.

The proof is given in Section 3.4.1.

From Theorem 3.1, we see that if $\hat{\boldsymbol{w}}_{SZ1}(c_s;\bar{\boldsymbol{x}},S)$ dominates $\hat{\boldsymbol{w}}_C(c_c;\bar{\boldsymbol{x}},S)$, its positive-part Stein-type estimator $\hat{\boldsymbol{w}}^+_{SZ1}(c_s;\bar{\boldsymbol{x}},S) = c_s\tau^{-1}S^{-1}\{(I-Y)+[1-r(\hat{\zeta}^2)/\hat{\zeta}^2]^+Y\}\bar{\boldsymbol{x}}$ also dominates $\hat{\boldsymbol{w}}_C(c_c;\bar{\boldsymbol{x}},S)$. However, the question of whether the positive-part Stein-type estimator $\hat{\boldsymbol{w}}^+_{SZ1}(c_s;\bar{\boldsymbol{x}},S)$ improves upon $\hat{\boldsymbol{w}}_{SZ1}(c_s;\bar{\boldsymbol{x}},S)$ has not been settled. The same remark applies to the theorems in the following.

For the case in which a prior information suggests that $\boldsymbol{\mu}$ is close to the subset $\{\boldsymbol{\mu}: Z\boldsymbol{\mu}=\boldsymbol{a}\}$ instead of $\mathcal{N}(Z)$, where $\boldsymbol{a}$ is a fixed non-zero vector, it is technically difficult to show the general dominance of the estimators of the form similar to $\hat{\boldsymbol{w}}_{SZ1}$. However, we can show the following theorem.

Theorem 3.2. *Let Z be a non-random $\ell\times p$ matrix of rank $Z=\ell$, $\tilde{\boldsymbol{x}}=\bar{\boldsymbol{x}}-SZ'(ZSZ')^{-1}(Z\bar{\boldsymbol{x}}-\boldsymbol{a})$, and $\hat{\zeta}^2=(\bar{\boldsymbol{x}}-\tilde{\boldsymbol{x}})'S^{-1}(\bar{\boldsymbol{x}}-\tilde{\boldsymbol{x}})$. Let $2<\ell\le p$ and $N>p+4$. If $c_c\ge c_s=(N-\ell-1)(N-p-2)N^{-1}(N-2)^{-1}$ and d is a constant that satisfies $0<d<2(\ell-2)(N-p-2)^{-1}$, then $\hat{\boldsymbol{w}}_C(c_c;\bar{\boldsymbol{x}},S)=c\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$ is dominated by the following Stein-type estimator for the mean-variance optimal portfolio weights:*

$$\hat{\boldsymbol{w}}_{SZ2}(c_s;\bar{\boldsymbol{x}},S)=\frac{c_s}{\tau}S^{-1}\left[\tilde{\boldsymbol{x}}+\left(1-\frac{d}{\hat{\zeta}^2}\right)(\bar{\boldsymbol{x}}-\tilde{\boldsymbol{x}})\right].$$

The proof is given in Section 3.4.1.

Example 3.1. *Black-Litterman Approach.* Black and Litterman (1992) have proposed the combination of the equilibrium risk premiums and an investor's views on the expected excess returns for risky assets. In their approach, the equilibrium risk premiums $\boldsymbol{x}_e$ are assumed to follow $N(\boldsymbol{\mu},\lambda\Sigma)$, where λ is a positive constant, and the investor's views are represented as $Z\boldsymbol{\mu}=\boldsymbol{a}+\boldsymbol{\varepsilon}$, where Z is an $\ell\times p$ ($\ell\le p$) matrix of rank $Z=\ell$, $\boldsymbol{a}$ is an $\ell\times 1$ vector, and $\boldsymbol{\varepsilon}$ is an $\ell\times 1$ vector of error terms distributed as $N(\boldsymbol{0},\Omega)$. Here, Ω represents the uncertainty of the investor. Based on the mixed estimation (cf., Theil 1971), an estimator of $\boldsymbol{\mu}$ is given

as $\hat{\boldsymbol{\mu}}_{BL} = \boldsymbol{x}_e - \lambda\Sigma Z'(\Omega + \lambda Z\Sigma Z')^{-1}(Z\boldsymbol{x}_e - \boldsymbol{a})$. Furthermore, since it is difficult to determine all elements of Ω according to the views of the investor, Ω is replaced by $\gamma Z\Sigma Z'$ or by the diagonal matrix with the same diagonal elements as $\gamma Z\Sigma Z'$, where γ is a positive constant. By replacing $\boldsymbol{x}_e$, Σ, and Ω by $\bar{\boldsymbol{x}}$, S, and $\gamma ZSZ'$, respectively, the estimator $\hat{\boldsymbol{\mu}}_{BL}$ reduces to $\bar{\boldsymbol{x}} - \gamma^* SZ'(ZSZ')^{-1}(Z\bar{\boldsymbol{x}} - \boldsymbol{a}) = \tilde{\boldsymbol{x}} + (1-\gamma^*)(\bar{\boldsymbol{x}} - \tilde{\boldsymbol{x}})$, where $\gamma^* = (\gamma\lambda^{-1} + 1)^{-1}$ and $\tilde{\boldsymbol{x}} = \bar{\boldsymbol{x}} - SZ'(ZSZ')^{-1}(Z\bar{\boldsymbol{x}} - \boldsymbol{a})$.

If we replace γ^* by $\min(1, d/[(Z\bar{\boldsymbol{x}} - \boldsymbol{a})'(ZSZ')^{-1}(Z\bar{\boldsymbol{x}} - \boldsymbol{a})])$, then we have the estimator given in Theorem 3.2 for the mean-variance optimal portfolio weights. Usually, in the practical application of the Black-Litterman approach, the investor must choose a pertinent value of γ^* reflecting the confidence in his/her views. If the investor chooses a value of d, instead of λ and γ, so as to satisfy the conditions given in Theorem 3.2, the estimator for the portfolio weights has a smaller risk than the estimator $\hat{\boldsymbol{w}}_C$.

Next, we give a dominance result for the estimator given by Equation (3.2). If $\mathcal{R}(B) = \mathcal{N}(Z)$, then $S^{-1/2}RS^{1/2}$ and $I - S^{-1/2}YS^{1/2}$ are the projection matrix onto $\mathcal{R}(S^{-1/2}B) = \mathcal{N}(ZS^{1/2})$. Thus, we can show that $R = I - Y$ if $\mathcal{R}(B) = \mathcal{N}(Z)$, and we have essentially the same estimator as that given in Theorem 3.1. Setting $Y = I - R$, we may state Theorem 3.1 as follows.

Theorem 3.1'. *Let B be a $p \times k$ non-random matrix of rank $B = k$, and set $R = B(B'S^{-1}B)^{-1} \times B'S^{-1}$. Let $p > k+2$ and $N > p+4$. If $c_c \geq c_s \geq (N-p+k-1)(N-p-2)N^{-1}(N-2)^{-1}$, $r(\cdot)$ is non-decreasing, and $0 \leq r(\cdot) \leq 2(p-k-2)(N-p-2)^{-1}$, then $\hat{\boldsymbol{w}}_C(c_c; \bar{\boldsymbol{x}}, S) = c_c\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$ is dominated by the following Stein-type estimator for the mean-variance optimal portfolio weights:*

$$\hat{\boldsymbol{w}}_{SB}(c_s; \bar{\boldsymbol{x}}, S) = \frac{c_s}{\tau} S^{-1}\left[R + \left(1 - \frac{r(\hat{\xi}^2)}{\hat{\xi}^2}\right)(I - R)\right]\bar{\boldsymbol{x}},$$

where $\hat{\xi}^2 = \bar{\boldsymbol{x}}'(I-R)'S^{-1}(I-R)\bar{\boldsymbol{x}}$.

If we set $B = \mathbf{1}$, then the estimator $\hat{\boldsymbol{w}}_{SB}(c; \bar{\boldsymbol{x}}, S)$ reduces to

$$\hat{\boldsymbol{w}}_{Sg}(c; \bar{\boldsymbol{x}}, S) = \frac{c}{\tau} S^{-1} \left[\hat{\mu}_g \mathbf{1} + \left(1 - \frac{r(\hat{\varphi}^2)}{\hat{\varphi}^2} \right) (\bar{\boldsymbol{x}} - \hat{\mu}_g \mathbf{1}) \right]$$

where $\hat{\mu}_g = \mathbf{1}' S^{-1} \bar{\boldsymbol{x}} (\mathbf{1}' S^{-1} \mathbf{1})^{-1}$ and $\hat{\varphi}^2 = (\bar{\boldsymbol{x}} - \hat{\mu}_g \mathbf{1})' S^{-1} (\bar{\boldsymbol{x}} - \hat{\mu}_g \mathbf{1})$. $\hat{\boldsymbol{w}}_{Sg}$ is the estimator that shrinks the sample mean toward its grand mean. The estimators proposed by Jorion (1986) and Kan and Zhou (2007) belong to this class.

Example 3.2. *Jorion's Bayes-Stein Estimator.* The estimator proposed by Jorion (1986) is $\tau^{-1} \hat{\Sigma}^{-1} \hat{\boldsymbol{x}}_{BS}$, where $\hat{\Sigma}$ is an estimator of Σ and

$$\hat{\boldsymbol{x}}_{BS} = \hat{\mu}_g \mathbf{1} + (1 - \hat{w})(\bar{\boldsymbol{x}} - \hat{\mu}_g \mathbf{1}) \quad \text{with} \quad \hat{w} = \frac{(p+2)(N-1)N^{-1}(N-p-2)^{-1}}{\hat{\varphi}^2 + (p+2)(N-1)N^{-1}(N-p-2)^{-1}}.$$

Although he used an estimator other than S, setting $\hat{\Sigma} = c_s^{-1} S$, Jorion's estimator reduces to $\hat{\boldsymbol{w}}_{Sg}(c_s; \bar{\boldsymbol{x}}, S)$ with $r(\hat{\varphi}^2) = \alpha_1 \hat{\varphi}^2 / (\hat{\varphi}^2 + \alpha_2)$ and $\alpha_1 = \alpha_2 = (p+2)(N-1)N^{-1}(N-p-2)^{-1}$. Since $\hat{\boldsymbol{w}}_{Sg}(c_s; \bar{\boldsymbol{x}}, S)$ with $r(\hat{\varphi}^2) = \alpha_1 \hat{\varphi}^2 / (\hat{\varphi}^2 + \alpha_2)$ improves upon $\hat{\boldsymbol{w}}_C(c_c; \bar{\boldsymbol{x}}, S)$ when $0 < \alpha_1 \leq 2(p-3)(N-p-2)^{-1}$ and $\alpha_2 \geq 0$, from Theorem 3.1, we see that when $N(p-8) + p + 2 \geq 0$ and $c_c \geq c_s \geq (N-p)(N-p-2)N^{-1}(N-2)^{-1}$, Jorion's estimator improves upon the classical estimator.

Example 3.3. *Kan and Zhou's three-fund rule estimator.* One of the estimators provided by Kan and Zhou (2007) is

$$\hat{\boldsymbol{w}}_{KZ3} = \frac{1}{\tau} \frac{(N-p-1)(N-p-4)}{N(N-2)} \left(\frac{\tilde{\varphi}^2(\hat{\varphi}^2)}{\tilde{\varphi}^2(\hat{\varphi}^2) + p/N} S^{-1} \bar{\boldsymbol{x}} + \frac{p/N}{\tilde{\varphi}^2(\hat{\varphi}^2) + p/N} \hat{\mu}_g S^{-1} \mathbf{1} \right),$$

where

$$\tilde{\varphi}^2(\hat{\varphi}^2) = \frac{(N-p-1)\hat{\varphi}^2 - (p-1)}{N} + \frac{2(\hat{\varphi}^2)^{(p-1)/2}(1+\hat{\varphi}^2)^{-(N-2)/2}}{N B_{\hat{\varphi}^2/(1+\hat{\varphi}^2)}((p-1)/2, (N-p+1)/2)}.$$

The estimator $N\tilde{\varphi}^2$ of $N\varphi^2$ improves upon the unbiased estimator $(N-p-1)\hat{\varphi}^2 - (p-1)$ under squared-error loss (cf., Kubokawa et al. 1993). If we set $c = (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$,

$r(\hat{\varphi}^2) \equiv r_{KZ3}(\hat{\varphi}^2) = \alpha_1\hat{\varphi}^2[\tilde{\varphi}^2(\hat{\varphi}^2) + \alpha_2/N]^{-1}$, $\alpha_1 = p/N$, and $\alpha_2 = p$, then the estimator $\hat{\boldsymbol{w}}_{Sg}(c; \bar{\boldsymbol{x}}, S)$ reduces to the estimator $\hat{\boldsymbol{w}}_{KZ3}$. If $N > p+2$, $p > 3$, $c \geq (N-p)(N-p-2)N^{-1}(N-2)^{-1}$, $0 < \alpha_1 < 2(p-3)N(N-p-2)^{-1}$, and $\alpha_2 \geq p-1$, then $\hat{\boldsymbol{w}}_{Sg}(c; \bar{\boldsymbol{x}}, S)$ with $r = r_{KZ3}$ improves upon $\hat{\boldsymbol{w}}_C(c; \bar{\boldsymbol{x}}, S)$, because we can show that $r_{KZ3}(\cdot)$ is non-decreasing under the above conditions (cf., Example 2.2 in Chapter 2). For the specific choice of Kan and Zhou's estimator, $\alpha_1 = p/N$ and $\alpha_2 = p$ satisfy the conditions, but $c = (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$ does not.

The three-fund rule estimator is a weighted average of the sample tangency portfolio $S^{-1}\bar{\boldsymbol{x}} \times(\bar{\boldsymbol{x}}'S^{-1}\mathbf{1})^{-1}$ and the sample global minimum variance portfolio $S^{-1}\mathbf{1}(\mathbf{1}'S^{-1}\mathbf{1})^{-1}$ and it is shrinking the sample tangency portfolio toward the sample global minimum variance portfolio. Kan and Zhou (2007) have proposed the three-fund rule estimator as a candidate which improves the two-fund rule estimator (cf., Example 2.2 in Chapter 2). The two-fund rule estimator is a scalar multiple of the sample tangency portfolio that is on the *ex post* minimum-variance frontier. They have suggested to combine the sample tangency portfolio with another portfolio on the same *ex post* minimum-variance frontier in order to reduce the estimation errors. They have chosen the sample global minimum variance portfolio as the one. They have suggested that the risk of the three-fund rule estimator is smaller than that of the two-fund rule estimator. However, if the true optimal portfolio weights of risky assets are close to $\mathbf{0}$, the two-fund rule estimator will perform better than the three-fund rule estimator. It depends on the accuracy of the investor's subjective information which estimator is superior and is to be chosen.

3.3. Case in which there are linear constraints on portfolio weights

We now consider the case in which we have linear constraints $A'\boldsymbol{w} = \boldsymbol{b}$, where A is a $p \times q$ matrix of rank $A = q$. We derive the estimator shrinking toward a linear subspace under linear constraints. We define the orthogonal matrix Q such that

$$QAA'Q' = D, \quad D = \begin{pmatrix} D_1 & O \\ O & O \end{pmatrix}, \quad Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}, \tag{3.3}$$

where D is a diagonal matrix with eigenvalues of AA' on the diagonal, D_1 is a $q \times q$ diagonal matrix with positive diagonal elements, Q_1 is a $q \times p$ matrix, and Q_2 is a $(p-q) \times p$ matrix. Since $F_1(A, \Sigma) = Q_2'(Q_2 \Sigma Q_2')^{-1} Q_2$ (cf., Mori 2004, Lemma 3.1), $\boldsymbol{w}_A(\boldsymbol{\mu}, \Sigma)$ given in Equation (1.5) is written as

$$\boldsymbol{w}_A(\boldsymbol{\mu}, \Sigma) = \frac{1}{\tau} Q_2'(Q_2 \Sigma Q_2')^{-1} Q_2 \boldsymbol{\mu} + F_2(A, \Sigma)\boldsymbol{b}. \tag{3.4}$$

We assume that we have a prior information that $\boldsymbol{\mu}$ is close to $\mathcal{R}(B)$. We set $\tilde{\boldsymbol{\mu}} = Q_2\boldsymbol{\mu}$, $\tilde{\Sigma} = Q_2 \Sigma Q_2'$, and $\tilde{B} = Q_2 B$, and assume that the matrix $B'F_1(A, \Sigma)B = B'Q_2'(Q_2 \Sigma Q_2')^{-1} Q_2 B = \tilde{B}'\tilde{\Sigma}^{-1}\tilde{B}$ is non-singular. We will discuss the condition that guarantees this later. Then, we have the generalized least squares estimator of $\tilde{\boldsymbol{\mu}}$ based on $Q_2\bar{\boldsymbol{x}}$ as $\tilde{B}(\tilde{B}'\tilde{\Sigma}^{-1}\tilde{B})^{-1}\tilde{B}'\tilde{\Sigma}^{-1}Q_2\bar{\boldsymbol{x}}$ when Σ is known. Thus, setting $R^* = B\left[B'F_1(A, \Sigma)B\right]^{-1} B'F_1(A, \Sigma)$, we may construct an estimator of $\tilde{\boldsymbol{\mu}}$ that shrinks the sample estimator $Q_2\bar{\boldsymbol{x}}$ toward $\tilde{B}(\tilde{B}'\tilde{\Sigma}^{-1}\tilde{B})^{-1}\tilde{B}'\tilde{\Sigma}^{-1}Q_2\bar{\boldsymbol{x}}$ as

$$\begin{aligned} &\tilde{B}\left(\tilde{B}'\tilde{\Sigma}^{-1}\tilde{B}\right)^{-1}\tilde{B}'\tilde{\Sigma}^{-1}Q_2\bar{\boldsymbol{x}} + \left(1 - \frac{r(\tilde{\xi}^2_{F_1})}{\tilde{\xi}^2_{F_1}}\right)\left[I - \tilde{B}\left(\tilde{B}'\tilde{\Sigma}^{-1}\tilde{B}\right)^{-1}\tilde{B}'\tilde{\Sigma}^{-1}\right]Q_2\bar{\boldsymbol{x}}, \\ =&Q_2\left[R^*\bar{\boldsymbol{x}} + \left(1 - \frac{r(\tilde{\xi}^2_{F_1})}{\tilde{\xi}^2_{F_1}}\right)(I - R^*)\bar{\boldsymbol{x}}\right], \end{aligned} \tag{3.5}$$

where

$$\begin{aligned} \tilde{\xi}^2_{F_1} &= \bar{\boldsymbol{x}}'Q_2'[I - \tilde{B}(\tilde{B}'\tilde{\Sigma}^{-1}\tilde{B})^{-1}\tilde{B}'\tilde{\Sigma}^{-1}]'\tilde{\Sigma}^{-1}[I - \tilde{B}(\tilde{B}'\tilde{\Sigma}^{-1}\tilde{B})^{-1}\tilde{B}'\tilde{\Sigma}^{-1}]Q_2\bar{\boldsymbol{x}} \\ &= \bar{\boldsymbol{x}}'(I - R^*)'F_1(A, \Sigma)(I - R^*)\bar{\boldsymbol{x}}. \end{aligned}$$

By substituting Equation (3.5) into $Q_2\boldsymbol{\mu}$ in Equation (3.4), we have the following estimator for $\boldsymbol{w}$ when Σ is known:

$$\hat{\boldsymbol{w}} = \frac{1}{\tau}F_1(A,\Sigma)\left[R^*\bar{\boldsymbol{x}} + \left(1 - \frac{r(\tilde{\xi}^2_{F_1})}{\tilde{\xi}^2_{F_1}}\right)(I - R^*)\bar{\boldsymbol{x}}\right] + F_2(A,\Sigma)\boldsymbol{b}. \tag{3.6}$$

By replacing Σ with $c_s^{-1}S$ in Equation (3.6), we obtain the estimator given in Theorem 3.3. Note that the condition $\mathcal{R}(A) \cap \mathcal{R}(B) = \{\mathbf{0}\}$ is necessary to guarantee that $B'F_1(A,S)B$ is non-singular.

Theorem 3.3. *Let B be a $p \times k$ non-random matrix of rank $B = k$ and assume that $\mathcal{R}(A) \cap \mathcal{R}(B) = \{\mathbf{0}\}$. Set $R_{F_1} = B(B'F_1(A,S)B)^{-1}B'F_1(A,S)$. Let $p > q + k + 2$ and $N > \max(p + 1, p-q+4)$. If $c_c \geq c_s \geq (N-p+q+k-1)(N-p+q-2)N^{-1}(N-2)^{-1}$, $r(\cdot)$ is non-decreasing and $0 \leq r(\cdot) \leq 2(p-q-k-2)(N-p+q-2)^{-1}$, then $\hat{\boldsymbol{w}}_{C,A}(c_c;\bar{\boldsymbol{x}},S) = c_c\tau^{-1}F_1(A,S)\bar{\boldsymbol{x}} + F_2(A,S)\boldsymbol{b}$ is dominated by the following Stein-type estimator for the mean-variance optimal portfolio weights:*

$$\hat{\boldsymbol{w}}_{SB,A}(c_s;\bar{\boldsymbol{x}},S) = \frac{c_s}{\tau}F_1(A,S)\left[R_{F_1} + \left(1 - \frac{r(\hat{\xi}^2_{F_1})}{\hat{\xi}^2_{F_1}}\right)(I - R_{F_1})\right]\bar{\boldsymbol{x}} + F_2(A,S)\boldsymbol{b},$$

where $\hat{\xi}^2_{F_1} = \bar{\boldsymbol{x}}'(I - R_{F_1})'F_1(A,S)(I - R_{F_1})\bar{\boldsymbol{x}}$.

The proof is given in Section 3.4.2.

Here, we present some remarks on Theorem 3.3. Although we have assumed that $B'F_1(A,S)B$ is non-singular, we can apply Theorem 3.3 by replacing its inverse with its generalized inverse even when it is singular. If we set $K = [I - S^{-1/2}A(A'S^{-1}A)^{-1}A'S^{-1/2}]S^{-1/2}B$, then $B'F_1(A,S)B = K'K$. Thus, by setting $(K'K)^-$ to be a generalized inverse of $K'K$, we have $F_1(A,S)R_{F_1} = S^{-1/2}K(K'K)^-K'S^{-1/2}$, where $K(K'K)^-K'$ is the projection matrix onto $\mathcal{R}(K)$ and does not depend on the choice of $(K'K)^-$. Since $\hat{\boldsymbol{w}}_{SB,A}$ depends on R_{F_1} only through $F_1(A,S)R_{F_1}$, Theorem 3.3 holds even when $B'F_1(A,S)B$ is singular, if we set the conditions on c and $r(\cdot)$ appropriately. Since rank Q_2B = rank (B) − dim$[\mathcal{R}(A) \cap \mathcal{R}(B)]$, by setting $k' = k - \dim[\mathcal{R}(A) \cap \mathcal{R}(B)]$, it is sufficient to replace the conditions on c_s and $r(\cdot)$ as follows:

$c_s \geq (N-p+q+k'-1)(N-p+q-2)N^{-1}(N-2)^{-1}$ and $0 \leq r(\cdot) \leq 2(p-q-k'-2)(N-p+q-2)^{-1}$.

Note that if $\mathcal{R}(B) \subseteq \mathcal{R}(A)$, the estimator $\hat{\boldsymbol{w}}_{SB,A}(c_s; \bar{\boldsymbol{x}}, S)$ reduces to that shrinking toward the origin. Since $F_1(A,S)B = O$ if $\mathcal{R}(B) \subseteq \mathcal{R}(A)$, the estimator $\hat{\boldsymbol{w}}_{SB,A}(c_s; \bar{\boldsymbol{x}}, S)$ reduces to $c_s\tau^{-1}F_1(A,S)[1 - r(\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}})/(\bar{\boldsymbol{x}}'F_1(A,S)\bar{\boldsymbol{x}})]\bar{\boldsymbol{x}} + F_2(A,S)\boldsymbol{b}$.

When we have a constraint $\mathbf{1}'\boldsymbol{w} = 1$, the effectiveness of applying $\hat{\mu}_g\mathbf{1} + (1 - r(\hat{\varphi}^2)/\hat{\varphi}^2)(\bar{\boldsymbol{x}} - \hat{\mu}_g\mathbf{1})$ with $\hat{\varphi}^2 = (\bar{\boldsymbol{x}} - \hat{\mu}_g\mathbf{1})'S^{-1}(\bar{\boldsymbol{x}} - \hat{\mu}_g\mathbf{1})$ as an estimator of $\boldsymbol{\mu}$ has been shown by simulation and empirically (e.g., Jorion 1986). However, this can be shown analytically. Substituting the estimator into $\hat{\boldsymbol{w}}_{C,A}$ instead of $\bar{\boldsymbol{x}}$, we have the following estimator:

$$\frac{c_s}{\tau}S^{-1}\left(1 - \frac{r(\hat{\varphi}^2)}{\hat{\varphi}^2}\right)(\bar{\boldsymbol{x}} - \hat{\mu}_g\mathbf{1}) + \frac{S^{-1}\mathbf{1}}{\mathbf{1}'S^{-1}\mathbf{1}}.$$

This is exactly the estimator $\hat{\boldsymbol{w}}_{SB,A}(c_s; \bar{\boldsymbol{x}}, S)$ when we have the constraint $\mathbf{1}'\boldsymbol{w} = 1$ and $B = \mathbf{1}$. Note that when we have the constraint $\mathbf{1}'\boldsymbol{w}$ is equal to a constant and $B = \mathbf{1}$, the estimator $\hat{\boldsymbol{w}}_{SB,A}(c_s; \bar{\boldsymbol{x}}, S)$ reduces to that shrinking toward the origin, as discussed above.

Finally, we note that obtaining results similar to Theorems 3.1 and 3.2 when there are linear constraints on portfolio weight is technically difficult. This problem is left for future research.

3.4. Proofs of Theorems 3.1, 3.2, and 3.3

In the proofs given below, for simplicity, we assume that the function $r(\cdot)$ is differentiable. If $r(\cdot)$ is not differentiable, the proofs proceed by applying Riemann integration and replacing the term $r'(x)\mathrm{d}x$ by $\mathrm{d}r(x)$.

3.4.1. Proofs of Theorems 3.1 and 3.2

Setting $\tilde{\boldsymbol{x}} = \bar{\boldsymbol{x}} - SZ'(ZSZ')^{-1}(Z\bar{\boldsymbol{x}} - \boldsymbol{a})$, $\hat{\zeta}^2 = (\bar{\boldsymbol{x}} - \tilde{\boldsymbol{x}})'S^{-1}(\bar{\boldsymbol{x}} - \tilde{\boldsymbol{x}})$, and

$$\hat{\boldsymbol{w}}_{SZ}(c_s; \bar{\boldsymbol{x}}, S) = \frac{c}{\tau}S^{-1}\left[\tilde{\boldsymbol{x}} + \left(1 - \frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\right)(\bar{\boldsymbol{x}} - \tilde{\boldsymbol{x}})\right],$$

and assuming that $c_c = c_s = c$, we first evaluate the risk difference between $\hat{\boldsymbol{w}}_C(c; \bar{\boldsymbol{x}}, S) = c\tau^{-1}S^{-1}\bar{\boldsymbol{x}}$ and $\hat{\boldsymbol{w}}_{SZ}(c; \bar{\boldsymbol{x}}, S)$. Note that when $\boldsymbol{a} = \boldsymbol{0}$, $\hat{\boldsymbol{w}}_{SZ}$ is the estimator $\hat{\boldsymbol{w}}_{SZ1}$ given in Theorem 3.1 and that when $\boldsymbol{a} \neq \boldsymbol{0}$ and $r(\cdot) = d$ (const), $\hat{\boldsymbol{w}}_{SZ}$ is the estimator $\hat{\boldsymbol{w}}_{SZ2}$ given in Theorem 3.2. We have the loss difference between $\hat{\boldsymbol{w}}_C(c; \bar{\boldsymbol{x}}, S)$ and $\hat{\boldsymbol{w}}_{SZ}(c; \bar{\boldsymbol{x}}, S)$ under the loss function given in Equation (1.2) as

$$\begin{aligned} 2\tau\Delta L =& 2c\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}(cS^{-1}\bar{\boldsymbol{x}} - \Sigma^{-1}\boldsymbol{\mu})'\Sigma Z'(ZSZ')^{-1}(Z\bar{\boldsymbol{x}} - \boldsymbol{a}) \\ & - c^2\left(\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\right)^2 (Z\bar{\boldsymbol{x}} - \boldsymbol{a})'(ZSZ')^{-1}Z\Sigma Z'(ZSZ')^{-1}(Z\bar{\boldsymbol{x}} - \boldsymbol{a}). \end{aligned} \tag{3.7}$$

Setting $W = \Sigma^{-1/2}S\Sigma^{-1/2}$, $\boldsymbol{y} = \Sigma^{-1/2}\bar{\boldsymbol{x}}$, $\boldsymbol{\eta} = \Sigma^{-1/2}\boldsymbol{\mu}$, and $\tilde{Z} = Z\Sigma^{1/2}$, Equation (3.7) is written as

$$\begin{aligned} 2\tau\Delta L =& 2c^2\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\boldsymbol{y}'W^{-1}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y} - \boldsymbol{a}) - 2c\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\boldsymbol{\eta}'\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y} - \boldsymbol{a}) \\ & - c^2\left(\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\right)^2 (\tilde{Z}\boldsymbol{y} - \boldsymbol{a})'(\tilde{Z}W\tilde{Z}')^{-1}\tilde{Z}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y} - \boldsymbol{a}), \end{aligned} \tag{3.8}$$

where $\hat{\zeta}^2 = (\tilde{Z}\boldsymbol{y} - \boldsymbol{a})'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y} - \boldsymbol{a})$.

Next, we present lemmas that are used to evaluate the expectation of Equation (3.8). Here, we define the orthogonal matrix P such that

$$P\tilde{Z}'\tilde{Z}P' = \Lambda = \begin{pmatrix} \Lambda_1 & O \\ O & O \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix},$$

where Λ is a diagonal matrix with eigenvalues of $\tilde{Z}'\tilde{Z}$ on its diagonal, Λ_1 is an $\ell \times \ell$ diagonal matrix with positive diagonal elements, P_1 is an $\ell \times p$ matrix, and P_2 is a $(p - \ell) \times p$ matrix. Setting $\tilde{\boldsymbol{y}} = P_1\boldsymbol{y}$, $\tilde{\boldsymbol{\eta}} = P_1\boldsymbol{\eta}$, $V_{ij} = P_iWP_j'$ for $P = (P_1' \ P_2')'$, $L = P_1\tilde{Z}'$, $\tilde{\boldsymbol{a}} = (L')^{-1}\boldsymbol{a}$, and $\omega = (\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{a}})'V_{11}(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{a}})[(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{a}})]^{-1}$, we see that $\tilde{\boldsymbol{y}} \sim N_\ell(\tilde{\boldsymbol{\eta}}, P_1\Sigma P_1')$, $N\omega^{-1} \sim \chi^2_{N-\ell}$, and ω is independent of $\tilde{\boldsymbol{y}}$. Using the fact that $(\tilde{Z}W\tilde{Z}')^{-1} = L^{-1}V_{11}^{-1}(L')^{-1}$, we can see that $\hat{\zeta}^2 = (\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{a}}) = (\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{a}})\omega$, and we have the following lemmas.

Lemma 3.1. *If the expectation of each term of the following equation exists, then*

$$E\left[\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\boldsymbol{y}'W^{-1}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y}-\boldsymbol{a})\right]=\frac{N-\ell-2}{N-p-2}E\left[\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\tilde{\boldsymbol{y}}'V_{11}^{-2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\right].$$

Proof. Note that $N\omega^{-1}\sim\chi^2_{N-\ell}$, $\tilde{\boldsymbol{y}}\sim N(\tilde{\boldsymbol{\eta}},I/N)$ because $\boldsymbol{y}\sim N(\boldsymbol{\eta},I/N)$, and ω and $\tilde{\boldsymbol{y}}$ are independent. Since $(L')^{-1}\tilde{Z}\boldsymbol{y}=(L')^{-1}\tilde{Z}P'P\boldsymbol{y}=(L')^{-1}(\tilde{Z}P_1'\ \tilde{Z}P_2')(P_1'\ P_2')'\boldsymbol{y}=(I\ O)(P_1'\ P_2')'\boldsymbol{y}=P_1\boldsymbol{y}=\tilde{\boldsymbol{y}}$, and $W^{-1}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}=[P_1'(V_{11}^{-1}+V_{11}^{-1}V_{21}'V_{22\cdot1}^{-1}V_{21}V_{11}^{-1})-P_2'V_{22\cdot1}^{-1}V_{21}V_{11}^{-1}]V_{11}^{-1}(L')^{-1}$, we have $\hat{\zeta}^2=(\tilde{Z}\boldsymbol{y}-\boldsymbol{a})'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y}-\boldsymbol{a})=(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})$, where $\tilde{\boldsymbol{a}}=(L')^{-1}\boldsymbol{a}$. Furthermore, we have $\hat{\zeta}^2=(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega$ from the definition of ω, and setting $V_{22\cdot1}=V_{22}-V_{21}V_{11}^{-1}V_{21}'$, we have

$$\begin{aligned}W^{-1}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}&=P'V^{-1}P\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}=P'V^{-1}\begin{pmatrix}L\\O\end{pmatrix}L^{-1}V_{11}^{-1}(L')^{-1}\\&=[P_1'(V_{11}^{-1}+V_{11}^{-1}V_{21}'V_{22\cdot1}^{-1}V_{21}V_{11}^{-1})-P_2'V_{22\cdot1}^{-1}V_{21}V_{11}^{-1}]V_{11}^{-1}(L')^{-1}.\end{aligned}$$

Thus, we have

$$\begin{aligned}&\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\boldsymbol{y}'W^{-1}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y}-\boldsymbol{a})\\&=\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\left[\tilde{\boldsymbol{y}}'(V_{11}^{-1}+V_{11}^{-1}V_{21}'V_{22\cdot1}^{-1}V_{21}V_{11}^{-1})-\boldsymbol{y}'P_2'V_{22\cdot1}^{-1}V_{21}V_{11}^{-1}\right]V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}}).\end{aligned}\tag{3.9}$$

Let $vec(X)$ denote the operation that stacks the columns of a matrix X into a vector. Since $V_{11}\sim W_\ell(N-1,I/N)$, $\mathrm{vec}(V_{11}^{-1/2}V_{21}')\sim N_{\ell\times(p-\ell)}(\boldsymbol{0},I/N)$, and $V_{22\cdot1}\sim W_{p-\ell}(N-\ell-1,I/N)$, and they are all independent (cf., Muirhead 1982, Theorem 3.2.10), we have the expectation of the second term of Equation (3.9) as

$$E\left\{\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\left[-\boldsymbol{y}'P_2'V_{22\cdot1}^{-1}E(V_{21}V_{11}^{-1/2})V_{11}^{-1/2}\right]V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\right\}=0.$$

Denote the column vectors of $V_{11}^{-1/2}V_{12}$ by $\boldsymbol{v}_j$, $j=1,2,\ldots,p-\ell$, then $\boldsymbol{v}_j$, $j=1,2,\ldots,p-\ell$ are all independently normally distributed with mean $\boldsymbol{0}$ and covariance matrix I/N. Since $E[V_{22\cdot1}^{-1}]=N(N-p-2)^{-1}I$ and $E[V_{11}^{-1/2}V_{12}V_{12}'V_{11}^{-1/2}]=E[\sum_{h=1}^{p-\ell}\boldsymbol{v}_\ell\boldsymbol{v}_\ell']=(p-\ell)N^{-1}I$, we have

the expression of the expectation of the first term of Equation (3.9) as

$$E\left\{\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\tilde{\boldsymbol{y}}'\left\{V_{11}^{-1}+V_{11}^{-1/2}E[V_{11}^{-1/2}V_{21}'E[V_{22\cdot1}^{-1}]V_{21}V_{11}^{-1/2}]V_{11}^{-1/2}\right\}V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\right\}$$
$$=\frac{N-\ell-2}{N-p-2}E\left[\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\tilde{\boldsymbol{y}}'V_{11}^{-2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\right].$$

□

Lemma 3.2. *Let $\boldsymbol{\alpha}$ be a non-random $\ell\times 1$ vector. If the expectation of each term of the following equations exists, then*

(i) $$E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}}))\frac{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]=\frac{N-2}{N-\ell-1}E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)\omega\right],$$

(ii) $$E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}}))\frac{\boldsymbol{\alpha}'V_{11}^{-2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]$$
$$=\frac{N-2}{N-\ell-1}E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)\frac{\boldsymbol{\alpha}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right],$$

(iii) $$E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}}))\frac{\boldsymbol{\alpha}'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]$$
$$=E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)\frac{\boldsymbol{\alpha}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right],$$

(iv) $$E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}}))^2\frac{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{[(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})]^2}\right]$$
$$=\frac{N-2}{N-\ell-1}E\left[\frac{r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)^2}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right].$$

Proof. We can show (i) through (iv) by applying arguments similar to the proofs of Lemma 2.1 in Section 2.4.1 of Chapter 2. □

Lemma 3.3. *Let $r(\cdot)$ be differentiable, and denote its derivative by $r'(\cdot)$. If the expectation of each term of the following equations exists, then*

(i) $$(N-\ell-2)E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)\omega\right]$$
$$=NE\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)\right]+2E\left[r'((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega^2\right],$$

(ii) $E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)\frac{(\tilde{\boldsymbol{\eta}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]$
$= E\left[r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)\right] - \frac{2}{N}E\left[r'((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)\omega\right]$
$-\frac{\ell-2}{N}E\left[\frac{r((\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega)}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right].$

Proof. We can show (i) and (ii) by applying arguments similar to the proofs of Lemma 2.2 in Section 2.4.1 of Chapter 2. □

Using Lemmas 3.1 through 3.3, we evaluate the expected loss given by Equation (3.8). Using Lemmas 3.1, 3.2 (i), 3.2 (ii), and 3.3 (i), the expectation of the first term of Equation (3.8) is written as

$$
\begin{aligned}
&E\left[\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\boldsymbol{y}'W^{-1}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y}-\boldsymbol{a})\right] \\
&= \frac{N-\ell-2}{N-p-2}\left\{E\left[\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\right] + E\left[\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\tilde{\boldsymbol{a}}'V_{11}^{-2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\right]\right\} \\
&= \frac{(N-\ell-2)(N-2)}{(N-p-2)(N-\ell-1)}\left\{E\left[r(\hat{\zeta}^2)\omega\right] + E\left[r(\hat{\zeta}^2)\frac{\boldsymbol{a}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]\right\} \\
&= \frac{(N-2)}{(N-p-2)(N-\ell-1)} \\
&\quad\times\left\{NE[r(\hat{\zeta}^2)] + 2E[r'(\hat{\zeta}^2)(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega^2] + (N-\ell-2)E\left[r(\hat{\zeta}^2)\frac{\tilde{\boldsymbol{a}}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]\right\}.
\end{aligned}
$$

From Lemmas 3.2 (iii) and 3.3 (ii), we have the expectation of the second term of Equation (3.8) as

$$
\begin{aligned}
&E\left[\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\boldsymbol{\eta}'\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y}-\boldsymbol{a})\right] = E\left[\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\tilde{\boldsymbol{\eta}}'V_{11}^{-1}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\right] = E\left[r(\hat{\zeta}^2)\frac{\tilde{\boldsymbol{\eta}}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right] \\
&= E[r(\hat{\zeta}^2)] - \frac{2}{N}E[r'(\hat{\zeta}^2)\omega] - \frac{\ell-2}{N}E\left[\frac{r(\hat{\zeta}^2)}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right] + E\left[r(\hat{\zeta}^2)\frac{\tilde{\boldsymbol{a}}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right].
\end{aligned}
$$

Finally, using the fact that $(\tilde{Z}W\tilde{Z}')^{-1}\tilde{Z}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1} = L^{-1}V_{11}^{-2}(L')^{-1}$ and Lemma 3.2 (iv), we

have the expectation of the third term of Equation (3.8) as

$$E\left[\left(\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\right)^2(\tilde{Z}\boldsymbol{y}-\boldsymbol{a})'(\tilde{Z}W\tilde{Z}')^{-1}\tilde{Z}\tilde{Z}'(\tilde{Z}W\tilde{Z}')^{-1}(\tilde{Z}\boldsymbol{y}-\boldsymbol{a})\right]$$
$$=E\left[\left(\frac{r(\hat{\zeta}^2)}{\hat{\zeta}^2}\right)^2(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'V_{11}^{-2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\right]=\frac{N-2}{N-\ell-1}E\left[\frac{r(\hat{\zeta}^2)^2}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right].$$

Therefore, we have the expectation of the loss difference given in Equation (3.8) as

$$\begin{aligned}2\tau E[\Delta L]=2c&\left(c\frac{N(N-2)}{(N-p-2)(N-\ell-1)}-1\right)E[r(\hat{\zeta}^2)]\\&+2c\frac{\ell-2}{N}E\left[\frac{r(\hat{\zeta}^2)}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]-c^2\frac{N-2}{N-\ell-1}E\left[\frac{r(\hat{\zeta}^2)^2}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]\\&+2c^2\frac{(N-\ell-2)(N-2)}{(N-p-2)(N-\ell-1)}E\left[r(\hat{\zeta}^2)\frac{\tilde{\boldsymbol{a}}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]\\&-2cE\left[r(\hat{\zeta}^2)\frac{\tilde{\boldsymbol{a}}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\right]+C,\end{aligned}\tag{3.10}$$

where

$$C=2c^2\frac{2(N-2)}{(N-p-2)(N-\ell-1)}E[r'(\hat{\zeta}^2)(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})\omega^2]+\frac{4c}{N}E[r'(\hat{\zeta}^2)\omega]\geq 0.$$

First, we consider the case in which $\boldsymbol{a}=\boldsymbol{0}$. We show that if

$$c\geq(N-\ell-1)(N-p-2)N^{-1}(N-2)^{-1}\quad\text{and}\quad 0\leq r(\hat{\zeta}^2)\leq 2(\ell-2)(N-p-2)^{-1},$$

$\hat{\boldsymbol{w}}_{SZ}(c;\bar{\boldsymbol{x}},S)$ improves upon $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$. We see that the fourth and fifth terms on the right-hand side of Equation (3.10) are zero when $\boldsymbol{a}=\boldsymbol{0}$. Note that since both $r(\hat{\zeta}^2)$ and $-1/(\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}})$ are non-decreasing functions of $\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}}$, $E[-r(\hat{\zeta}^2)/(\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}})]\leq E[r(\hat{\zeta}^2)]\ E[-1/(\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}})]$ (cf., Lehmann and Casella 1998, Lemma 6.6). Thus, we have

$$\begin{aligned}&E\left[\left(2\frac{\ell-2}{N}-c\frac{N-2}{N-\ell-1}r(\hat{\zeta}^2)\right)\frac{r(\hat{\zeta}^2)}{\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}}}\right]\\&\geq E\left[\left(2\frac{\ell-2}{N}-c\frac{N-2}{N-\ell-1}\frac{2(\ell-2)}{N-p-2}\right)\frac{r(\hat{\zeta}^2)}{\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}}}\right]\\&\geq 2\frac{\ell-2}{N}\left(c\frac{N(N-2)}{(N-\ell-1)(N-p-2)}-1\right)E\left[r(\hat{\zeta}^2)\right]E\left[-\frac{1}{\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}}}\right].\end{aligned}$$

Thus, we have

$$2\tau E\left[\Delta L\right] \geq 2c\left(c\frac{N(N-2)}{(N-\ell-1)(N-p-2)}-1\right)\left(1-\frac{\ell-2}{N}E\left[\frac{1}{\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}}}\right]\right)E\left[r(\hat{\zeta}^2)\right].$$

Since $N\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}}$ has the non-central χ^2 distribution with ℓ degrees of freedom and non-centrality parameter $\boldsymbol{\mu}'\Sigma^{-1/2}P_1'P_1\Sigma^{-1/2}\boldsymbol{\mu}$, we have $E[(\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}})^{-1}] = E[N(\ell-2+2X)^{-1}]$, where X is a random variable having the Poisson distribution with mean $(N/2)\boldsymbol{\mu}'\Sigma^{-1/2}P_1'P_1\Sigma^{-1/2}\boldsymbol{\mu}$. Thus, we see that $2\tau E\left[\Delta L\right] \geq 0$. Note that from the expression of the risk function of $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$ (cf., Kan and Zhou 2007, Equation (23)), the risk of $\hat{\boldsymbol{w}}_C$ is finite if $N > p+4$, and that this condition is also necessary for the finiteness of the risk of $\hat{\boldsymbol{w}}_S$. Furthermore, we see that the risk of $\hat{\boldsymbol{w}}_C(c_1;\bar{\boldsymbol{x}},S)$ is larger than that of $\hat{\boldsymbol{w}}_C(c_2;\bar{\boldsymbol{x}},S)$ when $c_1 > c_2 \geq (N-p-1)(N-p-4)N^{-1}(N-2)^{-1}$. Therefore, when we allow c in $\hat{\boldsymbol{w}}_C(c;\bar{\boldsymbol{x}},S)$ and $\hat{\boldsymbol{w}}_{SZ}(c;\bar{\boldsymbol{x}},S)$ to take different values and set them as c_c and c_s, respectively, we see that if $c_c \geq c_s \geq (N-p-1)(N-p-2)N^{-1}(N-2)^{-1}$, $\hat{\boldsymbol{w}}_{SZ}(c_s;\bar{\boldsymbol{x}},S)$ improves upon $\hat{\boldsymbol{w}}_C(c_c;\bar{\boldsymbol{x}},S)$. Thus, we have Theorem 3.1.

Next, we consider the case in which $\boldsymbol{a} \neq \boldsymbol{0}$. The sum of the fourth and fifth terms of Equation (3.10) is written as

$$2c^2\frac{(N-\ell-2)(N-2)}{(N-p-2)(N-\ell-1)}E\left[r(\hat{\zeta}^2)\frac{\tilde{\boldsymbol{a}}'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{a}})}\left(c\frac{(N-\ell-2)(N-2)\omega}{(N-p-2)(N-\ell-1)}-1\right)\right].$$

Since $\tilde{\boldsymbol{y}}$ and ω are independent and $E[\omega] = N(N-\ell-2)^{-1}$, we can easily see that the sum of the fourth and fifth terms is zero when $c = (N-\ell-1)(N-p-2)N^{-1}(N-2)^{-1}$ and $r(\cdot) = const.$ From arguments similar to those presented for the case in which $\boldsymbol{a} = \boldsymbol{0}$, we see that sum of the first, second, and third terms of Equation (3.10) is non-negative when $0 \leq r(\hat{\zeta}^2) \leq 2(\ell-2)(N-p-2)^{-1}$, and thus we have Theorem 3.2.

3.4.2. Proof of Theorem 3.3

First, we show that the condition $\mathcal{R}(A) \cap \mathcal{R}(B) = \{\boldsymbol{0}\}$ guarantees that $B'F_1(A,S)B$ is non-singular. Let $Q' = (Q_1',\ Q_2')$ be the orthogonal matrix defined by Equation (3.3). Since rank

$(B'F_1(A,S)B) = \text{rank}\ (Q_2B)$, $B'F_1(A,S)B$ is non-singular if and only if rank (Q_2B) = rank B. In general, rank (XY) = rank Y if and only if $\mathcal{N}(X) \cap \mathcal{R}(Y) = \{\mathbf{0}\}$. Since $\mathcal{N}(Q_2) = \mathcal{R}(Q_1') = \mathcal{R}(A)$, we see that rank (Q_2B) = rank B if and only if $\mathcal{R}(A) \cap \mathcal{R}(B) = \{\mathbf{0}\}$.

Next, assuming $\mathcal{R}(A) \cap \mathcal{R}(B) = \{\mathbf{0}\}$, we give the expression of the loss difference between $\hat{\boldsymbol{w}}_{C,A}(c;\bar{\boldsymbol{x}},S)$ and $\hat{\boldsymbol{w}}_{SB,A}(c;\bar{\boldsymbol{x}},S)$ under the loss function given in Equation (1.2). Then, we obtain the loss difference as

$$\begin{aligned}
2\tau\Delta L =& 2c\frac{r(\hat{\xi}_{F_1}^2)}{\hat{\xi}_{F_1}^2}\left[c\Sigma F_1(A,S)\bar{\boldsymbol{x}} - \boldsymbol{\mu}\right]' F_1(A,S)(I - R_{F_1})\bar{\boldsymbol{x}} \\
& - c^2\left(\frac{r(\hat{\xi}_{F_1}^2)}{\hat{\xi}_{F_1}^2}\right)^2 \bar{\boldsymbol{x}}'(I - R_{F_1})'F_1(A,S)\Sigma F_1(A,S)(I - R_{F_1})\bar{\boldsymbol{x}} \\
& + 2\tau c\frac{r(\hat{\xi}_{F_1}^2)}{\hat{\xi}_{F_1}^2}\bar{\boldsymbol{x}}'(I - R_{F_1})'F_1(A,S)\Sigma F_2(A,S)\boldsymbol{b}. \qquad (3.11)
\end{aligned}$$

Since $F_1(A,S)\Sigma F_2(A,S) = \Sigma^{-1/2}F_1(\Sigma^{-1/2}A, W)F_2(\Sigma^{-1/2}A, W)$, we can show that the expectation of the third term of Equation (3.11) is $\mathbf{0}$ by an argument similar to that used in the proof of Lemma 5.2 of Mori (2004). Thus, we evaluate only the expectation of the first and second terms of Equation (3.11). Setting $U_{22} = Q_2SQ_2'$, we have $F_1(A,S) = Q_2'U_{22}^{-1}Q_2$ and $F_1(A,S)(I - R_{F_1}) = (I - R_{F_1})'F_1(A,S) = Q_2'F_1(Q_2B, U_{22})Q_2$. Therefore, setting $\boldsymbol{y}^* = Q_2\bar{\boldsymbol{x}}$ and $\boldsymbol{\eta}^* = Q_2\boldsymbol{\mu}$, the first and second terms of Equation (3.11) are written as

$$\begin{aligned}
& 2c\frac{r(\hat{\xi}_{F_1}^2)}{\hat{\xi}_{F_1}^2}\left[c(Q_2\Sigma Q_2')U_{22}^{-1}\boldsymbol{y}^* - \boldsymbol{\eta}^*\right]' F_1(Q_2B, U_{22})\boldsymbol{y}^* \\
& - c^2\left(\frac{r(\hat{\xi}_{F_1}^2)}{\hat{\xi}_{F_1}^2}\right)^2 (\boldsymbol{y}^*)'F_1(Q_2B, U_{22})(Q_2\Sigma Q_2')F_1(Q_2B, U_{22})\boldsymbol{y}^*, \qquad (3.12)
\end{aligned}$$

and $\hat{\xi}_{F_1}^2 = \bar{\boldsymbol{x}}'(I - R_{F_1})'F_1(A,S)(I - R_{F_1})\bar{\boldsymbol{x}} = (\boldsymbol{y}^*)'F_1(Q_2B, U_{22})\boldsymbol{y}^*$. Choosing a matrix Z^* that satisfies the condition $\mathcal{N}(Z^*) = \mathcal{R}(Q_2B)$, we have $U_{22}^{-1/2}Q_2B[B'Q_2'U_{22}^{-1}Q_2B]^{-1}B'Q_2'U_{22}^{-1/2} = I - U_{22}^{1/2}(Z^*)'\left[Z^*U_{22}(Z^*)'\right]^{-1}Z^*U_{22}^{1/2}$, and thus we see that $F_1(Q_2B, U_{22}) = (Z^*)'[Z^*U_{22}(Z^*)']^{-1}Z^*$.

Therefore, Equation (3.12) is written as

$$2c\frac{r(\hat{\xi}_{F_1}^2)}{\hat{\xi}_{F_1}^2}\left[cU_{22}^{-1}\boldsymbol{y}^* - (Q_2\Sigma Q_2')^{-1}\boldsymbol{\eta}^*\right](Q_2\Sigma Q_2')(Z^*)'[Z^*U_{22}(Z^*)']^{-1}Z^*\boldsymbol{y}^*$$
$$- c^2\left(\frac{r(\hat{\xi}_{F_1}^2)}{\hat{\xi}_{F_1}^2}\right)^2(\boldsymbol{y}^*)'(Z^*)'[Z^*U_{22}(Z^*)']^{-1}Z^*(Q_2\Sigma Q_2')(Z^*)'[Z^*U_{22}(Z^*)']^{-1}Z^*\boldsymbol{y}^*, \qquad (3.13)$$

and $\hat{\xi}_{F_1}^2 = (\boldsymbol{y}^*)'(Z^*)'[Z^*U_{22}(Z^*)']^{-1}Z^*\boldsymbol{y}^*$. Since, independently, $\boldsymbol{y}^* \sim N(\boldsymbol{\eta}^*, Q_2\Sigma Q_2'/N)$ and $U_{22} \sim W_{p-q}(N-1,\, Q_2\Sigma Q_2'/N)$, we see that the first and second terms of Equation (3.13) are of essentially the same form as in Equation (3.7). Therefore, from Theorem 3.1, we obtain Theorem 3.3. Here, note that if $N > \max(p+1, p-q-4)$, the risk of $\hat{\boldsymbol{w}}_{C,A}(c;\bar{\boldsymbol{x}},S)$ is finite and that this condition is also necessary for the finiteness of the risk of $\hat{\boldsymbol{w}}_{SB,A}(c;\bar{\boldsymbol{x}},S)$.

Chapter 4

Concluding Remarks

We have evaluated analytically the Stein-type estimators for the mean-variance optimal portfolio weights when the covariance matrix is unknown and is estimated by the sample estimator. Mori (2004) demonstrated analytically a dominance result of a Stein-type estimator for the mean-variance optimal portfolio weights, which shrinks toward the origin. However, we have presented dominance results for a broader class of estimators. Furthermore, we have also presented the dominance results of the estimators that shrink toward an arbitrary fixed point and a linear subspace. From these dominance results, we have clarified the conditions for the estimators proposed previously to have smaller risks than the classical estimator. We have also presented general dominance results for the case in which there are linear constraints on portfolio weights, as discussed by Mori (2004).

In Chapter 2, we have considered the estimators that shrink toward the origin or a fixed point. We have found that the conditions for the shrinkage estimators to have smaller risks than the classical estimator differs from that for the Stein-type estimators of the mean vector. Although the condition $p > 2$ on the number of risky assets p is common, the range of the function $r(\cdot)$ in the shrinkage estimators is wider than that for the Stein-type estimators of the mean vector. We have demonstrated that some previously proposed estimators belong to

our class and have clarified the conditions for the estimators to dominate the classical estimator. Furthermore, we have proposed a new estimator that uses a prior information concerning the Sharpe ratio and have shown that the estimator has a much smaller risk than the classical estimator when the true value of the Sharpe ratio is close to that suggested by the prior information.

In Chapter 2, we have also given the dominance results for the case with linear constraints on portfolio weights. In this case, the Stein-type estimators dominate the classical estimator when $p > q + 2$ and $0 \leq r_A(\cdot) \leq 2(p - q - 2)(N - p + q - 2)^{-1}$, where q is the number of linear constraints and N is the number of observations. Therefore, we have found that when there are linear constraints, the effective range of the number of risky assets and the function $r(\cdot)$ is narrower than that for the case with no constraints.

In Chapter 3, we have presented the dominance results for the estimators that shrink toward a linear subspace and have obtained the conditions for Black and Litterman's (1992) estimator, Jorion's (1986) Bayes-Stein estimator, and Kan and Zhou's (2007) three-fund rule estimator to have smaller risks than the classical estimator. To apply Black and Litterman's estimator effectively, we need a pertinent prior distribution of the mean vector. However, in practice, it is difficult for portfolio managers to construct a reasonable prior distribution. It is expected that results presented in the present thesis will be helpful for portfolio managers to determine parameters in the prior distribution. Although Jorion's Bayes-Stein estimator is very popular in finance, there have been no studies that have addressed its effectiveness analytically. Since this estimator belongs to the proposed class of estimators that shrink toward a linear subspace, we have found the conditions for this estimator to have smaller risk. Kan and Zhou (2007) proposed the three-fund rule estimator, which is a shrinkage estimator toward the grand mean. Although its form is somewhat complicated, we have also presented the dominance result for this estimator. Kan and Zhou (2007) anticipated that the risk of the three-fund rule estimator

is smaller than that of the two-fund rule estimator. Note, however, that if the true optimal portfolio weights of risky assets are close to $\mathbf{0}$, then the two-fund rule estimator performs better than the three-fund rule estimator.

In the present thesis, we have studied not only the case with no constraints on the portfolio weights but also the case with linear constraints on portfolio weights. However, in practice, constraints other than linear equality constraints may be imposed on portfolio weights, such as linear inequality constraints on portfolio weights or an upper bound constraint on the variance of the optimal portfolio. The Stein-type estimators obtained for the case with no constraints or the case with linear constraints may not be effective when there are constraints other than linear equality constraints. Thus, better estimators should also be explored for these cases.

In the present thesis, we have estimated the covariance matrix by the sample covariance matrix. However, Ledoit and Wolf (2003, 2004), Jorion (1986), Frost and Savarino (1986), and Jobson, Korkie, and Ratti (1979) have adopted another type of estimator for the covariance matrix. Recently, Ledoit and Wolf's (2003, 2004) estimators have been applied in numerous studies in the finance literature. However, their estimators are not intended to improve under the loss function used in the estimation problem of the optimal portfolio weights. The effectiveness of these estimators has also been shown only by numerical simulations or empirical studies. We should address the problem of improving the classical estimator by using such estimators of the covariance matrix.

Finally, we note that the results of the present thesis have been obtained under the normality and independence assumptions. These assumptions are rather restrictive, and we hope that similar results will be obtained for a broader class of distributions.

Bibliography

Alexander, G. J. & Resnick, B. G. (1985). More on estimation risk and simple rules for optimal portfolio selection. *The Journal of Finance*, 40(1), 125–133.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd Ed. John Wiley.

Baranchik, A. J. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. *Stanford University Technical Report No. 51.*

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics*, 41(2), 642–645.

Bawa, V. S., Brown, S. J., & Klein, R. W. (1979). *Estimation Risk and Optimal Portfolio Choice*. North-Holland.

Berger, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *The Annals of Statistics*, 4(1), 223–226.

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

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

Brandt, M. (2009). Portfolio Choice Problems. In: *Handbook of Financial Econometrics*. Y. Ait-Sahalia & L. P. Hansen, eds. Elsevier Science.

Brown, S. J. (1976). Optimal portfolio choice under uncertainty. *Ph.D. Dissertation, University of Chicago*.

Brown, S. J. (1978). The portfolio choice problem: Comparison of certainty equivalence and optimal Bayes portfolios. *Communications in Statistics-Simulation and Computation*, 7(4), 321–334.

Brown, S. J. (1979). The effect of estimation risk on capital market equilibrium. *The Journal of Financial and Quantitative Analysis*, 14(2), 215–220.

Casella, G. & Hwang, J. T. (1987). Employing vague prior information in the construction of confidence sets. *Journal of Multivariate Analysis*, 21(1), 79–104.

Chen, S. N. & Brown, S. J. (1983). Estimation risk and simple rules for optimal portfolio selection. *The Journal of Finance*, 38(4), 1087–1093.

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

Efron, B. & Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *The Annals of Statistics*, 4(1), 11–21.

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2007). *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons, Inc.

Frost, P. A. & Savarino, J. E. (1986). An empirical Bayes approach to efficient portfolios. *Journal of Financial and Quantitative Analysis*, 21(3), 293–305.

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

Golosnoy, V. & Okhrin, Y. (2007). Multivariate shrinkage for optimal portfolio weights. *The European Journal of Finance*, 13(5), 441–458.

Grauer, R. R. & Hakansson, N. H. (1995). Stein and CAPM estimators of the means in asset allocation. *International Review of Financial Analysis*, 4(1), 35–46.

Grauer, R. R. & Hakansson, N. H. (2001). Applying portfolio change and conditional performance measures: The case of industry rotation via the dynamic investment model. *Review of Quantitative Finance and Accounting*, 17(3), 237–265.

Ingersoll, J. E. (1987). *Theory of Financial Decision Making*. Rowman & Littlefield.

James, W. & Stein, C. (1961). Estimation with quadratic loss. (In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability), 361–379.

Jobson, J. D., Korkie, B., & Ratti, V. (1979). Improved estimation for Markowitz portfolios using James-Stein type estimators. (In Proceedings of the American Statistical Association, Business and Economic Statistics Section), 279–284.

Jobson, J. D. & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371), 544–554.

Jobson, J. D. & Korkie, B. (1981). Putting Markowitz theory to work. *Journal of Portfolio Management*, 7(4), 70-74.

Jorion, P. (1985). International portfolio diversification with estimation risk. *The Journal of Business*, 58(3), 259–278.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3), 272–292.

Jorion, P. (1991). Bayesian and CAPM estimators of the means: Implications for portfolio selection. *Journal of Banking and Finance*, 15(3), 717–727.

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

Kashima, H. (2001). An improvement of the parameter certainty equivalence method in portfolio selection. *Asia-Pacific Financial Markets*, 8(1), 35–43.

Kashima, H. (2005). An application of a minimax Bayes rule and shrinkage estimators to the portfolio selection problem under the Bayesian approach. *Statistical Papers*, 46(4), 523–540.

Kinkawa, T. & Shinozaki, N. (2009). Estimation of the optimal portfolio weights by shrinking the mean vector towards a linear subspace. *Statistics: A Journal of Theoretical and Applied Statistics*, to appear.

Kinkawa, T. & Shinozaki, N. (2010). Dominance of a class of Stein type estimators for optimal portfolio weights when the covariance matrix is unknown. *Asia-Pacific Financial Markets*, 17(1), 19–50.

Klein, R. W. & Bawa, V. S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3), 215–231.

Klein, R. W. & Bawa, V. S. (1977). The effect of limited information and estimation risk on optimal portfolio diversification. *Journal of Financial Economics*, 5(2), 89–111.

Kubokawa, T., Robert, C. P., & Saleh, A. K. Md. E. (1993). Estimation of noncentrality parameters. *Canadian Journal of Statistics*, 21(1), 45–57.

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

Ledoit, O. & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 99(2), 365–411.

Lehmann, E. L. & Casella, G. (1998). *Theory of Point Estimation*, 2nd Ed. Springer.

Lin, P. & Tsai, H. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *The Annals of Statistics*, 1(1), 142–145.

Lindley, D. V. (1962). Discussion of Professor Stein's paper, confidence sets for the mean of a multivariate normal distribution. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 24, 265–296.

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

Michaud, R. O. (1989). The Markowitz optimization enigma: Is optimized optimal ? *Financial Analysts Journal*, 45(1) 21–42.

Michaud, R. O. (1998). *Efficient Asset Management*. Harvard Business School Press, Boston, MA.

Mori, H. (2004). Finite sample properties of estimators for the optimal portfolio weight. *Journal of the Japan Statistical Society*, 35(1), 27–46.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley.

Okhrin, Y. & Schmid, W. (2007). Comparison of different estimation techniques for portfolio selection. *AStA—Advances in Statistical Analysis*, 91(2), 109–127.

Saleh, A. K. Md. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications*. John Wiley.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1), 119–138.

Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1), 49–58.

Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *The Annals of Mathematical Statistics*, 42(1), 385–388.

Theil, H. (1971). *Principles of Econometrics*. John Wiley and Sons, New York.

Wang, Z. (2005). A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies*, 18(2), 673–705.

Acknowledgements

I especially would like to express my deepest gratitude to my supervisor, Professor Nobuo Shinozaki for his support and elaborated guidance throughout my research. I would also like to thank Professors Norio Hibiki, Junichi Imai, and Kunio Shimizu for valuable comments on this dissertation. I am very grateful to the members of Shinozaki laboratory for their support and invaluable discussion. I am also grateful to my colleagues at Nomura Asset Management Company Ltd. for their support and understanding. Finally I would like to thank my parents for their endless love, understanding, support and encouragement throughout my study.