The best estimation for high-dimensional Markowitz mean-variance optimization

Bai, Zhidong and Li, Hua and Wong, Wing-Keung

10 January 2013

The Best Estimation for High-Dimensional Markowitz Mean-Variance Optimization

Zhidong Bai
KLASMOE and School of Mathematics and Statistics
Northeast Normal University
Department of Statistics and Applied Probability
and Risk Management Institute
National University of Singapore

Hua Li
School of Sciences, Chang Chun University
Department of Statistics and Applied Probability
National University of Singapore

Wing-Keung Wong
Department of Economics
Hong Kong Baptist University

January 18, 2013

Corresponding author: Wing-Keung Wong, Department of Economics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. Tel: (852)-3411-7542, Fax: (852)-3411-5580, Email: awong@hkbu.edu.hk

Acknowledgments
The third author would like to thank Professors Robert B. Miller and Howard E. Thompson for their continuous guidance and encouragement. This research is partially supported by Northeast Normal University, the National University of Singapore, Chang Chun University, Hong Kong Baptist University, and the Research Grants Council (RGC) of Hong Kong.

The Best Estimation for High-Dimensional Markowitz Mean-Variance Optimization

Abstract The traditional (plug-in) return for the Markowitz mean-variance (MV) optimization has been demonstrated to seriously overestimate the theoretical optimal return, especially when the dimension to sample size ratio p/n is large. The newly developed bootstrap-corrected estimator corrects the overestimation, but it incurs the "under-prediction problem," it does not do well on the estimation of the corresponding allocation, and it has bigger risk. To circumvent these limitations and to improve the optimal return estimation further, this paper develops the theory of spectral-corrected estimation. We first establish a theorem to explain why the plug-in return greatly overestimates the theoretical optimal return. We prove that under some situations the plug-in return is $\sqrt{\gamma}$ times bigger than the theoretical optimal return, while under other situations, the plug-in return is bigger than but may not be $\sqrt{\gamma}$ times larger than its theoretic counterpart where $\gamma = \frac{1}{1-y}$ with y being the limit of the ratio p/n.

Thereafter, we develop the spectral-corrected estimation for the Markowitz MV model which performs much better than both the plug-in estimation and the bootstrap-corrected estimation not only in terms of the return but also in terms of the allocation and the risk. We further develop properties for our proposed estimation and conduct a simulation to examine the performance of our proposed estimation. Our simulation shows that our proposed estimation not only overcomes the problem of "over-prediction," but also circumvents the "under-prediction," "allocation estimation," and "risk" problems. Our simulation also shows that our proposed spectral-corrected estimation is stable for different values of sample size n, dimension p, and their ratio p/n. In addition, we relax the normality assumption in our proposed estimation so that our proposed spectral-corrected estimators could be obtained when the returns of the assets being studied could follow any distribution under the condition of the existence of the fourth moments.

Keywords: G11; C13

JEL Classification: Markowitz mean-variance optimization, Optimal Return, Optimal Portfolio Allocation, Large Random Matrix, Bootstrap Method.

1 Introduction

This paper aims to develop the best estimation for the problem of the high-dimensional Markowitz mean-variance (MV) portfolio optimization. Our proposed estimation may not be the best estimation, but we believe our approach at least enables academics and practitioners to get closer to obtaining the best estimation for the high-dimensional MV Markowitz optimization problem. We first discuss the literature on the issue.

The conceptual framework of the classical MV portfolio optimization was set forth by Markowitz in 1952. Since then, modeling Markowitz MV portfolio optimization theory is one of the most important topics to be empirically and theoretically studied by academics and practitioners. It is a milestone in modern finance theory, including optimal portfolio construction, asset allocation, utility maximization, and investment diversification. Given a set of assets, it enables investors to find the best allocation of wealth incorporating their preferences as well as their expectations of returns and risks. It provides a powerful tool for investors to allocate their wealth efficiently.

Although several procedures for computing optimal return estimates (e.g., Sharpe, 1967, 1971; Stone, 1973; Elton, Gruber, and Padberg, 1976, 1978: Markowitz and Perold, 1981; Perold, 1984; Carpenter et al., 1991; Jacobs, Levy, and Markowitz, 2005) have been put forth entirely since the 1960s, academics and practitioners still have doubts about the performance of the estimates. The portfolio formed by using the classical MV approach always results in extreme portfolio weights that fluctuate substantially over time and perform poorly in the sample estimation as well as in the out-of-sample forecasting. Several studies recommend disregarding the results, or abandoning the approach. For example, Frankfurter, Phillips, and Seagle (1971) find that the portfolio selected according to the Markowitz MV criterion is not as effective as an equally weighted portfolio. Michaud (1989) documents the MV optimization to be one of the outstanding puzzles in modern finance that has yet to meet with widespread acceptance by the investment community. He calls this puzzle the "Markowitz optimization enigma" and calls the MV optimizers "estimation-error maximizers." Simaan (1997) has found MV-optimized portfolios to be unintuitive, thereby making their estimates do more harm than good. Furthermore, Zellner and Chetty (1965), Brown (1978), and Kan and Zhou (2006) show that the Bayesian rule under a diffuse prior outperforms the MV optimization.

To investigate the reasons why the MV optimization estimate is so far away from its theoretic counterpart, different studies provide different observations and suggestions. So far, all believe that it is because the "optimal" return is formed by a combination of returns from an extremely large number of assets (e.g., McNamara, 1998). This is particularly troublesome because optimization routines are often characterized as error maximization algorithms. Small changes in the inputs can lead to large changes in the estimation (e.g., Frankfurter, Phillips, and Seagle, 1971). For the necessary input parameters, some studies (e.g., Michaud, 1989; Chopra, Hensel, and Turner, 1993; Jorion, 1992; Hensel and Turner, 1998) suggest that the estimation of the covariance matrix plays an important role in the problem. For instance, Jorion (1985) and others suggest that the main difficulty concerns the extreme weights that often arise when constructing sample efficient portfolios that are extremely sensitive to changes in asset means. Others suggest that the estimation of the correlation matrix plays an important role. For example, Laloux, Cizeau, Bouchaud, and Potters (1999) find that Markowitz's portfolio optimization scheme is not adequate because its lowest eigenvalues dominating the smallest risk portfolio are dominated by noise. Thus, how to use the Markowitz optimization procedure efficiently depends on whether the expected return and the covariance matrix can be estimated accurately.

Many studies have improved the estimate of the classical Markowitz MV approach by using different approaches. For example, by introducing the notion of "factors" influencing stock prices, Sharpe (1964) formulates the single-index model to simplify both the informational and computational complexity of the general model. Ross (1976) uses the arbitrage pricing theory and the multi-factor model to formulate the excessive returns of assets. Konno and Yamazaki (1991) propose a mean-absolute deviation portfolio optimization to overcome the difficulties associated with the classical Markowitz model, but Simaan (1997) finds that the estimation errors for the mean-absolute deviation portfolio model are still very severe, especially in small samples. Manganelli (2004) works with univariate portfolio GARCH models to provide a solution to the curse of dimensionality associated with multivariate generalized autoregressive conditionally heteroskedastic estimation. In addition, Wong, Carter, and Kohn (2003) impose some constraints on the correlation matrix to capture the essence of the real correlation structure while Ledoit and Wolf (2004) use shrinkage and the eigen-method to construct a better estimate. On the other hand, Jacobs, Levy, and Markowitz (2005) present fast algorithms for calculating MV efficient frontiers when the investor can sell securities short as well as buy them long, and when a factor and/or scenario model of covariance is assumed.

To improve the optimal return estimation, Bai, Liu, and Wong (2009,2009a) first prove that the traditional return estimate is always larger than its theoretical value with a fixed rate depending on the ratio of the dimension to the sample size p/n. They call this problem "over-prediction." In this paper we explore the issue further. We will look for reasons why the classical MV optimal return estimation is far away from the real return by adopting random matrix theory. We find that the estimation of getting the optimal return and the corresponding asset allocation (we call it plug-in estimators) by plugging the sample mean and the sample covariance matrix is highly unreliable because (a) the estimate contains substantial estimation error and (b) in the optimization step the estimation becomes "over-predicted." We also develop a theorem to explain why the plug-in return greatly overestimates the theoretical optimal return. For example, we prove that under some situations the plug-in return is $\sqrt{\gamma}$ times bigger than the theoretical optimal return, while, under other situations, the plug-in return is bigger than but may not be $\sqrt{\gamma}$ times larger than its theoretic counterpart where $\gamma = \frac{1}{1-y}$ with y being the limit of the ratio p/n.

To obtain a better optimal return estimator, Bai, Liu, and Wong (2009,2009a) propose a new method called the bootstrap-corrected estimation to reduce the error of over-prediction by using the bootstrap approach. They claim that their bootstrap-corrected estimator circumvents the "over-prediction" problem. Leung, Ng, and Wong (2012) extend their work by providing a closed form of the estimation. Nonetheless, to check how good an estimate of MV portfolio optimization is, one should not only care about how good the estimation of the return, but also about how good the estimation of the corresponding allocation is and how big their risk is. In this paper we find that the bootstrap-corrected estimation does not outperform the plug-in estimation for both the allocation and the risk, and sometimes it is even worse. We call the former the "allocation estimation" problem and the latter the "risk" problem. In addition, our simulation shows that although the bootstrap-corrected estimation could overcome the "over-prediction" problem, it incurs the "under-prediction" problem. Thus, looking for the best MV portfolio optimization estimation that could solve all of the defects in the MV portfolio optimization – the "over-prediction," "under-prediction," "allocation estimation," and "risk" problems – is still a very important outstanding problem.

In this paper we aim to develop a new estimator that could overcome all four defects. To do so, we modify the key point estimation – the eigenvalue of the covariance matrix. By doing so, we provide a more accurate covariance matrix estimator and, thereafter, develop the corresponding optimal estimators for both return and allocation. We establish some properties for the estimation and conduct simulation. Our simulation results show that our method not only solves the over-prediction and under-prediction problems, but also substantially reduces the estimation error of both the return and the allocation and reduces its risk. Our simulation also shows that our proposed spectral-corrected estimation is stable for different values of sample size n, dimension p, and their ratio p/n. In addition, we relax the normality assumption in our proposed estimation so that our proposed spectral-corrected estimators could be obtained for the problem of the high-dimensional Markowitz MV portfolio optimization when the returns of the assets being studied could follow any distribution under the condition of the existence of the fourth moments. Thus, our proposed estimation should be a very promising method for the Markowitz portfolio optimization procedure.

The rest of the paper is organized as follows. In Section 2, we will present the problem of Markowitz's MV portfolio optimization. In Section 3, we will discuss the theory of the large dimensional random matrix that could be used to solve the Markowitz portfolio optimization problem. In Section 4, we will first introduce the traditional plug-in and newly developed bootstrap-corrected estimators and, thereafter, develop the theory of the spectral-corrected estimators for the optimal return and its asset allocation. We will conduct a simulation in Section 5 to compare the performance of our proposed spectral-corrected estimators with those of the plug-in and bootstrap-corrected estimators. Section 6 provides the summary and conclusion and suggests some possible directions for further research.

2 Markowitz's Mean-Variance Principle

To distinguish the well-known results in the literature from the ones derived in this paper, all cited results will be called *Propositions* and our derived results will be called *Theorems*. We first discuss Markowitz's MV optimization principle.

The pioneering work of Markowitz (1952, 1959) on the MV portfolio optimization procedure is a milestone in modern finance. It provides a powerful tool for efficiently allocating wealth to different investment alternatives. This technique incorporates investors' preferences and expectations of returns and risks for all assets considered, as well as diversification effects, which reduce the overall portfolio risk. According to the theory, portfolio optimizers respond to the uncertainty of an investment by selecting portfolios that maximize profit, subject to achieving a specified level of calculated risk or, equivalently, minimize variance, subject to obtaining a predetermined level of expected gain (Markowitz, 1952, 1959, 1991; Kroll, Levy, and Markowitz, 1984). More precisely, we suppose that there are p-branch of assets whose returns are denoted by $\mathbf{r} = (\mathbf{r}_1, \cdots, \mathbf{r}_p)^{\mathbf{T}}$ with mean vector $\mu = (\mu_1, \cdots, \mu_p)^{\mathbf{T}}$ and covariance matrix $\Sigma = (\sigma_{ij})$. In addition, we assume that an investor will invest capital C on the p-branch of assets such that she wants to allocate her investable wealth to the assets to attain one of the following:

a. to maximize return subject to a given level of risk, or

b. to minimize risk for a given level of expected return.

Since the above two cases are equivalent, we consider only the first one in this paper. Without loss of generality, we assume $C = 1$ and her investment plan to be $\mathbf{c} = (c_1, \cdots, c_p)^T$. Hence, we have $\Sigma_{i=1}^p c_i \leq 1$ in which the strict inequality corresponds to the fact that the investor could invest only part of her wealth. Her anticipated return, R, will then be $\mathbf{c}^T\mu$ with risk $\mathbf{c}^T\Sigma\mathbf{c}$. In this paper, we further assume that short selling is allowed, and hence, any component of $\mathbf{c}$ could be positive as well as negative. Thus, the above maximization problem can be reformulated as:

$$\max \mathbf{c}^T\mu, \text{ subject to } \mathbf{c}^T\mathbf{1} \leq 1 \text{ and } \mathbf{c}^T\Sigma\mathbf{c} \leq \sigma_0^2 \tag{2.1}$$

where $\mathbf{1}$ represents the p-dimensional vector of ones and σ_0^2 is a given level of risk. We call $R = \max \mathbf{c}^T\mu$ satisfying (2.1) the **optimal return** and $\mathbf{c}$ its corresponding **allocation plan**. One could obtain the solution of (2.1) from the following proposition:

Proposition 2.1. *For the optimization problem shown in (2.1), the optimal return, R, and its corresponding investment plan,* $\mathbf{c}$*, are obtained as follows:*

a. *If*

$$\frac{\mathbf{1}^T\Sigma^{-1}\mu\sigma_0}{\sqrt{\mu^T\Sigma^{-1}\mu}} < 1, \tag{2.2}$$

then the optimal return, R, and corresponding investment plan, $\mathbf{c}$*, will be*

$$R = \sigma_0\sqrt{\mu^T\Sigma^{-1}\mu} \tag{2.3}$$

and

$$\mathbf{c} = \frac{\sigma_0}{\sqrt{\mu^T\Sigma^{-1}\mu}}\Sigma^{-1}\mu\ . \tag{2.4}$$

b. *If*

$$\frac{\mathbf{1}^T\Sigma^{-1}\mu\sigma_0}{\sqrt{\mu^T\Sigma^{-1}\mu}} > 1\ , \tag{2.5}$$

then the optimal return, R, and corresponding investment plan, $\mathbf{c}$*, will be*

$$R = \frac{\mathbf{1}^T\Sigma^{-1}\mu}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} + b\left(\mu^T\Sigma^{-1}\mu - \frac{\left(\mathbf{1}^T\Sigma^{-1}\mu\right)^2}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}\right) \tag{2.6}$$

and

$$\mathbf{c} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} + b\left(\Sigma^{-1}\mu - \frac{\mathbf{1}^T\Sigma^{-1}\mu}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1}\right)\ , \tag{2.7}$$

where

$$b = \sqrt{\frac{\mathbf{1}^T\Sigma^{-1}\mathbf{1}\sigma_0^2 - 1}{\mu^T\Sigma^{-1}\mu\mathbf{1}^T\Sigma^{-1}\mathbf{1} - \left(\mathbf{1}^T\Sigma^{-1}\mu\right)^2}}\ .$$

The set of efficient feasible portfolios for all possible levels of portfolio risk forms the MV efficient frontier. For any given level of risk, Proposition 2.1 seems to provide investors a unique optimal return and its corresponding MV-optimal investment plan, and thus, it seems to provide a good solution to Markowitz's MV optimization procedure. Some may think that the problem is straightforward and the problem has been solved completely. Nonetheless, in reality, this is not the case because the estimation of the optimal return and its corresponding investment plan is a difficult task. We will discuss the issue in the next section.

3 Large Dimensional Random Matrix Theory

The large dimensional random matrix theory (LDRMT) traces back to the development of quantum mechanics in the 1940s. Because of its rapid development in theoretical investigations and its wide applications, it has attracted growing attention in many areas, including signal processing, wireless communications, economics and finance, as well as mathematics and statistics. Whenever the dimension of the data is large, the classical limiting theorems are no longer suitable because the statistical efficiency will be substantially reduced. Hence, academics have to search for alternative approaches to conduct such data analysis and the LDRMT has been found to the right for this purpose. The main advantage of adopting the LDRMT is its ability to investigate the limiting spectrum properties of random matrices when the dimension increases proportionally with the sample size. This turns out to be a powerful tool in dealing with large dimensional data analysis.

We incorporate the LDRMT to analyze the high dimensional MV optimization problem. In the analysis, the sample covariance matrix plays an important role in analyzing this type of data. Let $\mathbf{x}_k = (x_{1k}, \cdots, x_{pk})^T$ $(k = 1, 2, \cdots, n)$ be i.i.d. random vectors with mean vector μ, covariance matrix Σ, and the sample covariance matrix

$$S = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \overline{\mathbf{x}})(\mathbf{x}_k - \overline{\mathbf{x}})^T \tag{3.1}$$

in which the sample mean $\overline{\mathbf{x}} = \sum_{k=1}^{n} \mathbf{x}_k/n$ is the estimate of the mean vector μ.

The major difficulty in the estimation of optimal return is well recognized to be the inadequacy of using the inverse of the estimated covariance to estimate the inverse of the covariance matrix; see, for example, Laloux, Cizeau, Bouchaud, and Potters (1999). To present and thereafter circumvent this problem, in this paper we first introduce some fundamental definitions and theoretical results for the LDRMT. To do so, we first define the *empirical spectral distribution* for the sample covariance matrix as follows:

Definition 3.1. ***(Empirical Spectral Distribution, ESD)*** *Suppose that the sample covariance matrix S defined in (3.1) is a $p \times p$ matrix with eigenvalues $\{\lambda_j : j = 1, 2, \cdots, p\}$. If all eigenvalues are real, the empirical spectral distribution function, F^S, of the eigenvalues $\{\lambda_j\}$ for the sample covariance matrix, S, is*

$$F^S(x) = \frac{1}{p} \sharp\{j \leq p : \lambda_j \leq x\}, \tag{3.2}$$

where $\sharp E$ is the cardinality of the set E.

One of the main problems in LDRMT is to investigate the convergence of the ESD for the sequence $F_n = F^{S_n}$ for a given sequence of random matrices $\{S_n\}$. The limit distribution F of F_n, which is usually nonrandom, is called the limiting spectral distribution (LSD) of the sequence of $\{S_n\}$. Here, we first introduce one of the most powerful tools—the well-known Stieltjes transform as follows:

Definition 3.2. ***(Stieltjes transform)*** *The Stieltjes transform of a measure F is*

$$m(z) = \int \frac{1}{x - z} dF(x), \quad z \in \mathbb{C}^+,$$

where $\mathbb{C}^+ \doteq \{z : z \in \mathbb{C}, \Im(z) > 0\}$ is the set of complex numbers with a positive imaginary part.

Applying the Stieltjes transform, the convergence of the ESD F_n could be reduced to the convergence of m_n under some mild conditions where

$$m_n = \int \frac{1}{x - z} dF_n(x) = \frac{1}{n} \sum_{i=1}^{p} \frac{1}{\lambda_i - z} = \frac{1}{n} \mathrm{tr}(S_n - zI)^{-1}. \tag{3.3}$$

From (3.3), one could easily find that the Stieltjes transform connects the ESD of the covariance matrix and its eigenvalues.

As as accompaniment to the sample covariance matrix S_n, we refer to $\underline{S}_n = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \overline{\mathbf{x}})^T(\mathbf{x}_k - \overline{\mathbf{x}})$ as the companion matrix of S_n. It is obvious that both $\underline{S}_n$ and S_n have identical nonzero eigenvalues, and therefore, we obtain

$$\underline{F}_n(x) = (1 - \frac{p}{n})\delta_0 + \frac{p}{n} F_n(x),$$

where $\underline{F}_n$ and F_n are, respectively, the ESDs of $\underline{S}_n$ and S_n. Taking the Stieltjes transform on both sides of the equation above, we get

$$m_n(z) = -\frac{1 - p/n}{z} + \frac{p}{n} m_n(z).$$

We denote $\underline{F}_n$, $\underline{m}_n$, $\underline{F}$, and $\underline{m}$ as the companion versions of their corresponding spectral distributions and Stieltjes transforms. In the development of the theory for covariance matrices, one remarkable work is Silverstein (1995), who studies the behavior of the LSD for a sample covariance matrix by connecting it with the LSD of the corresponding population covariance matrix as shown in the following proposition:

Proposition 3.1. *[Sliverstein (1995)]* *Suppose that* $\mathbf{y}_k = (y_{1k}, y_{2k}, \cdots, y_{pk})^T$ $(k = 1, 2, \cdots, n)$ *are i.i.d. random vectors with zero mean and identity covariance matrix. Assume that* Σ_n *is a* $p \times p$ *nonrandom Hermitian and nonnegative definite matrix and the empirical distribution* F^{Σ_n} *converges almost surely to a probability distribution function* H *on* $[0, \infty]$ *as* $n \to \infty$*. Set* $\mathbf{x}_k = \mu + \Sigma^{1/2}\mathbf{y}_k$*. If* $p = p(n)$ *with* $p/n \to y > 0$ *as* $n \to \infty$*, then the ESD* F^{S_n} *converges in distribution almost surely to a nonrandom distribution function* F*, whose companion Stieltjes transform* $\underline{m}(z)$ *is the unique solution from*

$$z = -\frac{1}{\underline{m}} + y \int \frac{tdH(t)}{1 + t\underline{m}} . \tag{3.4}$$

Although Proposition 3.1 does not provide explicit expressions of H and F, the expressions of most of their analytic behaviors can be derived from applying equation (3.4), especially when some important properties only involve the equation on the real line (Silverstein and Choi, 1995). The following proposition is one of them:

Proposition 3.2. *[Silverstein and Choi (1995)]* *For LSD* F*, we let* S_F *denote its support and* S_F^c *denote the complement of its support. If* $u \in S_{\underline{F}}^c$*, then* $\underline{m} = \underline{m}(u)$ *satisfies:*

a. $\underline{m} \in \mathbb{R}\backslash\{0\}$,

b. $(-\underline{m})^{-1} \in S_H^c$, *and*

c. $dz/d\underline{m} > 0$.

Conversely, if $\underline{m}$ *satisfies (a)-(c), then* $u = z(\underline{m}) \in S_{\underline{F}}^c$.

Suppose that a sequence of sample covariance matrices have LSD F with support S_F. Since S_F is a closed subset of the real field $\mathbb{R}$, $1/(x - u_0)$ is bounded in S_F for any $u_0 \in S_F^c$. Define the generalized Stieltjes transform (GST) of F to be

$$m(u) = \int \frac{1}{(x-u)} dF(x), \; u \in S_F^c ,$$

we can then express the companion GST of F (denoted by $\underline{m}(u)$) as:

$$\underline{m}(u) = -\frac{1-y}{u} + y \int \frac{1}{x-u} dF(x) , \; \forall u \in S_F^c\backslash\{0\} , \tag{3.5}$$

where y is the limit ratio of population size to sample size p/n. We state the following proposition, which is useful in the estimation of the high-dimensional Markowitz MV optimization:

Proposition 3.3. *[Li, Chen, Qin, Yao, and Bai (2013)] Under the conditions of Proposition 3.1, we denote* $\underline{m}_n(u)$ *and* $\underline{m}(u)$ *as the companion GST of* F^{B_n} *and its limit* F*. In addition, we let* $U = \liminf_{n\to\infty} S^c_{F_n} \backslash \{0\}$ *and its interior be* $\overset{\circ}{U}$*. Then, for any* $u \in \overset{\circ}{U}$*, we have*

a. $\underline{m}_n(u)$ *converges to* $\underline{m}(u)$ *almost surely;*

b. $\underline{m}(u)$ *is a solution to equation:*

$$u(\underline{m}) = -\frac{1}{\underline{m}} + y \int \frac{t}{1 + t\underline{m}} dH(t) \; ; \tag{3.6}$$

c. *under the restriction of* $du/d\underline{m} > 0$*, the solution is unique;*

d. *for any interval* $[a, b]$ *with* $0 < a < b$*,* H *is uniquely determined by* $\{(u, \underline{m}) : \underline{m} \in [a, b]\}$*; and*

e. *if* H *has finite support and* $[a, b]$ *is an increasing interval of* $u(\underline{m})$*, then* H *is uniquely determined by* $\{(u, \underline{m}) : \underline{m}[a, b]\}$.

Applying Propositions 3.1 to 3.3, we obtain a method to estimate the eigenvalues of the population covariance matrix. We will discuss the theory in the next section.

4 Markowitz Mean-Variance Optimization Estimation

In this section, we first introduce the traditional plug-in and newly developed bootstrap-corrected estimators. Thereafter, we will develop the spectral-corrected estimators for the optimal return and its asset allocation. The plug-in estimators are intuitively constructed by plugging the sample mean and sample covariance matrix into the formula of the theoretic optimal return as shown in Proposition 2.1, whereas the bootstrap-corrected estimators are constructed by employing the bootstrap estimation technique. In this paper we propose the spectral-corrected estimators for the estimation in which the covariance matrix is estimated by the LDRMT. This is a key technique of improving the performance of our proposed estimators. The details are given in the following subsections.

4.1 Plug-In Estimator

Proposition 2.1 provides the solution for the optimization problem stated in (2.1). In practice, the parameters μ and Σ are unknown. A simple and natural way to estimate μ and Σ is to use the corresponding sample mean $\bar{\mathbf{x}}$ and sample covariance matrix S, respectively. Thereafter, by plugging the sample mean $\overline{\mathbf{x}}$ and the sample covariance matrix S into the formulae of the asset allocation $\mathbf{c}$ in Proposition 2.1, we obtain the estimates:

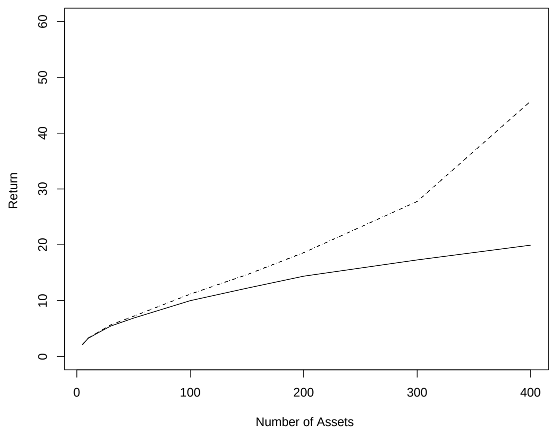
$$\begin{aligned} \widehat{R}_p &= \hat{\mathbf{c}}_p^T \overline{\mathbf{x}}\,, \\ \hat{\mathbf{c}}_p &= \begin{cases} \frac{S^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T S^{-1}\overline{\mathbf{x}}}} & \text{if} \quad \frac{\sigma_0 1^T S^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T S^{-1}\overline{\mathbf{x}}}} < 1, \\ \frac{S^{-1}1}{1^T S^{-1} 1} + \hat{b}_p (S^{-1}\overline{\mathbf{x}} - \frac{1^T S^{-1}\overline{\mathbf{x}}}{1^T S^{-1} 1} S^{-1}\mathbf{1}) & \text{if} \quad \frac{\sigma_0 1^T S^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T S^{-1}\overline{\mathbf{x}}}} > 1, \end{cases} \end{aligned} \tag{4.1}$$

for the optimal return and its corresponding allocation in which

$$\hat{b}_p = \sqrt{\frac{1^T S^{-1} 1 \sigma_0^2 - 1}{\overline{\mathbf{x}}^T S^{-1} \overline{\mathbf{x}} 1^T S^{-1} 1 - (1^T S^{-1} \overline{\mathbf{x}})^2}}\,.$$

For simplicity, we call $\widehat{R}_p$ the "**plug-in return**" and $\hat{\mathbf{c}}_p$ the "**plug-in allocation**." The "plug-in" return, $\widehat{R}_p$, has been used as the traditional return estimator after Markowitz introduce the MV portfolio optimization theory. This procedure is very simple but academics and practitioners have found that this estimate could do more harm than good and its estimate is not even as effective as an equally weighted portfolio estimate (e.g., Frankfurter, Phillips, and Seagle, 1971). In addition, Bai, Liu, and Wong (2009,2009a) have shown that the traditional return estimate is always larger than its theoretical value when n and p are large and the ratio of the dimension to sample size p/n is not small. They call this problem "over-prediction." Readers may also refer to Figure 1 for how severe the "over-prediction" is when p and n are large. We note that although $\overline{\mathbf{x}}$ is a good estimate of μ and $\hat{\mathbf{c}}_p$ is close to $\mathbf{c}$ (see Section 5 for the findings), $\widehat{R}_p = \hat{\mathbf{c}}_p \overline{\mathbf{x}}$ is not a good estimate of $\mathbf{c}\mu$. This is because in the expression of $\hat{c}_p$, the eigenvalues of S are working on the p entries of a vector with $\overline{x}$. So, when we compare them one by one and use the norm of the two-vector difference, it is not very big. But when we compute the return, we actually sum the inverse of the eigenvalues of S. So it is natural to get an $\hat{R}_p$ that is much larger than R even though $||\hat{c}_p - c||$.

Figure 1: Empirical and theoretical optimal returns for different numbers of assets



Solid line—the theoretical optimal return (R);
Dashed line—the plug-in return ($\widehat{R}_p$).

In this paper we establish the following theorem to explain the "over-prediction" phenomenon by analyzing the limiting behaviors of $\overline{\mathbf{x}}^T S_n^{-1} \overline{\mathbf{x}}$, $\mathbf{1}^T S_n^{-1} \overline{\mathbf{x}}$, and $\mathbf{1}^T S_n^{-1} \mathbf{1}$:

Theorem 4.1. *Suppose that*

a. $\mathbf{Y}_p = (\mathbf{y}_1, \cdots, \mathbf{y}_n) = (y_{i,j})_{p,n}$ *in which* $y_{i,j}$ $(i = 1, \cdots, p, j = 1, \cdots, n)$ *are i.i.d. random variables with* $Ey_{ij} = 0$, $E|y_{ij}|^2 = 1$, $E|y_{ij}|^4 < \infty$, *and* $\mathbf{x}_k = \Sigma_p^{1/2} \mathbf{y}_k$ *for each n and for* $k = 1, 2, \cdots, n$*;*

b. $\Sigma_p = U_p \Lambda_p U_p^*$ *is nonrandom Hermitian and nonnegative definite with its spectral norm bounded in p where*

$$\Lambda_p = diag(\underbrace{\lambda_1, \cdots, \lambda_1}_{p_1}, \underbrace{\lambda_2, \cdots, \lambda_2}_{p_2}, \underbrace{\cdots}_{\cdots}, \underbrace{\lambda_L, \cdots, \lambda_L}_{p_L}),$$

$\lambda_1 > \lambda_2 > \cdots > \lambda_L$, *and* $U_p = (U_{p_1}, U_{p_2}, \cdots, U_{p_L})$*; and*

c. *for any* $a_p, b_p \in \mathbb{C}^p = \{\mathbf{x} \in \mathbb{C}^p\}$, $\lim_{p\to\infty} \frac{p}{n} = y \in (0, \infty)$, *and* $a_p^T U_{p_i} U_{p_i}^T b_p = d_i$, $i = 1, 2, \cdots, L$.

Then, as $p, n \to \infty$, *we have*

$$a_p{}^T S_n^{-1} b_p \longrightarrow \frac{1}{(1-y)} a_p{}^T \Sigma^{-1} b_p$$

where $S_n = \frac{1}{n} \Sigma^{1/2} X_p X_p^T \Sigma^{1/2}$.

Applying Theorem 4.1, we obtain the following theorem for the plug-in return:

Theorem 4.2. *Under the conditions stated in Theorem 4.1, as $p, n \to \infty$ and $p/n \to y$, the plug-in return $\widehat{R}_p = \hat{\mathbf{c}}_p^T \overline{\mathbf{x}}$ could be expressed as:*

$$\widehat{R}_p \quad \cong \quad \begin{cases} \widehat{R}_p^{(1)} = \sqrt{\frac{\mu^T \Sigma^{-1} \mu}{1-y}} & \text{if } \frac{1}{1-y} \frac{\sigma_0 \mathbf{1}^T \Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} \mu}} < 1 \ \text{(Condition 1)}, \\ \widehat{R}_p^{(2)} = \frac{\mathbf{1}^T \Sigma^{-1} \mu}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} + \tilde{b} \left(\mu^T \Sigma^{-1} \mu - \frac{\mathbf{1}^T \Sigma^{-1} \mu}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \mathbf{1}^T \Sigma^{-1} \mu \right) & \text{if } \frac{1}{1-y} \frac{\sigma_0 \mathbf{1}^T \Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} \mu}} > 1 \ \text{(Condition 2)}, \end{cases}$$

where $\gamma = 1/(1-y)$ and

$$\tilde{b} = \sqrt{\frac{\mathbf{1}^T \Sigma^{-1} \mathbf{1} \sigma_0^2 - \sqrt{1-y}}{\mu^T \Sigma^{-1} \mu \mathbf{1}^T \Sigma^{-1} \mathbf{1} - (\mathbf{1}^T \Sigma^{-1} \mu)^2}} .$$

Obviously $\widehat{R}_p > R$ when $n, p \to \infty$ and $p/n \to y \in (0, 1)$. However, when y is close to zero, $\widehat{R}_p$ is close to the theoretical optimal return. This property is illustrated by Table 5 and Figure 1. There are two problems for the plug-in estimation: one problem is that the conditions of $\widehat{R}_p$ are not the same as those of the theoretical return. Obviously, *Condition 1* in Theorem 4.1 implies that the condition in (2.2) and *Condition 2* in Theorem 4.1 include two situations: the first one is that $1 - y < \frac{\sigma_0 \mathbf{1}^T \Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} \mu}} < 1$ belongs to the condition in (2.2), and $\frac{\sigma_0 \mathbf{1}^T \Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} \mu}} > 1$ belongs to the condition in (2.5). This means that the plug-in estimation may select $\widehat{R}_p^{(1)}$ as the return when (2.5) is correct. The other problem is that $\widehat{R}_p^{(1)}$ is $\sqrt{\gamma}$ times bigger than the real optimal return, while $\widehat{R}_p^{(2)}$ is bigger than but may not be $\sqrt{\gamma}$ times bigger than the theoretical optimal return.

4.2 Bootstrap-Corrected Estimation

To circumvent this limitation, Bai, Liu, and Wong (2009, 2009a) propose a bootstrap technique to circumvent the limitation of the "plug-in" estimators. They use the parametric approach of the bootstrap methodology to avoid possible singularity of the covariance matrix estimation in the bootstrap sample. We describe the details of this procedure as follows: First, a resample $\chi^* = \{\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*\}$ is drawn from the p-variate normal distribution with mean $\overline{\mathbf{x}}$ and covariance matrix S defined in equation (3.1). Then, invoking Markowitz's optimization procedure again on the resample χ^*, we obtain the "bootstrapped plug-in allocation," $\hat{\mathbf{c}}_p^*$, and the "bootstrapped plug-in return," $\hat{R}_p^* = \hat{\mathbf{c}}_p^{*T} \overline{\mathbf{x}}^*$, where $\overline{\mathbf{x}}^* = \sum_1^n \mathbf{x}_k^*/n$. Before we carry on the discussion, we first state the following proposition, which is one of the basic theoretical foundations for Markowitz's optimization estimation:

Proposition 4.1. *Assume that $\mathbf{y}_1, \cdots, \mathbf{y}_n$ are n independent random p-vectors of i.i.d. entries with zero mean and identity variance. Suppose that $\mathbf{x}_k = \boldsymbol{\mu} + \mathbf{z}_k$ with $\mathbf{z}_k = \Sigma^{\frac{1}{2}} \mathbf{y}_k$ where $\boldsymbol{\mu}$ is an unknown p-vector and Σ is an unknown $p \times p$ covariance matrix. Also, we assume that the* *entries of* $\mathbf{y}_k$*'s have finite fourth moments and as* $p, n \to \infty$ *and* $p/n \to y \in (0, 1)$*, we have*

$$\frac{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}{n} \longrightarrow a_1 \,, \quad \frac{\boldsymbol{1}^T\Sigma^{-1}\boldsymbol{1}}{n} \longrightarrow a_2 \,, \quad \text{and} \quad \frac{\boldsymbol{1}^T\Sigma^{-1}\boldsymbol{\mu}}{n} \longrightarrow a_3,$$

satisfying $a_1a_2 - a_3^2 > 0$*. Then, with probability 1, we have*

$$\lim_{n\to\infty} \frac{\widehat{R}_p}{\sqrt{n}} = \begin{cases} \sqrt{\gamma a_1} > \lim\limits_{n\to\infty} \dfrac{R^{(1)}}{\sqrt{n}} = \sqrt{a_1} & \text{when} \quad a_3 < 0, \\ \sigma_0\sqrt{\frac{\gamma(a_1a_2-a_3^2)}{a_2}} > \lim\limits_{n\to\infty} \dfrac{R^{(2)}}{\sqrt{n}} = \sigma_0\sqrt{\dfrac{a_1a_2 - a_3^2}{a_2}} & \text{when} \quad a_3 > 0\,, \end{cases}$$

where $R^{(1)}$ *and* $R^{(2)}$ *are the returns for the two cases given in Proposition 2.1, respectively,* $\gamma = \int_a^b \frac{1}{x}dF_y(x) = \frac{1}{1-y} > 1$, $a = (1 - \sqrt{y})^2$*, and* $b = (1 + \sqrt{y})^2$.

Applying this proposition, one could conclude that when n is large enough, one could obtain $\widehat{R}_p \simeq \sqrt{\gamma}R$. We note that the relation $A_n \simeq B_n$ means that $A_n/B_n \to 1$ in the limiting procedure and we say that A_n and B_n are *proportionally similar* to each other in the sequel. If B_n is a sequence of parameters, we shall say that A_n is *proportionally consistent* with B_n. As the relationship $\widehat{R}_p^* \simeq \sqrt{\gamma}\widehat{R}_p$ is its dual conclusion, one could then obtain the following equation:

$$\sqrt{\gamma}(R - \widehat{R}_p) \simeq \widehat{R}_p - \widehat{R}_p^* \,. \tag{4.2}$$

Applying the bootstrap-corrected approach to equation (4.2), we could construct the estimate

$$\widehat{R}_b = \widehat{R}_p + \frac{1}{\sqrt{\gamma}}(\widehat{R}_p - \widehat{R}_p^*) \tag{4.3}$$

of the optimal return. In addition, rewriting (4.2), we get

$$\sqrt{\gamma}\left(\mathbf{c}^T\mu - \hat{\mathbf{c}}^T\overline{\mathbf{x}}\right) \simeq \hat{\mathbf{c}}_p^T\overline{\mathbf{x}} - \hat{\mathbf{c}}_p^{*T}\overline{\mathbf{x}}^*$$

and obtain the estimate

$$\hat{\mathbf{c}}_b = \hat{\mathbf{c}}_p + \frac{1}{\sqrt{\gamma}}(\hat{\mathbf{c}}_p - \hat{\mathbf{c}}_p^*) \tag{4.4}$$

of the corresponding allocation. For simplicity, we call $\widehat{R}_b$ the "**bootstrap-corrected return**" and $\hat{\mathbf{c}}_b$ the "**bootstrap-corrected allocation**."

The main advantage of the bootstrap-corrected estimation is that its return estimate is consistent with the optimal return, and thus, it circumvents the over-prediction problem of the plug-in return estimate. Hence, one may believe that the bootstrap-corrected estimation is the best estimation for the MV portfolio optimization. Nonetheless, to check how good an estimate of MV portfolio optimization is, one should not only care about how good the estimation of the return is, but also about how good the estimation of the corresponding allocation is and how big their risk is.[1] According to our simulation in Section 5, we find that the bootstrap-corrected estimation does not even outperform the plug-in estimation in both allocation and risk and sometimes it could be even worse. We call the former the "allocation estimation" problem and the latter the "risk" problem. Moreover, our simulation, we find that, yes, the bootstrap-corrected estimation does overcome the "over-prediction" problem but it incurs an "under-prediction" problem. The "under-prediction" is not too serious when the dimension to sample size ratio ($y = p/n$) is not large but it becomes very serious when y is large. Thus, the bootstrap-corrected estimation is not the best MV portfolio optimization. Thus, looking for the best MV portfolio optimization estimation that could solve all of the defects in the MV portfolio optimization – the "over-prediction," "under-prediction," "allocation estimation," and "risk" problems – is still a very important outstanding problem. It is our objective in this paper to obtain an estimation that circumvents all four defects.

4.3 Spectral-Corrected Estimators

In this section, we will first discuss how to estimate the eigenvalues of the population covariance matrix, and thereafter, we will develop the theory of the spectral-corrected estimators, which will circumvent all the four defects—the over-prediction phenomenon, the under-prediction problem, the allocation estimation problem, and the problem of big risk. We will discuss the details in the following subsections.

4.3.1 Estimation of the eigenvalues of the population covariance and the population covariance matrix

Letting $(s_j)_{1\le j\le p}$ be the p eigenvalues of the population covariance matrix Σ, we consider the spectral distribution (S.D.) H of Σ such that

$$H(x) = \frac{1}{p}\sum_{j=1}^{p} \delta_{s_j}(x) , \tag{4.5}$$

in which δ_b is the Dirac point measure at b. It is obvious that the estimation of the eigenvalues of Σ could be converted to the estimation of the S.D. of H as shown in (4.5).

Bai, Chen, and Yao (2010) provide a method to estimate the S.D. of H, when the population spectrum is of finite support. They prove that their proposed estimate is consistent and asymptotically Gaussian when the size k of the limiting support is fixed and known. In addition, when the order k of the model is unknown, they incorporate a cross-validation procedure in their estimation method to select the unknown model dimension. They also construct the moment

[1] Readers may refer to equation (5.3) for the definition of risk.

relationship between the limits of ESD and the **population spectral distribution** (PSD), and then develop the moment estimation. In addition, by using the equations of the limiting spectral distribution of the sample covariance matrix and by adopting the Stietjes transform tools, Li, Chen, Qin, Yao, Bai (2013) develop a series of new techniques to provide consistent estimation for the population spectrum distribution. We state the steps to estimate H, the eigenvalues of the population covariance matrix, as follows:

Step 1: Set $B = \frac{1}{n}XX^T$;

Step 2: compute eigenvalues of matrix B, denoted as $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$;

Step 3: put B in formula (3.6) to obtain

$$\underline{m}(u) = -\frac{1-y}{u} + y\int \frac{1}{x-u}dF^B(x),\ \forall u \in A \equiv (-\infty, \lambda_1) \cup (\lambda_p, +\infty) \setminus \{0\};$$

Step 4: given $\{u_1, u_2, \cdots, u_I\} \subset A$, we get $\{\underline{m}_1, \cdots, \underline{m}_I\} = \{\underline{m}(u_1), \cdots, \underline{m}(u_I)\}$; and

Step 5: compute $\widehat{H}$ such that

$$\widehat{H} = \arg\min_H \sum_{i=1}^{I} \left(u(\underline{m}_i, H) - u_i\right)^2. \tag{4.6}$$

Then, the S.D. H of Σ can be estimated by $\widehat{H}$ as shown in (4.6).

From the estimation of the S.D. H of Σ in the above steps, we obtain the eigenvalue estimators $\hat{a}_1 \ge \hat{a}_2 \ge \cdots \ge \hat{a}_p$. According to the spectral theory, we have

$$S = V\widetilde{\Lambda}V^T, \tag{4.7}$$

where $\widetilde{\Lambda} = \mathrm{diag}(\tilde{\lambda}_1, \cdots, \tilde{\lambda}_p)$ with $\tilde{\lambda}_1 \ge \tilde{\lambda}_2 \ge \cdots \ge \tilde{\lambda}_p$ and the column vectors of V are the orthogonal eigenvectors of S with respect to $\tilde{\lambda}_1, \cdots, \tilde{\lambda}_p$. Suppose that $\widehat{\Lambda} = \mathrm{diag}\{\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p\}$ in which $\hat{a}_1 \ge \hat{a}_2 \ge \cdots \ge \hat{a}_p$ are the estimations of the eigenvalues for matrix Σ; we put $\widehat{\Lambda}$ in equation (4.7) and obtain the **spectral-corrected covariance**

$$\widehat{\Sigma}_s = V\widehat{\Lambda}V^T. \tag{4.8}$$

The spectral-corrected covariance in (4.8) could be used in the development of the "best" optimal estimation. We will discuss the issue in the following subsections.

4.3.2 Estimation of the optimal return and allocation

After estimating the spectral-corrected covariance $\widehat{\Sigma}_s$ from (4.8) and from the steps discussed in Section 4.3.1, one could plug the sample mean vector $\overline{\mathbf{x}}$ and the spectral-corrected covariance $\widehat{\Sigma}_s$ into the formulae of the asset allocation $\mathbf{c}$ in Proposition 2.1 to obtain

$$\hat{\mathbf{c}}_s = \begin{cases} \dfrac{\sigma_0 \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}} & \text{if } \dfrac{\sigma_0 \mathbf{1}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}} < 1, \\ \dfrac{\widehat{\Sigma}_s^{-1}\mathbf{1}}{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\mathbf{1}} + \hat{b}_s\left(\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}} - \dfrac{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\mathbf{1}}\widehat{\Sigma}_s^{-1}\mathbf{1}\right) & \text{if } \dfrac{\sigma_0 \mathbf{1}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}} > 1, \end{cases} \tag{4.9}$$

where

$$\hat{b}_s = \sqrt{\frac{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\mathbf{1}\sigma_0^2 - 1}{\overline{\mathbf{x}}^T\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}\mathbf{1}^T\widehat{\Sigma}_s^{-1}\mathbf{1} - (\mathbf{1}^T\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}})^2}}\,.$$

Since the estimator $\widehat{\Sigma}_s$ is obtained by estimating the eigenvalues of the population covariance, we call $\hat{\mathbf{c}}_s$ the **spectral-corrected allocation**. The corresponding return can be estimated by

$$\hat{R}_s = \hat{\mathbf{c}}_s^T\overline{\mathbf{x}}$$

which we call the **spectral-corrected return**. It can also be expressed as

$$\widehat{\mathbf{R}}_s = \begin{cases} \sigma_0\sqrt{\overline{\mathbf{x}}^T\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}} & \text{if } \dfrac{\sigma_0 \mathbf{1}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}} < 1\,, \\ \dfrac{\overline{\mathbf{x}}'\widehat{\Sigma}_s^{-1}\mathbf{1}}{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\mathbf{1}} + \hat{b}_s\left(\overline{\mathbf{x}}'\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}} - \dfrac{\left(\mathbf{1}^T\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}\right)^2}{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\mathbf{1}}\right) & \text{if } \dfrac{\sigma_0 \mathbf{1}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}} > 1\,. \end{cases} \tag{4.10}$$

In addition, the risk of the spectral-corrected allocation can be defined as

$$\begin{aligned} Risk_{\mathbf{c}}^s &= \hat{\mathbf{c}}_s^T\Sigma\hat{\mathbf{c}}_s \\ &= \begin{cases} \dfrac{\sigma_0^2\overline{\mathbf{x}}'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\overline{\mathbf{x}}^T\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}} & \text{if } \dfrac{\sigma_0 \mathbf{1}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}} < 1\,, \\ \left[\mathbb{A}^T + \hat{b}_s\left(\mathbb{B}^T + \mathbb{C}^T\right)\right]\Sigma\left[\mathbb{A} + \hat{b}_s\left(\mathbb{B} + \mathbb{C}\right)\right] & \text{if } \dfrac{\sigma_0 \mathbf{1}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}} > 1\,, \end{cases} \end{aligned} \tag{4.11}$$

which we call**spectral-corrected risk**. Here $\mathbb{A} = \frac{\widehat{\Sigma}_s^{-1}\mathbf{1}}{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\mathbf{1}}$, $\mathbb{B} = \widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}$ and $\mathbb{C} = \frac{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\mathbf{1}^T\widehat{\Sigma}_s^{-1}\mathbf{1}}\widehat{\Sigma}_s^{-1}\mathbf{1}$.

4.3.3 The limiting behavior of the spectral-corrected return

In the previous two subsections, we developed the theory for the construction of the spectral-corrected estimation. Now, we turn to comparing the performance of the spectral-corrected estimation with that of the plug-in and bootstrap-corrected estimations. Does the spectral-corrected return get closer to the theoretical optimal return? Does the spectral-corrected allocation also get closer to the theoretical optimal allocation? Is the spectral-corrected risk smaller and bounded by an acceptable level? In this subsection and the next subsection we will explore the answers of the above questions.

We start our discussion with $\hat{R}_s$. From equation (4.10), we know that $\bar{\mathbf{x}}'\widehat{\Sigma}_s^{-1}\bar{\mathbf{x}}$, $\mathbf{1}'\widehat{\Sigma}_s^{-1}\bar{\mathbf{x}}$, and $\mathbf{1}'\widehat{\Sigma}_s^{-1}\mathbf{1}$ are the main components in the formula of the spectral-corrected return. Thus, we only need to study the limit of ${a_p}^T\widehat{\Sigma}_s^{-1}b_p$ that enables us to get the limits of the above-mentioned items under some regularity conditions. This is because both a_p and b_p could be $\bar{\mathbf{x}}/||\bar{\mathbf{x}}||$ and $\mathbf{1}/\sqrt{p}$, and thus, studying the limit of ${a_p}^T\widehat{\Sigma}_s^{-1}b_p$ is as good as studying the limits of $\frac{\bar{\mathbf{x}}'}{||\bar{\mathbf{x}}||}\widehat{\Sigma}_s^{-1}\frac{\bar{\mathbf{x}}}{||\bar{\mathbf{x}}||}$, $\frac{\mathbf{1}'}{\sqrt{p}}\widehat{\Sigma}_s^{-1}\frac{\bar{\mathbf{x}}'}{||\bar{\mathbf{x}}||}$, and $\frac{\mathbf{1}'}{\sqrt{p}}\widehat{\Sigma}_s^{-1}\frac{\mathbf{1}'}{\sqrt{p}}$. To do so, we first establish the following theorem:

Theorem 4.3. *If*

a. $\mathbf{Y}_p = (\mathbf{y}_1, \cdots, \mathbf{y}_n) = (y_{i,j})_{p,n}$ *in which* $y_{i,j}$ $(i = 1, \cdots, p, j = 1, \cdots, n)$ *are i.i.d. random variables with* $Ey_{ij} = 0$, $E|y_{ij}|^2 = 1$, $E|y_{ij}|^4 < \infty$, *and* $\mathbf{x}_k = \Sigma_p^{1/2}\mathbf{y}_k$ *for each n and for* $k = 1, \cdots, n$*;*

b. $\Sigma_p = U_p\Lambda_pU_p^T$ *is nonrandom Hermitian and nonnegative definite with its spectral norm bounded in p where*

$$\Lambda_p = diag(\underbrace{\lambda_1, \cdots, \lambda_1}_{p_1,}, \underbrace{\lambda_2, \cdots, \lambda_2}_{p_2,}, \underbrace{\cdots,}_{\cdots,} \underbrace{\lambda_L, \cdots, \lambda_L}_{p_L}), \tag{4.12}$$

$\lambda_1 > \lambda_2 > \cdots > \lambda_L$, $U_p = (U_{p_1}, U_{p_2}, \cdots, U_{p_L})$, *and* $\lim_{p_i\to\infty}\frac{p_i}{n} = y_i \in (0, \infty)$*; and*

c. *for the sample covariance matrix* $S_n = V_p\widetilde{\Lambda}V_p^T$ *expressed in the form as shown in equation (3.1), the limiting spectral distribution is spectral separated,*

then, for any pair of vector sequences $\{a_p\}, \{b_p\} \in \mathbb{C}^p$ *satisfying* $a_p^TU_{p_i}U_{p_i}^Tb_p = d_i$ $(i = 1, 2, \cdots, L)$, *we have*

$${a_p}^TB_p^{-1}b_p \longrightarrow \sum_{k=1}^{L}\frac{d_k}{\lambda_k}\sum_{j=1}^{L}\frac{\lambda_k(u_j-\lambda_j)}{\lambda_j(u_j-\lambda_k)} \doteq \varsigma_{a_p,b_p} \ \ a.s. \ ,$$

as $p, n \to \infty$ *and* $p/n \to y$, *where* $B_p = V_p\Lambda_pV_p^T$, u_j *is the solution of* $1 + y\int\frac{t}{u-t}dH(t) = 0$ *for any* $j = 1, \cdots, L$ *with* $\lambda_1 > u_1 > \lambda_2 > \cdots > \lambda_L > u_L > 0$.

Applying both Theorem 4.3 and the consistent properties of the spectral estimation (Li, Chen, Qin, Yao, Bai (2013)), we obtain the following theorem:

Theorem 4.4. *Under the conditions stated in Theorem 4.3, as* $n, p \to \infty$ *and* $p/n \to y$, *we have*

$${a_p}^T\widehat{\Sigma}_s^{-1}b_p \longrightarrow \sum_{k=1}^{L}\frac{d_k}{\lambda_k}\sum_{j=1}^{L}\frac{\lambda_k(u_j-\lambda_j)}{\lambda_j(u_j-\lambda_k)} \doteq \varsigma_{a_p,b_p} \ \ a.s. \tag{4.13}$$

where ς_{a_p,b_p} *is the limit of* ${a_p}^T\widehat{\Sigma}_s^{-1}b_p$.

We note that ς_{a_p,b_p} is a function of d_i, λ_i, and u_i ($i = 1, \cdots, L$) in which d_i, λ_i, and u_i ($i = 1, \cdots, L$) are given in the conditions of Theorem 4.3. For ς_{a_p,b_p}, it is interesting to find the following result:

$$\left|{a_p}^T\Sigma^{-1}b_p\right| < \left|\varsigma_{a_p,b_p}\right| < \left|\frac{{a_p}^T\Sigma^{-1}b_p}{1-y}\right| \tag{4.14}$$

for any pair of unit vectors a_p and b_p, in which $\frac{{a_p}^T\Sigma^{-1}b_p}{1-y}$ is the limit of ${a_p}^T S_n^{-1} b_p$ as $p, n \to \infty$ and $p/n \to y$ according to Theorem 4.1. In this paper we will evaluate the performance of the spectral-corrected method by simulation and exhibit the simulation results in Tables 1 and 2. These tables report the values of ${a_p}^T S_n^{-1} b_p$, ${a_p}^T \widehat{\Sigma}_s^{-1} b_p$, and ${a_p}^T\Sigma^{-1}b_p$ for a pair of random bounded vectors a_p and b_p. From these tables, we notice that

$$\left|{a_p}^T\Sigma^{-1}b_p\right| < \left|{a_p}^T\widehat{\Sigma}_s^{-1}b_p\right| < \left|{a_p}^T S_n^{-1} b_p\right|. \tag{4.15}$$

We also note that the limits of the middle and right terms in equation (4.14) are the corresponding terms in equation (4.15), because $\left|{a_p}^T\widehat{\Sigma}_s^{-1}b_p\right| \to |\varsigma_{a_p,b_p}|$ and $\left|{a_p}^T S_n^{-1} b_p\right| \to \frac{\left|{a_p}^T\Sigma^{-1}b_p\right|}{(1-y)}$ as $p, n \to \infty$ and $p/n \to y$.

When we compare the standard deviations (s.d.'s) of the terms in (4.14), we find that ${a_p}^T\widehat{\Sigma}_s^{-1}b_p$ is much stabler than ${a_p}^T S_n^{-1} b_p$ for any y. When y increases from 0.1 to 0.9, the performance of both ${a_p}^T\widehat{\Sigma}_s^{-1}b_p$ and ${a_p}^T S_n^{-1} b_p$ gets worse, but the performance of ${a_p}^T\widehat{\Sigma}_s^{-1}b_p$ improves greatly by comparison with ${a_p}^T S_n^{-1} b_p$, not only because the mean of the former is closer to the theoretical value, but also the s.d. of the former is smaller and the estimation is more stable than that of S. In addition, our simulation shows that the inequalities in (4.14) hold. Thus, we recommend that academics and practitioners use the spectral-corrected estimation in their analysis. To obtain further analysis, we first establish the following theorem:

Theorem 4.5. *Under the conditions stated in Theorem 4.3, if* $\left(\frac{\mathbf{1}}{\sqrt{p}}, \frac{\mathbf{1}}{\sqrt{p}}\right)$, $\left(\frac{\mathbf{1}}{\sqrt{p}}, \frac{\mu}{||\mu||}\right)$, *and* $\left(\frac{\mu}{||\mu||}, \frac{\mu}{||\mu||}\right)$ *belong to*

$$\left\{(\upsilon_1, \upsilon_2) : \upsilon_1^T U_{p_i} U_{p_i}^T \upsilon_2 = d_i \in \mathbb{R}, i = 1, \cdots, L, \max\{||\upsilon_1||, ||\upsilon_2||\} \le M(>0)\right\},$$

$\sigma_0 = \xi_{\sigma_0}/\sqrt{p}$, $||\mu||/\sqrt{p} = \xi_\mu + o(1)$, *then, as* $p, n \to \infty$ *and* $p/n \to y$, *we have*

a.

$$\frac{\mathbf{1}'\widehat{\Sigma}_s^{-1}\mathbf{1}}{p} \longrightarrow \varsigma_{\mathbf{1},\mathbf{1}}\,, \quad \frac{\mathbf{1}'\widehat{\Sigma}_s^{-1}\mu}{\sqrt{p}||\mu||} \longrightarrow \varsigma_{\mathbf{1},\mu}\,, \quad \textit{and} \quad \frac{\mu'\widehat{\Sigma}_s^{-1}\mu}{||\mu||^2} \longrightarrow \varsigma_{\mu,\mu}\,, \tag{4.16}$$

b.

$$\frac{\mathbf{1}'\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{\sqrt{p}||\overline{\mathbf{x}}||} \longrightarrow \varsigma_{\mathbf{1},\mu} \quad \textit{and} \quad \frac{\overline{\mathbf{x}}'\widehat{\Sigma}_s^{-1}\overline{\mathbf{x}}}{||\overline{\mathbf{x}}||^2} \longrightarrow \varsigma_{\mu,\mu}\,, \tag{4.17}$$

where $\varsigma_{\mathbf{1},\mathbf{1}}$, $\varsigma_{\mathbf{1},\mu}$, *and* $\varsigma_{\mu,\mu}$ *are defined in (4.13).*

Now, we turn to analyzing the limit of the spectral-corrected return $\widehat{R}_s$ defined in (4.10). Suppose $\sigma_0 = \xi_{\sigma_0}/\sqrt{p}$. As $p, n \to \infty$ and $p/n \to y$, we first obtain the limit of the condition stated in (4.10) as follows:

$$\frac{\sigma_0 \mathbf{1}^T \widehat{\Sigma}_s^{-1} \overline{\mathbf{x}}}{\sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1} \overline{\mathbf{x}}}} = \frac{\xi_{\sigma_0} \frac{\mathbf{1}^T}{\sqrt{p}} \widehat{\Sigma}_s^{-1} \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|}}{\sqrt{\frac{\overline{\mathbf{x}}^T}{\|\overline{\mathbf{x}}\|} \widehat{\Sigma}_s^{-1} \frac{\overline{\mathbf{x}}^T}{\|\overline{\mathbf{x}}\|}}} \longrightarrow \xi_{\sigma_0} \frac{\varsigma_{\mathbf{1},\mu}}{\varsigma_{\mu,\mu}} \ .$$

For the spectral-corrected return stated in (4.10), the first value of $\widehat{R}_s$ possesses the following limit property:

$$\sigma_0 \sqrt{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1} \overline{\mathbf{x}}} = \xi_{\sigma_0} \sqrt{\frac{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1} \overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|^2} \cdot \frac{\|\overline{\mathbf{x}}\|^2}{p}} \longrightarrow \xi_{\sigma_0} \xi_\mu \sqrt{\varsigma_{\mu,\mu}} \ \text{ as } \ p, n \to \infty \text{ and } p/n \to y.$$

The second value of $\widehat{R}_s$ in (4.10) becomes

$$\begin{aligned}
\widehat{R}_s &= \frac{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1} \mathbf{1}}{\mathbf{1}^T \widehat{\Sigma}_s^{-1} \mathbf{1}} + \hat{b}_s \left(\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1} \overline{\mathbf{x}} - \frac{\left(\mathbf{1}^T \widehat{\Sigma}_s^{-1} \overline{\mathbf{x}}\right)^2}{\mathbf{1}^T \widehat{\Sigma}_s^{-1} \mathbf{1}} \right) \\
&= \frac{\|\overline{\mathbf{x}}\|}{\sqrt{p}} \frac{\frac{\overline{\mathbf{x}}^T}{\|\overline{\mathbf{x}}\|} \widehat{\Sigma}_s^{-1} \frac{\mathbf{1}}{\sqrt{p}}}{\frac{\mathbf{1}}{\sqrt{p}}^T \widehat{\Sigma}_s^{-1} \frac{\mathbf{1}}{\sqrt{p}}} + \hat{b}_s \|\overline{\mathbf{x}}\|^2 \left(\frac{\overline{\mathbf{x}}^T}{\|\overline{\mathbf{x}}\|} \widehat{\Sigma}_s^{-1} \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|} - \frac{\left(\frac{\mathbf{1}}{\sqrt{p}}^T \widehat{\Sigma}_s^{-1} \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|} \right)^2}{\frac{\mathbf{1}}{\sqrt{p}}^T \widehat{\Sigma}_s^{-1} \frac{\mathbf{1}}{\sqrt{p}}} \right).
\end{aligned}$$

Here, as $p, n \to \infty$ and $p/n \to y$, we have

$$\begin{aligned}
\hat{b}_s \|\overline{\mathbf{x}}\|^2 &= \|\overline{\mathbf{x}}\|^2 \sqrt{\frac{\mathbf{1}^T \widehat{\Sigma}_s^{-1} \mathbf{1} \sigma_0^2 - 1}{\overline{\mathbf{x}}^T \widehat{\Sigma}_s^{-1} \overline{\mathbf{x}} \mathbf{1}^T \widehat{\Sigma}_s^{-1} \mathbf{1} - (\mathbf{1}^T \widehat{\Sigma}_s^{-1} \overline{\mathbf{x}})^2}} \\
&= \frac{\|\overline{\mathbf{x}}\|}{\sqrt{p}} \sqrt{\frac{\frac{\mathbf{1}}{\sqrt{p}}^T \widehat{\Sigma}_s^{-1} \frac{\mathbf{1}}{\sqrt{p}} \xi_{\sigma_0} - 1}{\frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|}^T \widehat{\Sigma}_s^{-1} \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|} \frac{\mathbf{1}}{\sqrt{p}}^T \widehat{\Sigma}_s^{-1} \frac{\mathbf{1}}{\sqrt{p}} - (\frac{\mathbf{1}}{\sqrt{p}}^T \widehat{\Sigma}_s^{-1} \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|})^2}} \\
&\longrightarrow \xi_\mu \sqrt{\frac{\varsigma_{\mathbf{1},\mathbf{1}} \xi_{\sigma_0} - 1}{\varsigma_{\mu,\mu} \varsigma_{\mathbf{1},\mathbf{1}} - (\varsigma_{\mathbf{1},\mu})^2}} \ .
\end{aligned}$$

Thus, as $p, n \to \infty$ and $p/n \to y$, we obtain

$$\widehat{R}_s \longrightarrow \xi_\mu \frac{\varsigma_{\mathbf{1},\mu}}{\varsigma_{\mathbf{1},\mathbf{1}}} + \xi_\mu \sqrt{\frac{\varsigma_{\mathbf{1},\mathbf{1}} \xi_{\sigma_0} - 1}{\varsigma_{\mu,\mu} \varsigma_{\mathbf{1},\mathbf{1}} - (\varsigma_{\mathbf{1},\mu})^2}} \left(\varsigma_{\mu,\mu} - \frac{(\varsigma_{\mathbf{1},\mu})^2}{\varsigma_{\mathbf{1},\mathbf{1}}} \right) \ . \tag{4.18}$$

According to the above analysis, we obtain the following theorem:

Theorem 4.6. *Under the conditions and definitions stated in Theorem 4.5, as $n, p \to \infty$ and $p/n \to y$, we have*

$$\widehat{R}_s \longrightarrow \begin{cases} \xi_{\sigma_0} \xi_\mu \sqrt{\varsigma_{\mu,\mu}} & \text{if} \quad \xi_{\sigma_0} \varsigma_{\mathbf{1},\mu} / \varsigma_{\mu,\mu} < 1 \ , \\ \xi_\mu \frac{\varsigma_{\mathbf{1},\mu}}{\varsigma_{\mathbf{1},\mathbf{1}}} + \xi_\mu \sqrt{\frac{\varsigma_{\mathbf{1},\mathbf{1}} \xi_{\sigma_0} - 1}{\varsigma_{\mu,\mu} \varsigma_{\mathbf{1},\mathbf{1}} - (\varsigma_{\mathbf{1},\mu})^2}} \left(\varsigma_{\mu,\mu} - \frac{(\varsigma_{\mathbf{1},\mu})^2}{\varsigma_{\mathbf{1},\mathbf{1}}} \right) & \text{if} \quad \xi_{\sigma_0} \varsigma_{\mathbf{1},\mu} / \varsigma_{\mu,\mu} > 1 \ . \end{cases}$$

In this paper we hypothesize the conjecture that $\widehat{R}_s$ is *proportionally consistent* with the theoretical optimal return R defined in (2.3) or (2.6) under some regularity conditions. The results in Theorem 4.6 help us to check this conjecture. To complete the work, we establish the limit of the theoretical optimal return as shown in the following theorem:

Theorem 4.7. *Under the conditions of Theorem 4.5, as $p, n \to \infty$ and $p/n \to y$, we have*

a. *the limits of*

$$\frac{\mathbf{1}'\Sigma^{-1}\mathbf{1}}{p}, \quad \frac{\mathbf{1}'\Sigma^{-1}\mu}{\sqrt{p}||\mu||}, \quad \text{and} \quad \frac{\mu'\Sigma^{-1}\mu}{||\mu||^2}$$

exist, and

b. *the theoretical optimal return R satisfies*

$$R \longrightarrow \begin{cases} \xi_{\sigma_0}\xi_\mu \sqrt{\varsigma^0_{\mu,\mu}} & \text{if} \quad \xi_{\sigma_0}\varsigma^0_{\mathbf{1},\mu}/\varsigma^0_{\mu,\mu} < 1, \\ \xi_\mu \frac{\varsigma^0_{\mathbf{1},\mu}}{\varsigma^0_{\mathbf{1},\mathbf{1}}} + \xi_\mu \sqrt{\frac{\varsigma^0_{\mathbf{1},\mathbf{1}}\xi_{\sigma_0}-1}{\varsigma^0_{\mu,\mu}\varsigma^0_{\mathbf{1},\mathbf{1}}-(\varsigma_{\mathbf{1},\mu})^2}\left(\varsigma^0_{\mu,\mu} - \frac{(\varsigma^0_{\mathbf{1},\mu})^2}{\varsigma^0_{\mathbf{1},\mathbf{1}}}\right)} & \text{if} \quad \xi_{\sigma_0}\varsigma^0_{\mathbf{1},\mu}/\varsigma^0_{\mu,\mu} > 1, \end{cases}$$

where $\varsigma^0_{\mathbf{1},\mathbf{1}}$, $\varsigma^0_{\mathbf{1},\mu}$, and $\varsigma^0_{\mu,\mu}$ are the corresponding limits in (a).

From Table 1, we find that $(\varsigma_{\mathbf{1},\mathbf{1}}, \varsigma_{\mathbf{1},\mu}, \varsigma_{\mu,\mu})$ is very close to $(\varsigma^0_{\mathbf{1},\mathbf{1}}, \varsigma^0_{\mathbf{1},\mu}, \varsigma^0_{\mu,\mu})$. Thus, Theorems 4.6 and 4.7 and our simulation results support the conjecture that $\widehat{R}_s$ is *proportionally consistent* with the theoretical optimal return R under some regularity conditions.

4.3.4 The limiting behavior of the spectral-corrected risk

In this paper, we also hypothesize the conjecture that the spectral-corrected risk $Risk^s_{\mathbf{c}}$ (defined in equation (4.11)) is close to the *Risk* of the theoretical optimal return under some regularity conditions. To examine this conjecture, in this section we will study the limiting behavior of the spectral-corrected risk. To do so, from (4.11), we only need to examine the limiting behavior of ${a_p}^T\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b_p$ as stated in the following theorem:

Theorem 4.8. *Suppose that the projections on each U_j $(j = 1, \cdots, L)$ subspace of vectors a_p and b_p only have finite nonzero entries. Then, under the same conditions of Theorem 4.3, we have*

$$ {a_p}^T B_p^{-1}\Sigma B_p^{-1} b_p \longrightarrow \sum_{k=1}^{L} \frac{d_k}{\lambda_k}\left(\sum_{j=1}^{L}\frac{\lambda_k(u_j-\lambda_j)}{\lambda_j(u_j-\lambda_k)}\right)^2 \doteq \varrho_{a_p,b_p} \quad a.s. \tag{4.19}$$

From Theorem 4.8, we notice that ϱ_{a_p,b_p} depends only on the information of d_k, λ_k, and u_k $(k = 1, \cdots, L)$ about the population. Since it is difficult to obtain the theoretical result for the comparison of ${a_p}^T S_n^{-1}\Sigma S_n^{-1} b_p$, ${a_p}^T\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b_p$, and ${a_p}^T\Sigma^{-1}b_p$ for each pair of the uniform bounded vector a_p, b_p, in this paper we conduct a simulation for the comparison and report the results in Tables 3 and 4. Table 3 shows that, compared with ${a_p}^T S_n^{-1} \Sigma S_n^{-1} b_p$ and ${a_p}^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} b_p$, the limit of ${a_p}^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} b_p$ is much closer to the real value ${a_p}^T \Sigma^{-1} b_p$ for any y. From the results in Table 4, one could easily observe that ${a_p}^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} b_p$ converges. Thus, we establish the following theorem for the spectral-corrected risk $Risk_{\mathbf{c}}^s$:

Theorem 4.9. *Under the conditions of Theorem 4.5, if $p, n \to \infty$ and $p/n \to y$, then*

a. *the limits of*

$$\frac{\mathbf{1}^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} \mathbf{1}}{p}, \frac{\mathbf{1}^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} \mu}{\sqrt{p}||\mu||}, \quad and \quad \frac{\mu^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} \mu}{||\mu||^2},$$

exist and they are denoted by $\varrho_{\mathbf{1},\mathbf{1}}$, $\varrho_{\mathbf{1},\mu}$ and $\varrho_{\mu,\mu}$, and

b.

$$\frac{\mathbf{1}^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} \mathbf{1}}{p} \longrightarrow \varrho_{\mathbf{1},\mathbf{1}}, \frac{\mathbf{1}^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} \overline{X}}{\sqrt{p}||\overline{X}||} \longrightarrow \varrho_{\mathbf{1},\mu}, \frac{\overline{X}^T \widehat{\Sigma}_s^{-1} \Sigma \widehat{\Sigma}_s^{-1} \overline{X}}{||\overline{X}||^2} \longrightarrow \varrho_{\mu,\mu}.$$

In addition, we have

c. *when $\xi_{\sigma_0} \varsigma_{\mathbf{1},\mu} / \varsigma_{\mu,\mu} < 1$,*

$$p \cdot Risk_{\mathbf{c}}^s \to \xi_{\sigma_0} \frac{\varrho_{\mu,\mu}}{\varsigma_{\mu,\mu}} \quad a.s., \quad and$$

d. *when $\xi_{\sigma_0} \varsigma_{\mathbf{1},\mu} / \varsigma_{\mu,\mu} > 1$, $p \cdot Risk_{\mathbf{c}}^s$ almost surely converges to*

$$\frac{\varrho_{\mathbf{1},\mathbf{1}}}{\varsigma_{\mathbf{1},\mathbf{1}}} + \frac{\varrho_{\mathbf{1},\mathbf{1}}}{\xi_\mu \varsigma_{\mathbf{1},\mathbf{1}}} \sqrt{\frac{\varsigma_{\mathbf{1},\mathbf{1}} \xi_{\sigma_0} - 1}{\varsigma_{\mu,\mu} \varsigma_{\mathbf{1},\mathbf{1}} - (\varsigma_{\mathbf{1},\mu})^2}} + \sqrt{\frac{\varsigma_{\mathbf{1},\mathbf{1}} \xi_{\sigma_0} - 1}{\varsigma_{\mu,\mu} \varsigma_{\mathbf{1},\mathbf{1}} - (\varsigma_{\mathbf{1},\mu})^2}} \left(\varrho_{\mu,\mu} - 2 \frac{\varsigma_{\mathbf{1},\mu} \varrho_{\mathbf{1},\mu}}{\varsigma_{\mathbf{1},\mathbf{1}}} + \left(\frac{\varsigma_{\mathbf{1},\mu}}{\varsigma_{\mathbf{1},\mathbf{1}}} \right)^2 \varrho_{\mathbf{1},\mathbf{1}} \right).$$

We note that in Theorem 4.9, if we suppose that $\sigma_0 = \frac{\xi_{\sigma_0}}{\sqrt{p}}$, then we have $p \cdot Risk \to \xi_{\sigma_0}$ as $p, n \to \infty$ and $p/n \to y$. We also note that the limit of $p \cdot Risk_{\mathbf{c}}^s$ is not equal to that of $p \cdot Risk$. However, it is closer to that of $p \cdot Risk$ than the other two risks.

In addition, from Table 4, we observe that $(\varrho_{\mathbf{1},\mathbf{1}}, \varrho_{\mathbf{1},\mu}, \varrho_{\mu,\mu})$ is very close to $(\varsigma_{\mathbf{1},\mathbf{1}}^0, \varsigma_{\mathbf{1},\mu}^0, \varsigma_{\mu,\mu}^0)$. Thus, $Risk_{\mathbf{c}}^s$ is close to the theoretical risk. Theorems 4.6, 4.7, and 4.9 and our simulation results support our conjecture that $Risk_{\mathbf{c}}^s$ is close to the $Risk$ of the theoretical optimal return under some regularity conditions.

5 Simulation Study

In this section, we will conduct simulation to compare (1) how good the performance of the spectral-corrected return $\widehat{R}_s$ is in comparison with that of the plug-in return $\widehat{R}_p$ and bootstrap-corrected return $\widehat{R}_b$, (2) how good the performance of the spectral-corrected allocation $\hat{\mathbf{c}}_s$ in comparison with that of the plug-in allocation $\hat{\mathbf{c}}_p$ and bootstrap-corrected allocation $\hat{\mathbf{c}}_b$, and (3) what the risks of the plug-in return $\widehat{R}_p$, bootstrap-corrected return $\widehat{R}_b$, and spectral-corrected return $\widehat{R}_s$, and among them, which one is smallest.

In order to check how good the performance of the spectral-corrected return $\widehat{R}_s$ is in comparison with that of the plug-in return $\widehat{R}_p$ and bootstrap-corrected return $\widehat{R}_b$, we define

$$d_R^{\omega} = R_{\omega} - R \quad \text{with} \quad \omega = p, b, s \tag{5.1}$$

in which we call d_R^s the **spectral-corrected difference** for the return, which is the difference between the spectral-corrected optimal return estimate $\hat{R}_s$ and the theoretic optimal return R. The **plug-in difference** d_R^p and **bootstrap-corrected difference** d_R^b for the return are defined similarly as stated in (5.1).

To check how good the performance of the spectral-corrected allocation $\hat{\mathbf{c}}_s$ is in comparison with that of the plug-in allocation $\hat{\mathbf{c}}_p$ and bootstrap-corrected allocation $\hat{\mathbf{c}}_b$, we define

$$d_{\mathbf{c}}^{\omega} = ||\hat{\mathbf{c}}_{\omega} - \mathbf{c}|| \quad \text{with} \quad \omega = p, b, s \tag{5.2}$$

in which we call $d_{\mathbf{c}}^s$ the **spectral-corrected normed difference** for the allocation, which is the normed difference between the spectral-corrected optimal allocation estimate $\hat{\mathbf{c}}_s$ and the theoretic optimal allocation $\mathbf{c}$. The **plug-in normed difference** $d_{\mathbf{c}}^p$ and the **bootstrap-corrected normed difference** $d_{\mathbf{c}}^b$ are defined similarly as stated in (5.2).

Among the risks of the plug-in return $\widehat{R}_p$, bootstrap-corrected return, and spectral-corrected return $\widehat{R}_s$, to check which one is the smallest, we define

$$Risk_{\mathbf{c}}^{\omega} = \hat{\mathbf{c}}_{\omega}' \Sigma \hat{\mathbf{c}}_{\omega}, \quad \text{with} \quad \omega = p, b, s \tag{5.3}$$

in which we call $Risk_{\mathbf{c}}^b$, $Risk_{\mathbf{c}}^p$, and $Risk_{\mathbf{c}}^s$ the **plug-in risk**, **bootstrap-corrected risk**, and **spectral-corrected risk**, respectively. We will also compare $d_{\mathbf{c}}^{\omega}$, d_R^{ω}, and $risk_{\mathbf{c}}^{\omega}$ for $\omega = p, b, s$ with those for the theoretical optimal return R. They are d_R^R, $d_{\mathbf{c}}^{\mathbf{c}}$, and $Risk_{\mathbf{c}}^{\mathbf{c}}$ such that

$$d_R^R = R - R = 0 \ , \ d_{\mathbf{c}}^{\mathbf{c}} = ||\mathbf{c} - \mathbf{c}|| = 0 \ , \ \textit{and} \ Risk_{\mathbf{c}}^{\mathbf{c}} = \mathbf{c}' \Sigma \mathbf{c} = 1 \ . \tag{5.4}$$

Given a p-dimension nonzero vector $\mu = (\mu_1, \cdots, \mu_p)^{\mathbf{T}}$ and a positive definite matrix $\Sigma = (\sigma_{ij})$, which is assumed to be a diagonal matrix for simplicity, we state the simulation procedure as follows:

Step 1: For each round of N times simulation, we will first fix p and choose $\mu = (\mu_1, \cdots, \mu_p)^{\mathbf{T}}$ in which each μ_i is generated from $U(-1, 1)$. We will then select $\lambda = (\lambda_1, ..., \lambda_p)$, and $Weight = \left(\frac{p_1}{p}, ..., \frac{p_L}{p}\right)$. Thereafter, we set $\Sigma = \Lambda_p$ in which Λ_p is defined in equation (4.12).[2] We will fix p, μ, and λ for each round of simulation.

[2] Using λ and $Weight$ as described here is suitable to all the simulation conducted in this paper.

Step 2: Generate n vectors of returns $\mathbf{r} = (\mathbf{r}_1, \cdots, \mathbf{r}_p)^{\mathbf{T}}$ for the p-branch of assets from a population with mean μ and covariance matrix Σ.

Step 3: Compute the real optimal allocation $\mathbf{c}$ from (2.4) or (2.7) and return R from (2.3) or (2.6).

Step 4: Compute $\hat{\mathbf{c}}_\omega$ and $\hat{R}_\omega$ for $\omega = p, b, s$.

Step 5: Compute d_R^ω, $d_\mathbf{c}^\omega$, and $risk_\omega$ for $\omega = p, b, s$.

Step 6: Repeat Steps 2 to 5 N times.

Step 7: Compute the means and standard deviations of $\hat{R}_\omega$, d_R^ω, $d_\mathbf{c}^\omega$, and $risk_\omega$ for $\omega = p, b, s$ for each set of μ, λ, and $Weight$.

We conduct the simulation according to the above steps for each set of μ, λ and $Weight$ and exhibit in Table 5 the means and standard deviations of $\hat{R}_\omega$, d_R^ω, $d_\mathbf{c}^\omega$, and $risk_\omega$ for $\omega = p, b, s$. We also display R, d_R^R, $d_\mathbf{c}^\mathbf{c}$, and $Risk_\mathbf{c}^\mathbf{c}$ for the theoretical optimal return R in the table for comparison. In the three panels of Table 5, p is fixed and y increases from 0.1 to 0.9 for each given p. Here, we compare the performance of the plug-in, bootstrap-corrected, and spectral-corrected estimations under three different PSDs.

We first compare the performance of the plug-in return $\widehat{R}_p$, bootstrap-corrected return $\widehat{R}_b$, and spectral-corrected return $\widehat{R}_s$. From Panels A, B, and C of Table 5, one could notice that the performance of the plug-in return $\widehat{R}_p$ is not good even for $y = 0.1$ and the mean of the plug-in return is always higher than the real theoretical return R for any y and for any PSD, and thus, the plug-in difference d_R^p for the return is always positive, with d_R^p increasing sharply as y increases. This shows how serious the "over-prediction" problem is when one uses the plug-in return $\widehat{R}_p$. However, the s.d. of d_R^p (or $\widehat{R}_p$) is not too bad for $y = 0.1$ but it becomes worse when y increases. From the table, we find that when $y = 0.9$, the mean of $\widehat{R}_p$ is higher than twice the value of R and the s.d. is so big that we are not surprised that academics have commented that employing $\widehat{R}_p$ could do more harm than good.

We turn to examining the performance of the bootstrap-corrected return $\widehat{R}_b$. From Table 5, we find that the performance of $\widehat{R}_b$ is reasonably good for small values of y, say, for $y \leq 0.2$. Its performance becomes worse when y increases but its performance is still better than that of $\widehat{R}_p$. From Table 5, we find that the mean of $\widehat{R}_b$ always underestimates the theoretical optimal return because d_R^b is always negative. We call this the "under-prediction" problem. We observe that the absolute value of d_R^b is less than 10% of R when $y \leq 0.6$ in Panel A, when $y \leq 0.5$ in Panel B, and when $y \leq 0.4$ in Panel C of Table 5. Nonetheless, the absolute value of d_R^b is more than 30% of R when $y = 0.8$ and more than 80% of R when $y = 0.9$ in Panel A, more than 39% of R when $y = 0.8$ and more than 94% of R when $y = 0.9$ in Panel B, and more than 55% of R

when $y = 0.8$ and more than 115% of R when $y = 0.9$ in Panel C of Table 5. This shows that $\widehat{R}_b$ does circumvent the "over-prediction" problem but it incurs an "under-prediction" problem, especially for large values of y. In addition, from the table, we find that the s.d. of d_R^b (or $\widehat{R}_b$) is higher than that of d_R^b (or $\widehat{R}_p$) uniformly for any value of y and for any PSD. Thus, we conclude that the bootstrap-corrected return $\widehat{R}_b$ is still far from the ideal estimator for the optimal return R.

We now turn to examining the performance of our proposed spectral-corrected return $\widehat{R}_s$. From Table 5, we find that the mean of $\widehat{R}_s$ is still smaller than R, and thus, there is still an "under-prediction" problem for the spectral-corrected return. However, from the table, we find that $\widehat{R}_s$ is so close to its theoretical optimal return R that d_R^s is as small as 0.01% of R and less than 1.6% of R uniformly for any value of y from 0.1 to 0.9 and for any PSD. Thus, the "under-prediction" problem is very minimal if there even is one. In addition, from the table we find that the s.d. of d_R^s (or the s.d. of $\widehat{R}_s$) is so small that it is as small as 1.3% of the value of R and uniformly less than 6.4% of the value of R. The s.d. of d_R^s ($\widehat{R}_s$) is uniformly much smaller than those of d_R^p and d_R^b ($\widehat{R}_p$ and $\widehat{R}_b$). Moreover, from Table 5, we find that the mean of d_R^p (d_R^b) is as much as 12040 (10055) times d_R^s (d_R^s) while the s.d. of d_R^p (d_R^b) is as much as 257 (382) times d_R^s (d_R^s). Thus, we claim that our proposed spectral-corrected return $\widehat{R}_s$ could be the best estimator for the high-dimensional Markowitz MV portfolio optimization. If it is not, at least our proposed estimator enables academics and practitioners to get closer to obtaining the best estimation for the high-dimensional MV Markowitz optimization problem, and thus, we recommend that academics and practitioners use our proposed spectral-corrected return $\widehat{R}_s$ in their estimation. In addition, our simulation also shows that the estimation of our proposed spectral-corrected return $\widehat{R}_s$ and its standard deviation are stable for different values of sample size n, dimension p, and their ratio p/n but not for $\widehat{R}_p$ or $\widehat{R}_b$.

We turn to checking the "allocation estimation" problem by examining $d_{\mathbf{c}}^{\omega}$ defined in (5.2) for $\omega = p, b, s$. We first examine the performance of the plug-in allocation $\hat{\mathbf{c}}_p$. From Table 5, we find that although the plug-in estimation has a very serious "over-prediction" problem, it does not have any "allocation estimation" problem or at least the "allocation estimation" problem is not serious because $d_{\mathbf{c}}^p$ is doing very well. From the table, we find that the mean of $d_{\mathbf{c}}^p$ is smaller than 0.1 for any y and for any PSD except the value at $y = 0.9$ in Panel C of Table 5, in which case it is 0.13, which is still very small. In addition, most of its s.d.'s are smaller than 0.1 with the maximum of 0.23 at $y = 0.9$ in Panel C of Table 5, which is still very small. So, we conclude that the plug-in estimators do not have an "allocation estimation" problem or at least the "allocation estimation" problem is not serious.

On the contrary, although the bootstrap-corrected estimation is not serious for small values of y, the problem is serious for large values of y. From Table 5, we find that the mean of $d_{\mathbf{c}}^b$ is less than 0.1 only for $y \leq 0.6$ in Panels A and B and for $y \leq 0.4$ in Panel C of Table 5. However, the mean of $d^b_{\mathbf{c}}$ increases as y increases and it is higher than 1.2 (1.5, 2) for $y = 0.9$ in Panel A (B, C) of Table 5. This is unacceptably high. In addition, the s.d. of $d^s_{\mathbf{c}}$ is higher than 0.48 for $y \geq 0.7$ in all panels, higher than 3 for $y = 0.9$ in all panels and as high as 4.83 for $y = 0.9$ in Panel C of Table 5. This is also unacceptably high. Thus, we conclude that the "allocation estimation problem" is very serious for the bootstrap-corrected estimation for any large value of y.

On the other hand, from Table 5, we find that sometimes the spectral-corrected allocation $\hat{\mathbf{c}}_s$ does perform better than the plug-in allocation but, in general, the spectral-corrected allocation does not perform as well as the plug-in allocation. Nonetheless, the spectral-corrected allocation $\hat{\mathbf{c}}_s$ performs reasonably well because (1) nearly all of the means of $d^s_{\mathbf{c}}$ are smaller than those of $d^b_{\mathbf{c}}$ (except when $y = 0.5$ and 0.6 in Panel C of Table 5 in which case the difference is still very minimal); (2) all of the s.d.'s of $d^s_{\mathbf{c}}$ are smaller than those of $d^b_{\mathbf{c}}$; (3) the means of $d^s_{\mathbf{c}}$ are less than 0.1 when $y \leq 0.7$ (0.6, 0.4) in Panel A (B, C) and the biggest $d^s_{\mathbf{c}}$ is still smaller than 0.26, which is only 84% of the largest value of the mean of $d^b_{\mathbf{c}}$; and (4) the largest s.d. of $d^s_{\mathbf{c}}$ is still less than 0.35, which is only 46% of the largest value of the s.d. of $d^b_{\mathbf{c}}$. In addition, our simulation also shows that the estimation of our proposed spectral-corrected allocation $\hat{\mathbf{c}}_s$ is stable because $d^s_{\mathbf{c}}$ and its standard deviation are stable for different values of sample size n, dimension p, and their ratio p/n but not for $\hat{\mathbf{c}}_b$. Thus, we conclude that there is no "allocation estimation" problem for the spectral-corrected estimation or at least the "allocation estimation" problem is not serious.

Last, we study the risk problem for the three allocation estimations. We first study the risk problem for the plug-in estimation. From Table 5, we find that the risk problem is not serious for the plug-in estimation for any small value of y because the mean of $risk^p_{\mathbf{c}}$ is about 23% bigger than the theoretical risk when $y = 0.1$ and it is still less than 2 for $y = 0.2$. However, when y increases, the mean of $risk^p_{\mathbf{c}}$ increases sharply and it is around twice as big as the theoretical risk when $y = 0.3$, and 10 times as big as the theoretical risk when $y = 0.7$ and it is more than 86 (83,77) times bigger than the theoretical risk when $y = 0.9$ in Panel A (B, C) of Table 5. The s.d. of $risk^p_{\mathbf{c}}$ could be higher than 79. Since both the mean and the s.d. are unacceptably high for any large value of y, we conclude that the risk problem is serious for the plug-in estimation for any large value of y.

We turn to examining the risk problem for the bootstrap-corrected estimation. From Table 5, we find that the risk problem for the bootstrap-corrected estimation is even more serious than the plug-in estimation because (1) the mean of $risk^b_{\mathbf{c}}$ is uniformly higher than that of the mean of $risk^p_{\mathbf{c}}$ for any value of y and for any PSD; (2) the s.d. of $risk^b_{\mathbf{c}}$ is higher than that of $risk^p_{\mathbf{c}}$ for more than half (14) of the cases; (3) when y increases, $risk^b_{\mathbf{c}}$ increases even more sharply than $risk^p_{\mathbf{c}}$; and (4) the mean and s.d. of $risk^b_{\mathbf{c}}$ are as high as 151 and 170 (156 and 177, 166 and 188) for $y = 0.9$ in Panel A (B and C), respectively. Thus, we conclude that the risk problem for the bootstrap-corrected estimation is even more serious than that for the plug-in estimation.

Finally, we examine the risk problem for our proposed spectral-corrected estimation. From Table 5, we find that there is NO risk problem for the spectral-corrected estimation because (1) when $y = 0.1$, $risk_{\mathbf{c}}^{s}$ is only around 7% (with s.d. around 0.03) bigger than the theoretical risk for all panels; (2) when y increases, $risk_{\mathbf{c}}^{s}$ still increases but the speed is so slow that it is negligibl; (3) the mean of the $risk_{\mathbf{c}}^{s}$ is still less than 2 for $y \leq 0.7$ in Panel A, $y \leq 0.9$ in Panels B and C; (4) the mean of the $risk_{\mathbf{c}}^{s}$ is only 2.13 in Panel A, 1.82 in Panel B, 2.11 and 1.37 in Panel C for $y = 0.9$; and the s.d. of the $risk_{\mathbf{c}}^{s}$ is as small as 0.03 for $y = 0.1$ in all panels, increases when y increases, and is as high as 0.58, 0.48, and 0.28 for Panels A, B, and C for $y = 0.9$. In addition, our simulation also shows that the estimation of $risk_{\mathbf{c}}^{s}$ in our spectral-corrected estimation is stable because the estimate of $risk_{\mathbf{c}}^{s}$ and its standard deviation are stable for different values of sample size n, dimension p, and their ratio p/n but not for $risk_{\mathbf{c}}^{p}$ or $risk_{\mathbf{c}}^{b}$. Thus, we conclude that there is NO risk problem for the spectral-corrected estimation. Based on the above analysis, we conclude that the spectral-corrected estimation could be the best estimation for the problem of the high-dimensional Markowitz MV portfolio optimization or at least our approach enables academics and practitioners to get closer to obtaining the best estimation for the problem.

6 Conclusions

The purpose of this paper is to solve the "Markowitz optimization enigma" by developing a new covariance estimation to capture the essence of the portfolio selection. By using large dimensional data analysis, we first theoretically prove that the plug-in return, obtained by plugging the sample mean and sample covariance into the formulae of the optimal return, is always larger than its theoretically optimal value under more general conditions when the number of assets is large. We note that Bai, Liu, and Wong (2009, 2009a) have also proved that the plug-in return is always larger than its theoretically optimal value but they only show that the plug-in return is $\sqrt{\gamma}$ times bigger than the theoretical optimal return, while, in this paper, we develop more exact and generalizable results. For example, we prove that under some situations the plug-in return is $\sqrt{\gamma}$ times bigger than the theoretical optimal return, while under other situations the plug-in return is bigger than but may not be $\sqrt{\gamma}$ times bigger than the theoretical optimal return.

In the Markowitz MV portfolio optimization problem, the key problem actually is how to estimate the population covariance matrix accurately. In this paper, we introduce the spectral-corrected covariance matrix to correct the sample covariance matrix and derive some very important theoretical results. We construct the spectral-corrected covariance $\widehat{\Sigma}_s$ as the estimation of the population covariance matrix and provide the limiting behavior of the $a'\widehat{\Sigma}_s b$ for different bounded vectors a and b when p goes to infinity with n increasing proportionally. Our simulations do demonstrate that $a'\widehat{\Sigma}_s b$ estimates $a'\Sigma b$ very well. According to the theory we developed in this paper, we built up the spectral-corrected estimation that performs much better than both the plug-in and the bootstrap-corrected estimations, not only for the return but also for the allocation and the risk. Since our approach is easy to operate and implement in practice, the entire efficient frontier of our estimates can be constructed analytically. Thus, our proposed estimator facilitates the Markowitz MV optimization procedure, making it implementable and practically useful. In addition, the essence of the portfolio analysis problem could be adequately captured by our proposed approach. This greatly enhances the practical uses of the Markowitz mean-variance optimization procedure.

Since our model includes the situation in which one of the assets is a riskless asset, the separation theorem holds, and thus, our proposed return estimate is the optimal combination of the riskless asset and the optimal risky portfolio. We further note that the other assets listed in our model could be common stocks, preferred shares, bonds, and other types of assets so that the optimal return estimate proposed in our paper actually represents the optimal return for the best combination of riskless rate asset, bonds, stocks, and other assets. So, using the spectral-corrected estimation will be a very good investment strategy for the best combination of riskless rate asset, bonds, stocks, and other assets.

We remark that the returns being studied in the MV optimization procedure are usually assumed to be normally distributed. However, many studies (see, e.g., Fama, 1963, 1965; Clark, 1973; Blattberg and Gonedes, 1974; Fielitz and Rozelle, 1983; Fong and Wong, 2006) conclude that the normality assumption in the distribution of a security or portfolio return is violated. We further note that another contribution of our proposed approach is that we relax the normality assumption in the underlying distribution for the return being studied in the MV optimization procedure. More precisely, we relax the condition to the existence of the fourth moments. Thus, our proposed spectral-corrected estimators could be obtained for the problem of the high-dimensional Markowitz MV portfolio optimization when the returns of the assets being studied could follow any distribution under the condition of the existence of the fourth moments.

Last, we note that although we have developed many important theoretical results in this paper, there are still some results for which we should conduct simulations to check their relationships. Thus, further research could include developing such relationships theoretically. We also note that the theory developed in this paper could be applied to many related theories. For example, Korkie and Turtle (2002) and Bai, Liu, and Wong (2009, 2009a) have established a theory for the optimal return of self-financing portfolios. Academics could easily apply the estimation approach developed in this paper to extend their theory. In addition, although we claim that our estimation could be the best estimation, it might still be possible to get even better one(s). Thus, further research could also include improving our estimation further and developing even better estimations. For example, El Karoui's (2008) algorithm of estimating the population eigenvalues of large dimensional covariance matrices and the nonlinear shrinkage estimation of large-dimensional covariance matrices and their inverses developed by Ledoit and Wolf (2012) could be extended further to fit some weaker conditions. If this could be done, extensions could also include incorporating their covariance estimation to develop a new estimate for the high-dimensional Markowitz MV portfolio optimization.

Table 1: Comparison of $a'S_n^{-1}b$, $a'\widehat{\Sigma}_s^{-1}b$, $\lim_{p\to\infty,p/n\to y} a'\widehat{\Sigma}_s^{-1}b$, and $a'\Sigma^{-1}b$.

Panel A: $\lambda = (25, 10, 5, 1)$, $Weight = (0.25, 0.25, 0.25, 0.25)$.							
y	$a'S_n^{-1}b$	$s.d.\ of\ a'S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}b$	$s.d.\ of\ a'\widehat{\Sigma}_s^{-1}b$	$\frac{a'\Sigma^{-1}b}{1-y}$	$\varsigma_{a,b}$	$a'\Sigma^{-1}b$
0.1	2.0667	0.1308	1.8832	0.0938	2.066	1.8857	1.86
0.2	2.3315	0.2095	1.9175	0.1330	2.325	1.9153	1.86
0.3	2.6678	0.3085	1.9482	0.1644	2.657	1.9497	1.86
0.4	3.1142	0.4673	1.9840	0.2065	3.1	1.9896	1.86
0.5	3.7495	0.7119	2.0253	0.2459	3.72	2.0370	1.86
0.6	4.7594	1.0897	2.0822	0.2783	4.65	2.0953	1.86
0.7	6.4346	1.8411	2.1402	0.3138	6.2	2.1661	1.86
0.8	9.6998	3.7428	2.2027	0.3458	9.3	2.2479	1.86
0.9	20.638	14.465	2.2479	0.4005	18.6	2.3540	1.86
Panel B: $\lambda = (10, 5, 1)$, $Weight = (0.4, 0.3, 0.3)$.							
y	$a'S_n^{-1}b$	$s.d.\ of\ a'S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}b$	$s.d.\ of\ a'\widehat{\Sigma}_s^{-1}b$	$\frac{a'\Sigma^{-1}b}{1-y}$	$\varsigma_{a,b}$	$a'\Sigma^{-1}b$
0.1	1.8914	0.1124	1.7159	0.0783	1.888	1.7161	1.7
0.2	2.1294	0.1921	1.7348	0.1149	2.125	1.7348	1.7
0.3	2.4432	0.3064	1.7574	0.1527	2.428	1.7567	1.7
0.4	2.8605	0.4222	1.7829	0.1719	2.833	1.7823	1.7
0.5	3.4308	0.5982	1.8105	0.1938	3.4	1.8126	1.7
0.6	4.3315	1.0416	1.8452	0.2431	4.25	1.8498	1.7
0.7	5.9039	1.6676	1.8846	0.2519	5.666	1.8943	1.7
0.8	8.9074	3.4104	1.9236	0.2736	8.5	1.9444	1.7
0.9	19.060	11.968	1.9514	0.2913	17	2.0066	1.7

Panel C: $\lambda = (5, 3, 1)$, $Weight = (0.4, 0.3, 0.3)$.							
y	$a'S_n^{-1}b$	$s.d.\ of\ a'S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}b$	$s.d.\ of\ a'\widehat{\Sigma}_s^{-1}b$	$\frac{a'\Sigma^{-1}b}{1-y}$	$\varsigma_{a,b}$	$a'\Sigma^{-1}b$
0.1	2.5216	0.1528	2.3017	0.1102	2.5185	2.3016	2.2666
0.2	2.8384	0.2550	2.3396	0.1563	2.8333	2.3421	2.2666
0.3	3.2562	0.4079	2.3862	0.2061	3.2380	2.3892	2.2666
0.4	3.8107	0.5633	2.4343	0.2265	3.7777	2.4435	2.2666
0.5	4.5773	0.8110	2.4757	0.2483	4.5333	2.5066	2.2666
0.6	5.7787	1.3933	2.5069	0.2810	5.6666	2.5809	2.2666
0.7	7.8695	2.2318	2.5382	0.2793	7.5555	2.6643	2.2666
0.8	11.881	4.5272	2.5699	0.2882	11.333	2.7502	2.2666
0.9	25.446	16.054	2.5890	0.2989	22.666	2.8458	2.2666

Note: $p = 100$ is the dimension of the population, $y = p/n$, $N = 10000$ is the number of simulation, λ is the vector with the different eigenvalues of the population covariance matrix, and $Weight$ is the weight vector of the corresponding eigenvalues over the dimension p. Entries of a and b are generated from the uniform distribution on $(-1, 1)$. For easy comparison, we normalize a and b such that $a'\widehat{\Sigma}b$ is fixed. Readers may refer to footnote 2 in the text on how to use λ and $Weight$ in the simulation.

Table 2: Comparison of $a'S_n^{-1}b$, $a'\widehat{\Sigma}_s^{-1}b$, $\lim_{p\to\infty,p/n\to y} a'\widehat{\Sigma}_s^{-1}b$ and $a'\Sigma^{-1}b$.

Panel A: $y = 0.2$, $N = 10000$, $\lambda = (10, 5, 1)$, $Weight = (0.4, 0.3, 0.3)$.							
p	$a'S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}b$	$\frac{a'\Sigma^{-1}b}{1-y}$	$\varsigma_{a,b}$	$a'\Sigma^{-1}b$
50	2.1370	0.2663	1.7351	0.1533	2.125	1.7348	1.7
100	2.1309	0.1927	1.7347	0.1069	2.125	1.7348	1.7
150	2.1276	0.1472	1.7336	0.0851	2.125	1.7348	1.7
200	2.1264	0.1236	1.7345	0.0715	2.125	1.7348	1.7
250	2.1281	0.1102	1.7343	0.0635	2.125	1.7348	1.7
300	2.1266	0.1015	1.7350	0.0585	2.125	1.7348	1.7
Panel B: $y = 0.5$, $N = 10000$, $\lambda = (10, 5, 1)$, $Weight = (0.3, 0.3, 0.4)$.							
p	$a'S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}b$	$\frac{a'\Sigma^{-1}b}{1-y}$	$\varsigma_{a,b}$	$a'\Sigma^{-1}b$
50	3.5076	0.9089	1.8167	0.2754	3.4	1.8126	1.7
100	3.4564	0.5949	1.8106	0.1823	3.4	1.8126	1.7
150	3.4349	0.4785	1.8099	0.1496	3.4	1.8126	1.7
200	3.4258	0.3999	1.8099	0.1278	3.4	1.8126	1.7
250	3.4157	0.3678	1.8098	0.1149	3.4	1.8126	1.7
300	3.4124	0.3181	1.8087	0.1003	3.4	1.8126	1.7
Panel C: $y = 0.8$, $N = 10000$, $\lambda = (10, 5, 1)$, $Weight = (0.3, 0.3, 0.4)$.							
p	$a'S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}b$	$\frac{a'\Sigma^{-1}b}{1-y}$	$\varsigma_{a,b}$	$a'\Sigma^{-1}b$
50	10.2870	6.2017	1.9282	0.3798	8.5	1.9481	1.7
100	9.2256	3.6183	1.9169	0.2676	8.5	1.9444	1.7
150	9.0281	2.4383	1.9138	0.2019	8.5	1.9456	1.7
200	8.8168	2.1477	1.9175	0.1793	8.5	1.9444	1.7
250	8.8046	1.9380	1.9177	0.1659	8.5	1.9451	1.7
300	8.7166	1.6673	1.9160	0.1444	8.5	1.9444	1.7

Note:p is the dimension of the population, $y = p/n$, N is the number of simulation, λ is the vector with the different eigenvalues of the population covariance matrix, and $Weight$ is the weight vector of the corresponding eigenvalues over the dimension p. Entries of a and b are generated from the uniform distribution on $(-1, 1)$. For easy comparison, we normalize a and b such that $a'\widehat{\Sigma}b$ is fixed. Readers may refer to footnote 2 in the text on how to use λ and $Weight$ in the simulation.

Table 3: Comparison of $a'S_n^{-1}\Sigma S_n^{-1}b$, $a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$, $\lim_{p\to\infty} a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$, and $a'\Sigma^{-1}b$.

Panel A: $\lambda = (25, 10, 5, 1)$, $Weight = (0.25, 0.25, 0.25, 0.25)$.						
y	$a'S_n^{-1}\Sigma S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}\Sigma S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$\varrho_{a,b}$	$a'\Sigma^{-1}b$
0.1	2.5540	0.3291	2.0609	0.2072	1.9146	1.86
0.2	3.6663	0.6767	2.3280	0.3295	1.9848	1.86
0.3	5.5031	1.3299	2.6309	0.4554	2.0756	1.86
0.4	8.7847	2.8049	3.0013	0.6495	2.1930	1.86
0.5	15.385	6.2912	3.4631	0.8811	2.3474	1.86
0.6	31.376	16.011	4.1253	1.2105	2.5585	1.86
0.7	78.560	51.428	4.9380	1.6136	2.8447	1.86
0.8	272.04	268.23	5.8136	2.0421	3.2148	1.86
0.9	2874.0	6453.6	6.6451	2.5092	3.7593	1.86
Panel A: $\lambda = (10, 5, 1)$, $Weight = (0.4, 0.3, 0.3)$.						
y	$a'S_n^{-1}\Sigma S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}\Sigma S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$\varrho_{a,b}$	$a'\Sigma^{-1}b$
0.1	2.3411	0.2827	1.8597	0.1710	1.7354	1.7
0.2	3.3440	0.6207	2.0528	0.2763	1.7835	1.7
0.3	5.0434	1.3168	2.2779	0.4040	1.8483	1.7
0.4	8.1018	2.5581	2.5480	0.5107	1.9339	1.7
0.5	14.027	5.3002	2.8901	0.6743	2.0473	1.7
0.6	28.479	15.271	3.3483	0.9860	2.2014	1.7
0.7	72.610	47.662	3.8413	1.1533	2.4057	1.7
0.8	250.61	232.78	4.3277	1.3515	2.6607	1.7
0.9	2695.7	5616.6	4.7573	1.5257	3.0163	1.7
Panel A: $\lambda = (5, 3, 1)$, $Weight = (0.4, 0.3, 0.3)$.						
y	$a'S_n^{-1}\Sigma S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}\Sigma S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$\varrho_{a,b}$	$a'\Sigma^{-1}b$
0.1	3.1210	0.3839	2.5079	0.2419	2.3459	2.2666
0.2	4.4565	0.8244	2.7755	0.3769	2.4587	2.2666
0.3	6.7186	1.7533	3.1020	0.5570	2.6135	2.2666
0.4	10.786	3.4074	3.4696	0.6975	2.8173	2.2666
0.5	18.729	7.1874	3.8066	0.8334	3.0817	2.2666
0.6	38.021	20.461	4.0860	0.9681	3.4268	2.2666
0.7	96.768	63.820	4.3398	1.0042	3.8566	2.2666
0.8	333.82	307.84	4.5702	1.0590	4.3472	2.2666
0.9	3617.4	7589.3	4.7502	1.1209	4.9539	2.2666

Note: $p = 100$ is the dimension of the population, $y = p/n$, $N = 10000$ is the number of simulation, λ is the vector with the different eigenvalues of the population covariance matrix, and $Weight$ is the weight vector of the corresponding eigenvalues over the dimension p. Entries of a and b are generated from the uniform distribution on $(-1, 1)$. For easy comparison, we normalize a and b such that $a'\widehat{\Sigma}b$ is fixed. Readers may refer to footnote 2 in the text on how to use λ and $Weight$ in the simulation.

Table 4: Comparison of $a'S_n^{-1}\Sigma S_n^{-1}b$, $a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$, $\lim_{p\to\infty} a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$, and $a'\Sigma^{-1}b$.

Panel A: $y = 0.2$, $N = 10000$, $\lambda = (10, 5, 1)$, $Weight = (0.4, 0.3, 0.3)$.						
p	$a'S_n^{-1}\Sigma S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}\Sigma S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$\varrho_{a,b}$	$a'\Sigma^{-1}b$
50	3.3818	0.8707	2.0566	0.3698	1.7835	1.7
100	3.3510	0.6234	2.0536	0.2564	1.7835	1.7
150	3.3382	0.4742	2.0508	0.2051	1.7835	1.7
200	3.3295	0.3975	2.0513	0.1713	1.7835	1.7
250	3.3350	0.3535	2.0511	0.1528	1.7835	1.7
300	3.3292	0.3276	2.0520	0.1406	1.7835	1.7
Panel B: $y = 0.5$, $N = 10000$, $\lambda = (10, 5, 1)$, $Weight = (0.4, 0.3, 0.3)$.						
p	$a'S_n^{-1}\Sigma S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}\Sigma S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$\varrho_{a,b}$	$a'\Sigma^{-1}b$
50	15.1436	8.7435	2.9442	0.9984	2.0473	1.7
100	14.3526	5.3885	2.8950	0.6342	2.0473	1.7
150	14.0760	4.2313	2.8892	0.5152	2.0473	1.7
200	13.9539	3.4708	2.8806	0.4413	2.0473	1.7
250	13.8223	3.2053	2.8752	0.3920	2.0473	1.7
300	13.7772	2.7557	2.8691	0.3422	2.0473	1.7
Panel C: $y = 0.8$, $N = 10000$, $\lambda = (10, 5, 1)$, $Weight = (0.4, 0.3, 0.3)$.						
p	$a'S_n^{-1}\Sigma S_n^{-1}b$	$s.d.\ of$ $a'S_n^{-1}\Sigma S_n^{-1}b$	$a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$s.d.\ of$ $a'\widehat{\Sigma}_s^{-1}\Sigma\widehat{\Sigma}_s^{-1}b$	$\varrho_{a,b}$	$a'\Sigma^{-1}b$
50	422.1066	808.6923	4.4565	2.0094	2.6806	1.7
100	283.3364	274.6084	4.3262	1.3453	2.6607	1.7
150	260.2747	165.8750	4.2826	1.0236	2.6673	1.7
200	242.0023	137.1476	4.2875	0.9057	2.6607	1.7
250	239.3207	118.9240	4.2909	0.8456	2.6647	1.7
300	230.9242	100.4345	4.2752	0.7308	2.6607	1.7

Note : p is the dimension of the population, $y = p/n$, N is the number of simulation, λ is the vector with the different eigenvalues of the population covariance matrix, and $Weight$ is the weight vector of the corresponding eigenvalues over the dimension p. Entries of a and b are generated from the uniform distribution on $(-1, 1)$. For easy comparison, we normalize a and b such that $a'\widehat{\Sigma}b$ is fixed. Readers may refer to footnote 2 in the text on how to use λ and $Weight$ in the simulation.

Table 5: Comparison of spectral-corrected estimation with the plug-in and Bootstrap-corrected estimations

Panel A: $\lambda = (25, 10, 5, 1)$, $w = (0.25, 0.25, 0.25, 0.25)$								
		Return			$d_{\mathbf{c}}$		Risk	
		mean	d_R^{ω}	s.d.	mean	s.d.	mean	s.d.
y	real	3.8190	0	0	0	0	1	0
0.1	plug-in	4.0197	0.2006	0.0924	5.71E-16	4.303E-16	1.2323	0.0609
	bootstrap	3.8071	-0.0119	0.1312	2.90E-05	1.552E-3	1.2452	0.0806
	spectral	3.8138	-0.0052	0.0503	4.78E-16	3.641E-16	1.0771	0.0312
0.2	plug-in	4.2539	0.4348	0.1482	2.91E-04	0.0130	1.5553	0.1219
	bootstrap	3.7960	-0.0230	0.2074	1.46E-03	0.0321	1.5848	0.1516
	spectral	3.8069	-0.0121	0.0742	4.84E-16	3.653E-16	1.1675	0.0536
0.3	plug-in	4.5373	0.7183	0.2235	2.15E-03	0.0341	2.0276	0.2342
	bootstrap	3.7727	-0.0463	0.3165	8.35E-03	0.0836	2.0751	0.2609
	spectral	3.7973	-0.0217	0.0948	4.05E-04	0.0181	1.2729	0.0797
0.4	plug-in	4.8701	1.0511	0.3401	5.06E-03	0.0464	2.7319	0.4441
	bootstrap	3.7381	-0.0809	0.5096	2.09E-02	0.1353	2.8165	0.4297
	spectral	3.7857	-0.0333	0.1128	2.72E-03	0.0474	1.3939	0.1121
0.5	plug-in	5.2814	1.4623	0.5721	8.12E-03	0.0498	3.8820	0.9076
	bootstrap	3.6502	-0.1688	0.9054	4.04E-02	0.2040	4.0793	0.7797
	spectral	3.7800	-0.0390	0.1343	1.37E-02	0.1060	1.5416	0.1637
0.6	plug-in	5.8286	2.0095	0.8879	6.34E-03	0.0351	6.0203	1.8452
	bootstrap	3.5030	-0.3160	1.3923	6.22E-02	0.2751	6.5127	1.6391
	spectral	3.7679	-0.0511	0.1640	3.95E-02	0.1787	1.7010	0.2492
0.7	plug-in	6.5938	2.7747	1.4396	6.01E-03	0.0277	10.6988	4.3778
	bootstrap	3.2346	-0.5844	2.1844	1.31E-01	0.4856	12.1496	4.3399
	spectral	3.7626	-0.0564	0.1891	7.65E-02	0.2453	1.8649	0.3548
0.8	plug-in	7.6161	3.7970	2.4100	2.30E-02	0.0729	22.22	12.515
	bootstrap	2.5653	-1.2537	3.5775	0.3009	0.9693	28.768	15.926
	spectral	3.7605	-0.0585	0.2130	1.09E-01	0.2884	2.0102	0.4625
0.9	plug-in	9.9073	6.0882	4.7808	0.0820	0.1790	86.581	78.657
	bootstrap	0.7019	-3.1171	7.0065	1.2398	3.4164	151.27	170.23
	spectral	3.7585	-0.0604	0.2449	1.51E-01	0.3329	2.1382	0.5822

Panel B: $\lambda = (10, 5, 1)$, $w = (0.4, 0.3, 0.3)$								
		Return			$d_{\mathbf{c}}$		Risk	
		mean	d_R^{ω}	s.d.	mean	s.d.	mean	s.d.
y	real	4.0247	0	0	0	0	1	0
0.1	plug-in	4.2379	0.2131	0.0981	5.53E-16	4.14E-16	1.2326	0.0611
	bootstrap	4.0140	-0.0107	0.1391	1.83E-04	0.0053	1.2439	0.0808
	spectral	4.0196	-0.0050	0.0541	4.65E-16	3.47E-16	1.0708	0.0312
0.2	plug-in	4.4835	0.4588	0.1619	1.79E-03	0.0322	1.5532	0.1270
	bootstrap	3.9983	-0.0263	0.2322	6.02E-03	0.0729	1.5798	0.1531
	spectral	4.0122	-0.0125	0.0789	7.94E-05	0.0079	1.1524	0.0520
0.3	plug-in	4.7775	0.7527	0.2618	5.20E-03	0.0502	2.0194	0.2572
	bootstrap	3.9629	-0.0617	0.3950	1.87E-02	0.1289	2.0667	0.2655
	spectral	4.0034	-0.0213	0.1008	1.31E-03	0.0317	1.2444	0.0759
0.4	plug-in	5.1088	1.0841	0.4302	1.04E-02	0.0635	2.6997	0.5118
	bootstrap	3.8888	-0.1359	0.6871	4.02E-02	0.1927	2.8007	0.4346
	spectral	3.9933	-0.0314	0.1196	9.12E-03	0.0833	1.3462	0.1075
0.5	plug-in	5.5241	1.4993	0.7044	1.16E-02	0.0556	3.8153	1.0081
	bootstrap	3.7612	-0.2635	1.1514	6.57E-02	0.2691	4.0675	0.7844
	spectral	3.9909	-0.0338	0.1410	3.05E-02	0.1508	1.4652	0.1629
0.6	plug-in	6.0615	2.0368	1.0906	8.43E-03	0.0375	5.8713	2.0261
	bootstrap	3.5352	-0.4895	1.7415	9.70E-02	0.3608	6.5447	1.7161
	spectral	3.9828	-0.0419	0.1678	6.58E-02	0.2178	1.5793	0.2406
0.7	plug-in	6.8264	2.8017	1.7075	8.79E-03	0.0336	10.393	4.6787
	bootstrap	3.1870	-0.8377	2.6091	1.95E-01	0.6379	12.336	4.5793
	spectral	3.9844	-0.0402	0.1908	1.11E-01	0.2759	1.6811	0.3263
0.8	plug-in	7.8668	3.8420	2.7378	3.11E-02	0.0859	21.589	12.998
	bootstrap	2.4225	-1.6022	4.0791	0.4158	1.1867	29.425	16.636
	spectral	3.9842	-0.0404	0.2094	1.40E-01	0.3054	1.7668	0.3981
0.9	plug-in	10.147	6.1229	5.2831	0.0989	0.1956	83.53	79.77
	bootstrap	0.2299	-3.7948	7.7471	1.5703	3.9876	156.9	177.2
	spectral	3.9903	-0.0343	0.2342	1.83E-01	0.3408	1.8290	0.4788

Panel C: $\lambda = (5,3,1)$, $w = (0.4, 0.3, 0.3)$								
		Return			$d_{\mathbf{c}}$		Risk	
		mean	d_R^{ω}	s.d.	mean	s.d.	mean	s.d.
y	real	4.3376	0	0	0	0	1	0
0.1	plug-in	4.5684	0.2307	0.1088	1.39E-03	0.0263	1.2319	0.0634
	bootstrap	4.3260	-0.0116	0.1572	5.29E-03	0.0597	1.2412	0.0824
	spectral	4.3266	-0.0110	0.0679	1.69E-04	0.0097	1.0673	0.0352
0.2	plug-in	4.8172	0.4795	0.2181	9.56E-03	0.0622	1.5382	0.1637
	bootstrap	4.2767	-0.0609	0.3580	3.13E-02	0.1565	1.5646	0.1711
	spectral	4.3122	-0.0254	0.0974	1.05E-02	0.0771	1.1367	0.0585
0.3	plug-in	5.0988	0.7612	0.4000	1.57E-02	0.0708	1.9699	0.3554
	bootstrap	4.1712	-0.1664	0.6833	5.67E-02	0.2135	2.0361	0.2906
	spectral	4.3022	-0.0354	0.1265	4.39E-02	0.1550	1.2044	0.0930
0.4	plug-in	5.4127	1.0751	0.6435	1.70E-02	0.0631	2.5992	0.6690
	bootstrap	4.0139	-0.3237	1.1002	8.44E-02	0.2710	2.7718	0.4480
	spectral	4.3006	-0.0370	0.1552	9.37E-02	0.2186	1.2601	0.1349
0.5	plug-in	5.8043	1.4666	0.9996	1.24E-02	0.0430	3.6282	1.2445
	bootstrap	3.7780	-0.5596	1.6819	1.16E-01	0.3661	4.0675	0.8115
	spectral	4.3104	-0.0272	0.1756	1.37E-01	0.2547	1.3051	0.1783
0.6	plug-in	6.3027	1.9650	1.4782	8.96E-03	0.0304	5.5166	2.3755
	bootstrap	3.4027	-0.9349	2.4206	1.68E-01	0.5082	6.6470	1.8593
	spectral	4.3161	-0.0215	0.1981	1.80E-01	0.2808	1.3257	0.2105
0.7	plug-in	7.0149	2.6772	2.2115	2.44E-02	0.0641	9.6467	5.2082
	bootstrap	2.8573	-1.4803	3.4346	3.15E-01	0.8413	12.79	5.0708
	spectral	4.3282	-0.0094	0.2110	2.16E-01	0.2961	1.3450	0.2420
0.8	plug-in	8.0686	3.7309	3.3101	6.00E-02	0.1273	20.1585	13.6538
	bootstrap	1.9350	-2.4026	4.9736	0.6151	1.5195	30.8030	18.0832
	spectral	4.3301	-0.0075	0.2216	2.33E-01	0.3022	1.3621	0.2642
0.9	plug-in	10.35	6.0201	6.0308	0.1374	0.2333	77.81	79.22
	bootstrap	-0.6901	-5.0278	8.9579	2.0820	4.8334	166.1	188.6
	spectral	4.3371	-0.0005	0.2342	2.53E-01	0.3086	1.3754	0.2839

Note: $p = 100$ is the number of the assets, $N = 10000$ is the number of simulations, λ is constructed by the different eigenvalues of Σ, and w is the corresponding weight vector of λ on the whole p eigenvalues of Σ. The results are also compared and those of the real counterpart, which are denoted as "real." Readers may refer to footnote 2 in the text on how to use λ and $Weight$ in the simulation. We also note that the s.d. of $\hat{R}_{\omega}$ is the same of that of d_R^{ω}.

References

Bai, Z.D., Chen, J. and Yao, J.-F. 2010. On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. Australian & New Zeland Journal of Statistics , 52, 423-437

Bai, Z.D., Hui, Y.C., Wong, W.K., Zitikis, R., 2012. Evaluating Prospect Performance: Making a Case for a Non-Asymptotic UMPU Test, Journal of Financial Econometrics 10(4), 703-732.

Bai, Z.D., Li, H., Liu, H.X., Wong, W.K., 2011. Test Statistics for Prospect and Markowitz Stochastic Dominances with Applications. Econometrics Journal 122, 1-26.

Bai, Z.D., Liu, H.X., Wong, W.K., 2009. Enhancement of the Applicability of Markowitz's Portfolio Optimization by Utilizing Random Matrix Theory. Mathematical Finance 19(4), 639-667.

Bai, Zhidong, Huixia Liu and Wing-Keung Wong, (2009a), On the Markowitz Mean-Variance Analysis of Self-Financing Portfolios, Risk and Decision Analysis 1(1), 35-42.

Bai, Z.D., Liu, H.X., Wong, W.K., 2011. Asymptotic Properties of Eigenmatrices of A Large Sample Covariance Matrix. Annals of Applied Probability 21(5), 1994-2015.

Bai, Z. D., Miao, B. Q. and Pan, G. M. (2007). On the asymptotics of eigenvectors of large sample covariance matrix. Ann. Prob. 35.(.4), 1532-1572.

Bai, Z.D. and Jack W. Silverstein, 2010. Spectral Analysis of Large Dimensional Random Matrices. Second edition, Since Press, Beijing.

Bai, Z.D., Wang, K.Y., Wong, W.K., 2011. Mean-Variance Ratio Test, A Complement to Coefficient of Variation Test and Sharpe Ratio Test, Statistics and Probability Letters, 81(8), 1078-1085.

Blattberrg, R.C. and N.J. Gonedes (1974). A comparson of stable and student distribution as statistical models for stock prices. J. Bus. 47 244-280.

Broll, U., Egozcue, M., Wong, W.K., Zitikis, R., 2010. Prospect Theory, Indifference Curves, and Hedging Risks. Applied Mathematics Research Express 2010(2), 142-153.

Broll, U., Wahl, J.E., Wong, W.K., 2006. Elasticity of Risk Aversion and International Trade. Economics Letters 91(1), 126-130.

Brown, S.J. (1978). The portfolio choice problem: comparison of certainty equivalence and optimal bayes portfolios. Commun. Stat. Simul. Comput. 7 321-334.

Chan, C.Y., de Peretti, C., Qiao, Z., Wong, W.K., 2012. Empirical Test of the Efficiency of UK Covered Warrants Market: Stochastic Dominance and Likelihood Ratio Test Approach. Journal of Empirical Finance 19(1), 162-174.

Chan, J.C.C., Kroese D.P., 2010. Efficient estimation of large portfolio loss probabilities in t-copula models. European Journal of Operational Research 205(2), 361-367.

Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices. Econometrica 37 135-155.

Carpenter, G. A., Grossberg, S. and Reynolds, J. H. (1991). ARTMAP: Supervised real-time learning ad classification of nonstationary data by a self-organizing neural network. Neural Netw. 4:565-588.

Chopra, V.K., Hensel, C.R. and Turner, A.L. (1993). Massaging mean-variance inputs: returns from alternative global investment strategies in the 1980s. Management Science. 39(7) 845-855.

Clark, E., Jokung, O., Kassimatis, K., 2011. Making inefficient market indices efficient. European Journal of Operational Research 209(1) 83-93.

Egozcue, M., Wong, W.K., 2010. Gains from Diversification: A Majorization and Stochastic Dominance Approach. European Journal of Operational Research 200(3), 893-900.

Egozcue, M., Fuentes García, L., Wong, W.K., Zitikis, R., 2011. Do Investors Like to Diversify? A Study of Markowitz Preferences. European Journal of Operational Research 215(1), 188-193.

El Karoui Noureddine, 2008, Spectrum estimation for large dimensional covariance matrices using random matrix theory, The Annals of Statistics, Volume 36, Number 6 (2008), 2757-2790.

Elton, e.J., M.J. Gruber and M.W. Padberg (1976). Simple criteria for optimal portfolio selection. *J. Finance* 31(5) 1341-1357.

Elton, E.J., M.J. Gruber and M.W. Padberg (1978). Simple criteria for optimal portfolio selection: tracing out the efficient frontier. J. Finance33(1) 296-302.

Fama, E.F. (1963). Mandelbrot and the stable paretian hypothesis. J. Bus. 36 420-429.

Fama, E.F. (1965). Portfolio Analysis in a stable paretian market. Manage. Sc. 11 401-419.

Fielitz, B.D. and J.P. Rozelle (1983). Stable distributions and mixtures of distributions hypotheses for common stock returns. J. Am. Stat. Assoc. 78 28-36.

Fong, Wai Mun and Wing-Keung Wong, (2006), The Modified Mixture of Distributions Model: A Revisit, Annals of Finance, 2(2), 167 - 178.

Frankfurter, G.M., Phillips, H.E., Seagle, J.P., 1971. Portfolio Selection: The Effects of Uncertain Means, Variances and Covariances. Journal of Financial and Quantitative Analysis 6, 1251-1262.

Fu, C., Lari-Lavassani, A., Li, X., 2010. Dynamic mean-variance portfolio selection with borrowing constraint. European Journal of Operational Research 200(1), 312-319.

Gasbarro, D., Wong, W.K., Zumwalt, J.K., 2007. Stochastic dominance analysis of iShares. European Journal of Finance 13, 89-101.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer Series in Statistics. Springer-Verlag, New York.

Hensel, C. R. and A. L. Turner. 1998. Making Superior Asset Allocation Decisions: A Practitioner's Guide." In Worldwide Asset and Liability Modeling, Chapter 4, edited by W. T. Ziemba and J. M. Mulvey. Cambridge, United Kingdom: Cambridge University Press.

Jacobs, B.J., Levy, K.N., Markowitz, H.M., 2005. Portfolio Optimization with Factors, Scenarios, and Realistic Short Positions. Operations Research 53, 586-599.

Jansen, D.W., Li, Q., Wang, Z., Yang, J., 2008. Fiscal Policy and Asset Markets: A Semiparametric Analysis. Journal of Econometrics 147(1), 141-150.

Jorion, P., 1985. International Portfolio Diversification with Estimation Risk. Journal of Business 58(3), 259-278.

Jorion, P. (1992). Portfolio optimization in practice. Financial Analysts Journal. **48** 68-74.

Kan, R., Zhou, G., 2007. Optimal Portfolio Choice With Parameter Uncertainty. Journal of Financial and Quantitative Analysis 42(3), 621-656.

Konno, H., Watanabe, H., 1996. Bond portfolio optimization problems and their applications to index tracking: a partial optimization approach, Journal of the Operations Research Society of Japan 39(3), 295-306.

Konno, H., Yamazaki, H., 1991. Mean-Absolute Deviation Portfolio Optimization Model and its Applications to Tokyo Stock Market. Management Science 37(5), 519-531.

Korkie Bob, Turtle Harry J., 2002. A Mean-Variance Analysis of Self-Financing Portfolios. Management Science, vol. 48, no. 3, pp. 427-443, 2002

Kroll, Yoram & Levy, Haim & Markowitz, Harry M, 1984. " Mean-Variance versus Direct Utility Maximization," Journal of Finance, American Finance Association, vol. 39(1), pages 47-61, March.

Laloux, L., Cizeau, P., Bouchaud, J.P., Potters, M., 1999. Noise Dressing of Financial Correlation Matrices. Physical Review Letters 83, 1467-1470.

Lam, K., Liu, T., Wong, W.K., 2010. A pseudo-Bayesian model in financial decision making with implications to market volatility, under- and overreaction. European Journal of Operational Research 203(1), 166-175.

Ledoit, O., Wolf, M., 2004. A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. Journal of Multivariate Analysis 88(2), 365-411.

Ledoit, O., Wolf, M., 2012, Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices, The Annals of Statistics, Vol. 40, No. 2, 1024-1060.

Leung, P.L.,Ng, H.Y., Wong, W.K., 2012. An Improved Estimation to Make Markowitz's Portfolio Optimization Theory Users Friendly and Estimation Accurate with Application on the US Stock Market Investment, European Journal of Operational Research 222, 85-95.

Leung, P.L., Wong, W.K., 2008. On Testing The Equality Of The Multiple Sharpe Ratios, With Application On The Evaluation Of Ishares. Journal of Risk 10(3), 1-16.

Li, W. M., Chen, J. Q., Qin, Y. L., Yao, J. F., Bai, Z. D., 2013, Estimation of the population spectral distribution from a large dimensional sample covariance matrix, working paper.

Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. Review of Financial Studies 1, 41-66.

Ma, C., Wong, W.K., 2010. Stochastic Dominance and Risk Measure: A Decision-Theoretic Foundation for VaR and C-VaR. European Journal of Operational Research 207(2), 927-935.

Manganelli Simone, 2004. "Asset Allocation by Variance Sensitivity Analysis," Journal of Financial Econometrics, Society for Financial Econometrics, vol. 2(3), pages 370-389.

Markowitz, H.M., 1952. Portfolio Selection. Journal of Finance 7, 77-91.

Markowitz, H.M., 1959. Portfolio Selection. New York: John Wiley and Sons, Inc.

Markowitz, H.M., and A.F. Perold (1981). Portfolio analysis with factors and scenarios. J. Finance 871-877.

Markowitz, Harry M, 1991. " Foundations of Portfolio Theory," Journal of Finance, American Finance Association, vol. 46(2), pages 469-77, June.

McNamara, J.R., 1998. Portfolio Selection Using Stochastic Dominance Criteria. Decision Sciences 29(4), 785-801.

Michaud, R.O., 1989. The Markowitz Optimization Enigma: Is 'Optimized' Optimal? Financial Analysts Journal 45, 31-42.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. New York: John Wiley and Sons, Inc.

Pan, G.M. Comparision between two types of large sample covariance matrices. Accepted, Annales de l'Institut Henri Poincaré

Peccati, L., Cifarelli, D.M., 2003. VaR as a Risk Measure for Multiperiod Static Inventory Models. International Journal of Production Economics 81-82(1), 375-384.

Perold, A.F., 1984. Large-Scale Portfolio Optimization. Management Science 30(10), 1143-1160.

Ross, S.A., 1976. The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory 13, 341-360.

Sharpe, W.F., 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. Journal of Finance 19, 425-442.

Sharpe, W.F., 1967. A Linear Programming Algorithm for Mutual Fund Portfolio Selection. Management Sciences 22, 499-510.

Sharpe, William F., 1971. A Linear Programming Approximation for the General Portfolio

Analysis Problem. Journal of Financial and Quantitative Analysis, Cambridge University Press, vol. 6(05), pages 1263-1275, December.

Simaan, Y., 1997. Estimation Risk in Portfolio Selection: The Mean Variance Model versus the Mean Absolute Deviation Model. Management Science 43(10), 1437-1446.

Siverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. J. Multivariate Anal. 5 331-339.

Silverstein, J.W. and Choi, S.I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. J. Multivariate Anal. 54 295-309.

Stone, B.K.(1973). A linear Programming formulation of the general portfolio selection problem. J. Finan. Quant. Anal. 8(4) 621-636.

Soyer, R., Tanyeri, K., 2006. Bayesian portfolio selection with multi-variate random variance models. European Journal of Operational Research 171(3) 977-990.

Wong, F., Carter, C.K., Kohn, R., 2003. Efficient Estimation of Covariance Selection Models. Biometrika 90, 809-830.

Wong, W.K., 2007. Stochastic Dominance and Mean-Variance Measures of Profit and Loss for Business Planning and Investment. European Journal of Operational Research 182(2), 829-843.

Wong, W.K., Chan, R., 2008. Markowitz and Prospect Stochastic Dominances. Annals of Finance 4(1), 105-129.

Wong, W.K., Ma, C., 2008. Preferences over Meyer's location-scale family. Economic Theory 37(1), 119-146.

Wong, W.K., Phoon, K.F., Lean, H.H., 2008. Stochastic dominance analysis of Asian hedge funds, Pacific-Basin Finance Journal 16(3), 204-223.

Yu, M., Takahashi, S., Inoue, H., Wang, S., 2010. Dynamic portfolio optimization with risk control for absolute deviation model. European Journal of Operational Research 201(2), 349-364.

Zellner, A. and V.K. Chetty (1965). Prediction and descision problems in regression models from the bayesian point of view. J. Am. Stat. Assoc. 60 608-616.

Zhao, Y., Ziemba, W.T., 2008. Calculating risk neutral probabilities and optimal portfolio policies in a dynamic investment model with downside risk control. European Journal of Operational Research 185(3), 1525-1540.

Zymler, S., Rustem, B., Kuhn, D., 2011. Robust portfolio optimization with derivative insurance guarantees. European Journal of Operational Research 210(2), 410-424.