*Mathematical Finance,* Vol. 19, No. 4 (October 2009), 639–667

# ENHANCEMENT OF THE APPLICABILITY OF MARKOWITZ'S PORTFOLIO OPTIMIZATION BY UTILIZING RANDOM MATRIX THEORY

ZHIDONG BAI

*Northeast Normal University and National University of Singapore*

HUIXIA LIU

*National University of Singapore*

WING-KEUNG WONG

*Hong Kong Baptist University*

The traditional estimated return for the Markowitz mean-variance optimization has been demonstrated to seriously depart from its theoretic optimal return. We prove that this phenomenon is natural and the estimated optimal return is always $\sqrt{\gamma}$ times larger than its theoretic counterpart, where $\gamma = \frac{1}{1-y}$ with $y$ as the ratio of the dimension to sample size. Thereafter, we develop new bootstrap-corrected estimations for the optimal return and its asset allocation and prove that these bootstrap-corrected estimates are proportionally consistent with their theoretic counterparts. Our theoretical results are further confirmed by our simulations, which show that 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.

KEY WORDS: optimal portfolio allocation, mean-variance optimization, large random matrix, bootstrap method.

## 1. INTRODUCTION

The pioneer work of Markowitz (1952, 1959) on the mean-variance (MV) portfolio optimization procedure is the milestone of modern finance theory for optimal portfolio construction, asset allocation, and investment diversification. In the procedure, 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,

The authors are grateful to Professor Dilip B. Madan and anonymous referees for substantive comments that have significantly improved this article. The authors would also like to show appreciation to Professor Oliver B. Linton and Professor Harry M. Markowitz for their helpful comments and thank participants at the 2007 International Symposium on Financial Engineering and Risk Management for their valuable comments. The research is partially supported by NSF grant No. 10871036 from China and the grant R-155-000-079-112 from Department of Statistics and Applied Probability, National University of Singapore, and the grants R-703-000-015-720 and R-155-000-096-646 from Risk Management Institute, National University of Singapore.

*Manuscript received June 2006; final revision received April 2008.*

Address correspondence to Wing-Keung Wong, Department of Economics, Hong Kong Baptist University, Kowloon Tong, Hong Kong; e-mail: awong@hkbu.edu.hk.

minimize variance subject to obtaining a predetermined level of expected gain (Markowitz 1952, 1959, 1991; Merton 1972; Kroll, Levy, and Markowitz 1984).

Despite the fact that the conceptual framework of the classical MV portfolio optimization had first been set forth by Markowitz more than half a century ago, and despite the fact that several procedures for computing the corresponding estimates (see, e.g., Sharpe 1967, 1971; Stone 1973; Elton, Gruber, and Padberg 1976, 1978; Markowitz and Perold 1981; Perold 1984) have produced substantial experimentation in the investment community for nearly four decades, there have been persistent doubts about the performance of the estimates. Instead of implementing nonintuitive decisions dictated by portfolio optimizations, it has long been known anecdotally that a number of experienced investment professionals simply disregard the results or abandon the entire approach, since many studies (see, e.g., Michaud 1989; Canner, Mankiw, and Weil 1997; Simaan 1997) have found the MV-optimized portfolios to be unintuitive, thereby making their estimates do more harm than good. For example, Frankfurter, Phillips, and Seagle (1971) find that the portfolio selected according to the Markowitz MV criterion is likely not as effective as an equally weighted portfolio, while Zellner and Chetty (1965), Brown (1978), and Kan and Zhou (2007) show that the Bayesian decision rule under a diffuse prior outperforms the MV optimization. Michaud (1989) notes that MV optimization is one of the outstanding puzzles in modern finance and that it has yet to meet with widespread acceptance by the investment community, particularly as a practical tool for active equity investment management. He terms this puzzle the "Markowitz optimization enigma" and calls the MV optimizers "estimation-error maximizers."

To investigate the reasons why the MV optimization estimate is so far away from its theoretic counterpart, different studies have produced a range of opinions and observations. So far, all believe that it is because the "optimal" return is formed by a combination of returns from an extremely large number of stocks (see, e.g., McNamara 1998). Jorion (1985), Best and Grauer (1991), and Britten-Jones (1999) suggest the main difficulty concerns the extreme weights that often arise when constructing sample efficient portfolios that are extremely sensitive to changes in asset means. Another school suggests that the estimation of the correlation matrix plays an important role in this problem. For example, Laloux et al. (1999) find that Markowitz's portfolio optimization scheme, which is based on a purely historical determination of the correlation matrix, is not adequate because its lowest eigenvalues dominating the smallest risk portfolio are dominated by noise.

Many studies have tried to show that the difficulty can be alleviated by using different approaches. For example, Pafka and Kondor (2004) impose some constraints on the correlation matrix to capture the essence of the real correlation structure. Although this is expected to improve the overall performance, it may certainly introduce some biases in the estimation. In addition, by introducing the notion of "factors" influencing the stock prices, Sharpe (1964), Cohen and Pogue (1967), and Perold (1984) formulate the single-index model to simplify both the informational and computational complexity of the general model. 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 both the mean-absolute deviation portfolio model and the classical Markowitz model are still very severe, especially in small samples.

Our paper complements the theoretical work of Markowitz by developing a new bias-corrected estimator to reliably capture the essence of portfolio selection. We will first explain and thereafter share the theoretical proofs of our discovery about the "Markowitz optimization enigma." Our findings support the idea from Laloux et al. (1999) and others on this issue that the empirical correlation matrix plays an important role in this problem.

We also find that the estimation of the optimal return is poor due to the poor estimation of the asset allocations.[1] When the dimension of the data is large, by the theory of the large-dimensional random matrix, it is well known that the sample covariance matrix is not an efficient estimator of the population covariance matrix.[2] Thus, plugging the sample mean and covariance matrix into the MV optimization procedure will result in a serious departure of the optimal return estimate and the corresponding portfolio allocation estimate from their theoretic counterparts when the number of the assets is large. In the remainder of this paper, this return estimate will be called the "plug-in" return, and its corresponding estimate for the asset allocation will be called the "plug-in" allocation.[3] We also prove that the plug-in return is always larger than its theoretical value with a fixed rate depending on the ratio of the dimension to sample size.[4] We call this phenomenon "overprediction." To circumvent this overprediction problem, we further propose a new method to reduce this error by incorporating the idea of bootstrap into the theory of large-dimensional random matrix. The principal consideration of bootstrap is that there is a similar relationship between the biases of the estimators based on the original sample and the resampled one. By doing this, we obtain a bootstrap-modified estimate that analytically corrects the overprediction and drastically reduces the error. We further theoretically prove that the bootstrap-corrected estimate of return and its corresponding allocation estimate are proportionally consistent with their counterpart parameters. Our simulation further confirms the consistency of our proposed estimates, implying that the essence of the portfolio analysis problem could be adequately captured by our proposed estimates. Our simulation also shows that our proposed method improves the estimation accuracy so substantially that its relative efficiency[5] could be as high as 139 times when compared with the traditional "plug-in" estimate for 300 assets with a sample size of 500. The relative efficiency will be much higher for bigger sample sizes and larger numbers of assets. Similar results are also obtained for its corresponding allocation estimate.

## 2. THEORY

This section studies the theoretic optimal solution for Markowitz's MV optimization procedure and introduces the theory of the large-dimensional random matrix to explain the "Markowitz optimization enigma," which holds that the Markowitz MV optimization procedure is impractical. In the next section, we invoke a new approach to make the optimization procedure more practically useful. 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*.

### 2.1. Optimal Solution

Suppose that there are $p$-branch assets, $\boldsymbol{S} = (s_1, \ldots, s_p)^T$, whose returns are denoted by $\boldsymbol{r} = (r_1, \ldots, r_p)^T$, with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$ and covariance matrix $\Sigma = (\sigma_{ij})$. In addition, we suppose that an investor will invest capital $C$ on the $p$-branch assets $\mathbf{S}$ such

[1] We note that Aït Sahalia and Brandt (2001) suggest focusing directly on the optimal portfolio weights first. This could provide an alternative solution to overcome the Markowitz optimization enigma.

[2] See, for example, Laloux et al. (1999).

[3] See, for example, Zellner and Chetty (1965), Brown (1978), and Kan and Zhou (2007).

[4] We note that Maller, Durand, and Lee (2005) also prove the maximum Sharpe ratio is biased upward for its population value.

[5] Readers may refer to equation (4.4) for the definition of "relative efficiency."

that she or he wants to allocate her or his investable wealth on the assets but obtain any of the following:

1. to maximize return subject to a given level of risk, or
2. to minimize her or his risk for a given level of expected return.

Since the above two problems are equivalent, we only look for a solution to the first problem in this paper.[6] Without loss of generality, we assume $C \leq 1$ and her or his investment plan to be $\boldsymbol{c} = (c_1, \ldots, c_p)^T$. Hence, we have $\sum_{j=1}^{p} c_j = C \leq 1$. Also, the mean and risk of her or his anticipated return will then be $\boldsymbol{c}^T\boldsymbol{\mu}$ and $\boldsymbol{c}^T \Sigma \boldsymbol{c}$, respectively. In this paper, we further assume that short selling is allowed, and hence any component of **c** could be negative. Thus, the above maximization problem can be reformulated as the following optimization problem:

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

where **l** represents the $p$-dimensional vector of ones and $\sigma_0^2$ is a given risk level.[7] We call $R$ satisfying equation (2.1) the **optimal return** and the solution **c** to the maximization the **optimal allocation plan**. One could easily extend the separation theorem (Cass and Stiglitz 1970) and the mutual fund theorem (Merton 1972) to obtain the analytical solution of equation (2.1)[8] from the following proposition:

PROPOSITION 2.1. *For the optimization problem shown in equation (2.1), the optimal return, R, and its corresponding investment plan,* **c**, *are obtained as follows:*

*1. If*

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

*then the optimal return, R, and corresponding investment plan,* **c**, *will be*

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

*and*

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

*2. If*

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

[6] Readers may refer to equation (2.1) for the formulation of the maximization problem. This problem is equivalent to

$$\sigma^2 = \min \mathbf{c}^T\Sigma\mathbf{c} \quad \text{subject to} \quad \mathbf{c}^T\mathbf{1} \leq 1 \quad \text{and} \quad R \leq \mathbf{c}^T\boldsymbol{\mu}. \tag{2.1a}$$

As the minimization problem stated in equation (2.1a) is the dual problem of the maximization problem in equation (2.1), we only discuss the maximization problem stated in equation (2.1) in this paper.

[7] We note that in this paper we study the optimal return. However, another direction of research is to study the optimal portfolio variance; see, for example, Pafka and Kondor (2003) and Papp et al. (2005).

[8] Many studies, for example, Ju and Pearson (1999) and Maller and Turkington (2002), use settings similar to the setting in equation (2.1) to obtain different solutions.

*then the optimal return, R, and corresponding investment plan,* **c**, *will be*

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

*and*

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

*where*

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

The proof of Proposition 2.1 can be found from many references in the literature; a short proof is also given in the Appendix.

REMARK 2.1. The intuition of the inequalities to distinguish the two cases of the solutions in Proposition 2.1 can be seen from the following: the maximization is taken in the intersection of the ellipsoid $\boldsymbol{c}^T\Sigma\boldsymbol{c} \le \sigma_0$ and the half space $\boldsymbol{c}^T\mathbf{1} \le 1$ (note that the intersection is not empty because the point $\boldsymbol{c} = \mathbf{0}$ belongs to both the half space and the ellipsoid). If the ellipsoid is completely contained in the half space, that is, the ellipsoid does not intersect with the hyperplane $\boldsymbol{c}^T\mathbf{1} = 1$, then the solution is the same as the maximization problem without the half space restriction. Hence, the solution is then given by the first case. Otherwise, the maximizer should be on the intersection of the ellipse $\boldsymbol{c}^T\Sigma\boldsymbol{c} = \sigma_0$ and the hyperplane $\boldsymbol{c}^T\mathbf{1} = 1$, since the target function $\boldsymbol{c}^T\boldsymbol{\mu}$ is a linear function in $\mathbf{c}$. The inequality $\frac{\mathbf{1}^T\Sigma^{-1}\mu\sigma_0}{\sqrt{\mu^{\mathrm{T}}\Sigma^{-1}\mu}} < 1$ could then be used to test whether the maximizer of $\boldsymbol{c}^T\boldsymbol{\mu}$ is in the ellipsoid, that is, whether $\mathbf{c} = \frac{\Sigma^{-1}\mu\sigma_0}{\sqrt{\mu^{\mathrm{T}}\Sigma^{-1}\mu}}$ is an inner point of the half space.

The set of efficient feasible portfolios for all possible levels of portfolio risk forms the MV efficient frontier. We note that in this paper we formulate the $p$-branch assets for this optimization problem in which the assets could be common stocks, preferred shares, bonds, and other types of assets. We also note that in this paper a riskless asset is assumed to be available for both borrowing and lending and that excess return is calculated by subtracting the return of this riskless asset from the total return. The return calculated in this paper could be set as the total return or the excess return. For any given level of risk, Proposition 2.1 seems to provide a unique optimal return with its corresponding MV optimal investment plan or asset allocation to represent the best investment alternative given the selected assets. Thus, it seems to provide a solution to Markowitz's MV optimization procedure. Nonetheless, one may expect the problem to be straightforward; however, this is not so, since the estimation of the optimal return and its corresponding investment plan is a difficult task. This issue will be discussed in the following sections.

## 2.2. Large-Dimensional Random Matrix Theory

The large dimensional random matrix theory (LDRMT) traces back to the development of quantum mechanics (QM) in the 1940s. Because of its rapid development in theoretic investigation and its wide application, it has since attracted growing attention in many areas, such as economics and finance, as well as mathematics and statistics. Wherever the dimension of data is large, the classical limit theorems are no longer suitable, since the statistical efficiency will be substantially reduced when they are employed. Hence, statisticians have to search for alternative approaches in such data analysis, and thus, the LDRMT is found to be useful. A major concern of the LDRMT is to investigate the limiting spectrum properties of random matrices where the dimension increases proportionally with the sample size. This turns out to be a powerful tool in dealing with large-dimensional data analysis.

We utilize the LDRMT to study MV optimization by analyzing the corresponding high-dimensional data. In the analysis, an estimation of the sample covariance matrix plays an important role in examining this type of data. However, the covariance matrix estimate is inherently limited to the MV framework and thus subject to the limitations of being a risk measure. For any practical use, it would therefore be necessary to have reliable estimates for the correlations of returns, which are usually obtained from historical return series data. If one estimates a $p \times p$ correlation matrix from $p$ time series of length $n$ each, one inevitably introduces estimation errors that can become so overwhelming that the whole applicability of the theory may become questionable for large $p$. Suppose that $\{x_{jk}\}$ for $j = 1, \ldots, p$ and $k = 1, \ldots, n$ is a set of double array real random variables that are independent and identically distributed (iid) with mean zero and variance $\sigma^2$. Let $\mathbf{x}_k = (x_{1k}, \ldots, x_{pk})^T$ and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$; the sample covariance matrix[9], $S$, of $p \times p$ dimension is then defined as

$$S = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^T, \tag{2.2}$$

where $\bar{\mathbf{x}} = \sum_{k=1}^{n} \mathbf{x}_k / n$.

It is widely recognized that the major difficulty in the estimation of optimal returns is the inadequacy of using the inverse of the estimated covariance to measure the inverse of the covariance matrix.[10] To circumvent this problem, we introduce some fundamental limit theorems (see, among others, Jonsson 1982, Bai and Yin 1993, Bai 1999, and Bai and Silverstein 1998, 1999, 2004) in the LDRMT to take care of the empirical spectral distribution of the eigenvalues for the sample covariance matrix. Suppose that the sample covariance matrix $S$ defined in equation (2.2) is a $p \times p$ matrix with eigenvalues $\{\lambda_j : j = 1, 2, \ldots, p\}$. Since all eigenvalues are real, the empirical spectral distribution function, $F^S$, of the eigenvalues $\{\lambda_j\}$ for the sample covariance matrix, $S$, is then defined as

$$F^S(x) = \frac{1}{p} \#\{j \le p : \lambda_j \le x\}. \tag{2.3}$$

[9] In the literature of LDRMT, the random variables are usually considered to be complex and the sample covariance matrix is defined by $S = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^*$, where $^*$ stands for the complex conjugate transpose of vectors or matrices. For the purposes of this paper and for simplicity, we limit the general results of LDRMT to the real case in all of our quoted results.

[10] See, for example, Laloux et al. (1999).

Here, $\#E$ is the cardinality of the set $E$. Before introducing theorems for the empirical spectral distribution function of the eigenvalues, we first define the *Marčenko-Pastur law* (MP law) as follows.

DEFINITION 2.1. Let $y$ be the limit of the dimension-to-sample-size ratio index, $p/n$, and $\sigma^2$ be the scale parameter. The MP law is defined as follows:

1. If $y \leq 1$, the MP law $F_y(x)$ is completely defined by the density function:

$$p_y(x) = \begin{cases} \dfrac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)}, & \text{if} \quad a < x < b \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

   where $a = \sigma^2(1-\sqrt{y})^2$ and $b = \sigma^2(1+\sqrt{y})^2$.
2. If $y > 1$,[11] then $F_y(x)$ has a point mass $1 - 1/y$ at the origin, and the remaining mass of $1/y$ is distributed over $(a, b)$ by the density $p_y$ defined in equation (2.4).

We note that if $\sigma^2 = 1$, the MP law is called the standard MP law. The MP law is named after Marčenko and Pastur because of their renowned work of 1967. We are now ready to introduce the following theorem for the empirical spectral distribution function of the sample covariance matrix:

PROPOSITION 2.2. *Suppose that $\{x_{jk}\}$ for $j = 1, \ldots, p$ and $k = 1, \ldots, n$ is a set of i.i.d. real random variables with mean zero and variance $\sigma^2$. If $p/n \to y \in (0, \infty)$ then, with probability one, the empirical spectral distribution function, $F^S$, defined in equation (2.3) tends to the MP law almost surely.*

Readers could refer to Bai (1999) for the proof of Proposition 2.2. This proposition shows that the eigenvalues in the covariance matrix behave undesirably. As indicated by Proposition 2.2, when the population covariance is an identity, that is, all the eigenvalues are 1, the eigenvalues of the sample covariance will then spread from $(1-\sqrt{y})^2$ to $(1+\sqrt{y})^2$. For example, if $n = 500$ and $p = 5$, that is, even the dimension-to-sample-size ratio is as small as $y = p/n = 0.01$, the eigenvalues of the sample covariance will then spread in the interval of (0.81,1.21). The larger the ratio, the wider the interval. For instance, for the same $n$ with $p = 300$, we have $y = 0.6$ and the interval for the eigenvalues of the sample covariance will then become (0.05, 3.14), a much wider interval. The spread of eigenvalues for the inverse of the sample covariance matrix will be more severe; for example, the spreading intervals for the inverses of the sample covariance matrices for the above-mentioned two cases will be (0.83, 1.23) and (0.32, 19.68), respectively.

The returns being studied in the MV optimization procedure are usually assumed to be independently and identically normal-distributed (Feldstein 1969; Hanoch and Levy 1969; Rothschild and Stiglitz 1970, 1971; Hakansson 1972). However, in reality, most of the empirical returns are not identically normal-distributed and they are not independent either. Nonetheless, some investors may choose to invest in assets with small correlations and thus the independence requirement may not be essential. However, the assumptions of identical distribution and normality may be violated in many cases; for example, see Fama (1963, 1965), Clark (1973), Blattberg and Gonedes (1974), Clark and Fielitz

[11] We note that in this paper, we only study the situation in which $y < 1$. However, as we cite the well-known results from the literature, for completeness, we include all other situations ($y = 1$ and $y > 1$) as stated in the original literature.

and Rozelle (1983). Thus, it is of practical interest to consider the situation in which the elements of matrix $\mathbf{X}$ depend on $n$, and for each $n$, they are independent but not necessarily identically or normally distributed. For this non-iid and nonnormality case, we introduce the following proposition for the empirical distribution function of the eigenvalues for the sample covariance matrix:

PROPOSITION 2.3. *Suppose that $\{x_{jk}\}$ for $j = 1, \ldots, p$ and $k = 1, \ldots, n$ is a set of independent real random variables with mean zero and variance $\sigma^2$ but not necessarily identically distributed. Let n be the sample size and p be the number of assets. We assume that $p/n \to y \in (0, \infty)$, and that for any $\eta > 0$, we have*

$$\frac{1}{\eta^2 np} \sum_{jk} E\left(\left|x_{jk}^{(n)}\right|^2 I_{(|x_{jk}^{(n)}| \geq \eta\sqrt{n})}\right) \to 0.$$

*Then, with probability one, the empirical distribution function, $F^S$, of the eigenvalues for S defined in equation (2.3) tends to the MP law defined in Definition 2.1 with the dimension-to-sample-size ratio index, y, and scale index, $\sigma^2$.*

Refer to Bai (1999) for a proof of Proposition 2.3. Obviously, Proposition 2.3 enables us to relax the iid assumption to the one with independent but not necessarily identically distributed entries in equation (2.2) in developing portfolio optimization theory. In practice, the theory also applies to the case where the asset observations are correlated as explained in Theorem 3.2. For example, if both the vectors $\boldsymbol{y}_k = \Sigma^{1/2}\, \boldsymbol{x}_k$ and the entries of $\boldsymbol{x}_k$'s defined in equation (2.2) satisfy the assumption of Proposition 2.3, then the entries of $\boldsymbol{y}_k$ are correlated. In this case, the sample covariance matrix for $\boldsymbol{y}_k$'s is $\Sigma^{1/2} S \Sigma^{1/2}$. We shall apply Propositions 2.2 and 2.3 to $S$ in $\Sigma^{1/2} S \Sigma^{1/2}$ in the development of our theory. However, for simplicity, we will keep the iid assumption in developing the theory in this paper.

In many cases, the integrands of integrals with respect to the empirical spectral distributions are unbounded at 0 and/or at infinity. As such, when using the limiting spectral distribution to find the limit of the linear spectral statistic, we must include the condition that the eigenvalues of the random matrices are bounded away from the points where the integrands are unbounded. To handle this situation, we introduce the following proposition of the extreme eigenvalues for any large-dimensional sample covariance matrix.

PROPOSITION 2.4. *Suppose that $\{x_{jk}\}$ for $j = 1, \ldots, p$ and $k = 1, \ldots, n$ is a set of double array of iid real random variables with mean zero, variance $\sigma^2$, and a finite fourth moment. S is the sample covariance matrix constructed by the n vectors $\{(x_{1k}, \ldots, x_{pk})^T;\ k = 1, \ldots, n\}$. If $p/n \to y \in (0, \infty)$, then, with probability one, the maximum eigenvalue of S tends to $b = \sigma^2(1 + \sqrt{y})^2$ and in addition,*

1. *if $y \leq 1$, the smallest eigenvalue of S tends to $a = \sigma^2(1 - \sqrt{y})^2$, and*
2. *if $y > 1$, the $p - n + 2^{nd}$ smallest eigenvalue of S tends to $a = \sigma^2(1 - \sqrt{y})^2$.*

The proof of this proposition can be found in Bai and Yin (1993)[12] and Bai (1999). If $p$ is fixed, applying the law of large numbers, one can easily show that the sample

[12] In Bai and Yin (1993), their Conclusion 2 is the $p - n + 1^{st}$ smallest eigenvalue approaching to $a$ because the sample covariance matrix in their paper is defined by $n^{-1} \sum \mathbf{x}_k \mathbf{x}_k^T$, which has one more rank than the covariance matrix defined in this paper.

covariance matrix will be close to the population covariance with high probability. However, according to the LDRMT, when the dimension $p$ is large, the sample covariance will no longer be an efficient estimator for the population covariance (see, e.g., Laloux et al. 1999). Moreover, the performance of the estimator will rapidly worsen with the increase of the dimension of the covariance matrix. This results in an unavoidable severe departure of its estimated optimal portfolio allocation from its theoretic counterpart and, thus, explains the "Markowitz optimization enigma" phenomenon that the "Markowitz optimal procedure" is not practically useful or that at least the procedure is far from satisfactory.

## 3. ESTIMATION

In this section, we first introduce the traditional plug-in estimators and then develop the bootstrap-corrected estimators for the optimal return and its asset allocation. In contrast to conventional MV portfolio analysis implemented by simply plugging in the sample means and sample covariance matrix into the formula of the theoretic optimal return, our proposed bootstrap estimator will be constructed by incorporating the bootstrap technique into the LDRMT. The former is found not to be a good estimator, while the latter is. We will discuss the theory and the properties for these estimators in the following subsections.

### 3.1. Plug-In Estimator

In the Markowitz MV optimization, we call the procedure of substituting the population mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ in the optimal return $R$ shown in equation (2.1) by their corresponding sample mean vector $\bar{\mathbf{x}}$ and the sample covariance matrix $S$ the "plug-in" procedure and call its estimator of the optimal return the "plug-in" return (estimate) to distinguish it from any attainable efficient return estimate, since this plug-in return is far from satisfactory. As a result, many academic researchers and practitioners have recommended not using the plug-in return estimate.

The poor estimation is actually due to the poor estimation of $\mathbf{c}$ by "plugging in" the sample mean vector $\bar{\mathbf{x}}$ and the sample covariance matrix $S$ into the formulae of the asset allocation $\mathbf{c}$ in Proposition 2.1 such that

$$\hat{\mathbf{c}}_p = \begin{cases} \dfrac{S^{-1}\bar{\mathbf{x}}}{\sqrt{\bar{\mathbf{x}}^T S^{-1}\bar{\mathbf{x}}}}, & \text{if} \quad \dfrac{\sigma_0 \mathbf{1}^T S^{-1}\bar{\mathbf{x}}}{\sqrt{\bar{\mathbf{x}}^T S^{-1}\bar{\mathbf{x}}}} < 1, \\ \dfrac{S^{-1}\mathbf{1}}{\mathbf{1}^T S^{-1}\mathbf{1}} + \hat{b}\left(S^{-1}\bar{\mathbf{x}} - \dfrac{\mathbf{1}^T S^{-1}\bar{\mathbf{x}}}{\mathbf{1}^T S^{-1}\mathbf{1}} S^{-1}\mathbf{1}\right), & \text{otherwise;} \end{cases} \tag{3.1}$$

where

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

The problem arises because $\hat{\mathbf{c}}_p$ differs from the optimal allocation $\mathbf{c}$ dramatically when the dimension $p$ of the covariance matrix is large. Thereafter, when one "plugs" $\hat{\mathbf{c}}_p$ into the optimal return $\boldsymbol{c}^T\boldsymbol{\mu}$ as $\hat{\mathbf{c}}_p^T\boldsymbol{\mu}$, one should not be surprised that $\hat{\mathbf{c}}_p^T\boldsymbol{\mu}$ is so far away from $\boldsymbol{c}^T\boldsymbol{\mu}$ that it is not practically useful. In this connection, we do not call $\hat{\mathbf{c}}_p^T\boldsymbol{\mu}$ an estimator of the optimal return $\boldsymbol{c}^T\boldsymbol{\mu}$. Instead, we call $\hat{\mathbf{c}}_p$ in equation (3.1) the plug-in allocation and

$$\hat{R}_p = \hat{\mathbf{c}}_p^T\boldsymbol{\mu} \tag{3.2}$$

the plug-in return, respectively. Thereafter, we substitute $\bar{\mathbf{x}}$ back to $\mu$ in equation (3.2) such that

$$\hat{\hat{R}}_p = \hat{\mathbf{c}}_p^T \bar{\mathbf{x}} \tag{3.3}$$

to get an estimate of the plug-in return. We note that although the sample covariance $S$ is not a good estimator of the true covariance $\Sigma$ when the dimension is large, the sample mean $\bar{\mathbf{x}}$ is still a good estimator of $\mu$. Thus, we expect to have $\hat{\mathbf{c}}_p^T \bar{\mathbf{x}} \simeq \hat{\mathbf{c}}_p^T \boldsymbol{\mu}$, and hence, $\hat{\hat{R}}_p$ is still a good estimator of $\hat{R}_p$. 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 $B_n$. Our simulation results shown in Figure 4.1 and Table 4.1 in the next section support this argument. We further prove theoretically that this argument is correct as stated in the following theorem.

THEOREM 3.1. *Under the general conditions as stated in Theorem 3.2 below, the estimator $\hat{\hat{R}}_p$ of the plug-in return $\hat{R}_p$ is asymptotically similar to $\hat{R}_p$, where $\hat{R}_p$ and $\hat{\hat{R}}_p$ are defined in equations (3.2) and (3.3), respectively.*

The proof of Theorem 3.1 is straightforward. Now, we explain the poor estimation in detail. In reality, the number of assets available to the investors is very large, but the dimension of the covariance is also huge. Thus, according to the above propositions and theorems, the eigenvalues of the estimated covariance matrix will then be widely spread. On the other hand, by the elementary law of large numbers, it is easy to prove that the trace of the sample covariance matrix is almost surely proportionally consistent with the trace of the population covariance matrix. Therefore, by Jessen's inequality,[13] the linear functional of the empirical spectral distribution of the sample covariance matrix with respect to any convex function is definitely larger than the counterpart with respect to the population covariance matrix. For the return in the MV optimization procedure, the corresponding function is $\boldsymbol{X}^{-1}$ ($\boldsymbol{X} > 0$), which is convex.[14] This confirms that the plug-in return is always larger than the theoretic optimal return. An intuitive interpretation for this phenomenon is that the inverse of small eigenvalues will become very large, causing the ill-conditioning properties when employing this plug-in MV optimization procedure. We call this phenomenon "overprediction."

REMARK 3.2. We use the term "overprediction" instead of the popularly used "estimation/measurement error" for the following reasons. First, the "error" does not attribute to the measurement error. That is, even when there is no measurement error, for example, even when the samples are exactly measured and recorded, the "error" in the estimation of the covariance matrix still exists. Hence, the plug-in return is still larger than the theoretical return. Second, the measurement error could be positive or negative and should have mean 0 in the long run. The phrase "overprediction" means the plug-in return is always larger than its theoretic counterpart, whether in the long run or short run.

[13] Jessen's inequality states that $\varphi(E\,\mathbf{x}) \le E\varphi(\mathbf{x})$ for any convex function $\varphi$ provided that $E\varphi(\boldsymbol{x})$ is finite, the inequality is strict unless $\varphi$ is linear.

[14] We note that Jessen's inequality holds for multivariate convex functions as well as matrix functions. It is easy to show that $\mathbf{x}^{-1}$ is a convex function in the domain of positive definite matrices. Thus, $\Sigma^{-1} = (ES)^{-1} \le E(S^{-1})$, which implies that $\boldsymbol{\mu}'\Sigma^{-1}\,\boldsymbol{\mu} \le E(\boldsymbol{\mu}'S^{-1}\boldsymbol{\mu})$.

To explain the overprediction phenomenon, in this paper we further theoretically prove that the plug-in return is not appropriate for practical use as shown in the following theorem.

THEOREM 3.2. *Assume that* $\mathbf{y}_1, \ldots, \mathbf{y}_n$ *are n independent random p-vectors of iid entries with mean zero and variance* 1. *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* $\Sigma$ *is an unknown* $p \times p$ *covariance matrix. Also, we assume that the entries of* $\mathbf{y}_k$*'s have finite fourth moments and assume that as* $p/n \to y \in (0, 1)$, *we have*

$$\frac{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}{n} \to a_1, \qquad \frac{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}{n} \to a_2, \qquad \frac{\mathbf{1}^T\Sigma^{-1}\boldsymbol{\mu}}{n}, \to a_3,$$

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

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

*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* $b = (1+\sqrt{y})^2$.

REMARK 3.3. From Proposition 2.1, the return takes the form $R^{(1)}$ if $\mathbf{1}^T\Sigma^{-1}\boldsymbol{\mu} < \sqrt{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}$. When $a_3 < 0$, for all large $n$, the condition for the first case holds, and hence, we obtain the limit for the first case. If $a_3 > 0$, the condition $\mathbf{1}^T\Sigma^{-1}\boldsymbol{\mu} < \sqrt{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}$ is eventually not true for all large $n$, and hence, the return takes the form $R^{(2)}$. When $a_3 = 0$, the case becomes very complicated. The return may attain the value in both cases and, hence, $\frac{\hat{\hat{R}}_p}{\sqrt{n}}$ may jump between the two limit points.

The computation of $\gamma$ will be presented in Appendix, and the proof of Theorem 3.2 can be easily derived from the following lemma.

LEMMA 3.1. *Under the assumptions of Theorem 3.2, we have*

(a)

$$\frac{\bar{\mathbf{x}}^T S^{-1}\bar{\mathbf{x}}}{n} \xrightarrow{\text{a.s.}} a_1 \cdot \gamma\ ,$$

(b)

$$\frac{\mathbf{1}^T S^{-1}\mathbf{1}}{n} \xrightarrow{\text{a.s.}} a_2 \cdot \gamma\ , \quad \textit{and}$$

(c)

$$\frac{\mathbf{1}^T S^{-1}\bar{\mathbf{x}}}{n} \xrightarrow{\text{a.s.}} a_3 \cdot \gamma$$

*where a.s. stands for "almost surely."*

The proof of Lemma 3.1 is given in Appendix. Nevertheless, from Theorems 3.1 and 3.2, we know that both the plug-in return, $\hat{R}_p$, and the estimator of the plug-in return, $\hat{\hat{R}}_p$, are always bigger than their theoretic optimal return, $R = \mathbf{c}^T\boldsymbol{\mu}$, defined in equation (2.1), and the difference is so big that both $\hat{R}_p$ and $\hat{\hat{R}}_p$ are not recommended for use in practice. Nonetheless, from Theorem 3.1, the estimator, $\hat{\hat{R}}_p$, of the plug-in return is found to be a good estimator of the plug-in return $\hat{R}_p$. Since $\hat{\hat{R}}_p$ is observable but not $\hat{R}_p$, we will use only $\hat{\hat{R}}_p$ in our computation, but we will use the notation $\hat{R}_p$ in this paper to represent both $\hat{\hat{R}}_p$ and $\hat{R}_p$ from now on if no confusion occurs.

We note that the returns being studied in the MV optimization procedure are usually assumed to be multivariate normally distributed. However, many studies (see, e.g., Fama 1963, 1965; Clark 1973; Blattberg and Gonedes 1974; Fielitz and Rozelle 1983) conclude that the normality assumption in the distribution of a security or portfolio return is violated. In this paper, we further relax the multivariate normality and multivariate stable distribution assumption to the existence of only finite fourth moment in Proposition 2.4 and Theorem 3.2 and to the existence of only finite second moment in Propositions 2.2 and 2.3. We also note that Michaud (1989) and others conclude that the estimator $\hat{R}_p$ in the MV optimization procedure generates large estimation error. Nonetheless, if the poor estimation is attributed to the estimation error, the estimate should be bigger or smaller than its theoretic value by chance. However, as we proved in Theorem 3.2, the estimate $\hat{R}_p$ is about $\sqrt{\gamma}$ times larger than the optimal return. Thus, this overprediction is a natural phenomenon, but not a random error.

The theoretical MV optimization procedure introduced by Markowitz more than half a century ago is expected to be a powerful tool, since it enables investors to efficiently allocate their wealth to different investment alternatives and reduce overall portfolio risk. However, this procedure has almost never been put into practice since its discovery because, so far, nobody has provided a good solution for overcoming the "Markowitz optimization enigma." Laloux et al. (1999) and others suggest that the determination of the correlation matrix and its eigenvalues could be the reason for the poor estimation. But so far, nobody has proved this phenomenon theoretically or provided a reasonably efficient estimator for the optimal return. In this connection, besides providing an explanation and thereafter proving the "Markowitz optimization enigma" as shown in the earlier part of this paper, this paper also aims to develop a good estimator of the optimal return to circumvent the overprediction problem. To achieve this, we incorporate both the LDRMT and the bootstrap technique to obtain a bootstrap estimate to correct the overprediction and to reduce the estimation error dramatically. We explain this approach in detail in the next subsection.

### 3.2. Bootstrap-Corrected Estimation

There are two main approaches to the bootstrap procedure: the nonparametric and parametric methods. Refer to Hall (1992) for more details. The basic idea in the nonparametric bootstrap technique is to use the empirical distribution to replace the unknown population distribution. Given a sample of $n$ iid random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ and a real-valued estimator $\theta(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ (denoted by $\hat{\theta}$) of the unknown parameter $\theta$, the bootstrap procedure is to construct another estimator $\hat{\theta}^*$ that is similar to the original estimator $\hat{\theta}$ based on a random sample $\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_n^*$ drawn from the empirical distribution function $F_n$ of the original sample.

Another approach is the parametric method of the bootstrap methodology described as follows: suppose that $\chi = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a sample drawn from the population $F_\theta$, where $\theta$ is the parameter to be estimated. Let $\hat{\theta} = \theta(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be an estimator of $\theta$. Then, a sample $\chi^* = \{\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*\}$ is drawn from the population $F_{\hat{\theta}}$. Thereafter, another estimator, $\hat{\theta}^* = \theta(\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*)$, of $\hat{\theta}$ can be constructed from the resample $\chi^*$. If the dimension of $\boldsymbol{x}_k$ is fixed, by the law of large numbers, $\hat{\theta}$ is close to $\theta$, and hence, the $F_{\hat{\theta}}$ is close to $F_\theta$ by the contiguity of distribution. As a result, the distribution of $\hat{\theta} - \theta$ will be similar to that of $\hat{\theta}^* - \hat{\theta}$. Repeating this resampling procedure, we can get as many iid bootstrap estimators $\hat{\theta}^*$ as desired. As such, we could use the empirical distribution of $\hat{\theta} - \hat{\theta}^*$ to approximate the unknown distribution of $\theta - \hat{\theta}$. Now, suppose that $\hat{\theta}$ is not a consistent estimator of $\theta$ but converges to a constant that is not equal to $\theta$ with probability one. When $n$ is large enough, we expect the relationship

$$\frac{\hat{\theta}^* - \hat{\theta}}{\hat{\theta} - \theta} \simeq \alpha \tag{3.4}$$

could still be held where $\alpha$ is a constant. Making use of this relationship, we obtain an estimate $\hat{\theta} - \frac{1}{\alpha}(\hat{\theta}^* - \hat{\theta})$. We expect it to be a consistent estimator for $\theta$.

In this paper, we will use the parametric approach of the bootstrap methodology to avoid possible singularity of the covariance matrix in the bootstrap sample. Now, we describe the procedure to construct a parametric bootstrap estimate from the estimate of the plug-in return, $\hat{R}_p$, defined in equation (3.2) as follows. First, draw a resample $\chi^* = \{\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*\}$ from the $p$-variate normal distribution with mean vector $\bar{\mathbf{x}}$ and covariance matrix $S$ defined in equation (2.2). Then, invoking Markowitz's optimization procedure again on the resample $\chi^*$, we obtain the **bootstrapped "plug-in" allocation**, $\hat{\mathbf{c}}_p^*$, and the **bootstrapped "plug-in" return**, $\hat{R}_p^*$, such that

$$\hat{R}_p^* = \hat{\mathbf{c}}_p^{*T} \bar{\mathbf{x}}^*, \tag{3.5}$$

where $\bar{\mathbf{x}}^* = \frac{1}{n} \sum_1^n \mathbf{x}_k^*$.

We note that the bootstrapped "plug-in" allocation $\hat{\mathbf{c}}_p^*$ will be different from the original "plug-in" allocation $\hat{\mathbf{c}}_p$ and, similarly, the bootstrapped "plug-in" return $\hat{R}_p^*$ is different from the "plug-in" return $\hat{R}_p$, but we expect that the relationship in equation (3.4) will still hold for $\hat{R}_p^*$ and $\hat{R}_p$ and for some $\alpha$. Thus, the **bootstrap-corrected return estimate** $\hat{R}_b$ can then be obtained such that

$$\hat{R}_b = \hat{R}_p + \frac{1}{\alpha}\left(\hat{R}_p - \hat{R}_p^*\right), \tag{3.6}$$

where $\hat{R}_p$ and $\hat{R}_p^*$ are defined in equations (3.3) and (3.5), respectively.

We expect that this estimate would perform well in that it will be much closer to the theoretic value of the optimal return, $R$, when $n$ is large. In this paper, it is our goal to look for the best $\alpha$ in equation (3.6). To achieve this, we introduce the following theorem.

THEOREM 3.3. *Under the conditions in Theorem 3.2 and using the bootstrapped plug-in procedure as described above, we have*

$$\sqrt{\gamma}(R - \hat{R}_p) \simeq \hat{R}_p - \hat{R}_p^*, \tag{3.7}$$

*where $\gamma$ is defined in Theorem 3.2, $R$ is the theoretic optimal return obtained from Proposition 2.1, $\hat{R}_p$ is the plug-in return estimate defined in equation (3.3) and obtained by using the original sample $\chi$, and $\hat{R}_p^*$ is the bootstrapped plug-in return estimate defined in equation (3.6) and obtained by using the bootstrapped sample $\chi^*$, respectively.*

In Theorem 3.2, the relationship $\hat{R}_p \simeq \sqrt{\gamma}\, R$ has been proved. As the relationship $\hat{R}_p^* \simeq \sqrt{\gamma}\, \hat{R}_p$ is its dual conclusion, the proof of Theorem 3.3 follows immediately. Now, we are ready to construct a consistent estimate for the optimal return, $R$, in accordance with the value of the ratio of dimension to sample size. From equation (3.7), we have

$$\sqrt{\gamma}\left(\mathbf{c}^T \boldsymbol{\mu} - \hat{\mathbf{c}}_p^T \bar{X}\right) \simeq \hat{\mathbf{c}}_p^T \bar{X} - \hat{\mathbf{c}}_p^{*T} \bar{X}^* .$$

Using this relationship, we construct the **bootstrap-corrected allocation**, $\hat{\mathbf{c}}_b$, and then construct the **bootstrap-corrected return estimate**, $\hat{R}_b$, as stated in the following theorem.

THEOREM 3.4. *Under the conditions in Theorem 3.2 and using the bootstrap correction procedure described above, the* ***bootstrap-corrected allocation***, $\hat{\mathbf{c}}_b$, *and* ***bootstrap-corrected return estimate***, $\hat{R}_b$, *are given by*

$$\begin{aligned} \hat{\mathbf{c}}_b &= \hat{\mathbf{c}}_p + \frac{1}{\sqrt{\gamma}}\left(\hat{c}_p - \hat{\mathbf{c}}_p^*\right), \\ \hat{R}_b &= \hat{R}_p + \frac{1}{\sqrt{\gamma}}\left(\hat{R}_p - \hat{R}_p^*\right), \end{aligned} \tag{3.8}$$

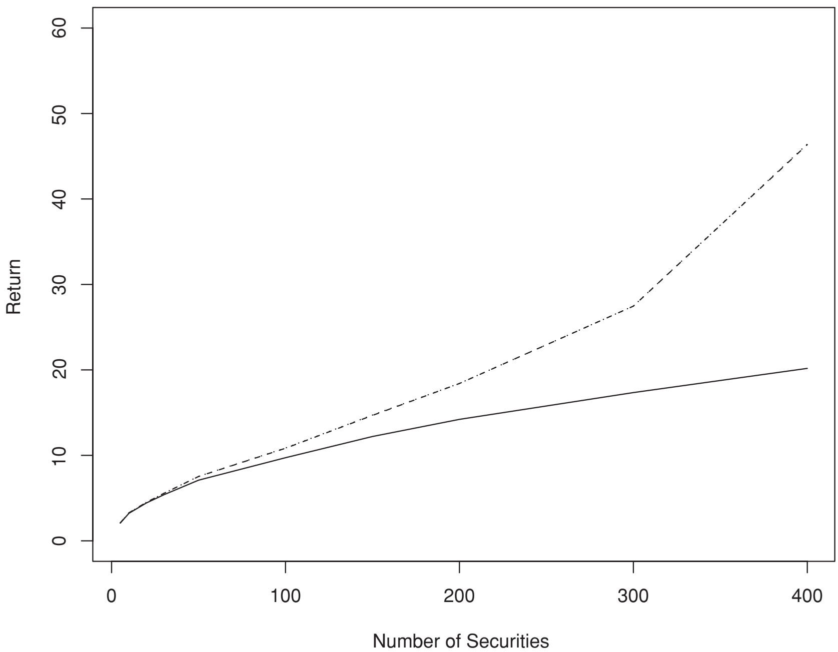
*where* $\hat{\mathbf{c}}_p$ *and* $\hat{\mathbf{c}}_p^*$ *are the plug-in and bootstrapped plug-in allocations defined in equations (3.3) and (3.5), respectively,* $\hat{R}_p$ *and* $\hat{R}_p^*$ *are their corresponding plug-in and bootstrapped plug-in returns defined in equations (3.3) and (3.5), respectively, and* $\gamma$ *is defined in Theorem 3.2.*

We note that Michaud (see Michaud 1998 and Meucci 2005) provides a bootstrap estimation approach to estimate the allocation and thereafter estimate the return. His resampling estimator of the allocation is equivalent to the $\hat{\mathbf{c}}_p^*$ in our paper. We also note that Michaud's method is to repeat 3,500 times and then take the average as the estimate of the allocation. Theoretically, we may interpret Michaud's estimator as the expected plug-in estimator when the training sample is considered as the population. Applying the analysis in our paper, we obtain $\hat{R}_p / R_p \to \sqrt{\gamma}$ $(= \sqrt{1/(1-y)}\,)$, which implies $\hat{R}_p^* / \hat{R}_p \to \sqrt{\gamma}$; therefore, $\hat{R}_p^* / R_p \to \gamma$. We have simulated our proposed estimate, plug-in estimate, and the estimate proposed by Michaud. Our result shows that ours is much better than Michaud's. The results are available on request.

In our simulation, the desired properties of consistency and efficiency are found for the bootstrap-corrected return estimate $\hat{R}_b$ for any number of assets, regardless of whether they are small or large. Figure 4.2 displayed in the next section shows the obvious merit of our new bootstrap-corrected optimal return estimate $\hat{R}_b$ defined in equation (3.8) over the plug-in return estimate $\hat{R}$ defined in equation (3.3). For instance, for $n = 500$ and $p = 200$ from the figure, we observe that $\hat{R}_b$ is close to the true optimal return $R$, but $\hat{R}$ is far away from the true optimal return. We will demonstrate the superiority of our estimates by conducting simulations as presented in the next section.

## 4. SIMULATION STUDY

In this section, we first demonstrate the overprediction problem by displaying the quantity of the overprediction increases as the dimension increases. Second, we present simulation results on comparisons between the bootstrap-corrected estimates and the plug-in estimates for the return and allocation by means of both mean square errors and relative efficiencies.




Solid line—the theoretic optimal return ($R$);
Dashed line—the plug-in return ($\hat{R}_p$);
Dotted line—the estimate of plug-in return ($\hat{\hat{R}}_p$).

Note: The dashed line and dotted line are coincidental in the entire range.
FIGURE 4.1. Empirical and theoretical optimal returns for different numbers of assets.

## 4.1. Overprediction

To illustrate the overprediction problem, for simplicity we generate $p$-branch standardized security returns from a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$ and identity covariance matrix $\Sigma = (I_{jk})$ in which $I_{jk} = 1$ when $j = k$ and $I_{jk} = 0$ otherwise. Given the level of risk with the known population mean vector, $\boldsymbol{\mu}$, and known population covariance matrix, $\Sigma$, we can compute the theoretic optimal allocation, $\mathbf{c}$, and, thereafter, compute the theoretic optimal return, $R$, for the portfolios. These values will then be used to compare the performance of all the estimators being studied in our paper. Using this data set, we apply the formula in equation (2.2) to compute the sample mean, $\bar{\mathbf{x}}$, and sample covariance, $S$, which, in turn, enables us to obtain the plug-in return, $\hat{R}_p$, and its corresponding plug-in allocation, $\hat{\mathbf{c}}_p$, by substituting $\bar{\mathbf{x}}$ and $S$ into $\boldsymbol{\mu}$ and $\Sigma$, respectively, in the formula of $\hat{R}_p$ and $\hat{\mathbf{c}}_p$ shown in Proposition 2.1. To illustrate the overprediction problem, we first plot the theoretic optimal returns, $R$, and the plug-in returns, $\hat{R}_p$, for different values of $p$ with the same sample size $n = 500$ in Figure 4.1. For further evaluation, we depict the simulation theoretic optimal returns, $R$, and the plug-in returns, $\hat{R}_p$, in Table 4.1 for two different cases: (a) for different values of $p$ with the same dimension-to-sample-size ratio $p/n$ (=0.5) and (b) for the same value of $p$ (=252[15]) but different dimension-to-sample-size ratios $p/n$.

From Figure 4.1 and Table 4.1, we find the following: (1) both the plug-in return $\hat{R}_p$ defined in equation (3.2) and its estimate $\hat{\hat{R}}_p$ defined in (3.3) are good estimates of the

[15] We choose this number to ensure that $n$ is an integer for the different ratios being chosen.

TABLE 4.1
Performance of $\hat{R}_p$ and $\hat{\hat{R}}_p$ over the Optimal Return $R$ for Different Values of $p$ and for Different Values of $p/n$

| $p$ | $p/n$ | $R$ | $\hat{R}_p$ | $\hat{\hat{R}}_p$ | $p$ | $p/n$ | $R$ | $\hat{R}_p$ | $\hat{\hat{R}}_p$ |
|---|---|---|---|---|---|---|---|---|---|
| 100 | 0.5 | 9.77 | 13.89 | 13.96 | 252 | 0.5 | 14.71 | 20.95 | 21.00 |
| 200 | 0.5 | 13.93 | 19.67 | 19.73 | 252 | 0.6 | 14.71 | 23.42 | 23.49 |
| 300 | 0.5 | 17.46 | 24.63 | 24.66 | 252 | 0.7 | 14.71 | 26.80 | 26.92 |
| 400 | 0.5 | 19.88 | 27.83 | 27.85 | 252 | 0.8 | 14.71 | 33.88 | 34.05 |
| 500 | 0.5 | 22.29 | 31.54 | 31.60 | 252 | 0.9 | 14.71 | 48.62 | 48.74 |

Note: The table compares the performance between $\hat{R}_p$ and $\hat{\hat{R}}_p$ for same $p/n$ ratio with different numbers of assets, $p$, and for same $p$ with different $p/n$ ratio, where $n$ is number of sample, $R$ is the optimal return defined in equation (2.1), $\hat{R}_p$ and $\hat{\hat{R}}_p$ are defined in equations (3.2) and (3.3), respectively.

theoretic optimal return $R$ when $p$ is small ($\leq 30$); (2) when $p$ is large ($\geq 60$), the difference between the theoretic optimal return $R$ and the plug-in return $\hat{R}_p$ (or $\hat{\hat{R}}_p$) becomes dramatically large; (3) the larger the $p$, the greater the difference; and (4) when $p$ is large, both the plug-in return $\hat{R}_p$ and its estimate $\hat{\hat{R}}_p$ are always larger than the theoretic optimal return, $R$, computed by using the true mean and covariance matrix. These confirm the "Markowitz optimization enigma" that the plug-in returns $\hat{R}_p$ should not be used in practice. In addition, Figure 4.1 and Table 4.1 confirm a fairly high congruence between $\hat{\hat{R}}_p$ and $\hat{R}_p$ for all values of $p$. Hence, in this paper we will use $\hat{R}_p$ to represent both $\hat{\hat{R}}_p$ and $\hat{R}_p$ if no confusion occurs.

## 4.2. Bootstrap Correction Method

In this section, our simulation is to show the superiority of both $\hat{R}_b$ and $\hat{\mathbf{c}}_b$ over their plug-in counterparts $\hat{R}_p$ and $\hat{\mathbf{c}}_p$. To this end, we first define the **bootstrap-corrected difference**, $d_b^R$, for the return as the difference between the bootstrap-corrected optimal return estimate $\hat{R}_b$ and the theoretic optimal return $R$; that is,

$$d_b^R = \hat{R}_b - R, \tag{4.1}$$

which will be used to compare the **plug-in difference**,

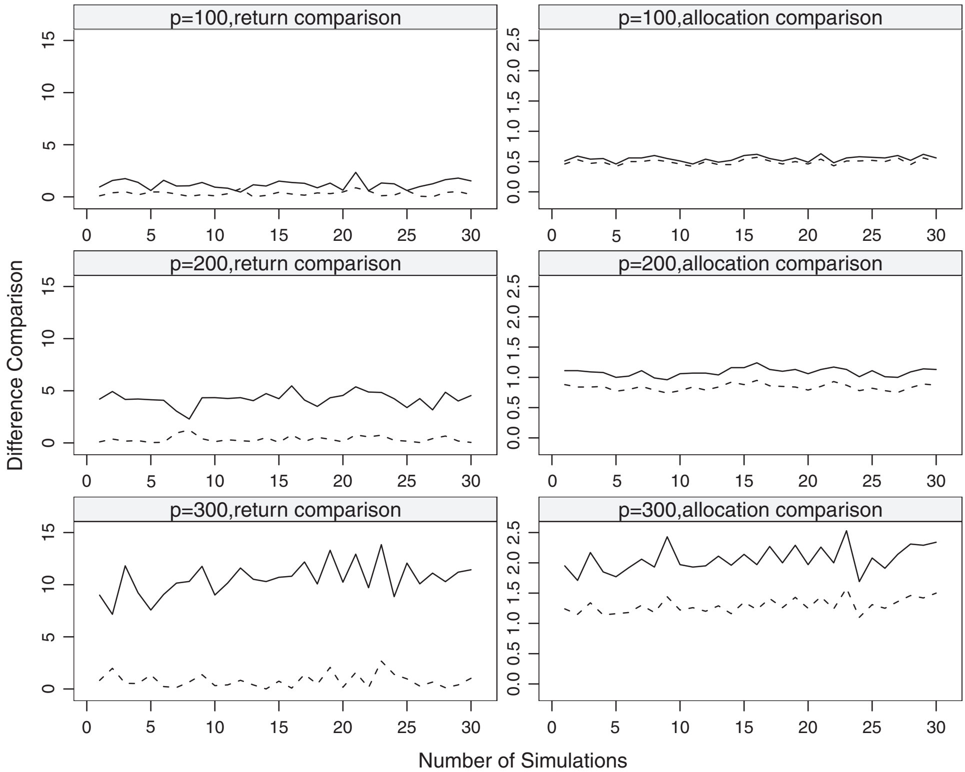
$$d_p^R = \hat{R}_p - R \tag{4.2}$$

for the return, where $\hat{R}_p$ and $\hat{R}_b$ are defined in equations (3.3) and (3.8), respectively.

To compare the bootstrapped allocation with the plug-in allocation, we define the **bootstrap-corrected difference norm**, $d_b^c$, and the **plug-in difference norm**, $d_p^c$, for the allocations to be

$$d_b^c = \|\hat{\mathbf{c}}_b - \mathbf{c}\| \quad \text{and} \quad d_p^c = \|\hat{\mathbf{c}}_p - \mathbf{c}\|, \tag{4.3}$$

where $d_b^c$ is the difference norm between the bootstrap-corrected allocation $\hat{\mathbf{c}}_b$ and the theoretic optimal allocation $\mathbf{c}$, while $d_p^c$ is the plug-in difference norm between the plug-in allocation $\hat{\mathbf{c}}_p$ and the theoretic optimal allocation $\mathbf{c}$. We then simulate 30 times to




Solid line—the absolute value of $d_p^c$ and $d_p^R$, respectively;
Dashed line—the absolute value of $d_b^c$ and $d_b^R$, respectively.

Note: The top, middle, and bottom two subfigures are the plot for $p = 100$, 200, and 300, respectively. The plots on the left are the plots for $d_p^R$ and $d_b^R$, while the plots on the right are the plots for $d_p^c$ and $d_b^c$, respectively. $d_b^R$, $d_p^R$, $d_b^c$, and $d_p^c$ are defined in equations (4.1)–(4.3), respectively.

FIGURE 4.2. Comparison between the empirical and corrected portfolio allocations and returns.

compute $d_x^R$ and $d_x^c$ for $x = p$ and $b$, $n = 500$ and $p = 100$, 200, and 300. The results are displayed in Table 4.1 and depicted in Figure 4.2.

From Table 4.2 and Figure 4.2, we find the desired property that $d_b^R$ ($d_b^c$) is much smaller than $d_p^R$ ($d_p^c$) in absolute value for all cases. This infers that the estimate obtained by utilizing the bootstrap-corrected method is much more accurate in estimating the theoretic value than that obtained by using the plug-in procedure. Furthermore, as $p$ increases, the two lines of $d_p^R$ and $d_b^R$ (or $d_p^c$ and $d_b^c$) on each level as shown in Figure 4.2 separate further, implying that the magnitude of improvement from $d_p^R$ ($d_p^c$) to $d_b^R$ ($d_b^c$) is remarkable.

To further illustrate the superiority of our estimate over the traditional plug-in estimate, we present in Table 4.3 the mean square errors (MSEs) of the different estimates for different $p$ and plot these values in Figure 4.3. In addition, we define their **relative efficiencies (REs)** for both allocations and returns to be

TABLE 4.2
Comparison between the Empirical and Corrected Portfolio Returns and Allocations

| | $p = 100$ and $n = 500$ | | | | $p = 200$ and $n = 500$ | | | | $p = 300$ and $n = 500$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | $d_p^R$ | $d_b^R$ | $d_p^c$ | $d_b^c$ | $d_p^R$ | $d_b^R$ | $d_p^c$ | $d_b^c$ | $d_p^R$ | $d_b^R$ | $d_p^c$ | $d_b^c$ |
| 1 | 0.95 | −0.10 | 0.51 | 0.46 | 4.20 | 0.10 | 1.11 | 0.88 | 9.00 | −0.82 | 1.95 | 1.24 |
| 2 | 1.57 | 0.39 | 0.59 | 0.53 | 4.93 | 0.37 | 1.11 | 0.84 | 7.16 | −1.98 | 1.71 | 1.15 |
| 3 | 1.75 | 0.49 | 0.54 | 0.47 | 4.17 | 0.17 | 1.09 | 0.84 | 11.81 | 0.54 | 2.17 | 1.34 |
| 4 | 1.39 | 0.19 | 0.55 | 0.49 | 4.21 | 0.22 | 1.08 | 0.85 | 9.23 | −0.52 | 1.85 | 1.14 |
| 5 | 0.61 | −0.47 | 0.46 | 0.42 | 4.14 | 0.03 | 1.00 | 0.77 | 7.57 | −1.35 | 1.77 | 1.16 |
| 6 | 1.59 | 0.47 | 0.56 | 0.50 | 4.09 | −0.06 | 1.02 | 0.80 | 9.06 | −0.23 | 1.92 | 1.18 |
| 7 | 1.04 | −0.27 | 0.56 | 0.50 | 3.05 | −0.95 | 1.11 | 0.85 | 10.15 | −0.14 | 2.06 | 1.30 |
| 8 | 1.06 | −0.05 | 0.60 | 0.53 | 2.29 | −1.24 | 0.99 | 0.79 | 10.31 | −0.67 | 1.93 | 1.18 |
| 9 | 1.38 | 0.21 | 0.55 | 0.50 | 4.33 | 0.40 | 0.96 | 0.74 | 11.76 | 1.37 | 2.43 | 1.44 |
| 10 | 0.93 | −0.10 | 0.51 | 0.46 | 4.34 | 0.12 | 1.06 | 0.78 | 9.01 | −0.32 | 1.97 | 1.22 |
| 11 | 0.83 | −0.30 | 0.46 | 0.42 | 4.26 | 0.30 | 1.07 | 0.84 | 10.16 | −0.39 | 1.93 | 1.26 |
| 12 | 0.46 | −0.79 | 0.54 | 0.50 | 4.34 | −0.21 | 1.07 | 0.79 | 11.60 | 0.83 | 1.95 | 1.20 |
| 13 | 1.16 | 0.00 | 0.49 | 0.45 | 4.05 | −0.14 | 1.04 | 0.84 | 10.52 | 0.39 | 2.11 | 1.29 |
| 14 | 1.04 | −0.14 | 0.52 | 0.45 | 4.73 | 0.51 | 1.16 | 0.92 | 10.30 | 0.00 | 1.96 | 1.16 |
| 15 | 1.51 | 0.44 | 0.60 | 0.54 | 4.24 | 0.05 | 1.16 | 0.88 | 10.71 | 0.75 | 2.14 | 1.35 |
| 16 | 1.38 | 0.27 | 0.62 | 0.57 | 5.47 | 0.77 | 1.24 | 0.95 | 10.81 | 0.08 | 1.97 | 1.23 |
| 17 | 1.30 | 0.16 | 0.55 | 0.51 | 4.11 | 0.14 | 1.13 | 0.86 | 12.18 | 1.40 | 2.27 | 1.41 |
| 18 | 0.87 | −0.38 | 0.51 | 0.46 | 3.50 | −0.54 | 1.10 | 0.85 | 10.07 | −0.46 | 2.00 | 1.26 |
| 19 | 1.32 | 0.31 | 0.56 | 0.50 | 4.33 | 0.34 | 1.13 | 0.84 | 13.30 | 2.07 | 2.29 | 1.43 |
| 20 | 0.64 | −0.47 | 0.49 | 0.46 | 4.55 | −0.10 | 1.06 | 0.79 | 10.24 | 0.16 | 1.97 | 1.25 |
| 21 | 2.35 | 0.87 | 0.63 | 0.54 | 5.38 | 0.77 | 1.13 | 0.85 | 12.93 | 1.58 | 2.26 | 1.44 |
| 22 | 0.59 | −0.57 | 0.48 | 0.43 | 4.88 | 0.58 | 1.17 | 0.93 | 9.71 | 0.13 | 2.00 | 1.24 |
| 23 | 1.33 | 0.11 | 0.56 | 0.51 | 4.84 | 0.74 | 1.13 | 0.87 | 13.84 | 2.67 | 2.53 | 1.58 |
| 24 | 1.25 | 0.17 | 0.58 | 0.51 | 4.24 | 0.25 | 1.01 | 0.78 | 8.85 | −1.40 | 1.69 | 1.10 |
| 25 | 0.61 | −0.60 | 0.57 | 0.52 | 3.39 | −0.17 | 1.11 | 0.82 | 12.07 | 0.97 | 2.08 | 1.31 |
| 26 | 0.99 | −0.06 | 0.56 | 0.50 | 4.25 | 0.04 | 1.01 | 0.78 | 10.08 | −0.28 | 1.91 | 1.25 |
| 27 | 1.25 | 0.01 | 0.60 | 0.56 | 3.17 | −0.41 | 1.00 | 0.75 | 11.11 | 0.66 | 2.14 | 1.36 |
| 28 | 1.65 | 0.42 | 0.52 | 0.45 | 4.86 | 0.66 | 1.09 | 0.83 | 10.29 | −0.12 | 2.31 | 1.46 |
| 29 | 1.80 | 0.51 | 0.62 | 0.56 | 4.02 | 0.18 | 1.14 | 0.89 | 11.20 | 0.39 | 2.29 | 1.42 |
| 30 | 1.53 | 0.22 | 0.56 | 0.50 | 4.54 | −0.05 | 1.13 | 0.87 | 11.43 | 1.02 | 2.34 | 1.50 |

Note: $d_b^R$, $d_p^R$, $d_b^c$, and $d_p^c$ are defined in equations (4.1)–(4.3), respectively. $k$ is the number of simulations.

$$RE_{p,b}^c = \frac{MSE(d_p^c)}{MSE(d_b^c)} \quad \text{and} \quad RE_{p,b}^R = \frac{MSE(d_p^R)}{MSE(d_b^R)} \tag{4.4}$$

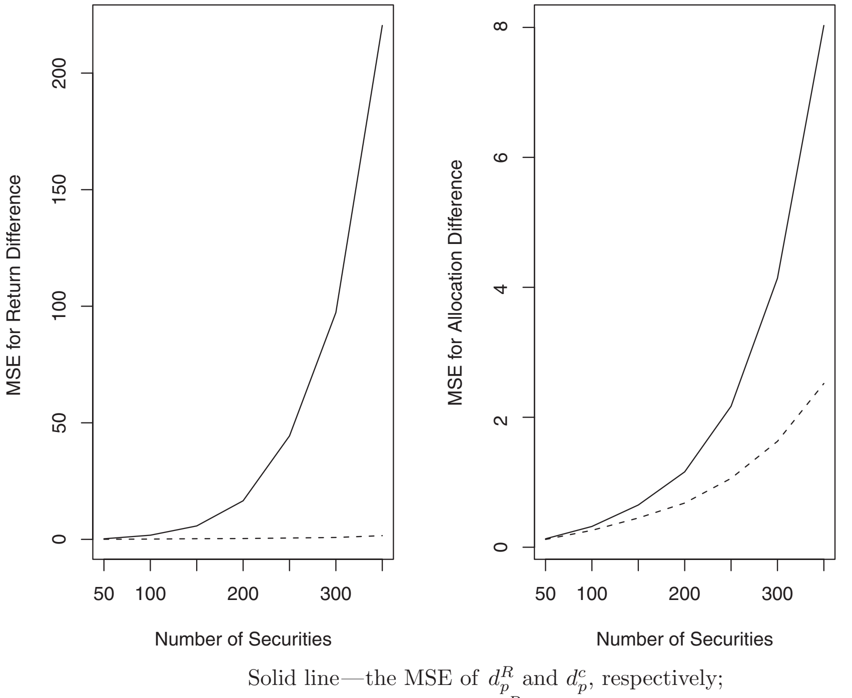
and report their values in Table 4.3.

Comparing the MSE of $d_b^R$ ($d_b^c$) with that of $d_p^R$ ($d_p^c$) in Table 4.3 and Figure 4.3, the MSEs of both $d_b^R$ and $d_b^c$ have been reduced dramatically from those of $d_p^R$ and $d_p^c$, indicating that our proposed estimates are superior. We find that the MSE of $d_b^R$ is only 0.04, improving 6.25 times from that of $d_p^R$ when $p = 50$. When the number of assets increases, the improvement becomes much more substantial. For example,

TABLE 4.3
MSE and Relative Efficiency Comparison

| $p$ | MSE($d_p^R$) | MSE($d_b^R$) | MSE($d_p^c$) | MSE($d_b^c$) | $RE_{p,b}^R$ | $RE_{p,b}^c$ |
|---|---|---|---|---|---|---|
| $p = 50$ | 0.25 | 0.04 | 0.13 | 0.12 | 6.25 | 1.08 |
| $p = 100$ | 1.79 | 0.12 | 0.32 | 0.26 | 14.92 | 1.23 |
| $p = 150$ | 5.76 | 0.29 | 0.65 | 0.45 | 19.86 | 1.44 |
| $p = 200$ | 16.55 | 0.36 | 1.16 | 0.68 | 45.97 | 1.71 |
| $p = 250$ | 44.38 | 0.58 | 2.17 | 1.06 | 76.52 | 2.05 |
| $p = 300$ | 97.30 | 0.82 | 4.14 | 1.63 | 118.66 | 2.54 |
| $p = 350$ | 220.43 | 1.59 | 8.03 | 2.52 | 138.64 | 3.19 |

Note: $p$ is number of assets. $d_p^R$, $d_b^R$, $RE_{p,b}^R$, and $RE_{p,b}^c$ are defined in equations (4.1), (4.2), and (4.4), respectively.




Solid line—the MSE of $d_p^R$ and $d_p^c$, respectively;
Dashed line—the MSE of $d_b^R$ and $d_b^c$, respectively.

Note: The plots on the left are the plots of the MSEs for $d_p^R$ and $d_b^R$, while the plots on the right are the plots of the MSEs for $d_p^c$ and $d_b^c$, respectively. $d_b^R$, $d_p^R$, $d_b^c$, and $d_p^c$ are defined in equations (4.1)–(4.4), respectively.
FIGURE 4.3. MSE comparison between the empirical and corrected portfolio allocations/returns.

when $p = 350$, the MSE of $d_b^R$ is only 1.59 but the MSE of $d_p^R$ is 220.43, improving 138.64 times from that of $d_p^R$. This is an unbelievable improvement. We note that when both $n$ and $p$ are bigger, the relative efficiency of our proposed estimate over the traditional plug-in estimate could be much larger. On the other hand, the improvement from $d_p^c$ to $d_b^c$ is also tremendous.

## 5. ILLUSTRATION

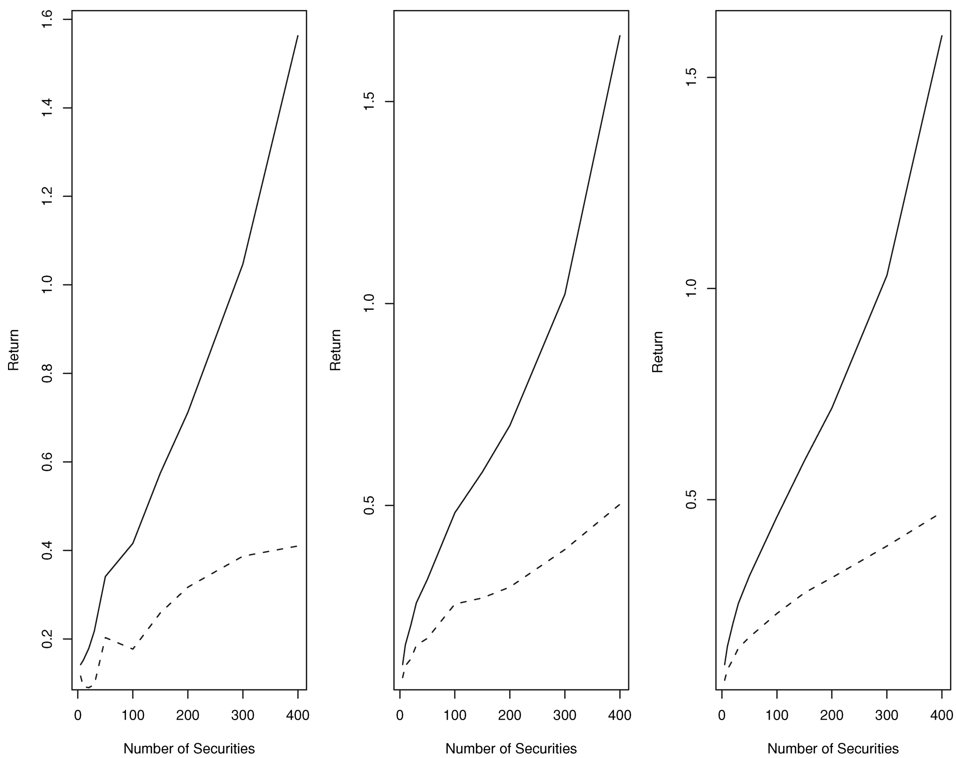
We illustrate the superiority of our approach by comparing the estimates of the bootstrap-corrected return and the plug-in return for daily S&P500 data. To match our simulation of $n = 500$ as shown in Table 4.1 and Figure 4.1, we choose 500 daily data backward from December 30, 2005, for all companies listed in the S&P500 as the database for our estimation. We then choose the number of assets ($p$) from 5 to 400, and for each $p$, we select $p$ stocks from the S&P500 database randomly without replacement and compute the plug-in return and the corresponding bootstrap-corrected return. We plot the plug-in returns and the corresponding bootstrap-corrected returns in Figure 5.1 and report these returns and their ratios in Table 5.1 for different $p$. We also repeat the procedure ($m$=) 10 and 100 times for checking. For each $m$ and for each $p$, we first compute the bootstrap-corrected returns and the plug-in returns. Thereafter, we compute their averages for both the bootstrap-corrected returns and the plug-in returns and plot these values in Panels 2 and 3 of Figure 5.1, respectively, for comparison with the results in Panel 1 for $m = 1$.

From Table 5.1 and Figure 5.1, we find that as the number of assets increases, (1) the values of the estimates from both the bootstrap-corrected returns and the plug-in returns for the S&P500 database increase, and (2) the values of the estimates of the plug-in returns increase much faster than those of the bootstrap-corrected returns and thus their differences become wider. These empirical findings are consistent with the theoretical discovery of the "Markowitz optimization enigma" that the estimated plug-in return is always larger than its theoretical value and their difference becomes larger when the number of assets is large.

TABLE 5.1
Plug-in Returns and Bootstrap-Corrected Returns

| | $m = 1$ | | | $m = 10$ | | | $m = 100$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $p$ | $\hat{R}_p$ | $\hat{R}_b$ | $\hat{R}_b/\hat{R}_p$ | $\hat{R}_p$ | $\hat{R}_b$ | $\hat{R}_b/\hat{R}_p$ | $\hat{R}_p$ | $\hat{R}_b$ | $\hat{R}_b/\hat{R}_p$ |
| 5 | 0.142 | 0.116 | 0.820 | 0.106 | 0.074 | 0.670 | 0.109 | 0.072 | 0.632 |
| 10 | 0.152 | 0.092 | 0.607 | 0.155 | 0.103 | 0.650 | 0.152 | 0.097 | 0.616 |
| 20 | 0.179 | 0.09 | 0.503 | 0.204 | 0.120 | 0.576 | 0.206 | 0.121 | 0.573 |
| 30 | 0.218 | 0.097 | 0.447 | 0.259 | 0.154 | 0.589 | 0.254 | 0.148 | 0.576 |
| 50 | 0.341 | 0.203 | 0.597 | 0.317 | 0.171 | 0.529 | 0.319 | 0.174 | 0.541 |
| 100 | 0.416 | 0.177 | 0.426 | 0.482 | 0.256 | 0.530 | 0.459 | 0.230 | 0.498 |
| 150 | 0.575 | 0.259 | 0.450 | 0.583 | 0.271 | 0.463 | 0.592 | 0.279 | 0.469 |
| 200 | 0.712 | 0.317 | 0.445 | 0.698 | 0.298 | 0.423 | 0.717 | 0.315 | 0.438 |
| 300 | 1.047 | 0.387 | 0.369 | 1.023 | 0.391 | 0.381 | 1.031 | 0.390 | 0.377 |
| 400 | 1.563 | 0.410 | 0.262 | 1.663 | 0.503 | 0.302 | 1.599 | 0.470 | 0.293 |

Note: $\hat{R}_p$ and $\hat{R}_b$ are defined in equations (3.3) and (3.8), respectively.




Solid line—Plug-in Return;
Dashed line—Bootstrap-Corrected Return.

Note: The left, middle, and right figures are the plots for $m = 1, 10$, and 100, respectively.
FIGURE 5.1. Comparison between the plug-in returns and bootstrap-corrected returns.

Comparing Figures 5.1 and 4.1 (or Tables 5.1 and 4.1), one will find that the shapes of the graphs of both the bootstrap-corrected returns and the corresponding plug-in returns are similar to those in Figure 4.1. This infers that our empirical findings based on the S&P500 are consistent with our theoretical and simulation results, which, in turn, confirm that our proposed bootstrap-corrected return performs better.

One may doubt the existence of bias in our sampling as we choose only one sample in the analysis. To circumvent this problem, we, in addition, repeat the procedure $m$ (=10, 100) times. For each $m$ and for each $p$, we compute the bootstrap-corrected returns and the plug-in returns and then compute their averages for both the bootstrap-corrected returns and the plug-in returns. Thereafter, we plot the averages of the returns in Figure 5.1 and report these averages and their ratios in Table 5.1 for $m = 10$ and 100. When comparing the values of the returns for $m = 10$ and 100 with $m = 1$, we find that the plots are basically of the similar values for each $p$ but become smoother, inferring that the sampling bias has been eliminated by increasing the number of $m$. The results for $m = 10$ and 100 are also consistent with the plot in Figure 4.1 in our simulation, inferring that our bootstrap-corrected return is a better estimate for the theoretical return in the sense that its value is much closer to the theoretical return when compared with the corresponding plug-in return.

## 6. CONCLUSION

Being of both theoretical and practical interest, the basic problem for MV analysis is identifying those combinations of assets that constitute attainable efficient portfolios. Unfortunately, there are problems that accompany any MV analysis. With this in mind, this paper sets out to solve this dilemma by developing a new optimal return estimate to capture the essence of portfolio selection. Since our approach is easy to operate and implement in practice, the whole 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.

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, bonds, stocks, and other assets.

For instance, the optimization problem can be formulated with short-sales restrictions, trading costs, liquidity constraints, turnover constraints, and budget constraints[16]; see, for example, Detemple and Rindisbacher (2005), Muthuraman and Kumar (2006), and Lakner and Nygren (2006). Each of these constraints leads back to a different model for determining the shape, composition, and characteristics of the efficient frontier and, thereafter, makes MV optimization a more flexible tool. For example, Xia (2005) investigates the problem with a nonnegative wealth constraint in a semimartingale model. Another direction for further research is to adopt the continuous-time multiperiod Markowitz's problem (see, e.g., Li and Ng 2000; Emmer, Klüppelberg, and Korn 2001; Xia and Yan 2006). Further research could also include conducting an extensive analysis to compare the performance of our estimators with other state-of-the-art estimators in the literature, for example, factor models or Bayesian shrinkage estimators.

Finally, we note 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) 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. In addition, we relax the condition to the existence of the second moments for some cases and to the fourth moments for some other cases. The returns could follow any distribution, and furthermore, they are not necessarily identically distributed in our proposed approach.

## APPENDIX

*Proof of Proposition 2.1.* Let $\tilde{\mathbf{1}} = \Sigma^{-1/2}\mathbf{1}$, $\tilde{\boldsymbol{\mu}} = \Sigma^{-1/2}\boldsymbol{\mu}$, and $\tilde{\mathbf{c}} = \Sigma^{1/2}c$, then the problem in equation (2.1) becomes

$$\max \tilde{\mathbf{c}}^T \tilde{\boldsymbol{\mu}}, \quad \text{subject to} \quad \tilde{\mathbf{c}}^T \tilde{\mathbf{1}} \le 1, \quad \text{and} \quad \tilde{\mathbf{c}}^T \tilde{\mathbf{c}} \le \sigma_0^2. \tag{A.1}$$

[16] We note that after imposing any of these additional constraints, there may not be any explicit solution. Even if an explicit solution can be found, it could be very complicated and the development of the theory could be very tedious. We will address this question in our further research.

Thereafter, we let

$$\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}} - P_{\tilde{\mathbf{1}}}\tilde{\boldsymbol{\mu}},$$

where $P_{\tilde{\mathbf{1}}}$ is a projection matrix. As

$$P_{\tilde{\mathbf{1}}} = \frac{\tilde{\mathbf{1}}\tilde{\mathbf{1}}^T}{\tilde{\mathbf{1}}^T\tilde{\mathbf{1}}},$$

we have

$$\hat{\boldsymbol{\mu}} = \left(I - \frac{\tilde{\mathbf{1}}\tilde{\mathbf{1}}^T}{\tilde{\mathbf{1}}^T\tilde{\mathbf{1}}}\right)\tilde{\boldsymbol{\mu}}.$$

Thus, $\tilde{\mathbf{c}}$ defined in equation (A.1) can then be decomposed as

$$\tilde{\mathbf{c}} = x\tilde{\mathbf{1}} + y\hat{\boldsymbol{\mu}} + \hat{\mathbf{z}},$$

where $\hat{\mathbf{z}} \perp (\tilde{\mathbf{1}}, \hat{\boldsymbol{\mu}})$. The problem in equation (A.1) then becomes

$$\max\ \tilde{\mathbf{c}}^T\tilde{\boldsymbol{\mu}} = \max(x\tilde{\mathbf{1}}^T\tilde{\boldsymbol{\mu}} + y\hat{\boldsymbol{\mu}}^T\tilde{\boldsymbol{\mu}}) \tag{A.2}$$

subject to

$$x\tilde{\mathbf{1}}^T\tilde{\mathbf{1}} \le 1, \quad \tilde{\mathbf{c}}^T\tilde{\mathbf{c}} = x^2\tilde{\mathbf{1}}^T\tilde{\mathbf{1}} + y^2\hat{\boldsymbol{\mu}}^T\hat{\boldsymbol{\mu}} + |\hat{\mathbf{z}}|^2 \le \sigma_0^2.$$

Obviously, to maximize the objective of equation (A.2), we have $\hat{\mathbf{z}} = 0$. In addition, if we consider the maximization problem under only the second restriction, $\tilde{\mathbf{c}}^T\tilde{\mathbf{c}} = x^2\tilde{\mathbf{1}}^T\tilde{\mathbf{1}} + y^2\hat{\boldsymbol{\mu}}^T\hat{\boldsymbol{\mu}} + |\hat{\mathbf{z}}|^2 \le \sigma_0^2$, the solution will be

$$\tilde{c} = \frac{\sigma_0\tilde{\boldsymbol{\mu}}}{\|\tilde{\boldsymbol{\mu}}\|} \quad \text{or, equivalently,} \quad c = \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}}\Sigma^{-1}\boldsymbol{\mu}. \tag{A.3}$$

Therefore, if

$$\mathbf{c}^T\mathbf{1} = \frac{\sigma_0\boldsymbol{\mu}^T\Sigma^{-1}\mathbf{1}}{\sqrt{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}} \le 1,$$

then equation (A.3) is the solution of optimal allocation to the maximization problem in equation (2.1). Otherwise, the solution of the maximization can be obtained by solving the equations

$$\max\left\{x\tilde{\mathbf{1}}^T\tilde{\boldsymbol{\mu}} + y\hat{\boldsymbol{\mu}}^T\tilde{\boldsymbol{\mu}}\right) \quad \text{subject to} \quad x\tilde{\mathbf{1}}^T\tilde{\mathbf{1}} = 1 \quad \text{and} \quad x^2\tilde{\mathbf{1}}^T\tilde{\mathbf{1}} + y^2\hat{\boldsymbol{\mu}}^T\hat{\boldsymbol{\mu}} = \sigma_0^2\right\}. \tag{A.4}$$

Applying the Lagrange method to solve equation (A.4), one could easily obtain the solutions as given by Proposition 2.1. □

*Proof of Lemma 3.1.* Before proceeding to the proof of Lemma 3.1, we first present two lemmas as follows.

LEMMA A.2. *Suppose* $\frac{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}{n} \to a$ *and* $S = \frac{1}{n}\sum_{k=1}^n(\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^T$, *where* $\boldsymbol{x}_k = \boldsymbol{\mu} + \boldsymbol{z}_k$ *with* $\mathbf{z}_k = \Sigma^{\frac{1}{2}}\mathbf{y}_k$ *where* $\boldsymbol{y}_k$*'s have iid entries with mean* 0 *and variance* 1 *and finite fourth* *moment. If* $p/n \to y \in (0, 1)$*, we have*

$$\frac{\boldsymbol{\mu}^T S^{-1} \boldsymbol{\mu}}{n} \xrightarrow{\text{a.s.}} a\gamma,$$

*where* $\gamma$ *with* $y = \lim_{n\to\infty}(p/n) \in (0, 1)$ *to be the constant defined in Theorem 3.2.*

*Proof of Lemma A.2.* Let $\tilde{\boldsymbol{\mu}} = \Sigma^{-1/2}\boldsymbol{\mu}$. Then, we have

$$\boldsymbol{\mu}^T S^{-1} \boldsymbol{\mu} = \|\tilde{\boldsymbol{\mu}}\|^2 \boldsymbol{\alpha}^T \tilde{S}^{-1} \boldsymbol{\alpha},$$

where $\boldsymbol{\alpha} = \tilde{\boldsymbol{\mu}}/\|\tilde{\boldsymbol{\mu}}\|$ is a nonrandom unit vector and

$$\tilde{S} = \Sigma^{-1/2} S \Sigma^{-1/2} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{y}_k - \bar{\mathbf{y}})(\mathbf{y}_k - \bar{\mathbf{y}})^T. \tag{A.5}$$

By Corollary 2 of Bai, Miao, and Pan (2007), we have

$$\lim_{n\to\infty} \boldsymbol{\alpha}^T \tilde{S}^{-1} \boldsymbol{\alpha} \xrightarrow{a.s.} \int_a^b x^{-1} dF_y(x) = \gamma. \tag{A.6}$$

Thereafter, under the condition of our assumption

$$\frac{1}{n}\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} = \frac{1}{n}\|\tilde{\boldsymbol{\mu}}\|^2 \to a\,,$$

the lemma is then proved. □

LEMMA A.3. *Suppose* $\frac{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}{n}$ *and* $\frac{\boldsymbol{\nu}^T\Sigma^{-1}\boldsymbol{\nu}}{n}$ *are uniformly bounded and* $\boldsymbol{\mu}^T\Sigma^{-1}\ \boldsymbol{\nu} = 0$*. Assume that S has the same distribution as given in Lemma A.2. If* $p/n \to y \in (0, 1)$*, then we have*

$$\frac{\boldsymbol{\mu}^T S^{-1} \boldsymbol{\nu}}{n} \xrightarrow{\text{a.s.}} 0.$$

Readers may refer to Bai, Liu, and Wong (2006) for the proof of Lemma A.3. Now, we come back to the proof of Lemma 3.1 as shown in the following:

*Proof of Lemma 3.1.* Let $\boldsymbol{x}_k = \boldsymbol{\mu} + \boldsymbol{z}_k$ where $\mathbf{z}_k = \Sigma^{\frac{1}{2}}\mathbf{y}_k$. We also let $\bar{\mathbf{z}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{z}_k$. It is obvious that

$$E\mathbf{z}_k = 0 \quad \text{and} \quad E(\sqrt{n}\bar{\mathbf{z}})(\sqrt{n}\bar{\mathbf{z}}^T) = \Sigma.$$

Thus,

$$\frac{\bar{\mathbf{x}}^T S^{-1} \bar{\mathbf{x}}}{n} = \frac{\bar{\mathbf{z}}^T S^{-1} \bar{\mathbf{z}}}{n} + \frac{2\boldsymbol{\mu}^T S^{-1} \bar{\mathbf{z}}}{n} + \frac{\boldsymbol{\mu}^T S^{-1} \boldsymbol{\mu}}{n}\,. \tag{A.7}$$

By Lemma A.2, we have

$$\frac{\boldsymbol{\mu}^T S^{-1} \boldsymbol{\mu}}{n} \xrightarrow{\text{a.s.}} \gamma a_1.$$

Hence, to show Part (a) of Lemma 3.1, it suffices to show

$$\text{(A.8)} \qquad \frac{\bar{\mathbf{z}}^T S^{-1} \bar{\mathbf{z}}}{n} \xrightarrow{\text{a.s.}} 0.$$

As

$$\bar{\mathbf{z}}^T S^{-1} \bar{\mathbf{z}} = (\bar{\mathbf{y}})^T \tilde{S}^{-1} (\bar{\mathbf{y}}) = (\bar{\mathbf{y}})^T \bar{S}^{-1} (\bar{\mathbf{y}}) - \frac{[(\bar{\mathbf{y}})^T \bar{S}^{-1} (\bar{\mathbf{y}})]^2}{1 + (\bar{\mathbf{y}})^T \bar{S}^{-1} (\bar{\mathbf{y}})},$$

where $\tilde{S}$ is defined in equation (A.5) and $\bar{S} = \bar{S}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{y}_k \mathbf{y}_k^T$. The assertion (A.8) is obviously implied by

$$\text{(A.9)} \quad \frac{\bar{\mathbf{y}}^T \bar{S}_n^{-1} \bar{\mathbf{y}}}{n} = \frac{1}{n^2} \left[ \sum_{k=1}^{n} \frac{\mathbf{r}_k^T \bar{S}_{nk}^{-1} \mathbf{r}_k}{1 + \mathbf{r}_k^T \bar{S}_{nk}^{-1} \mathbf{r}_k} + \sum_{k \neq j} \frac{\mathbf{r}_k^T \bar{S}_{nkj}^{-1} \mathbf{r}_j}{\left(1 + \mathbf{r}_k^T \bar{S}_{ni}^{-1} \mathbf{r}_k\right)\left(1 + \mathbf{r}_j^T \bar{S}_{nkj}^{-1} \mathbf{r}_j\right)} \right] \xrightarrow{\text{a.s.}} 0,$$

where $\bar{S}_{nkj} = \bar{S}_n - \mathbf{r}_k \mathbf{r}_k^T - \mathbf{r}_j \mathbf{r}_j^T$ for $k \neq j$. Obviously, the convergence (A.9) follows from

$$\text{(A.10)} \qquad \max_{kj} \left| \mathbf{r}_k^T \bar{S}_{nkj}^{-1} \mathbf{r}_j \right| \xrightarrow{\text{a.s.}} 0.$$

By equation (A.6), we conclude that

$$\text{(A.11)} \qquad \max_{k} \left| \mathbf{r}_k^T \bar{S}_{nkj}^{-1} \mathbf{r}_k \right| \big/ \|\mathbf{r}_k\|^2 \xrightarrow{\text{a.s.}} \gamma,$$

and by Lemma A.3,

$$\text{(A.12)} \qquad \max_{kj} \left| \mathbf{r}_k^T \bar{S}_{nkj}^{-1} \mathbf{r}_{jk} \right| \xrightarrow{\text{a.s.}} 0,$$

where $\boldsymbol{r}_{jk} = \boldsymbol{r}_j - \boldsymbol{r}_k \boldsymbol{r}_k^T \boldsymbol{r}_j / \|\boldsymbol{r}_k\|^2$ with $\boldsymbol{r}_{jk}^T \boldsymbol{r}_k = 0$. Note that here we have used the facts that $\|\mathbf{r}_k\|^2 \xrightarrow{\text{a.s.}} y$. Finally, the assertion (A.10) follows from equations (A.11) and (A.12) and the fact that $\mathbf{r}_k^T \mathbf{r}_j \xrightarrow{\text{a.s.}} 0$.

Consequently, we have proved

$$\frac{\bar{\mathbf{x}}^T S^{-1} \bar{\mathbf{x}}}{n} \xrightarrow{\text{a.s.}} a_1 \gamma,$$

which is Part (a) of the lemma.

Part (b) of Lemma 3.1 is proved by setting $\boldsymbol{\mu} = 1$ and applying Lemma A.2. To prove Part (c) of Lemma 3.1, we notice that

$$\frac{1^T S^{-1} \bar{\mathbf{x}}}{n} = \frac{1^T S^{-1} \boldsymbol{\mu}}{n} + \frac{1^T S^{-1} \bar{\mathbf{z}}}{n}.$$

As we have

$$\frac{|1^T S^{-1} \bar{\mathbf{z}}|^2}{n^2} \leq \frac{1^T S^{-1} 1}{n} \frac{\bar{\mathbf{z}}^T S^{-1} \bar{\mathbf{z}}}{n} \xrightarrow{\text{a.s.}} 0\,,$$

by applying the Cauchy-Schwarz inequality, to prove Part (c) of Lemma 3.1 it is equivalent to show that

$$\frac{1^T S^{-1} \boldsymbol{\mu}}{n} \xrightarrow{\text{a.s.}} a_3 \gamma.$$

Let $\boldsymbol{\mu} = \frac{1^T\Sigma^{-1}\boldsymbol{\mu}}{1^T\Sigma^{-1}1}1 + \hat{\boldsymbol{\mu}}$ where $\hat{\boldsymbol{\mu}}^T\Sigma^{-1}1 = 0$. By Lemma A.3, we conclude that

$$\frac{1^T S^{-1}\hat{\boldsymbol{\mu}}}{n} \xrightarrow{\text{a.s.}} 0.$$

Thus,

$$\lim_n \frac{1^T S^{-1}\boldsymbol{\mu}}{n} = \lim_n \frac{1^T\Sigma^{-1}\boldsymbol{\mu}1^T S^{-1}1}{1^T\Sigma^{-1}1n} = a_3\gamma, \quad a.s.$$

and thus Part (c) of Lemma 3.1 is obtained. □

*Calculus of* $\gamma$ in *Theorem 3.2.*

$$\begin{aligned}\gamma &= \int_a^b \frac{1}{x}\,dF_y(x)\\ &= \int_a^b \frac{1}{x}\frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}\,dx\\ &= \frac{1}{\pi}\int_0^{2\pi}\frac{1}{[1+y-\sqrt{y}(e^{i\theta}+e^{-i\theta})]^2}\left(\frac{e^{i\theta}-e^{-i\theta}}{2i}\right)^2 d\theta \qquad \text{(A.13)}\\ &= -\frac{1}{4\pi i}\oint_{|\xi|=1}\frac{1}{[1+y-\sqrt{y}(\xi+\xi^{-1})]^2}(\xi-\xi^{-1})^2\xi^{-1}d\chi\,.\end{aligned}$$

The function has three poles at $\xi_1 = 0$ and $\xi_{2,3} = \frac{(1+y)\pm(1-y)}{2\sqrt{y}}$. By calculating the residuals at these three poles, one can easily show that the integral (29) is equal to $\frac{1}{1-y}$, which is obviously bigger than 1 as $0 < y < 1$. Thus, $\gamma$ is obtained. □

## REFERENCES

AÏT SAHALIA, Y., and M. BRANDT (2001): Variable Selection for Portfolio Choice, *J. Finance* 56, 1297–1351.

BAI, Z. D. (1999): Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review, *Stat. Sinica* 9, 611–677.

BAI, Z. D., and J. W. SILVERSTEIN (1998): No Eigenvalues outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices, *Ann. Probab.* 26(1), 316–345.

BAI, Z. D., and J. W. SILVERSTEIN (1999): Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices, *Ann. Probab.* 27(3), 1536–1555.

BAI, Z. D., and J. W. SILVERSTEIN (2004): CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices, *Ann. Probab.* 32(1A), 553–605.

BAI, Z. D., and Y. Q. YIN (1993): Limit of the Smallest Eigenvalue of Large Dimensional Covariance Matrix, *Ann. Probab.* 21, 1275–1294.

BAI, Z. D., H. LIU, and W. K. WONG (2006): *Asymptotic Properties of Eigenmatrices of Large Sample Covariance Matrix*, Working Paper, Risk Management Institute, National University of Singapore.

BAI, Z. D., B. Q. MIAO, and G. M. PAN (2007): Asymptotics of Eigenvectors of Large Sample Covariance Matrix, *Ann. Probab.* 35(4), 1532–1572.

Best, M. J., and R. R. Grauer (1991): On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results, *Rev. Finan. Stud.* 4(2), 315–342.

Blattberg, R. C., and N. J. Gonedes (1974): A Comparison of Stable and Student Distribution as Statistical Models for Stock Prices, *J. Bus.* 47, 244–280.

Britten-Jones, M. (1999): The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights, *J. Finance* 54(2), 655–671.

Brown, S. J. (1978): The Portfolio Choice Problem: Comparison of Certainty Equivalence and Optimal Bayes Portfolios, *Commun. Stat. Simul. Comput.* 7, 321–334.

Canner, N., N. G. Mankiw, and D. N. Weil (1997): An Asset Allocation Puzzle, *Am. Econ. Rev.* 87(1), 181–191.

Cass, D., and J. E. Stiglitz (1970): The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds, *J. Econ. Theory* 2(2), 122–160.

Clark, P. K. (1973): A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, *Econometrica* 37, 135–155.

Cohen, K. J., and J. A. Pogue (1967): An Empirical Evaluation of Alternative Portfolio-Selection Models, *J. Bus.* 40(2), 166–193.

Detemple, J., and M. Rindisbacher (2005): Closed-Form Solutions for Optimal Portfolio Selection with Stochastic Interest Rate and Investment Constraints, *Math. Finance* 15(4), 539–568.

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. Finance* 33(1), 296–302.

Emmer, S., C. Klüppelberg, and R. Korn (2001): Optimal Portfolios with Bounded Capital at Risk, *Math. Finance* 11(4), 365–384.

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.

Feldstein, M. S. (1969): Mean Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection, *Rev. Econ. Stud.* 36(1), 5–12.

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.

Frankfurter, G. M., H. E. Phillips, and J. P. Seagle (1971): Portfolio Selection: The Effects of Uncertain Means, Variances and Covariances, *J. Finan. Quant. Anal.* 6, 1251–1262.

Hakansson, N. H. (1972): Mean Variance Analysis in a Finite World, *J. Finan. Quant. Anal.* 7(5), 1873–1880.

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

Hanoch, G., and H. Levy (1969): The Efficiency Analysis of Choices Involving Risk, *Rev. Econ. Stud.* 36, 335–346.

Jonsson, D. (1982): Some Limit Theorems for the Eigenvalues of Sample Covariance Matrix, *J. Multivariate Anal.* 12, 1–38.

Jorion, P. (1985): International Portfolio Diversification with Estimation Risk, *J. Bus.* 58(3), 259–278.

Ju, X., and N. Pearson (1999): Using Value-at-Risk to Control Risk Taking: How Wrong Can You Be? *J. Risk* 1(2), 5–36.

KAN, R., and G. ZHOU (2007): Optimal Portfolio Choice with Parameter Uncertainty, *J. Finan. Quant. Anal.* 42(3), 621–656.

KONNO, H., and H. YAMAZAKI (1991): Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market, *Manage. Sci.* 37(5), 519–531.

KROLL, Y., H. LEVY, and H. H. MARKOWITZ (1984): Mean-Variance versus Direct Utility Maximization, *J. Finance* 39, 47–61.

LAKNER, P., and L. M. NYGREN (2006): Portfolio Optimization with Downside Constraints, *Math. Finance* 16(2), 283–299.

LALOUX, L., P. CIZEAU, J. P. BOUCHAUD, and M. POTTERS (1999): Noise Dressing of Financial Correlation Matrices, *Phys. Revi. Lett.* 83, 1467–1470.

LI, D., and W. L. NG (2000): Optimal Dynamic Portfolio Selection: Multiperiod Mean-Variance Formulation, *Math. Finance* 10(3), 339–406.

MALLER, R. A., and D. A. TURKINGTON (2002): New Light on the Portfolio Allocation Problem, *Math. Methods Operations Res.* 56(3), 501–511.

MALLER, R., R. B. DURAND, and P. T. LEE (2005): Bias and Consistency of the Maximum Sharpe Ratio, *J. Risk* 7(4), 103–115.

MARČENKO, V. A., and L. A. PASTUR (1967): Distribution for Some Sets of Random Matrices, *Math. USSR–Sbornik* 1, 457–483.

MARKOWITZ, H. M. (1952): Portfolio Selection, *J. Finance* 7, 77–91.

MARKOWITZ, H. M. (1959): *Portfolio Selection*, New York: John Wiley and Sons.

MARKOWITZ, H. M. (1991): *Portfolio Selection: Efficient Diversification of Investment*, Cambridge, MA: Blackwell.

MARKOWITZ, H. M., and A. F. PEROLD (1981): Portfolio Analysis with Factors and Scenarios, *J. Finance* 36, 871–877.

MCNAMARA, J. R. (1998): Portfolio Selection Using Stochastic Dominance Criteria, *Decision Sci.* 29(4), 785–801.

MERTON, R. C. (1972): An Analytic Derivation of the Efficient Portfolio Frontier, *J. Finan. Quant. Anal.* 7, 1851–1872.

MEUCCI, A. (2005): *Risk and Asset Allocation*, Berlin: Springer.

MICHAUD, R. O. (1989): The Markowitz Optimization Enigma: Is "Optimized" Optimal? *Finan. Analy. J.* 45, 31–42.

MICHAUD, R. O. (1998): *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*, Cambridge, MA: Harvard Business School Press.

MUTHURAMAN, K., and S. KUMAR (2006): Multidimensional Portfolio Optimization with Proportional Transaction Costs, *Math. Finance* 16(2), 301–335.

PAFKA, S., and I. KONDOR (2003): Noisy Covariance Matrices and Portfolio Optimization II, *Physica A* 319, 387–396.

PAFKA, S., and I. KONDOR (2004): Estimated Correlation Matrices and Portfolio Optimization, *Physica A* 343, 623–634.

PAPP, G., S. PAFKA, M. A. NOWAK, and I. KONDOR (2005): Random Matrix Filtering in Portfolio Optimization, *Acta Physica Polonica B* 36(9), 2757–2765.

PEROLD, A. F. (1984): Large-Scale Portfolio Optimization, *Manage. Sci.* 30(10), 1143–1160.

ROTHSCHILD, M., and J. E. STIGLITZ (1970): Increasing Risk: I. A Definition, *J. Econ. Theory* 2, 225–243.

ROTHSCHILD, M., and J. E. STIGLITZ (1971): Increasing Risk: II. Its Economic Consequences, *J. Economic Theory* 3, 66–84.

SHARPE, W. F. (1964): Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *J. Finance* 19, 425–442.

SHARPE, W. F. (1967): A Linear Programming Algorithm for Mutual Fund Portfolio Selection, *Manage. Sci.* 22, 499–510.

SHARPE, W. F. (1971): A Linear Programming Approximation for the General Portfolio Analysis Problem, *J. Finan. Quant. Anal.* 6(5), 1263–1275.

SIMAAN, Y. (1997): Estimation Risk in Portfolio Selection: The Mean Variance Model versus the Mean Absolute Deviation Model, *Manage. Sci.* 43(10), 1437–1446.

STONE, B. K. (1973): A Linear Programming Formulation of the General Portfolio Selection Problem, *J. Finan. Quant. Anal.* 8(4), 621–636.

XIA, J. (2005): Mean-Variance Portfolio Choice: Quadratic Partial Hedging, *Math. Finance* 15(3), 533–538.

XIA, J., and J. A. YAN (2006): Markowitz Portfolios Optimization in an Incomplete Market, *Math. Finance* 16(1), 203–216.

ZELLNER, A., and V. K. CHETTY (1965): Prediction and Decision Problems in Regression Models from the Bayesian Point of View, *J. Am. Stat. Assoc.* 60, 608–616.