Shrinkage Method for Estimating Optimal Expected Return of Self-financing Portfolio

LIU, Yan

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Statistics

The Chinese University of Hong Kong
June 2011

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Thesis / Assessment Committee

Professor Minggao Gu (Chairman)

Professor Ngai Hang Chan (Thesis Supervisor)

Professor Xiaodan Fan (Committee Member)

Professor Shiqing Ling (External Examiner)

Abstract of thesis entitled:

Shrinkage Method for Estimating Optimal Expected Return of Self-financing Portfolio

Submitted by LIU, Yan

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in May 2011

By the seminal work of Markowitz in 1952, modern portfolio theory studies how to maximize the portfolio expected return for a given risk, or minimize the risk for a given expected return. Since these two issues are equivalent, this thesis only focuses on the study of the optimal expected return of a self-financing portfolio for a given risk.

The mean-variance portfolio optimization procedure requires two crucial inputs: the theoretical mean vector and the theoretical covariance matrix of the portfolio in one period. Since the traditional plug-in method using the sample mean vector and the sample covariance matrix of the historical data incurs substantial estimation errors, this thesis explores how the sample mean vector and the sample covariance matrix behave in the optimization procedure based on the idea of conditional expectation and finds that the effect of the sample mean vector is an additive process while the effect of the sample covariance matrix is a multiplicative process.

A new estimator for calculating the optimal expected return of a self-

financing portfolio is proposed, by considering the joint impact of the sample mean vector and the sample covariance matrix. A shrinkage covariance matrix is designed to substitute the sample covariance matrix in the optimization procedure, which leads to an estimate of the optimal expected return smaller than the plug–in estimate. The new estimator is also applicable for both $p < n$ and $p \geq n$. Simulation studies are conducted for two empirical data sets. The simulation results show that the new estimator is superior to the previous methods.

Finally, under certain assumptions, we extend our research in the framework of random matrix theory.

摘要

馬可維茨於 1952 年提出的現代投資組合理論主要研究如下問題：給定風險，極大化期望收益；或給定期望收益，極小化風險。因為這兩個問題是等價命題，本文主要針對自籌資投資組合對上述第一個命題進行了研究。

均值--方差投資組合優化理論需要兩個重要的輸入：一個時段內投資組合的理論均值及協方差矩陣。傳統的插值方法由於使用樣本均值及樣本協方差矩陣來代替理論值，受到較大的估計誤差的影響。本文將最優期望收益視作一個整體，利用條件期望研究樣本均值及樣本協方差矩陣在優化過程中對其的影響，發現這兩個量的估計誤差對於最優期望收益的作用機制分別為加和效應及乘積效應。

因此，在充分考慮到樣本均值及樣本協方差矩陣的交互效應的基礎上，本文針對自籌資投資組合的最優期望收益提出了一種新的估計方法。本文利用收縮方法構建了一個穩定的協方差矩陣來代替樣本協方差矩陣，並證明利用該矩陣得到的估計值小於插值法得到的估計值。而且，該估計量適用於 $p \le n$ 和 $p > n$ 兩種情況。本文對兩個經驗數據進行了模擬研究，模擬結果表明新的估計量與以往的估計方法相比，更接近真實值。

最後，文章在一個嚴格的數據假設下，利用隨機矩陣理論擴展了研究結果。

Acknowledgement

First of all, I would like to show my deepest gratitude to my two supervisors. Professor Ngai Hang Chan, whose help was not confined in academic field. He gave me a lot of advice which would benefit me for my whole life. His enthusiasm and passion in research made me to follow, but not feel pushed. And Professor Samuel Po Shing Wong, a scholar of individuality, helped me a lot when I was faced with difficulties. He ceaselessly came back to the university when I encountered problems even after he left for industry. They guided me to the beautiful world of statistics and helped me build the confidence in statistical research. Without them, this thesis would not have been possible. I really appreciate their efforts in the past three years. It is my honor and luck to have them, two brilliant and dedicated professors as my PhD supervisors.

Secondly, I would like to thank all the faculties and staffs in Department of Statistics. From you, I learned many statistical skills. I will definitely find a way to reciprocate the department in future. I would also like to thank The Chinese University of Hong Kong, who provided me many facilities in my study.

Further, I would express my appreciation to all my dear friends: Ms. Lin Yueqiong and Ms. Wang Man, who gave me many kind suggestions and conscientious help when I wrote this thesis; Ms. Lin Nan, who accompanied me and encouraged me for all my PhD days; Special acknowledgement also goes to my

other good friends Ms. Gao Fei, Ms. Li Han, Mr. Chen Pengcheng and the list goes on. You added too much joy to my PhD life. Thank you for your company.

Finally, I am forever indebted to my dear family. Your unconditional support and confidence in me is my greatest motivation in pursuing dreams. Your love helped me overcome so many difficulties. My husband Shen Song, gave me strong backup and understanding. Whenever I was frustrated, he would travel several hundred miles to visit me even during his busiest days.

This work is dedicated to my family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background Study

As a cornerstone of modern portfolio theory, the mean variance (MV) optimization procedure has always been a controversial topic since its establishment in 1952 by Markowitz [32, 33]. It has achieved great merits for its contribution to help building up a mathematical model for investing, and in the mean time, it has also received a lot of criticism for its impracticality.

Based on the assumptions that the asset returns are normally distributed, all the investors are rational and risk-averse and aim to maximize economic utility, the mean variance optimization procedure characterizes the asset allocation problem as a trade-off between risk and expected return, which are represented by the variance and mean of returns respectively. Specifically, there are two issues to study: maximizing the expected return given a fixed risk or minimizing the risk given a fixed expected return, both of which lead to the formulation of the efficient frontier, and from this, the investors can choose their preferred portfolios.

However, the application of the theory is not as helpful as one might think. People criticized it because they found that they could not pinpoint the best portfolio on the efficient frontier. In fact, most of the time what they got were even inferior to the equal weighting portfolios (see, for example, Frankfurter, Phillips and Seagle, 1971 [23]).

Regardless of the other reasons against the MV method, such as the efficiency of the market, or behavioral finance, in this thesis, we only concentrate on the problem of "error maximization" (Michaud, 1989 [37]) incurred by the two critical inputs, the sample mean vector and the sample covariance matrix, from a statistical perspective. Since the two optimization issues mentioned above are equivalent, and the optimal expected return is an important criterion in comparing different portfolios for the same risk, in this study, we only focus on the first issue; to provide a better estimator of the optimal expected return under the assumption that the dimension to sample size ratio p/n is not ignorable.

1.2 Modern Portfolio Theory

1.2.1 Self–Financing Portfolio

In this thesis, we choose to study the self-financing portfolios. A self-financing portfolio is a set of long and short investments such that the sum of their investment weights is zero (Korkie and Turtle, 2002 [27]), which means that the purchase of a new asset must be financed by the sale of an old one. Examples of self-financing portfolios are hedges, swaps, overlays, arbitrage portfolios, long/short portfolios.

1.2.2 Mean Variance Optimization Procedure

Suppose that there are totally p stocks with returns at time t given by $\boldsymbol{x} = (x_{t1}, x_{t2}, \ldots, x_{tp})^{\mathrm{T}}$, and $\boldsymbol{x}$ follows a p–dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Herein, $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_p)^{\mathrm{T}}$ is the vector of weights corresponding to the p stocks. In this thesis, short sales are allowed, which means that $\boldsymbol{\omega}^{\mathrm{T}}\mathbf{1} = 0$.

Since we only focus on the analysis of optimal expected return for a given risk in this study, the problem can be described as follows:

$$\max(P) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\mu},$$

$$\text{subject to } \begin{cases} \boldsymbol{\omega}^{\mathrm{T}}\Sigma\boldsymbol{\omega} & = \sigma_0^2, \\ \boldsymbol{\omega}^{\mathrm{T}}\mathbf{1} & = 0, \end{cases}$$

where P is the expected return of the portfolio and σ_0 characterizes the given risk.

By Lagrange multiplier method (see, for example, Markowitz, 1959 [33]) the optimized maximum expected return is:

$$P^* = \sigma_0 \sqrt{\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu})^2}{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}}}.$$

1.2.3 Plug–in Method

Since the true mean vector and the true covariance matrix of the returns are unknown, traditionally, the portfolio analysis proceeds in two steps:

(a) the sample mean $\hat{\boldsymbol{\mu}}$ and the sample covariance matrix S of returns are estimated from a time series of historical returns;

(b) then the MV problem is solved as if the sample estimates were true values.

This "certainty equivalence" viewpoint is also called the plug–in method. And the plug–in optimal expected return is:

$$\begin{aligned} \hat{P}^* &= \sigma_0 \sqrt{\hat{\boldsymbol{\mu}}^{\mathrm{T}} S^{-1} \hat{\boldsymbol{\mu}} - \frac{(\mathbf{1}^{\mathrm{T}} S^{-1} \hat{\boldsymbol{\mu}})^2}{\mathbf{1}^{\mathrm{T}} S^{-1} \mathbf{1}}} \\ &= \sigma_0 \sqrt{\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{-1}) \hat{\boldsymbol{\mu}}}. \end{aligned}$$

Herein, we define the function h as

$$h(S^{-1}) = S^{-1} - \frac{S^{-1}\mathbf{1}\mathbf{1}^{\mathrm{T}} S^{-1}}{\mathbf{1}^{\mathrm{T}} S^{-1} \mathbf{1}}.$$

The value σ_0 is assumed to be one since it does not affect the analysis. In this study, we analyze at $\hat{P}^{*2}$.

1.3 Literature Review

Many authors (see, for example, Frankfurter, Phillips and Seagle, 1971 [23]; Barry, 1974 [7]; Brown, 1976 [9]; Michaud, 1989 [37]) have realized the impact of this "certainty equivalence" problem. For example, in Frankfurter, Phillips and Seagle (1971) [23], an experiment was conducted in which the simultaneous effect of errors in estimating the means, variances and covariances of returns were taken into consideration. The experimental results demonstrated that the impact of the estimation errors was so severe that the usefulness of the portfolio theory was brought into question.

More recently, Michaud (1989) [37] summarized the previous work and attributed the inefficiency to the reason that MV optimizers were "estimation error maximizers" since they had the tendency to overweight the securities with large estimated returns, negative correlations and small variances, which were most likely to have large estimation errors.

Since the success of a portfolio selection rule depends on the qualify of the two inputs, the sample mean vector and the sample covariance matrix (see, for example, Bodie, Kane and Marcus, 2010 [8]), many new estimators of the mean and the covariance matrix with less estimation errors have been designed to deal with this problem (see, for example, Sharpe, 1963 [40]; Elton, Gruber and Padberg, 1979 [21]; Klein and Bawa, 1976 [26]; Jorion, 1986 [25]; Ledoit and Wolf, 2003, 2004a & 2004b [28, 29, 30]).

Index Models

A famous estimator of the covariance matrix is the single–index model proposed by Sharpe in 1963 [40], which is also known as the market model. It is defined as:

$$x_{ti} = \alpha_i + \beta_i x_{0t} + \epsilon_{ti},$$

where x_{ti} is the return of stock i at time t, x_{0t} is the return of the market index at time t, β_i measures the sensitivity of stock i to the market index, and $\epsilon_{ti} \sim N(0, \delta_i^2)$. The covariance matrix of the single–index model is:

$$F = \sigma_m^2 \boldsymbol{\beta}\boldsymbol{\beta}' + D,$$

where $D = \text{diag}(\delta_1^2, \ldots, \delta_p^2)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ and σ_m^2 is the variance of the market index.

Single–index model fully considers the property of the stock market and greatly reduces the number of parameters to be estimated. Elton and Gruber (1973) [20] proved that this model was more accurate than directly estimating the sample covariance matrix.

To capture more information of the market, further works extended the single–index model to multi–index models, such as Cohen and Pogue (1967)

[11] & Roll and Ross (1980) [38]. Then the portfolio problem was reformulated using the multi-index models (see, for example, Ross, 1978 [39]; Elton, Gruber and Padberg, 1979 [21]).

Bayesian Method

Barry (1974) [7], Brown (1976) [9], Klein and Bawa (1976) [26] advocated a Bayesian approach to efficient portfolio selection.

Using a diffuse prior, Brown [9] showed that the Bayes diffuse method outperformed the plug-in method in portfolio selections. However, the set of investment weights on the Bayes diffuse efficient frontier are the same as the set of investments weights on the plug-in efficient frontier, which hampers the application of Bayes diffuse method (see Barry, 1974 [7]).

Thereafter, people applied Bayesian approach using informative priors to the portfolio theory (see Klein and Bawa, 1976 [26]). However, the assertion of all the prior parameters still caused problems in parameter estimation.

Empirical Bayes Method

Stein (1956) [41] proved that for p–dimensional multivariate normal distribution, when $p \geq 3$, the sample mean vector was not admissible relative to a quadratic loss function. Then he proposed a so–called James–Stein estimator (Stein, 1961 [42]). The essence of this estimator is that it shrinks the maximum likelihood estimators towards a common value, which leads to a uniformly lower risk than the sample mean vector.

For further extensions of James–Stein estimator, see Efron and Morris (1971, 1972a, 1972b, 1973, 1975, 1976) [12, 13, 14, 15, 16, 17]. They modified

James-Stein estimator and proposed an empirical Bayes approach to obtain estimates that are more appropriate in practical situations.

Efron and Morris (1976) [18] also proposed empirical Bayes estimators of the inverse of a covariance matrix. Such estimators shrink the sample eigenvalues towards a central value, working in the same way as the James–Stein estimator, and they dominate any constant multiple of the inverse sample covariance matrix in the sense of the uniformly lower risk.

The advantage of the empirical Bayes method lies in that the prior of the parameters can be directly derived from the data. Under the assumption that the dimension to sample size ratio p/n goes to zero, Merton (1980) [35] demonstrated that the covariance matrix was stable across time. Based on this result, Jorion (1986) [25] applied Stein's shrinkage estimator of mean to portfolio theory and achieved better results.

Frost and Savarino (1986) [24] proposed an empirical Bayes approach to efficient portfolio selection. In their study, the informative prior was specified as "all the stocks are identical". The simulation results showed that the performance of the empirical Bayes method was substantially superior to the Bayes Diffuse investment rules.

Shrinkage Covariance Matrix

Recently, portfolio managers realize that p cannot be ignored relative to n; it can even be larger than n, which makes the sample covariance matrix S unstable or singular. And totally ignoring the covariance structure such as index models can also incur problems.

To circumvent this difficulty, Ledoit and Wolf (2003, 2004a, 2004b) [28, 29, 30] proposed a shrinkage covariance matrix under the weaker assumption that p/n was only bounded. It inherits the advantage of the unbiasedness of the sample covariance matrix and the combination of the highly structured shrinkage target makes it stable and invertible. In fact, this method is an enhancement of the empirical Bayes approach.

In Ledoit and Wolf (2003) [28], the shrinkage target was chosen as the single–index model. They proposed a consistent estimator for the shrinkage intensity. By applying the shrinkage covariance matrix to the NYSE and AMEX stock returns, they got significantly lower out–of–sample variance than existing estimators.

In Ledoit and Wolf (2004a) [29], the shrinkage target was chosen as the constant correlation model. It gave comparable results as the single–index model, but had less parameters to estimate so it was easy to implement.

In M. Wolf (2006) [43], the shrinkage method was compared with Michaud's resampling method (1998) [36]. And it also asked whether the combination of two methods could do better than either technique. However, there was no evidence showing that the combination outperformed the shrinkage method.

Random Matrix Theory

Bai, Liu and Wong (2009a, 2009b) [2, 3] offered a new idea on this issue from the view point of random matrix theory. Using the result of Bai, Miao and Pan (2007) [6], they proved that when $p/n \to y \in (0, 1)$, the plug-in optimal expected return over-predicts due to the over-dispersion of the eigenvalues of the sample covariance matrix. They also calculated the over-prediction ratio

based on the limit spectral distribution. This method, however, is inapplicable when p/n is substantial because the sample covariance matrix becomes singular.

1.4 Goal of the Study

The joint effect of the sample mean and the sample covariance matrix together has rarely been considered. Although some studies have considered the Bayes estimators of mean and covariance matrix together, their interactions usually complicate the issue.

Following Bai et. al (2009a, 2009b) [2, 3], this thesis focuses on the optimal expected return of a self-financing portfolio. It first exploits the joint effect of the two quantities and proposes a new estimator for evaluating the optimal expected return. To make the sample covariance matrix stable and invertible, shrinkage covariance matrices are combined with the proposed estimator.

1.5 Organization of the Thesis

This thesis consists of five chapters and is organized as follows.

In Chapter 1, we give the background of this study and the brief review of modern portfolio theory. The development of this field is also introduced in this chapter. Chapter 2 is the main part of this thesis. Using conditional expectation, the joint impact of the errors incurred by the sample mean and the sample covariance matrix is investigated. A new estimator of the optimal expected return is then proposed, which is applicable for both $p < n$ and $p \geq n$. In Chapter 3, simulation studies are conducted to compare the new estimator with the previous methods. The simulation results show that the

new estimator is better than others because it is much closer to the benchmark value. Chapter 4 consists of discussion and further research in the framework of random matrix theory. Chapter 5 concludes.

Chapter 2

Main Results

2.1 Introduction

In this chapter, a new estimator for estimating the optimal expected return of a self–financing portfolio is proposed. Since the sample mean vector and the sample covariance matrix both incur substantial estimation errors, especially when the number of stocks is very large and the sample size is correspondingly small, this chapter firstly investigates how these two inputs behave in estimating procedures using conditional expectation. Simulation studies show that both the inputs overestimate the optimal expected return.

Condition on $\hat{\boldsymbol{\mu}}$, a shrinkage covariance matrix is designed to substitute the sample covariance matrix in the optimization procedure, which leads to an estimate of the optimal expected return smaller than the plug–in estimate. The result is proved by the theories of partial ordering. Then the effect of $\hat{\boldsymbol{\mu}}$ is deducted. This method behaves well because it fully considers both the joint impact of the two inputs and the error patterns. More importantly, it also avoids the problem of over shrinkage.

2.2 Sample Mean and Sample Covariance Matrix

Using conditional expectation, $\mathrm{E}(\hat{P}^{*2})$ can be expressed as:

$$\begin{aligned}\mathrm{E}(\hat{P}^{*2}) &= \mathrm{E}[\mathrm{E}(\hat{P}^{*2}|S)] \\ &= \mathrm{E}[\mathrm{E}(\hat{P}^{*2}|\hat{\boldsymbol{\mu}})].\end{aligned}$$

This motivates us to study the impact of S and $\hat{\boldsymbol{\mu}}$ on $\mathrm{E}(\hat{P}^{*2})$ individually. Introduce the following notations:

$$P^{*2} = f(\boldsymbol{\mu}, \Sigma) = \boldsymbol{\mu}^{\mathrm{T}} h(\Sigma^{-1})\boldsymbol{\mu},$$

$$\hat{P}^{*2} = f(\hat{\boldsymbol{\mu}}, S) = \hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{-1})\hat{\boldsymbol{\mu}},$$

$$\hat{P}_1^{*2} = f(\hat{\boldsymbol{\mu}}, \Sigma) = \hat{\boldsymbol{\mu}}^{\mathrm{T}} h(\Sigma^{-1})\hat{\boldsymbol{\mu}},$$

$$\hat{P}_2^{*2} = f(\boldsymbol{\mu}, S) = \boldsymbol{\mu}^{\mathrm{T}} h(S^{-1})\boldsymbol{\mu}.$$

Herein, P^{*2} is calculated using the true mean vector and the true covariance matrix; $\hat{P}^{*2}$ is calculated by plugging in the sample mean vector and the sample covariance matrix; $\hat{P}_1^{*2}$ is calculated by only plugging in the sample mean vector, assuming that the covariance matrix is known; and the last one, $\hat{P}_2^{*2}$ is calculated assuming that the mean is known while the covariance matrix unknown.

To find the pattern of the errors incurred by using $\hat{\boldsymbol{\mu}}$ and S, simulation studies are conducted based on the data sets generated from the following case using multivariate normal distribution. For a given p and a given n, the true covariance matrix $\Sigma = I$; the true mean vector $\boldsymbol{\mu}$ is generated from a standard normal distribution and then it is regarded as a known value.

For each figure reported in this section (Figures 2.1–2.4), the variable i in the x-axis ranges from 1 to 20. Each i corresponds to a data set with dimension

p_i and sample size n_i. The relationship between p_i and n_i is:

$$\frac{p_i}{n_i} = c \ ,$$

where $p_i = pi$, $n_i = ni$ and c is a fixed constant. That is to say, in each Figure, the ratio between the dimension and the sample size is the same. The theoretical optimal expected return P^* is regarded as the benchmark value.

2.2.1 Impact of the Sample Mean

To study the impact of the sample mean vector on the optimal expected return, assume that the true covariance matrix Σ is known. We know that for a p–dimensional multivariate normal distribution with covariance matrix I and $p \geq 3$ (see, for example, Stein, 1956 [41]; T. W. Anderson, 2003 [1]), we have

$$\mathrm{E}(\hat{\boldsymbol{\mu}}^{\mathrm{T}}\hat{\boldsymbol{\mu}}) = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} + \frac{p}{n}.$$

Consider the deviation D of using the sample mean vector:

$$D = \mathrm{E}(\hat{P}_1^{*2}) - P^{*2}.$$

In Figure 2.1, the y-axis measures the difference D and the x-axis denotes the variable i. For each i, we simulate 20 data sets with the same covariance matrix and the same mean vector using multivariate normal distribution. For each data set, we have a value $\hat{P}_1^{*2}$. Then $\mathrm{E}(\hat{P}_1^{*2})$ is approximated by taking the sample average of these twenty values of $\hat{P}_1^{*2}$ as the estimate of the mean, see Figure 2.1.

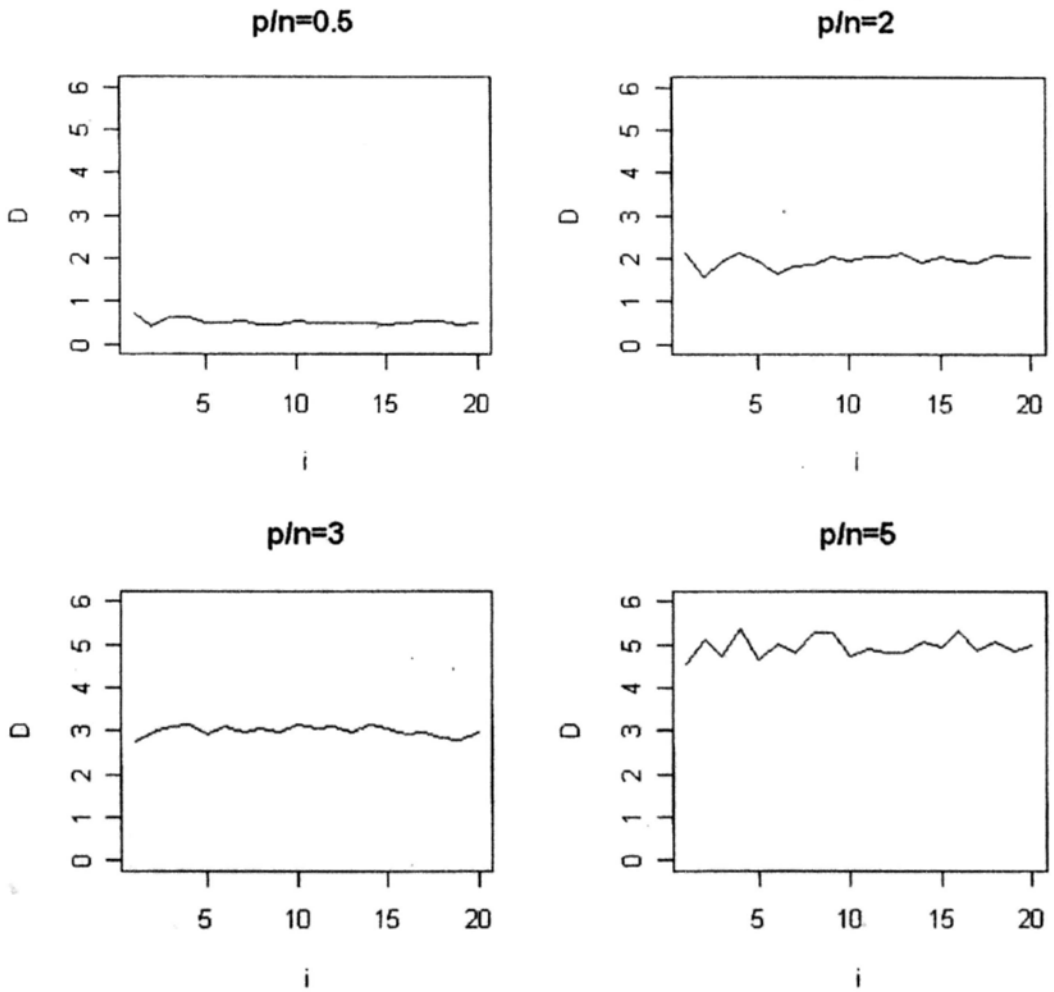

Figure 2.1: Difference Between $\mathrm{E}(\hat{P}_1^{*2})$ and P^{*2}.

From Figure 2.1, it is observed that the sample mean overestimates P^{*2} and the difference D fluctuates around the value of p/n. This fact is reflected by the following result.

Theorem 2.2.1. For a given S,

$$\mathrm{E}[\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{-1})\hat{\boldsymbol{\mu}}] = \frac{1}{n}\mathrm{tr}[h(S^{-1})\Sigma] + \boldsymbol{\mu}^{\mathrm{T}} h(S^{-1})\boldsymbol{\mu}.$$

Proof.

$$\begin{aligned}
\mathrm{E}[\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{-1})\hat{\boldsymbol{\mu}}] &= \mathrm{E}[\mathrm{tr}(\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{-1})\hat{\boldsymbol{\mu}})] \\
&= \mathrm{tr}\mathrm{E}[\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\mathrm{T}}]h(S^{-1}) \\
&= \mathrm{tr}[\mathrm{Var}(\hat{\boldsymbol{\mu}}) + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}]h(S^{-1}) \\
&= \mathrm{tr}[\frac{\Sigma}{n} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}]h(S^{-1}) \\
&= \frac{1}{n}\mathrm{tr}[h(S^{-1})\Sigma] + \boldsymbol{\mu}^{\mathrm{T}} h(S^{-1})\boldsymbol{\mu}.
\end{aligned}$$

In this case, S is known to be the real covariance matrix Σ so that $D = (p-1)/n$.

2.2.2 Impact of the Sample Covariance Matrix

In this part, $\boldsymbol{\mu}$ is known and only the impact of the sample covariance matrix is considered. The knowledge of the random matrix theory is employed to investigate the effect of S. However, some basic assumptions of the random matrix theory must be introduced first.

Suppose that $\{x_{jk},\ j = 1,\ldots,n,\ k = 1,\ldots,p\}$ is a set of double arrays of *i.i.d.* real random variables with mean 0 and variance σ^2. Let $\boldsymbol{x}_k = (x_{1k},\ldots,x_{pk})^{\mathrm{T}}$ and $X = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$. Then the sample mean vector and the sample covariance matrix are defined as:

$$\begin{aligned}
\bar{\boldsymbol{x}} &= \sum_{k=1}^{n} \frac{\boldsymbol{x}_k}{n}, \\
S &= \frac{1}{n-1}\sum_{k=1}^{n}(\boldsymbol{x}_k - \bar{\boldsymbol{x}})(\boldsymbol{x}_k - \bar{\boldsymbol{x}})^{\mathrm{T}}.
\end{aligned}$$

The empirical spectral distribution of the sample covariance matrix S is defined as:

$$F_S(x) = \frac{1}{p}\sum_{i=1}^{p} \mathbf{1}(\lambda_i \le x),$$

where λ_i is the ith smallest eigenvalue of S and

$$1(\lambda_i \leq x) = \begin{cases} 1 & \text{if } \lambda_i \leq x, \\ 0 & \text{otherwise.} \end{cases}$$

By Bai (1999) [5], if $p/n \to y \in (0, \infty)$, then with probability one, $F_S(x)$ converges to the M-P law $F_y(x)$ almost surely. The M-P law is defined as follows (see, for example, Bai and Silverstein, 2010 [4]; Marčenko and Pastur, 1967 [31]):

Definition 2.2.1. (*Marčenko-Pastur Law*) The density function of the limit spectral distribution $F_y(x)$ is given by:

$$f_y(x) = \begin{cases} \frac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)}, & if\ a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

It has a point mass $1 - 1/y$ at the origin if $y > 1$, where $a = \sigma^2(1 - \sqrt{y})^2$, $b = \sigma^2(1 + \sqrt{y})^2$, $p/n \to y \in (0, \infty)$. If $\sigma^2 = 1$, it is called the standard M-P law.

To see the impact of the sample covariance matrix on the optimal expected return, we need to introduce the following lemmas.

Lemma 2.2.1. (*see Lemma A.2 of Bai, Liu and Wong, 2009a*). Suppose that

$$\frac{\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}}{n} \to a_1,$$

and

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

where $\boldsymbol{x}_k = \boldsymbol{\mu} + \boldsymbol{z}_k$ with $\boldsymbol{z}_k = \Sigma^{1/2}\boldsymbol{y}_k$ and $\boldsymbol{y}_k$'s having i.i.d. entries with mean

0 and variance 1 and finite fourth moment. If $p/n \to y \in (0,1)$, then

$$\frac{\boldsymbol{\mu}^{\mathrm{T}} S^{-1} \boldsymbol{\mu}}{n} \to a_1 \gamma,$$

where γ is defined as:

$$\gamma = \int_a^b \frac{1}{x} dF_y(x) = \frac{1}{1-y} > 1,$$

and $a = (1-\sqrt{y})^2$, $b = (1+\sqrt{y})^2$ and $F_y(x)$ is the limit spectral distribution of the eigenvalues.

The proof can be referred to Corollary 2 of Bai, Miao and Pan (2007) [6] & Bai, Liu and Wong (2009a) [2].

Lemma 2.2.2. (*see Lemma 3.1 of Bai, Liu and Wong, 2009a*). Under the assumptions of Lemma 2.2.1, suppose that

$$\frac{\mathbf{1}^{\mathrm{T}} \Sigma^{-1} \mathbf{1}}{n} \to a_2,$$

then

$$\frac{\mathbf{1}^{\mathrm{T}} S^{-1} \mathbf{1}}{n} \xrightarrow{a.s.} a_2 \gamma.$$

By replacing $\boldsymbol{\mu}$ with $\mathbf{1}$, and applying Lemma 2.2.1, we obtain the following result.

Lemma 2.2.3. (*see Lemma 3.1 of Bai, Liu and Wong, 2009a*). Under the assumptions of Lemma 2.2.1, suppose that

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

then

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

The proof can be found in Bai, Liu and Wong (2009a) [2].

Combining Lemmas 2.2.1, 2.2.2 and 2.2.3, we conclude that for a fixed $\boldsymbol{\mu}$, P^* is overestimated by a ratio $\sqrt{\gamma}$, due to the over-dispersion of the eigenvalues of the sample covariance matrix.

Simulation studies are conducted to give an intuitive impression of the above results. Consider the ratio

$$k = \frac{\mathrm{E}(\hat{P}_2^{*2})}{P^{*2}}.$$

In Figure 2.2, the y-axis measures the ratio k and the x-axis denotes the variable i. Again, for each i, twenty data sets are simulated using the same covariance matrix and the mean vector. For each data set, we have a value of $\hat{P}_2^{*2}$, then $\mathrm{E}(\hat{P}_2^{*2})$ is approximated by taking the sample average of these twenty values of $\hat{P}_2^{*2}$ as the estimate of the mean.

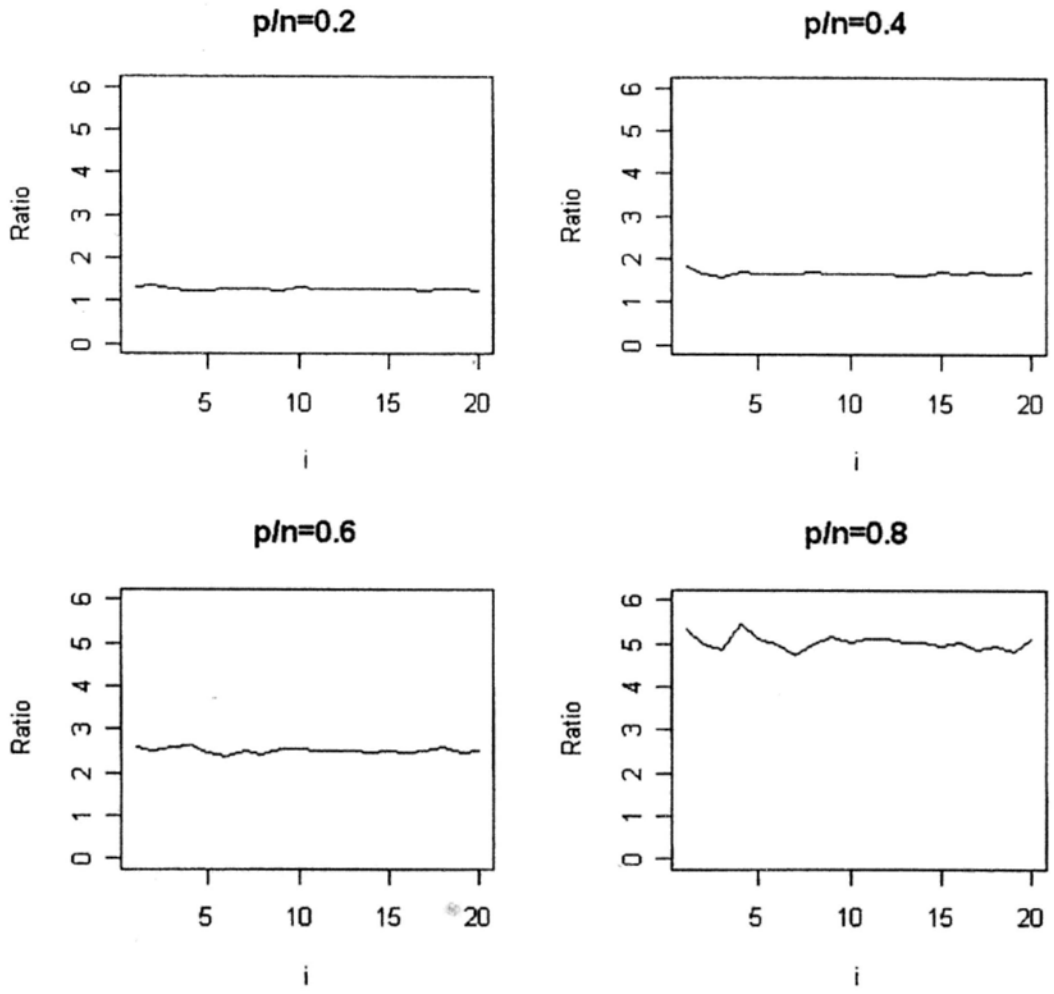

Figure 2.2: Ratio Between $\mathrm{E}(\hat{P}_2^{*2})$ and P^{*2}.

From Figure 2.2, it is seen that for a fixed $\boldsymbol{\mu}$, the overestimating ratio incurred by S is stable around the value γ as p and n increase with i.

2.2.3 Joint Impact

To consider the joint impact, we take the sample mean and the sample covariance matrix simultaneously. The plug-in optimal expected return $\hat{P}^*$ is plotted against the theoretical value P^*. The solid line in Figure 2.3 shows the behavior of $\hat{P}^*$ and the dash line denotes P^*.

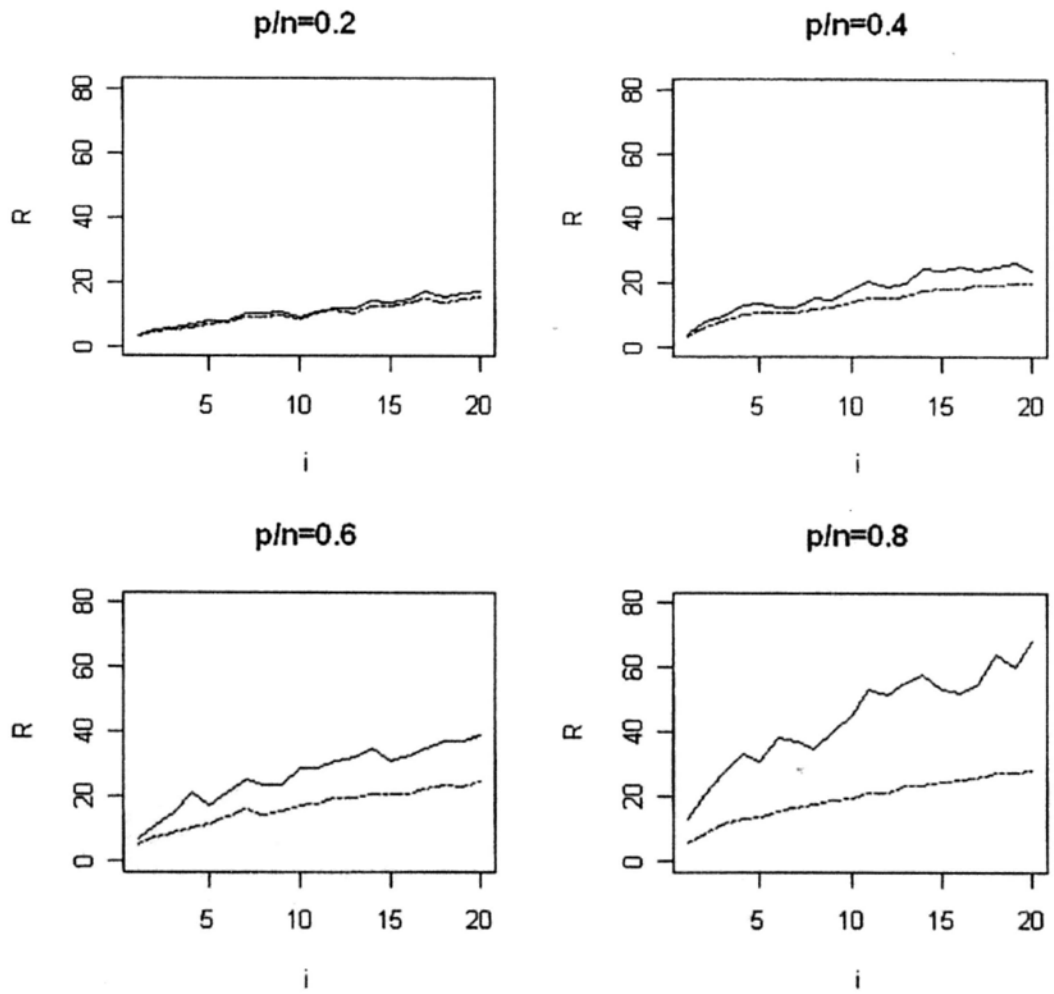

Figure 2.3: Plug-in Optimal Expected Return $\hat{P}^*$ vs. Benchmark Value P^*.

It is observed that as p/n grows larger, $\hat{P}^*$ deviates farther from P^*. However, there is no obvious pattern in Figure 2.3.

Consider the quantity

$$r = \frac{\mathrm{E}(\hat{P}^{*2}) - \gamma(p-1)/n}{\gamma P^{*2}}.$$

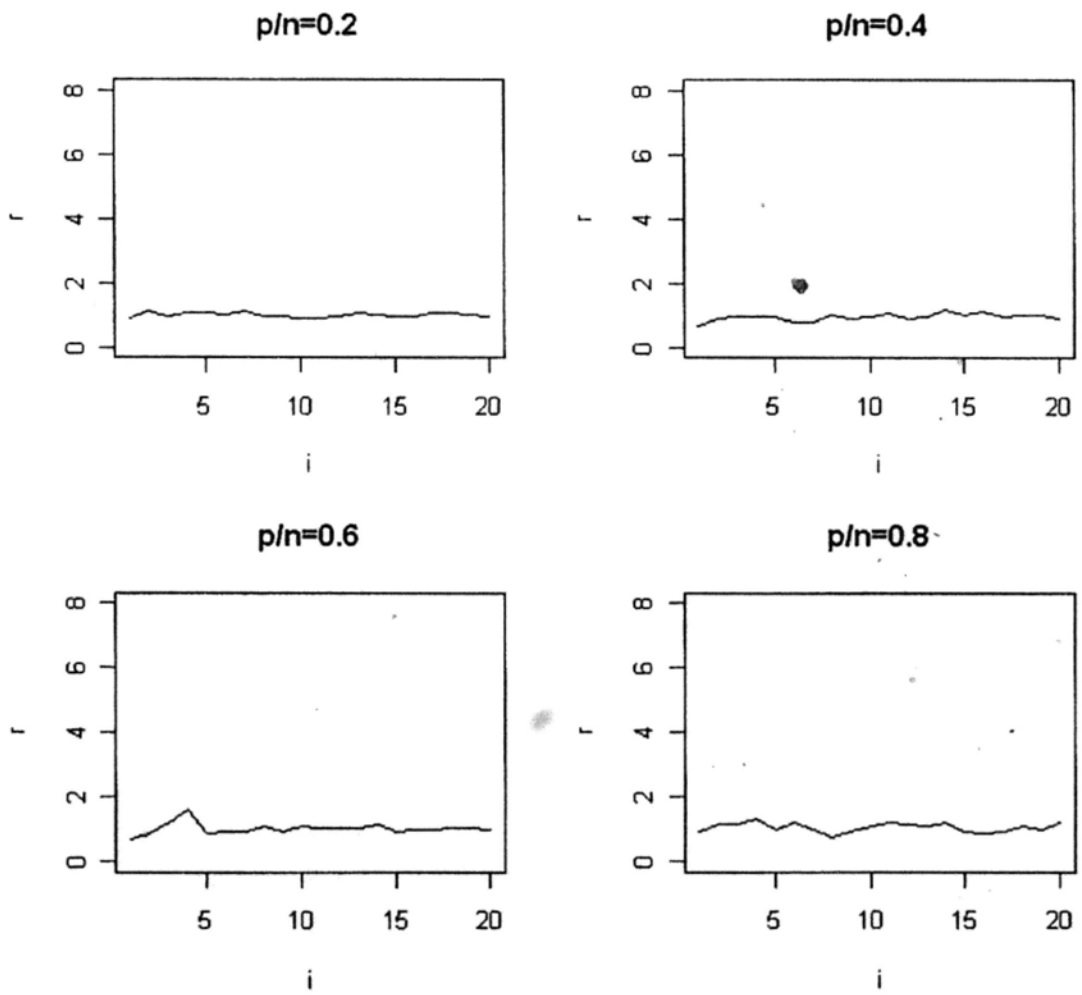

Figure 2.4: Ratio Between $\mathrm{E}(\hat{P}^{*2}) - \gamma(p-1)/n$ and γP^{*2}.

Since the ratio between $\mathrm{E}(\hat{P}^{*2}) - \gamma(p-1)/n$ and γP^{*2} fluctuates around 1, the value $(\mathrm{E}(\hat{P}^{*2}) - \gamma(p-1)/n)/\gamma$ is approximately the same as P^{*2}.

From these numerical studies, we conclude that the effect of the mean is an additive process while the effect of the sample covariance matrix is a multiplicative process. Both of them over predict the optimal expected return. It can also be argued that the error patterns incurred by $\hat{\mu}$ and S together cannot be ignored.

2.3 Estimating the Covariance Matrix

In this part, Ledoit and Wolf's shrinkage method (2003, 2004a, 2004b) [28, 29, 30] is employed to construct a shrinkage covariance matrix, which leads to an estimate with smaller errors than the plug-in optimal expected return. The advantage of using the shrinkage covariance matrix is proved by means of the notion partial ordering (see, for example, Marshall and Olkin, 1979 [34]).

2.3.1 Shrinkage Covariance Matrix

The shrinkage covariance matrix is a linear combination of the sample covariance matrix and a highly structured covariance matrix, which is estimated from the data set. The structured covariance matrix is also called the shrinkage target. The weight on the shrinkage target, which is called shrinkage intensity, is chosen based on the criterion of minimizing the risk function.

The shrinkage covariance matrix is not only invertible, but also well conditioned, which means that inverting it does not amplify the estimation errors. More importantly, it inherits the advantage of the unbiasedness of the sample covariance matrix.

To avoid the singularity of the sample covariance matrix, the Frobenius norm is chosen to define the loss function, which is also called the L_2 norm (see Ledoit and Wolf, 2004b [30]).

Definition 2.3.1 The L_2 norm of a $p \times p$ symmetric matrix Z with entries $z_{ij},\ i,j = 1,\ldots,p$ and eigenvalues $\lambda_i,\ i = 1,2,\ldots,p$ is defined by:

$$\|Z\|^2 = \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p} z_{ij}^2 = \frac{1}{p}\sum_{i=1}^{p}\lambda_i^2.$$

2.3.2 Shrinkage Target

Ledoit and Wolf (2003, 2004a, 2004b) [28, 29, 30] undertook three ways to specify the shrinkage target:

(a) the identity matrix multiplied by the mean of all the eigenvalues of the sample covariance matrix;

(b) the single-index model;

(c) the constant correlation model.

Alternatively, we utilize the largest eigenvalue of the sample covariance matrix to specify the shrinkage target, which is defined as:

$$F = \lambda_1 I,$$

where λ_1 denotes the largest eigenvalue of S. Then, the shrinkage covariance matrix is:

$$S^* = \alpha F + (1 - \alpha)S,$$

where α is the shrinkage intensity.

2.3.3 Computation of the Shrinkage Intensity

By the L_2 norm, the loss function of the shrinkage covariance matrix is:

$$L(\alpha) = \|\alpha F + (1 - \alpha)S - \Sigma\|^2.$$

Let s_{ij} denote the element of the sample covariance matrix S, σ_{ij} denote the element of the true covariance matrix, and f_{ij} denote the element of F.

The shrinkage intensity is estimated by minimizing the expectation of the loss function, which is called the risk function. The risk function is defined

and decomposed as follows:

$$\begin{aligned}
R(\alpha) &= \mathrm{E}(L(\alpha)) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{E}[\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij}]^2 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{E}[\alpha(f_{ij} - s_{ij}) + (s_{ij} - \sigma_{ij})]^2 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}[\alpha^2\mathrm{E}(f_{ij} - s_{ij})^2 + \mathrm{Var}(s_{ij}) + 2\alpha\mathrm{E}(f_{ij} - s_{ij})(s_{ij} - \sigma_{ij})] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}[\alpha^2\mathrm{E}(f_{ij} - s_{ij})^2 + \mathrm{Var}(s_{ij}) \\
&\qquad +2\alpha\mathrm{E}(f_{ij} - \mathrm{E}f_{ij} + \mathrm{E}f_{ij} - s_{ij})(s_{ij} - \sigma_{ij})] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}[\alpha^2\mathrm{E}(f_{ij} - s_{ij})^2 + \mathrm{Var}(s_{ij}) + 2\alpha\mathrm{Cov}(f_{ij}, s_{ij}) \\
&\qquad -2\alpha\mathrm{E}s_{ij}(s_{ij} - \sigma_{ij})] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}[\alpha^2\mathrm{E}(f_{ij} - s_{ij})^2 + (1-2\alpha)\mathrm{Var}(s_{ij}) + 2\alpha\mathrm{Cov}(f_{ij}, s_{ij})].
\end{aligned}$$

Differentiating $R(\alpha)$ with respect to α, we get

$$\begin{aligned}
R'(\alpha) &= \sum_{i=1}^{N}\sum_{j=1}^{N}[2\alpha\mathrm{E}(f_{ij} - s_{ij})^2 - 2\mathrm{Var}(s_{ij}) + 2\mathrm{Cov}(f_{ij}, s_{ij})], \\
R''(\alpha) &= \sum_{i=1}^{N}\sum_{j=1}^{N}[2\mathrm{E}(f_{ij} - s_{ij})^2].
\end{aligned}$$

Since $R''(\alpha) \geq 0$, the risk function has a minimum value. Setting $R'(\alpha) = 0$, we get the estimate of the shrinkage intensity α^* as:

$$\alpha^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Var}(s_{ij}) - \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}(f_{ij}, s_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{E}(f_{ij} - s_{ij})^2}.$$

2.3.4 Preliminaries of Partial Ordering

In this part, we will give preliminary results of *Löewner* partial ordering (see Wang, Wu and Jia, 2006), which will be useful in demonstrating the advantage of using the shrinkage covariance matrix. Suppose that A, B and C are all Hermitian matrices.

Definition 2.3.2. The *Löewner* partial ordering on the set of Hermitian matrices is defined as:

(a) $A \leq B$ iff $B - A$ is positive semidefinite;

(b) $A < B$ iff $B - A$ is positive definite.

Property. The partial ordering is:

(a) reflexive: $A \geq A$ for all A;

(b) antisymmetric: $A \geq B$ and $B \geq A$ are both true iff $A = B$;

(c) transitive: if $A \geq B$ and $B \geq C$ then $A \geq C$.

Any pair of Hermitian matrices satisfy precisely one of the following:

(a) none of the relations $A < B$, $A \leq B$, $A = B$, $A \geq B$, $A > B$ is true;

(b) only $A < B$ and $A \leq B$ are true;

(c) only $A \leq B$ is true;

(d) only $A = B$, $A \leq B$ and $A \geq B$ are true;

(e) only $A \geq B$ is true;

(f) only $A > B$ and $A \geq B$ are true.

Now, some preliminary lemmas are introduced to provide the foundation of the main theorem in this chapter.

Lemma 2.3.1. (*see Theorem 7.2.2 of Wang, Wu and Jia, 2006*) Suppose that A and B are both $n \times n$ Hermitian matrices, Q is $n \times k$ matrix. Then

$$A \geq B \Longrightarrow Q^*AQ \geq Q^*BQ.$$

Proof. By Definition 2.3.2, we know that for any $x \in R^n$, we have $x^*(A - B)x \geq 0$. Therefore, for any $x \in R^n$,

$$x^*(Q^*AQ - Q^*BQ)x = (Qx)^*(A - B)(Qx) \geq 0,$$

which completes the proof.

Lemma 2.3.2. (*see Theorem 7.2.3 of Wang, Wu and Jia, 2006*) $A > 0$ and $B > 0$, then

$$A \geq B \Longleftrightarrow \lambda_1(BA^{-1}) \leq 1.$$

Proof. By Lemma 2.3.1, we have

$$\begin{aligned} A \geq B &\Longleftrightarrow A^{-1/2}(A - B)A^{-1/2} \geq 0 \\ &\Longleftrightarrow I - A^{-1/2}BA^{-1/2} \geq 0 \\ &\Longleftrightarrow \lambda_i(I - A^{-1/2}BA^{-1/2}) \geq 0,\ i = 1, \ldots, n \\ &\Longleftrightarrow \lambda_1(A^{-1/2}BA^{-1/2}) \leq 1 \\ &\Longleftrightarrow \lambda_1(BA^{-1}) \leq 1, \end{aligned}$$

where $\lambda_i(A)$ denotes the ith largest eigenvalue of A.

From Lemma 2.3.2, we have the following proposition.

Proposition 2.3.1. (*see Proposition 7.2.2 of Wang, Wu and Jia, 2006*) If A and B are positive definite Hermitian matrices, then

$$A \geq B \Longleftrightarrow B^{-1} \geq A^{-1}.$$

Lemma 2.3.3. (*see Lemma 7.2.1 of Wang, Wu and Jia, 2006*) If $A \geq B \geq 0$, then

$$\mathscr{M}(B) \subset \mathscr{M}(A),$$

where $\mathscr{M}(A)$ denotes the column space of A.

Proof. From Definition 2.3.2, we know that for any vector $x \in R^n$, we have

$$A \geq B \Longleftrightarrow x^*Ax \geq x^*Bx.$$

If $x \in \mathscr{M}(A)^{\perp}$, where $\mathscr{M}(A)^{\perp}$ denotes the complementary space of $\mathscr{M}(A)$, then $x^*Ax = 0 \Longrightarrow x^*Bx = 0$, which means that $x \in \mathscr{M}(B)^{\perp}$. Thus we have

$$\mathscr{M}(A)^{\perp} \subset \mathscr{M}(B)^{\perp},$$

further,

$$\mathscr{M}(B) \subset \mathscr{M}(A).$$

Lemma 2.3.4. (*see Lemma 7.2.2 of Wang, Wu and Jia, 2006*) If A and B are positive semidefinite Hermitian matrices, then

$$A \geq B \Longleftrightarrow \mathscr{M}(B) \subset \mathscr{M}(A), \text{and } A(A-B)A \geq 0.$$

Proof. ($\Longrightarrow$) By Lemma 2.3.3 and the definition of positive semidefiniteness, the necessary condition is obviously satisfied.
($\Longleftarrow$) Herein let A^+ denote the Moore–Penrose inverse of matrix A. Since AA^+ is an orthogonal projection matrix to $\mathscr{M}(A)$, then from $\mathscr{M}(B) \subset \mathscr{M}(A)$, we get

$$AA^+(A-B)AA^+ = A - B.$$

Therefore, under the condition that $\mathscr{M}(B) \subset \mathscr{M}(A)$,

$$A \geq B \Longleftrightarrow x^* AA^+(A-B)AA^+x \geq 0, \text{ for any } x \in R^n.$$

$\mathscr{M}(AA^+) = \mathscr{M}(A)$, therefore,

$$x^* AA^+(A-B)AA^+x \geq 0, \text{ for any } x \in R^n$$

equates to

$$x^* A(A-B)Ax \geq 0, \text{ for any } x \in C^n,$$

which concludes the proof.

Lemma 2.3.5. (*see Theorem 1.7.4 of Wang, Wu and Jia, 2006*) If $B \neq 0$, $C \neq 0$, then BA^-C is independent of the choice of A^-, if and only if $\mathscr{M}(C) \subset \mathscr{M}(A)$, $\mathscr{M}(B') \subset \mathscr{M}(A')$.

Proof. ($\Longrightarrow$) From the assumption, there exist matrices X and Y, s.t. $C = AX$, $B' = A'Y$. Therefore,

$$BA^-C = Y'AA^-AX = Y'AX,$$

while the right side is independent of A^-.
($\Longleftarrow$) Any general inverse of A can be expressed as:

$$A^- = A^- + U - A^-AUAA^-, \text{ for all } U.$$

Therefore,

$$BA^-C = BA^-C + BUC - BA^-AUAA^-C, \text{ for all } U.$$

$$BUC - BA^-AUAA^-C = 0, \text{ for all } U.$$

Let $U = A^-AZ$, Z can be any matrix, then

$$BA^-AZ(C - AA^-C) = 0.$$

Therefore, $BA^-A = 0$ or $C = AA^-C$. If $BA^-A = 0$, then $BUC = 0$ for all U, which implies that $B = 0$ or $C = 0$, contrary to the assumption that $B \neq 0$ and $C \neq 0$. Therefore, $C = AA^-C$, $\mathscr{M}(C) \subset \mathscr{M}(A)$. Using the same method, we can prove that $\mathscr{M}(B') \subset \mathscr{M}(A')$.

By Lemma 2.3.5, we have the following proposition.

Proposition 2.3.2. (*see Proposition 1.7.1 of Wang, Wu and Jia, 2006*) If $B \neq 0$, then $B'A^-B$ is independent of A if and only if $\mathscr{M}(B) \subset \mathscr{M}(A)$.

Lemma 2.3.6. (*see Theorem 7.2.4 of Wang, Wu and Jia, 2006*) If A and B are Hermitian matrices, and $A \geq 0$, $B \geq 0$, then

$$A \geq B \Longleftrightarrow \mathscr{M}(B) \subset \mathscr{M}(A),\ \lambda_1(BA^-) \leq 1,$$

where $\lambda_1(BA^-) \leq 1$ is independent of the choice of A^-.

Proof. From Lemma 2.3.3, we can easily get

$$A \geq B \Longrightarrow \mathscr{M}(B) \subset \mathscr{M}(A).$$

Therefore, this is equivalent to proving

$$A \geq B \Longleftrightarrow \lambda_1(BA^+) \leq 1$$

under the condition $\mathscr{M}(B) \subset \mathscr{M}(A)$. By Lemma 2.3.4, the problem further becomes

$$A(A-B)A \geq 0 \Longleftrightarrow \lambda_1(BA^+) \leq 1$$

under the same condition. Do full rank decomposition to A: $A = LL^*$, herein L is $n \times k$ matrix, $k = r(A)$. Then

$$A(A-B)A \geq 0 \Longleftrightarrow L^*(A-B)L \geq 0.$$

Multiply $(L^*L)^{-1}$ on the left side and the right side of the above inequality, we get

$$\begin{aligned} L^*(A-B)L \geq 0 &\iff I - T^*BT \geq 0, \\ &\iff \lambda_1(T^*BT) \leq 0 \\ &\iff \lambda_1(BTT^*) \leq 0 \\ &\iff \lambda_1(BA^+) \leq 0, \end{aligned}$$

where $T = L(L^*L)^{-1}$ and $TT^* = L(L^*L)^{-2}L^* = A^+$.
Finally, we need to prove for any general inverse A^-,

$$\lambda_1(BA^-) = \lambda_1(BA^+).$$

Since $\mathscr{M}(B) \subset \mathscr{M}(A)$, by Proposition 2.3.2, $B^{1/2}A^-B^{1/2}$ is independent of the choice of A^-. We have, for all A^-,

$$B^{1/2}A^-B^{1/2} = B^{1/2}A^+B^{1/2}.$$

Therefore,

$$\lambda_1(BA^-) = \lambda_1(B^{1/2}A^-B^{1/2}) = \lambda_1(B^{1/2}A^+B^{1/2}) = \lambda_1(BA^+).$$

2.3.5 Theorem

Based on these results, we have the following theorem.

Theorem 2.3.1. For given $\hat{\boldsymbol{\mu}}$,

$$E(\hat{\boldsymbol{\mu}}^{\mathrm{T}}h(S^{-1})\hat{\boldsymbol{\mu}}) \geq E(\hat{\boldsymbol{\mu}}^{\mathrm{T}}h(S^{*-1})\hat{\boldsymbol{\mu}}),$$

where $S^* = \alpha^*\lambda_1 I + (1-\alpha^*)S$.

To prove Theorem 2.3.1, we have to prove the following theorems first.

Theorem 2.3.2. $S^{-1} \geq S^{*-1}$.

Proof. By Proposition 2.3.1,

$$S^{-1} \geq S^{*-1} \Longleftrightarrow S^{*} \geq S.$$

Factorize S as $Q\Lambda Q'$, where $QQ' = I$, $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$.

$$\begin{aligned} S^{*} - S &= \alpha\lambda_1 I + (1-\alpha)S - S \\ &= \alpha(\lambda_1 I - S) \\ &= \alpha Q(\lambda_1 I - \Lambda)Q' \\ &\geq 0. \end{aligned}$$

For notational convenience, in the following part, let $A = S^{-1}$ and $B = S^{*-1}$.

Theorem 2.3.3. $\mathscr{M}(h(B)) \subset \mathscr{M}(h(A))$.

Proof. If we can find a matrix X which makes $h(B) = h(A)X$ hold, then we can obtain the result. Since

$$(I - \frac{A\mathbf{1}\mathbf{1}^{\mathrm{T}}}{\mathbf{1}^{\mathrm{T}}A\mathbf{1}})(I - \frac{B\mathbf{1}\mathbf{1}^{\mathrm{T}}}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}}) = I - \frac{B\mathbf{1}\mathbf{1}^{\mathrm{T}}}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}},$$

let

$$X = A^{-1}(I - \frac{B\mathbf{1}\mathbf{1}^{\mathrm{T}}}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}})B,$$

the proof is completed.

Theorem 2.3.4. $\lambda_1[h(B)h(A)^{+}] \leq 1$.

Proof. By Lemma 2.3.2 and Lemma 2.3.6,

$$\lambda_1[h(B)h(A)^{+}] = \lambda_1[h(B)h(A)^{-}],$$

where $h(A)^-$ can be any general inverse of $h(A)$. One of the general inverses of $h(A)$ is

$$h(A)^- = (I - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}A}{\mathbf{1}^{\mathrm{T}}A\mathbf{1}})A^{-1}.$$

$$\begin{aligned}
\lambda_1[(h(B)h(A)^-)] &= \lambda_1[B(I - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}B}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}})(I - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}A}{\mathbf{1}^{\mathrm{T}}A\mathbf{1}})A^{-1}] \\
&= \lambda_1[B(I - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}B}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}})A^{-1}] \\
&= \lambda_1[A^{-1}B(I - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}B}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}})] \\
&\leq \lambda_1(A^{-1}B)\lambda_{max}(I - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}B}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}}) \\
&= \lambda_1(BA^{-1})\lambda_{max}(I - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}B}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}}).
\end{aligned}$$

Since $A \geq B$, by Lemma 2.3.2, $\lambda_1(BA^{-1}) \leq 1$. $I - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}B}{\mathbf{1}^{\mathrm{T}}B\mathbf{1}}$ is an idempotent matrix, therefore, the largest eigenvalue of it is 1. Thus we have

$$\lambda_1[(h(B)h(A)^+)] \leq 1.$$

Combining Theorem 2.3.2, Theorem 2.3.3 and Theorem 2.3.4, we have

$$h(S^{-1}) \geq h(S^{*-1}).$$

Here $\hat{\boldsymbol{\mu}}$ can be any vector, taking expectation of both sides, by Lemma 2.3.6, we have Theorem 2.3.1.

2.4 New Estimator

In this part, we introduce a new estimator of the optimal expected return. The joint impact of the sample mean vector and the sample covariance matrix are taken into consideration together. Consider the following theorem.

Theorem 2.3.5. $\mathrm{E}(\hat{P}^{*2})$ can be reduced in two steps as:

$$\begin{aligned}
\mathrm{E}(\hat{P}^{*2}) &= \mathrm{E}[\mathrm{E}(\hat{P}^{*2}|\hat{\boldsymbol{\mu}})] \\
&= \mathrm{E}[\mathrm{E}(\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{-1})\hat{\boldsymbol{\mu}}|\hat{\boldsymbol{\mu}})] \\
&\geq \mathrm{E}[\mathrm{E}(\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}|\hat{\boldsymbol{\mu}})] \\
&= \mathrm{E}(\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}) \\
&= \mathrm{E}[\mathrm{E}(\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}|S^{*})] \\
&= \frac{1}{n}\mathrm{E}[\mathrm{tr}h(S^{*-1})\Sigma] + \mathrm{E}[\boldsymbol{\mu}^{\mathrm{T}} h(S^{*-1})\boldsymbol{\mu}] \\
&> \mathrm{E}[\boldsymbol{\mu}^{\mathrm{T}} h(S^{*-1})\boldsymbol{\mu}]
\end{aligned}$$

2.4.1 New Estimator

Theorem 2.3.5 is employed to construct our new estimator. Define K as

$$K = \frac{1}{n}\mathrm{tr}[h(S^{*-1})\Sigma)].$$

The new estimator for P^{*} is:

$$\hat{P}^{*}_{new} = \begin{cases} \sqrt{\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}} - K}, & \text{if } \hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}} \geq K, \\ \sqrt{\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}}, & \text{otherwise.} \end{cases}$$

Herein, notice that $\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}$ cannot always be larger than K since we use one observation to replace the expectation. When $\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}$ is smaller than K, we can only use the shrinkage covariance matrix. Although the sample mean's effect is not taken into consideration for this case, the estimator is at least as good as Ledoit and Wolf's estimator.

2.4.2 Algorithm

In this part, we introduce how to estimate the values of the shrinkage intensity α^{*} and K.

Shrinkage Intensity α^*

According to our shrinkage target $F = \lambda_1 I$, recall that the theoretical estimate of α^* is:

$$\alpha^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(s_{ij}) - \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}(f_{ij}, s_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{E}(f_{ij} - s_{ij})^2}.$$

Therefore, to estimate α^*, the values $\mathrm{Var}(s_{ij})$, $\mathrm{Cov}(f_{ij}, s_{ij})$ and $\mathrm{E}(f_{ij} - s_{ij})^2$ need to be estimated.

Herein, the bootstrap method (see, for example, Efron and Tibshirani, 1993 [19]) is chosen to give numerical estimates of these three values. Since the sample covariance matrix S is always singular in our case, and the parametric resampling needs the sample covariance matrix to be invertible, it is therefore not a good choice. Consequently, the non-parametric resampling method is used to generate different data sets based on the observations.

Suppose that the number of resampling times is N. Each time, resampling is taken within each asset with replacement, and the kth data set is generated as follows:

$$\begin{pmatrix} x_{11}^{(k)} & x_{12}^{(k)} & \dots & x_{1p}^{(k)} \\ x_{21}^{(k)} & x_{22}^{(k)} & \dots & x_{2p}^{(k)} \\ & \dots & \dots & \\ x_{n1}^{(k)} & x_{n2}^{(k)} & \dots & x_{np}^{(k)} \end{pmatrix}$$

and the kth sample covariance matrix $S^{(k)}$ is:

$$\begin{pmatrix} s_{11}^{(k)} & s_{12}^{(k)} & \dots & s_{1p}^{(k)} \\ s_{21}^{(k)} & s_{22}^{(k)} & \dots & s_{2p}^{(k)} \\ & \dots & \dots & \\ s_{p1}^{(k)} & s_{p2}^{(k)} & \dots & s_{pp}^{(k)} \end{pmatrix}.$$

The kth shrinkage target is:

$$F^{(k)} = \lambda_1^{(k)} I,$$

where $\lambda_1^{(k)}$ is the largest eigenvalue of $S^{(k)}$.

Then we get the estimates of the three values as:

$$\begin{aligned}
\hat{\mathrm{Var}}(s_{ij}) &= \frac{1}{N-1}\sum_{k=1}^{N}(s_{ij}^{(k)} - \bar{s}_{ij})^2, \\
\hat{\mathrm{Cov}}(f_{ij}, s_{ij}) &= \frac{1}{N-1}\sum_{k=1}^{N}(s_{ij}^{(k)} - \bar{s}_{ij})(f_{ij}^{(k)} - \bar{f}_{ij}), \\
\hat{\mathrm{E}}(f_{ij} - s_{ij})^2 &= \frac{1}{N}\sum_{k=1}^{N}(f_{ij}^{(k)} - s_{ij}^{(k)})^2,
\end{aligned}$$

where $\bar{s}_{ij} = \sum_{k=1}^{N} s_{ij}^{(k)}/N$ and $\bar{f}_{ij} = \sum_{k=1}^{N} f_{ij}^{(k)}/N$.

Therefore, the estimate of α^* is:

$$\hat{\alpha}^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathrm{Var}}(s_{ij}) - \sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathrm{Cov}}(f_{ij}, s_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathrm{E}}(f_{ij} - s_{ij})^2}.$$

Impact of the Sample Mean Vector: K

Since the impact of the sample mean vector K depends on the true covariance matrix Σ, herein, for simplicity, we just replace Σ by the sample covariance matrix S. Therefore, the estimate of K is:

$$\hat{K} = \frac{1}{n}\mathrm{tr}[h(S^{*-1})S)].$$

In this expression, S^{*-1} is always invertible, therefore, there is no difficulty in estimating K. Thus far, we have given the estimates of the two unknown values of the new estimator. We will do simulations to check the performance of the new estimator in Chapter 3.

□ **End of chapter.**

Chapter 3

Simulation Studies

3.1 Introduction

In this chapter, simulation studies are conducted to show how the new estimator behaves when estimating the optimal expected return. It is compared with four different estimators: the benchmark value, the plug-in estimate, the bootstrap corrected estimate (Bai et al., 2009a, 2009b [2] [3]) and the shrinkage estimate (Ledoit and Wolf, 2004a [29]). The details are introduced in the second part of this chapter. We employ two different methods to construct data sets: one is to use the multi–index model while the other is to use historical stock returns of the American stock market to specify the true mean and the true covariance matrix, from which the empirical returns are simulated using the multivariate normal distribution.

Following the convention of financial analysis (see Chan, 2010 [10]), in this chapter, the return on a stock i at time t is asymptotically defined as:

$$x_{ti} = \log P_{ti} - \log P_{ti-1},$$

where P_{ti} denotes the price of stock i at time t.

3.2 Comparison Estimators

In this part, we mainly introduce the four comparison estimators and the corresponding algorithms.

1. Benchmark Value (r_{real})

The benchmark value is the theoretical optimal expected return computed using the true mean vector $\boldsymbol{\mu}$ and the true covariance matrix Σ of the data set.

2. Plug-in Estimate (r_{plug})

The plug-in estimate is computed using the sample mean vector $\hat{\boldsymbol{\mu}}$ and the sample covariance matrix S.

3. Shrinkage Estimate (r_{shrink})

The shrinkage estimate is computed by plugging in the sample mean $\hat{\boldsymbol{\mu}}$ and the shrinkage covariance matrix S^*. Following Ledoit and Wolf's work (2004a) [29], the shrinkage target is chosen as the constant correlation model, because it is easy to implement. The population constant correlation model is denoted as Φ and the sample constant correlation model is denoted as F. The constant correlation model is constructed as follows.

Suppose that the elements of the true covariance matrix Σ and the sample covariance matrix S are denoted by σ_{ij} and s_{ij} respectively. And the elements of Φ and F are ϕ_{ij} and f_{ij}. The population and sample correlation between two stocks i and j are defined as:

$$\begin{aligned} \rho_{ij} &= \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}, \\ r_{ij} &= \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}; \end{aligned}$$

and the averages of the population and the sample correlations are given by:

$$\bar{\rho} = \frac{2}{(p-1)p}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\rho_{ij} ,$$
$$\bar{r} = \frac{2}{(p-1)p}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}r_{ij} .$$

Then Φ and F are constructed as:

$$\phi_{ii} = \sigma_{ii} ,$$
$$\phi_{ij} = \bar{\rho}\sqrt{\sigma_{ii}\sigma_{jj}} ,$$
$$f_{ii} = s_{ii} ,$$
$$f_{ij} = \bar{r}\sqrt{s_{ii}s_{jj}} .$$

Under the assumption that p is fixed while n tends to infinity, the optimal shrinkage intensity δ behaves asymptotically like a constant over the sample size n (see Ledoit and Wolf (2004a) [29]), and the constant is given by:

$$\kappa = \frac{\pi - \rho}{\gamma},$$

where π, ρ and γ are defined as:

$$\pi = \sum_{i=1}^{p}\sum_{j=1}^{p}\text{AsyVar}\left[\sqrt{n}s_{ij}\right];$$
$$\rho = \sum_{i=1}^{p}\sum_{j=1}^{p}\text{AsyCov}\left[\sqrt{n}f_{ij}, \sqrt{n}s_{ij}\right];$$
$$\gamma = \sum_{i=1}^{p}\sum_{j=1}^{p}(\phi_{ij} - \sigma_{ij})^2;$$

where the notations AsyVar and AsyCov represent the asymptotic variance and the asymptotic covariance respectively. Since these three values are unknown, the specific algorithm for estimating them is given as follows (see Ledoit and Wolf (2004a) [29]).

The consistent estimator of π is given by:

$$\hat{\pi} = \sum_{i=1}^{p}\sum_{j=1}^{p}\hat{\pi}_{ij},$$

with

$$\hat{\pi}_{ij} = \frac{1}{n}\sum_{t=1}^{n}\{(x_{ti} - \bar{r}_{i.})(x_{tj} - \bar{r}_{j.}) - s_{ij}\}^2;$$

The consistent estimator of ρ is given by:

$$\hat{\rho} = \sum_{i=1}^{p}\hat{\pi}_{ii} + \sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p}\frac{\bar{r}}{2}\left(\sqrt{\frac{s_{jj}}{s_{ii}}}\hat{\theta}_{ii,ij} + \sqrt{\frac{s_{ii}}{s_{jj}}}\hat{\theta}_{jj,ij}\right),$$

with

$$\hat{\theta}_{ii,ij} = \frac{1}{n}\sum_{t=1}^{n}\{(x_{ti} - \bar{r}_{i.})^2 - s_{ii}\}\{(x_{ti} - \bar{r}_{i.})(x_{tj} - \bar{r}_{j.}) - s_{ij}\},$$

$$\hat{\theta}_{jj,ij} = \frac{1}{n}\sum_{t=1}^{n}\{(x_{tj} - \bar{r}_{j.})^2 - s_{jj}\}\{(x_{ti} - \bar{r}_{i.})(x_{tj} - \bar{r}_{j.}) - s_{ij}\};$$

and the consistent estimator of γ is given by:

$$\hat{\gamma} = \sum_{i=1}^{p}\sum_{j=1}^{p}(f_{ij} - s_{ij})^2.$$

Therefore, the consistent estimator of κ is:

$$\hat{\kappa} = \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}};$$

and the estimate of the optimal shrinkage intensity is given by:

$$\hat{\delta} = \max\left\{0, \min\left\{\frac{\hat{\kappa}}{n}, 1\right\}\right\}.$$

4. Bootstrap Corrected Estimate (r_{bs})

In Bai et al. (2009b) [3], an efficient estimator using the theory of random matrices was developed to solve the over prediction problem. The parametric bootstrap technique was employed in their study. The procedure is as follows.

(a) A resample $\chi^* = (X_1^*, \ldots, X_n^*)$ is drawn from the p-dimensional multivariate normal distribution with mean vector $\hat{\boldsymbol{\mu}}$ and covariance matrix S.

(b) The sample mean vector and the sample covariance matrix of the resample data set is calculated denoted by $\hat{\mu}_{bs}$ and S_{bs} respectively. Then, by applying the optimization procedure again, we obtain the bootstrapped plug-in estimate of the optimal expected return r^*_{plug}.

(c) The bootstrapped corrected return estimate is given by:

$$r_{bs} = r_{plug} + \frac{1}{\sqrt{\gamma}}(r_{plug} - r^*_{plug}).$$

3.3 Empirical Study 1

To make the data set more realistic, multi–index model (see, for example, Elton and Gruber, 1997 [22]) is employed to generate empirical stock returns. Firstly, we give a brief introduction to the multi–index model, then the data set is constructed based on the information drawn from ten 'blue-chip' stocks of Dow Jones Industrial Index.

3.3.1 Multi–index Model

In this study, we choose the number of indices as three. Then, the return on a stock i can be written as

$$x_{ti} = \alpha_i + \beta_{i1}x_{0t} + \beta_{i2}y_{0t} + \beta_{i3}z_{0t} + \epsilon_{ti},$$

where x_{0t}, y_{0t} and z_{0t} denote the rate of return on the three market indices; α_i is the component of the return of stock i that is independent of the market's performance; β_{i1}, β_{i2} and β_{i3} are constants which measure the expected change in x_{ti} given a change in x_{0t}, y_{0t} and z_{0t} respectively; and ϵ_{ti} is the idiosyncratic error term which is uncorrelated with x_{0t}, y_{0t} and z_{0t}.

Consider the statistical structure of the multi-index model. Assume

$$\begin{aligned} x_{0t} &\sim N(\mu_x, \sigma_x^2), \\ y_{0t} &\sim N(\mu_y, \sigma_y^2), \\ z_{0t} &\sim N(\mu_z, \sigma_z^2), \\ \epsilon_{it} &\sim N(0, \delta_{ii}^2). \end{aligned}$$

Since the correlations between market indices cannot be zero, suppose that

$$\begin{aligned} \rho_{xy} &= \text{Corr}(x_{0t}, y_{0t}), \\ \rho_{yz} &= \text{Corr}(y_{0t}, z_{0t}), \\ \rho_{xz} &= \text{Corr}(x_{0t}, z_{0t}); \end{aligned}$$

and the covariance structure of the error terms are

$$\begin{aligned} \text{Cov}(\epsilon_{it}, \epsilon_{jt}) &= 0, \\ \text{Cov}(\epsilon_{it}, \epsilon_{is}) &= 0. \end{aligned}$$

Therefore, for the covariance matrix of the multi-index model V, the elements on and off the diagonal are respectively

$$\begin{aligned} V_{ii} &= \beta_{i1}^2\sigma_x^2 + \beta_{i2}^2\sigma_y^2 + \beta_{i3}^2\sigma_z^2 + \delta_{ii}^2 + 2\beta_{i1}\beta_{i2}\rho_{xy}\sigma_x\sigma_y \\ &\quad +2\beta_{i1}\beta_{i3}\rho_{xz}\sigma_x\sigma_z + 2\beta_{i2}\beta_{i3}\rho_{yz}\sigma_y\sigma_z, \\ V_{ij} &= \beta_{i1}\beta_{j1}\sigma_x^2 + \beta_{i2}\beta_{j2}\sigma_y^2 + \beta_{i3}\beta_{j3}\sigma_z^2 + (\beta_{i1}\beta_{j2} + \beta_{i2}\beta_{j1})\rho_{xy}\sigma_x\sigma_y \\ &\quad +(\beta_{i1}\beta_{j3} + \beta_{i3}\beta_{j1})\rho_{xz}\sigma_x\sigma_z + (\beta_{i2}\beta_{j3} + \beta_{i3}\beta_{j2})\rho_{yz}\sigma_y\sigma_z. \end{aligned}$$

3.3.2 Constructing the Data Set

In this part, we use the multi-index model to construct data sets of empirical stock returns with different combinations of dimension p and sample size n. The procedure is as follows: firstly, generate the three market indices x_{0t}, y_{0t} and z_{0t} using normal distributions; secondly, generate the values of α, β_1, β_2,

β_3 and ϵ_{ti} also using normal distributions while generating the variance of ϵ_{ti} using chi–square distribution. Then we get the return of stock i at time t as

$$x_{ti} = \alpha_i + \beta_{i1}x_{0t} + \beta_{i2}y_{0t} + \beta_{i3}z_{0t} + \epsilon_{ti}.$$

To specify more realistic values of the parameters of the distributions mentioned above, prices of three market indices and ten 'blue-chip' stocks of Dow Jones Industrial Index over a time period from January 3, 2007 to December 31, 2010 are downloaded from the American stock market. The three market indices chosen are Dow–Jones Industrial, S & P 500 and NASDAQ Composite. And the ten stocks are listed below:

1. JP Morgan Chase & Co. Common St (JPM);
2. Bank of America Corporation Com (BAC);
3. Boeing Company (The) Common Sto (BA);
4. Wal-Mart Stores, Inc. Common St (WMT);
5. McDonald's Corporation Common S (MCD);
6. Inter Corporation (INTC);
7. Microsoft Corporation (MSFT);
8. E. I. du Pont de Nemours and Com (DD);
9. Coca-Cola Company (The) Common (KO);
10. General Electric Company Common (GE).

Consider the market indices first. The mean and variance of each real index, and the correlation structures are summarized in Table 3.1 and Table 3.2.

Table 3.1: Mean and Variance of Market Indices.

	x_{0t}	y_{0t}	z_{0t}
mean	$-3.22 * 10^{-5}$	$-5.10 * 10^{-5}$	$3.91 * 10^{-5}$
variance	$4.67 * 10^{-5}$	$5.65 * 10^{-5}$	$5.98 * 10^{-5}$

Table 3.2: Correlation between Market Indices.

ρ_{xy}	ρ_{yz}	ρ_{xz}
0.9874	0.9642	0.9430

Then the empirical returns of the market indices for a given sample size are simulated from a 3–dimensional multivariate normal distribution, with the mean vector and the covariance matrix specified by the above two tables.

Secondly, consider specifying the parameter values of the distributions of α, β_1, β_2, β_3 and ϵ_{ti}. Regression analysis is applied to estimate the values of α, β_1, β_2, β_3 and residual standard errors for each 'blue-chip' stock. The results are summarized in Table 3.3.

Table 3.3: Regression Analysis of Ten Stocks Using Multi-Index Model.

No.	stock	α	β_1	β_2	β_3	residual se
1	JPM	0.0002	-1.3435	3.6042	-0.7147	0.0108
2	BAC	-0.0002	-2.9595	6.6935	-1.9552	0.0153
3	BA	-0.0001	2.0844	-1.1487	0.2260	0.0065
4	WMT	0.0001	1.4873	-0.8392	0.0314	0.0048
5	MCD	0.0003	1.3654	-0.8665	0.2143	0.0046
6	INTC	0.0000	0.5941	-0.8572	1.3822	0.0058
7	MSFT	-0.0001	1.3333	-1.6611	1.3983	0.0056
8	DD	0.0001	0.4785	0.5383	0.1592	0.0054
9	KO	0.0002	1.3363	-0.7314	0.0724	0.0047
10	GE	-0.0001	-0.3495	1.9441	-0.4622	0.0077

The means and variances of α, β_1, β_2 and β_3 are summarized in Table 3.4, from which the empirical values are simulated.

Table 3.4: Mean and Variance of Parameters.

	α	β_1	β_2	β_3
mean	$4.00 * 10^{-5}$	0.4027	0.6676	0.0351
variance	$2.71 * 10^{-8}$	2.4034	7.1344	0.9477

Suppose that

$$\delta_{ii}^2 \sim \frac{6.13 * 10^{-5}}{n-1} \chi_{n-1}^2,$$

herein $n = 1005$, the total number of observations.

Once the values of α, β_1, β_2, β_3 and δ_{ii}^2 are simulated, they are regarded as nonrandom. Then the true mean vector and true covariance matrix V can be fixed. Thus far, we have generated all the empirical returns with the true mean vector and true covariance matrix known.

3.3.3 Simulation Results

To be general, we consider six combinations of p and n. When $p \leq n$, we have the theoretical explanation as given in Chapter 3. The new estimator is also applied to the cases of $p > n$ to get empirical results. Consider the following cases: $p = 30,\ n = 50$; $p = 50,\ n = 100$; $p = 50,\ n = 50$; $p = 100,\ n = 100$; $p = 100,\ n = 50$; $p = 150,\ n = 100$.

For each case, we simulate 100 data sets and for each data set i, we calculate the values of r_{real}^i, r_{new}^i, r_{shrink}^i, r_{plug}^i and r_{bs}^i. The average distance is used to characterize the deviations of the four estimates from the benchmark value. Define

$$
\begin{aligned}
d_1 &= \frac{1}{100}\sum_{i=1}^{100}\left|r_{new}^i - r_{real}^i\right|, \\
d_2 &= \frac{1}{100}\sum_{i=1}^{100}\left|r_{plug}^i - r_{real}^i\right|, \\
d_3 &= \frac{1}{100}\sum_{i=1}^{100}\left|r_{shrink}^i - r_{real}^i\right|, \\
d_4 &= \frac{1}{100}\sum_{i=1}^{100}\left|r_{bs}^i - r_{real}^i\right|.
\end{aligned}
$$

From Table 3.5, we see that r_{new} has the minimum average distance from the benchmark value in each combination of p and n. When $p < n$, r_{plug} is the

Table 3.5: Average distance of the Four Estimates.

	d_1	d_2	d_3	d_4
$p = 30,\ n = 50$	0.0960	1.0906	0.6736	0.4321
$p = 50,\ n = 100$	0.0617	0.8477	0.6123	0.3519
$p = 50,\ n = 50$	0.0852	—	1.0386	—
$p = 100,\ n = 100$	0.1097	—	1.1604	—
$p = 100,\ n = 50$	0.0786	—	2.1276	—
$p = 150,\ n = 100$	0.1947	—	1.8751	—

worst among all the estimators. r_{bs} performs better than r_{shrink}. When $p \geq n$, the plug-in estimate and the bootstrap corrected estimate cannot be applied. We also observe that when p/n is substantial, the shrinkage estimate becomes less accurate.

In Figures 3.1–3.12, the y–axis denotes the optimal expected return; and the x–axis represents the ith data set. The black solid line corresponds to the benchmark value r_{real}; the red dashed line corresponds to the new estimator r_{new}; the yellow dotted line corresponds to the shrinkage estimate r_{shrink}, the green dot–dashed line corresponds to the plug-in estimate r_{plug}; and the blue long–dashed line corresponds to the bootstrap corrected estimate r_{bs}.

From Figures 3.1–3.6, we find that the plug-in estimate is the worst among all the estimators. It over predicts the benchmark value heavily in each data set. When $p < n$, the bootstrap corrected estimate is closer to the benchmark value than the shrinkage estimate. However, it is not as stable as the shrinkage estimate over time, and it is not efficient when $p \geq n$. As p/n grows,

the shrinkage estimate deviates farther from the benchmark value. The new estimator not only is much more stable than the others in each case, but it is also always closest to the benchmark value.

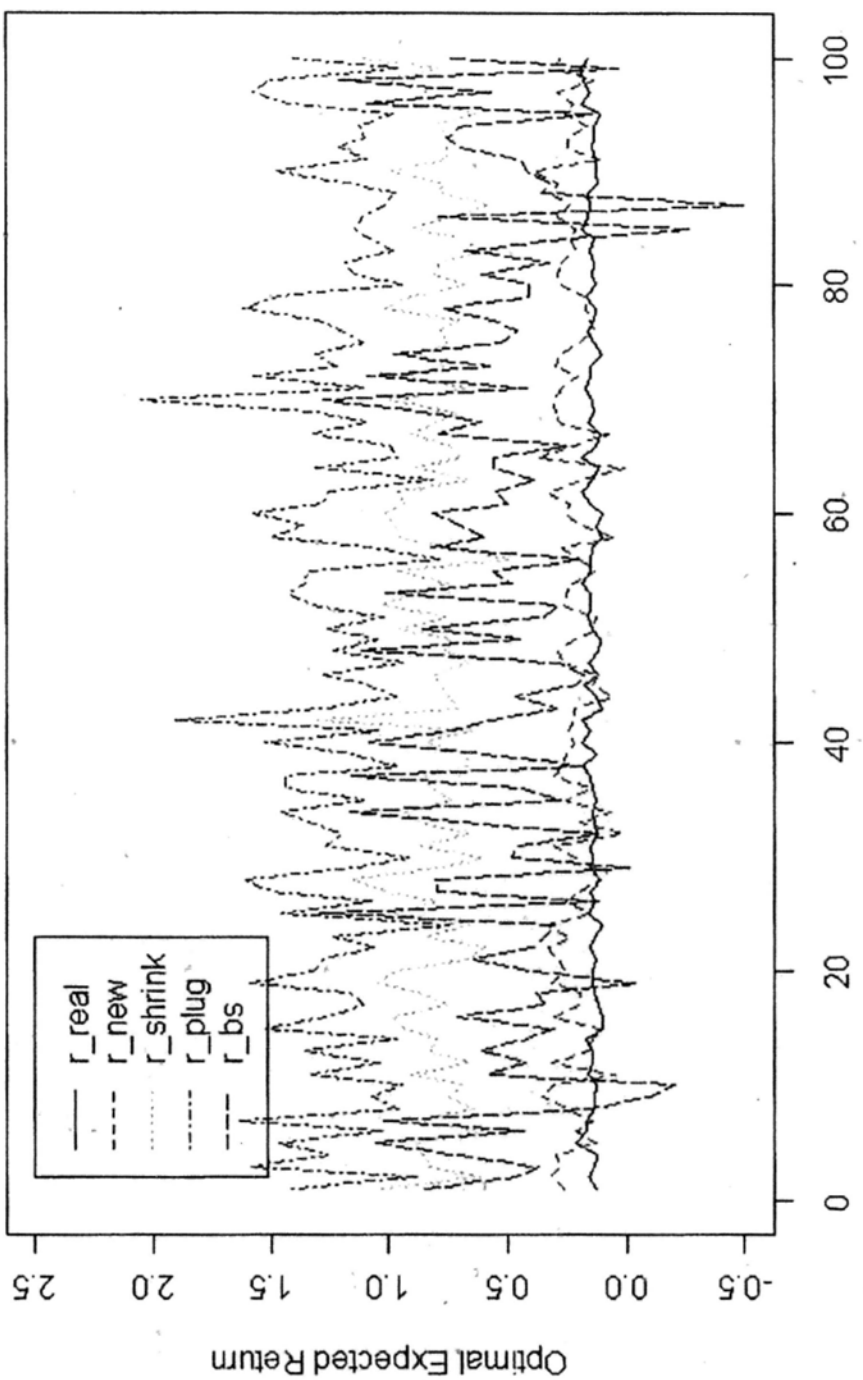

Figure 3.1: Model Comparison $p = 30$, $n = 50$.

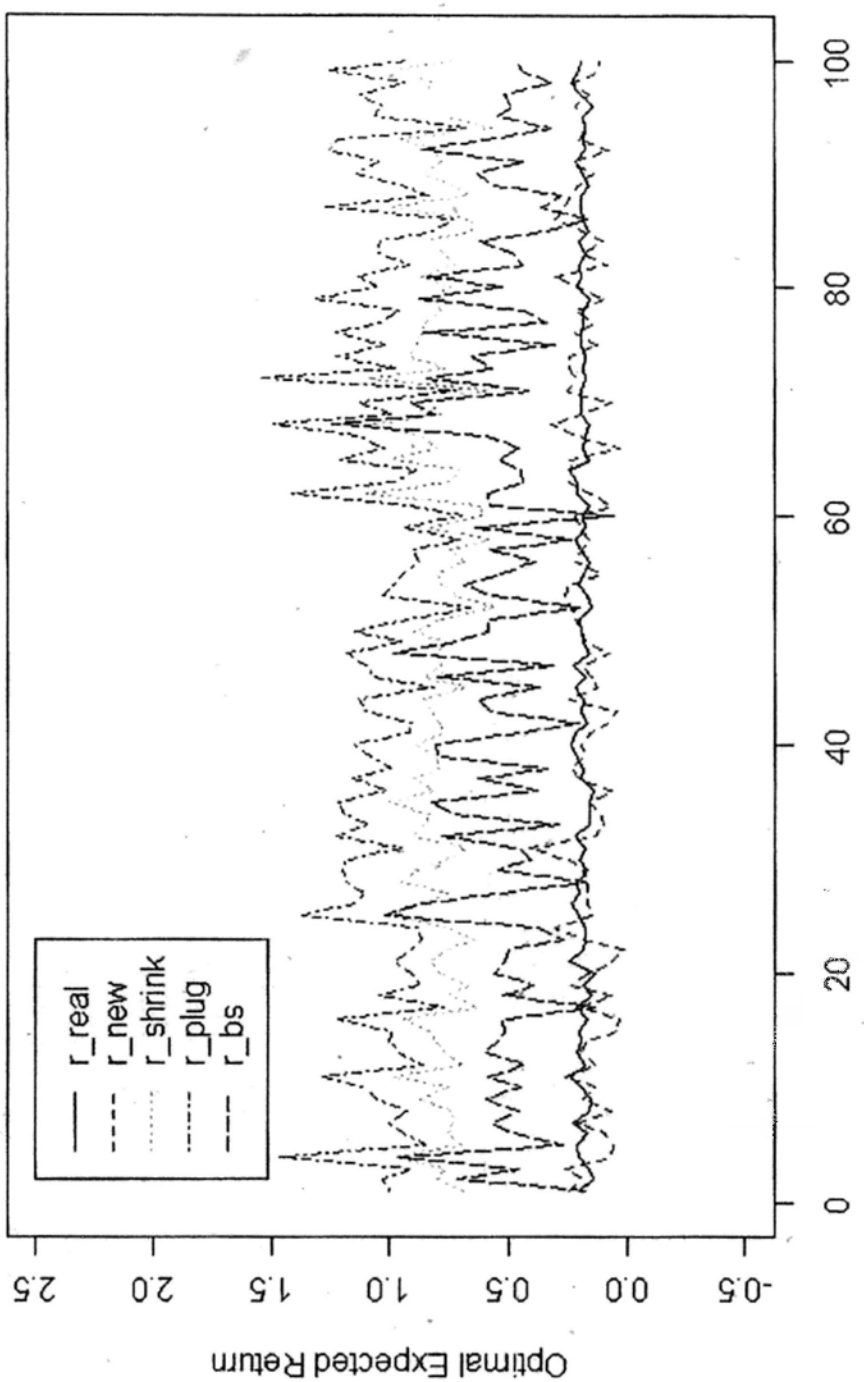

Figure 3.2: Model Comparison $p = 50$, $n = 100$.

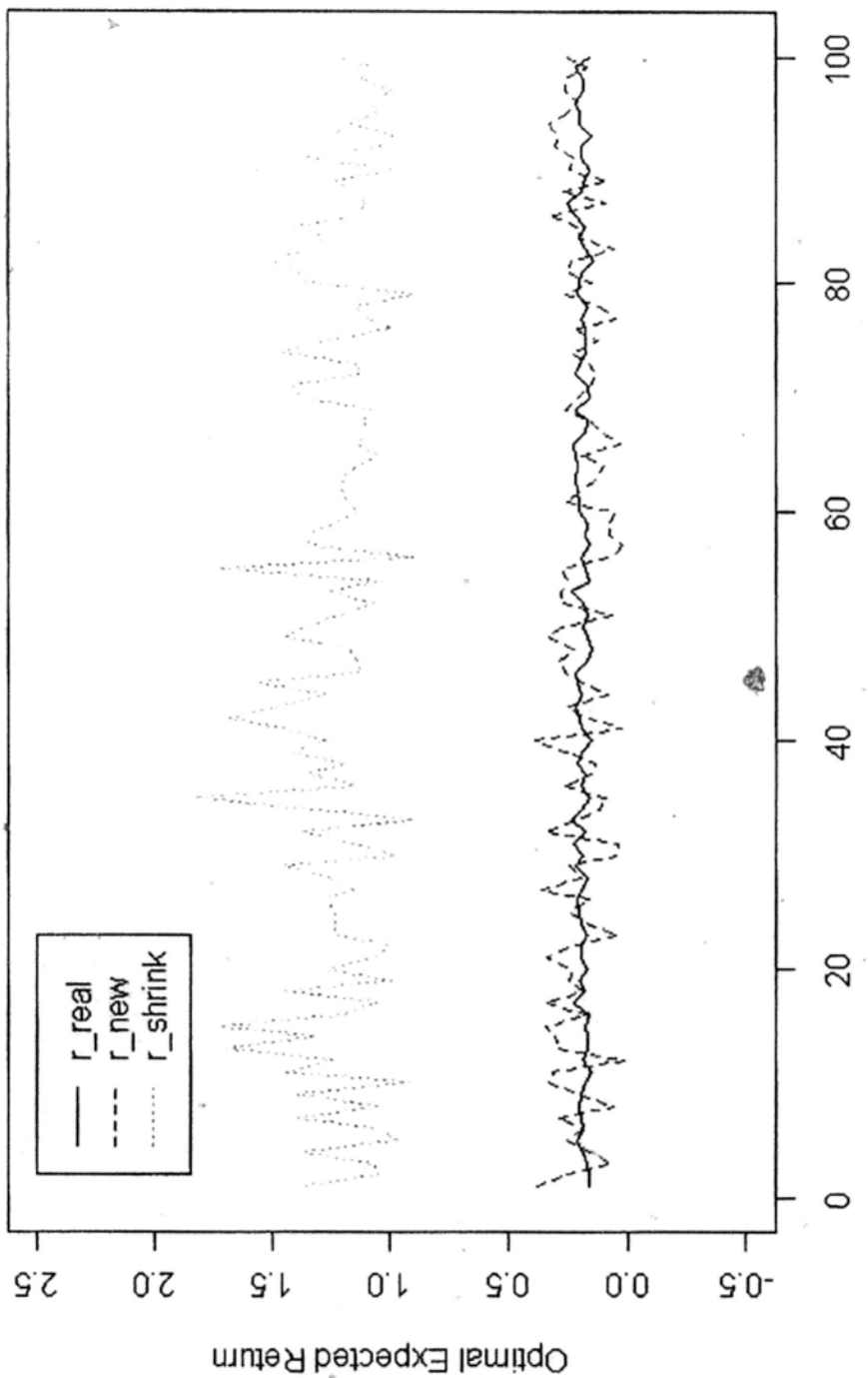

Figure 3.3: Model Comparison $p = 50$, $n = 50$.

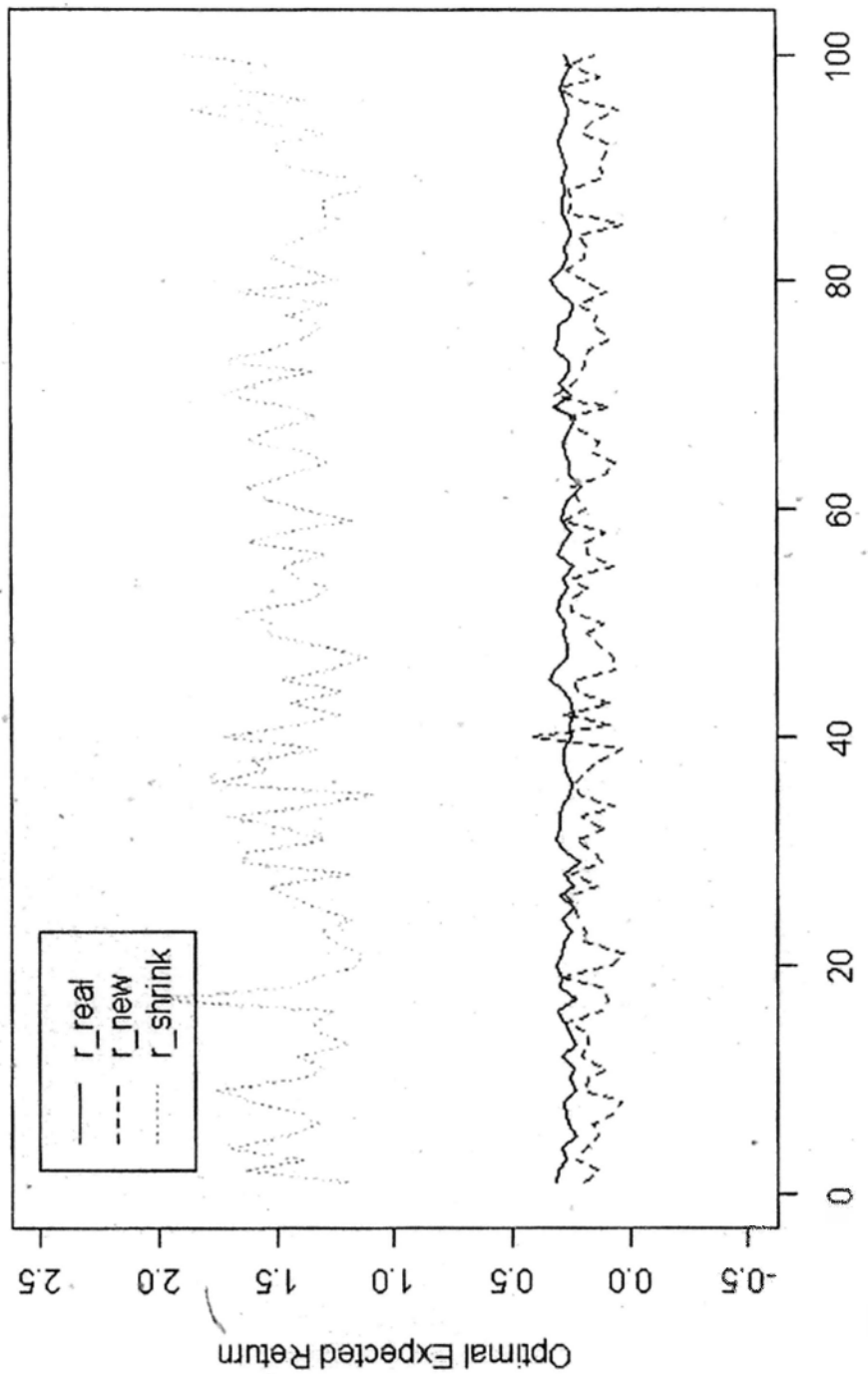

Figure 3.4: Model Comparison $p = 100$, $n = 100$.

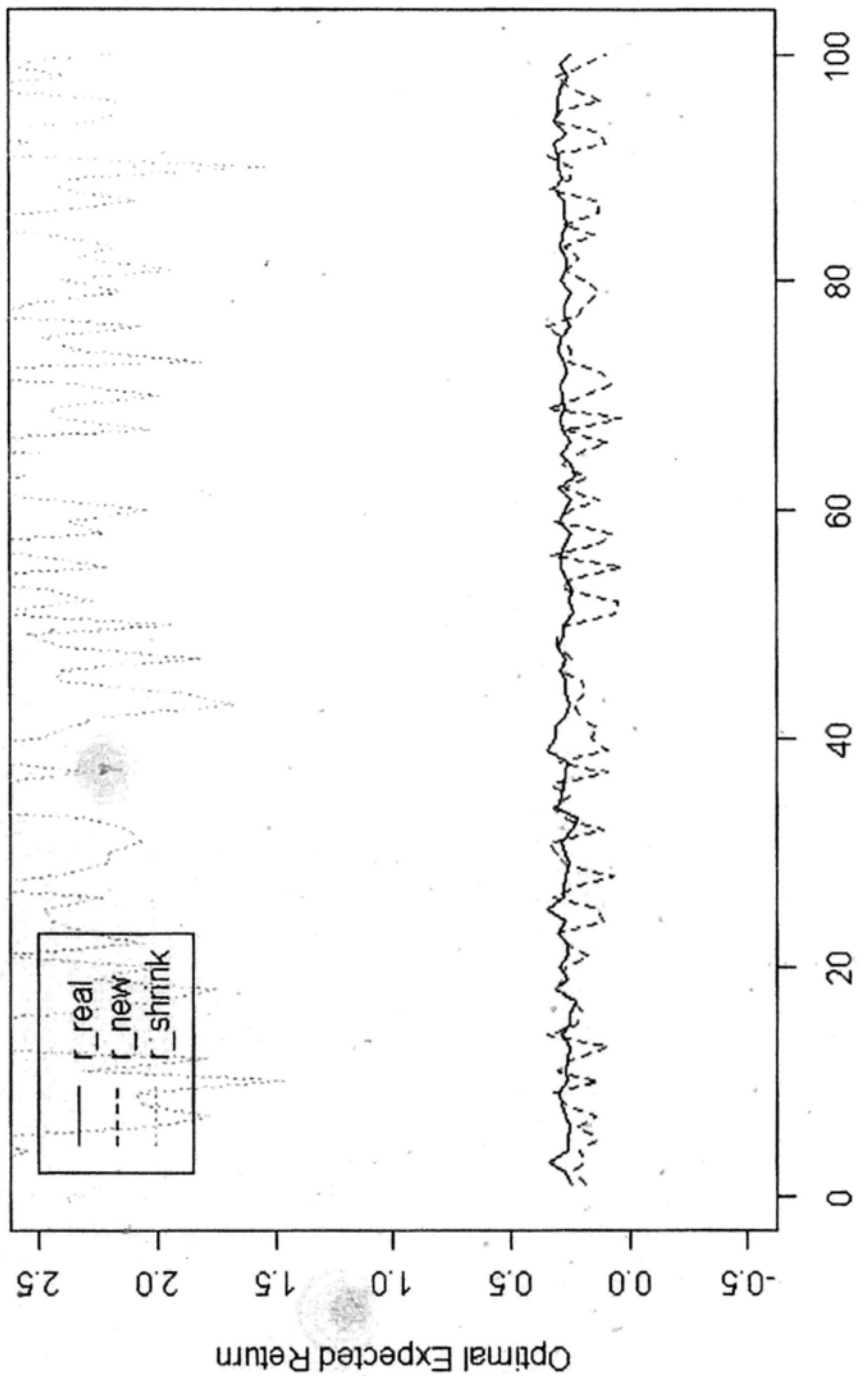

Figure 3.5: Model Comparison $p = 100$, $n = 50$.

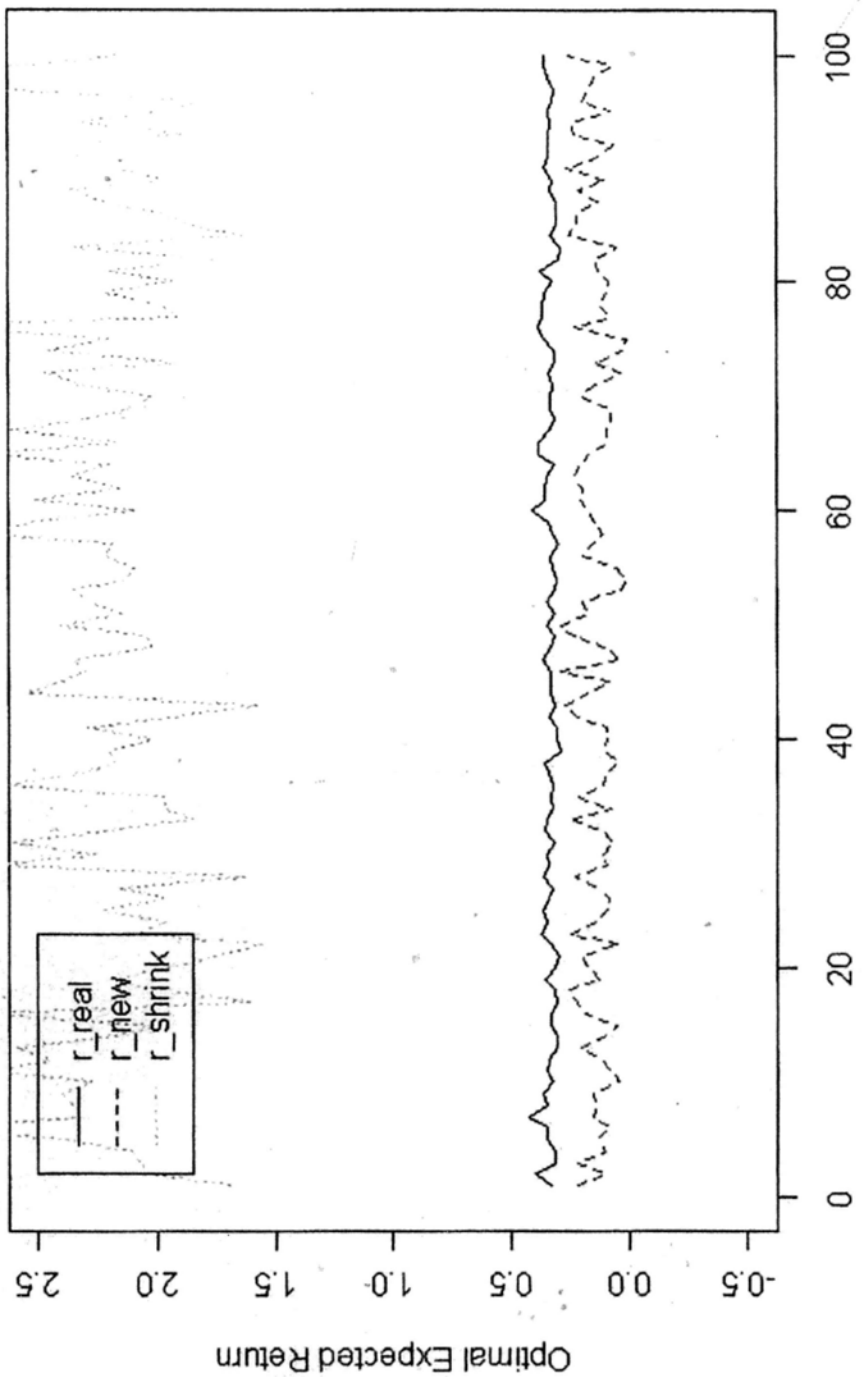

Figure 3.6: Model Comparison $p = 150$, $n = 100$.

3.4 Empirical Study 2

3.4.1 Constructing the Data Set

In this simulation study, we first choose three groups of stocks from the American stock market and download the historical stock prices of these three groups during the period from January 2, 2001 to December 31, 2010. The number of stocks of them are 30, 60 and 80 respectively. Then the sample mean vector and the sample covariance matrix of these three data sets are calculated and regarded as the true parameters from which the empirical returns are generated using multivariate normal distribution.

Since the market changes significantly across time, we fix the number of observations n as 50 days and 100 days. Therefore, we have six combinations of p and n. The six cases are $p = 30,\ n = 50$; $p = 30,\ n = 100$; $p = 60,\ n = 50$; $p = 60,\ n = 100$; $p = 80,\ n = 50$ and $p = 80,\ n = 100$ respectively.

3.4.2 Simulation Results

Again, for each case, we simulate 100 data sets and for each data set i, we calculate the values of r^i_{real}, r^i_{new}, r^i_{plug}, r^i_{shrink}, and r^i_{bs}. We still use d_1, d_2, d_3 and d_4 to measure the average distance between the four estimates and the benchmark value. Recall that d_1 corresponds to r_{new}, d_2 corresponds to r_{plug}, d_3 corresponds to r_{shrink} and d_4 corresponds to r_{bs}. Then we have Table 3.6.

Table 3.6: Average distance of the Four Estimates.

	d_1	d_2	d_3	d_4
$p = 30,\ n = 50$	0.2294	1.1195	0.6946	0.4903
$p = 30,\ n = 100$	0.1564	0.5513	0.4582	0.2711
$p = 60,\ n = 50$	0.2696	—	0.9739	—
$p = 60,\ n = 100$	0.1652	1.0914	0.6682	0.4040
$p = 80,\ n = 50$	0.2817	—	1.1584	—
$p = 80,\ n = 100$	0.1402	1.8457	0.7956	0.5714

From Table 3.6, we observe that r_{new} still has the minimum average distance among all the estimators. When the sample covariance matrix is not singular, r_{bs} is the second best estimator. r_{shrink} becomes worse when p/n grows larger.

Figures 3.7–3.12 demonstrate our conclusion. Moreover, r_{new} is much more stable than the other estimators.

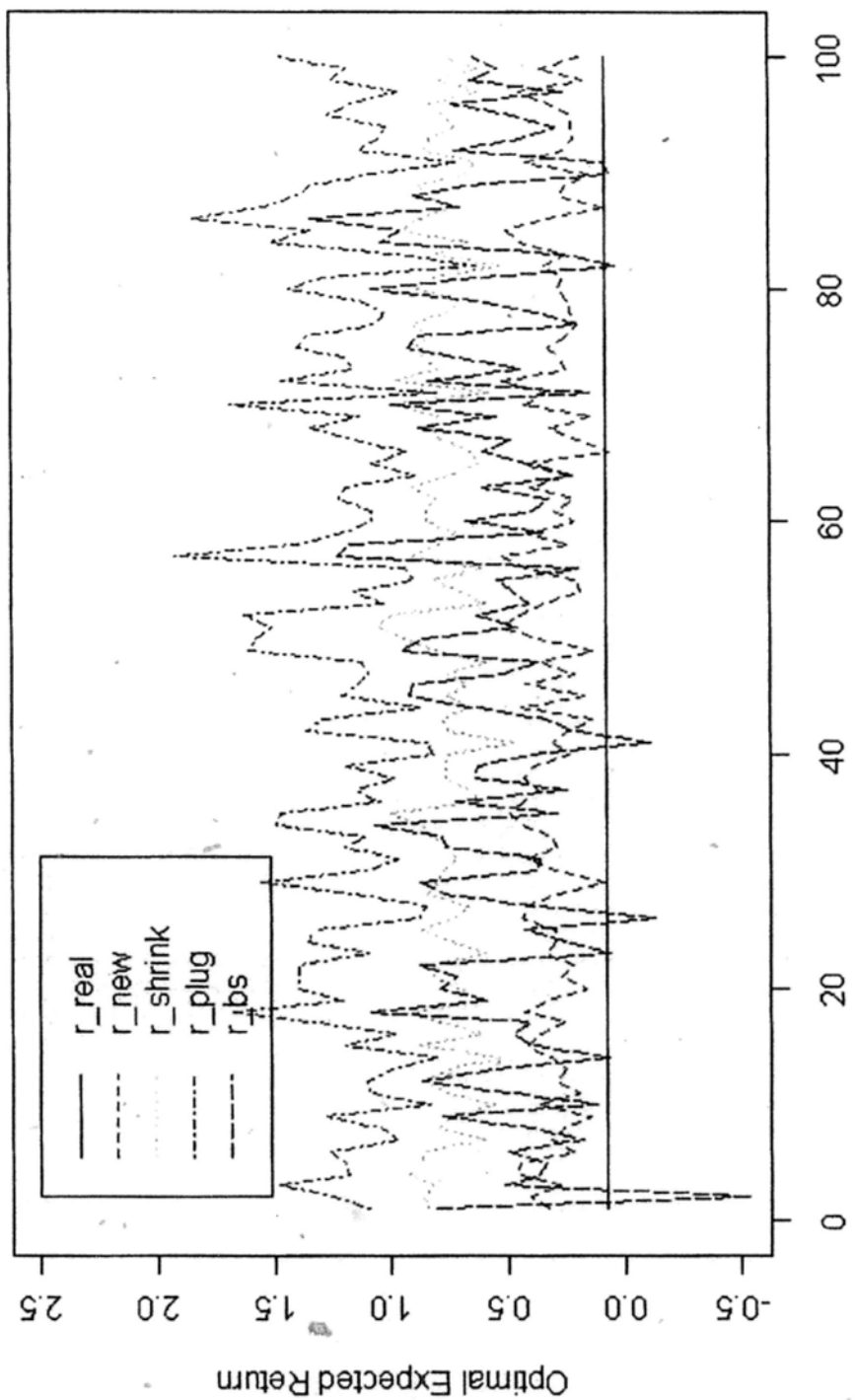

Figure 3.7: Model Comparison $p = 30$, $n = 50$.

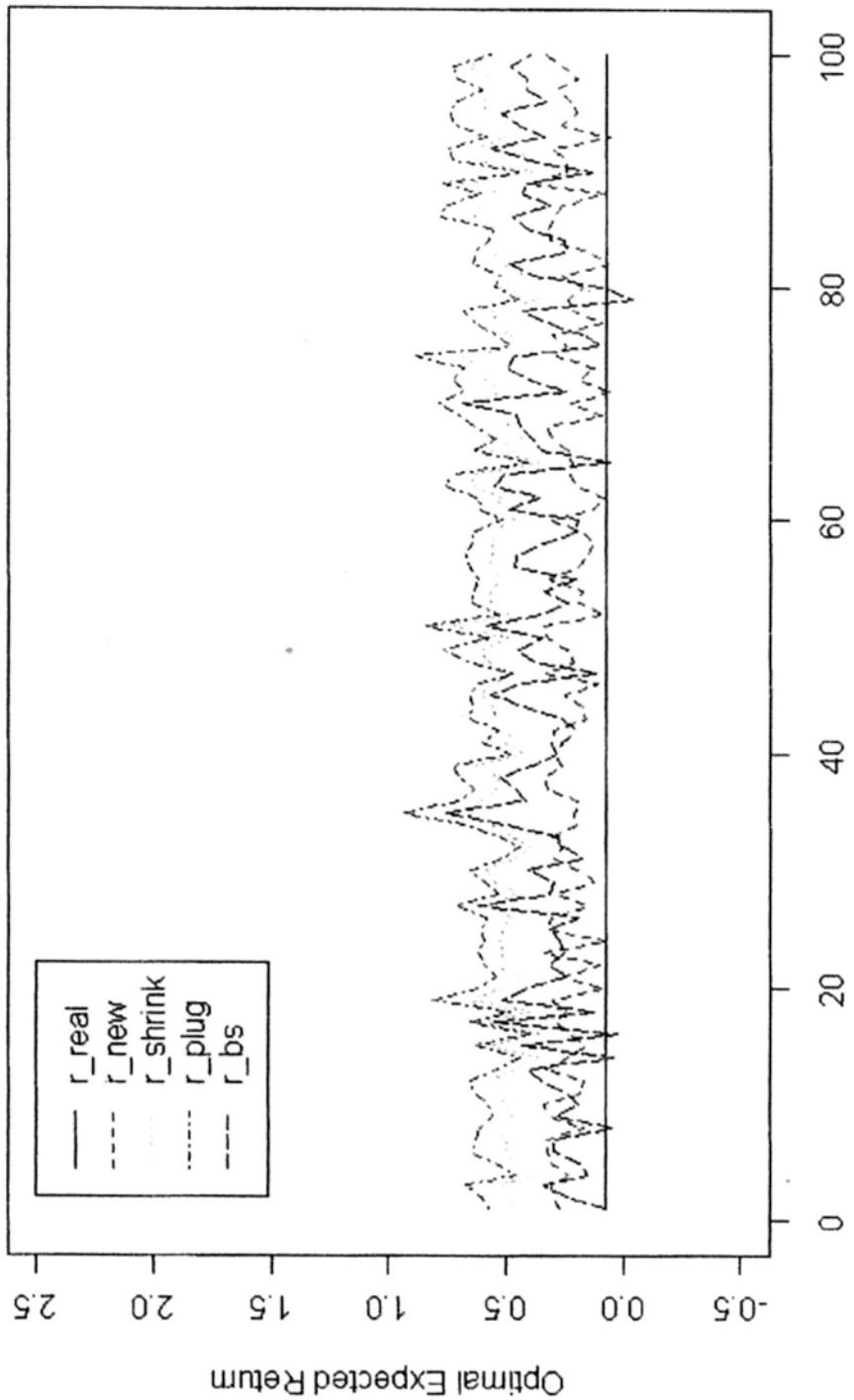

Figure 3.8: Model Comparison $p = 30$, $n = 100$.

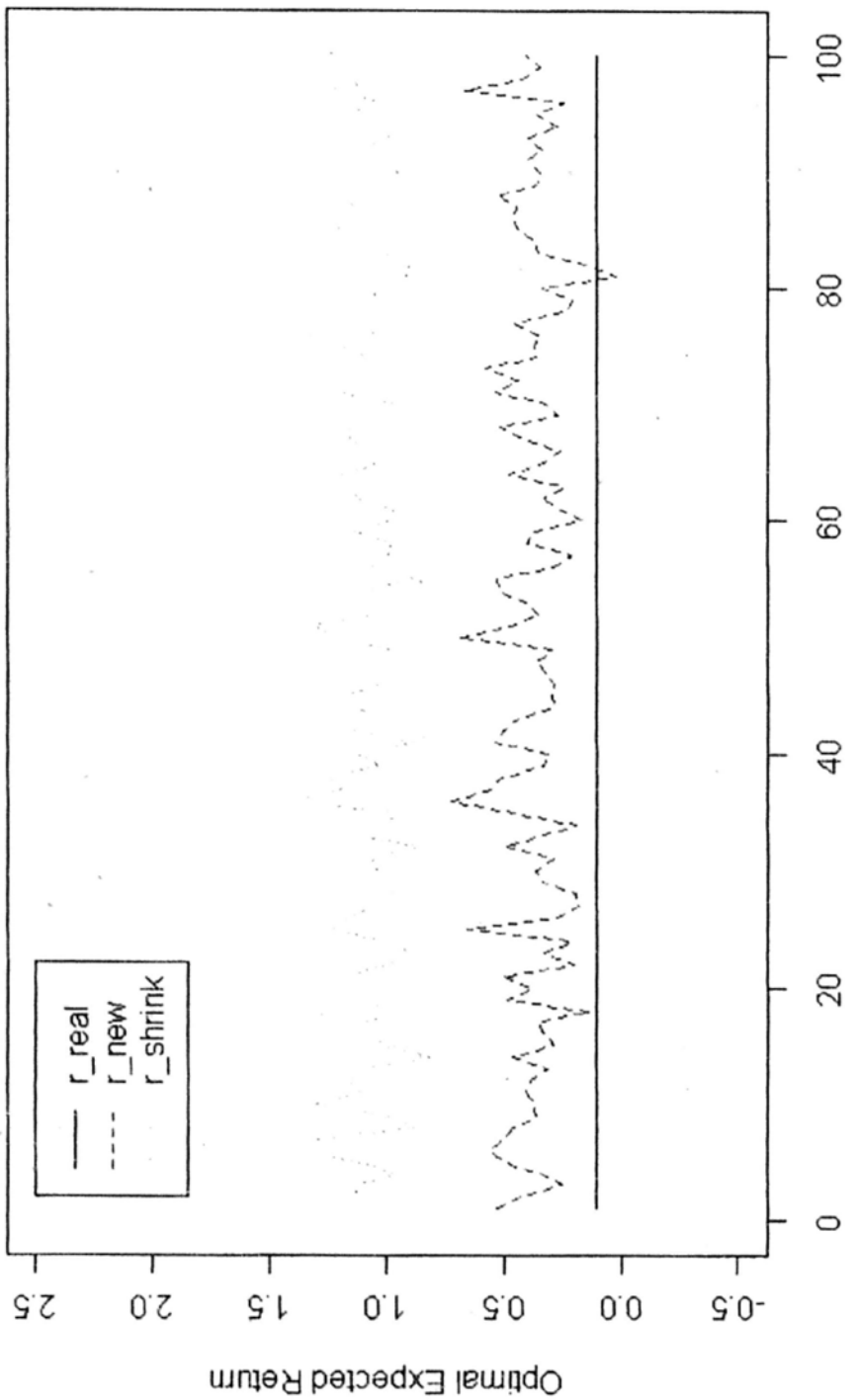

Figure 3.9: Model Comparison $p = 60$, $n = 50$.

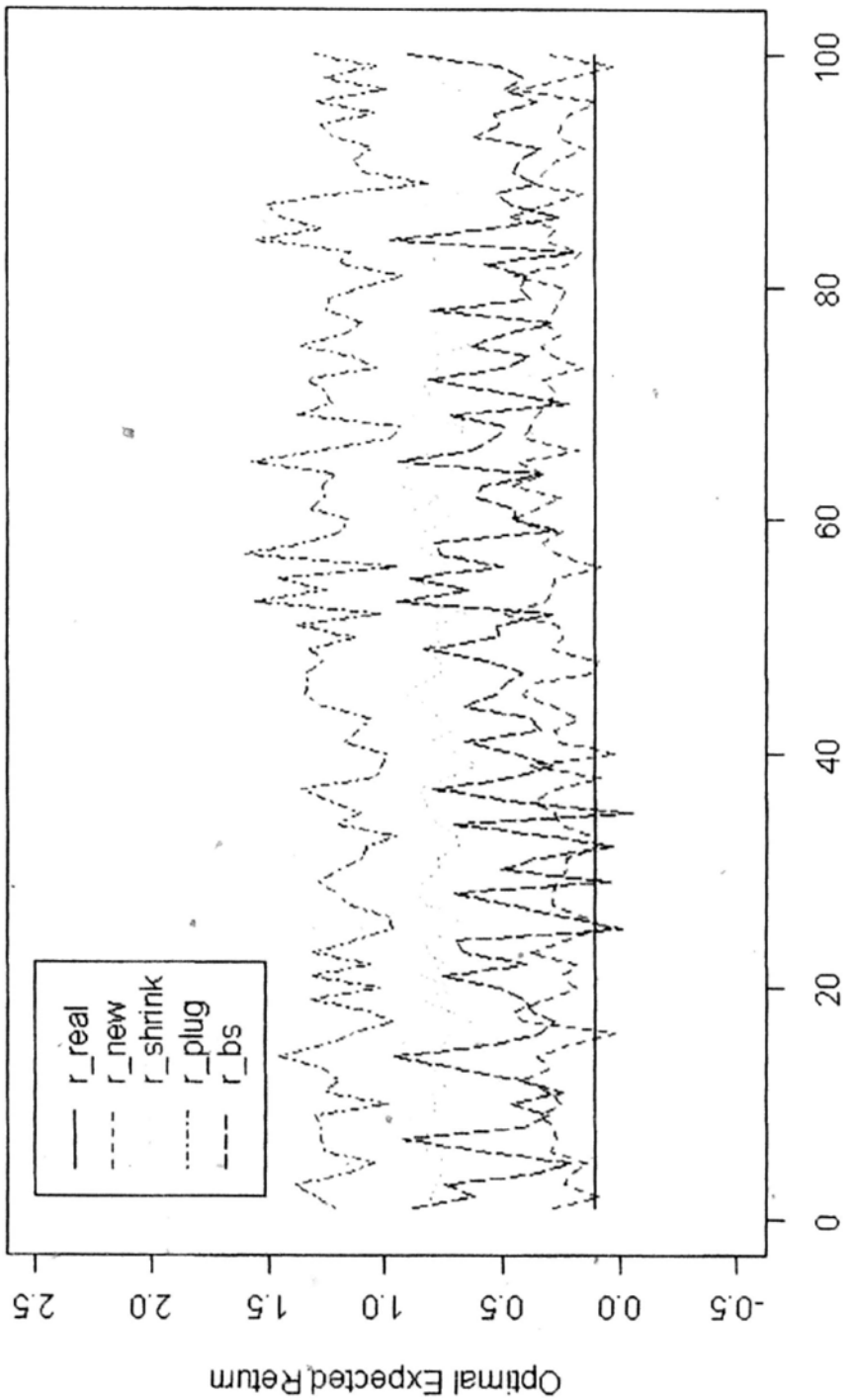

Figure 3.10: Model Comparison $p = 60$, $n = 100$.

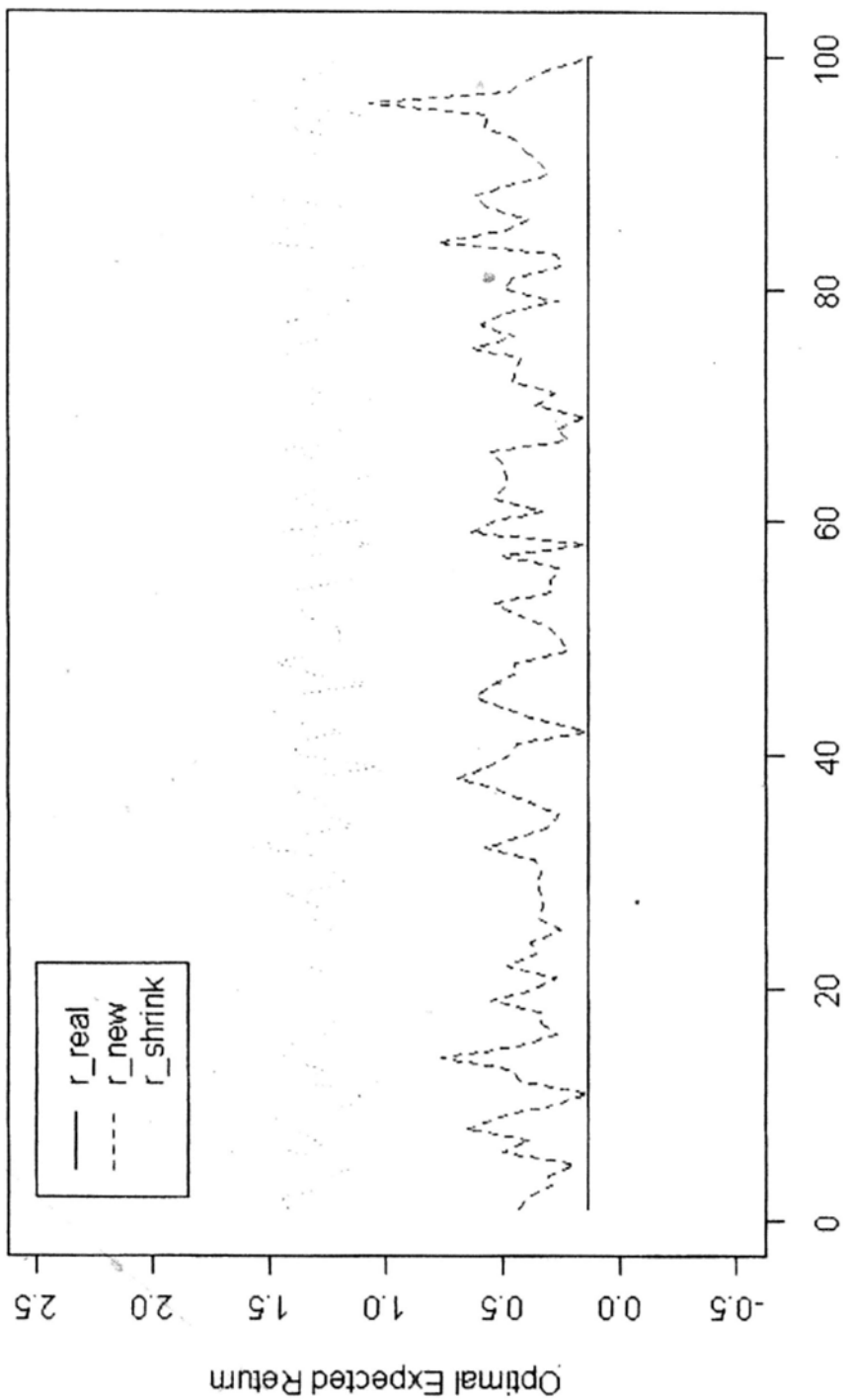

Figure 3.11: Model Comparison $p = 80$, $n = 50$.

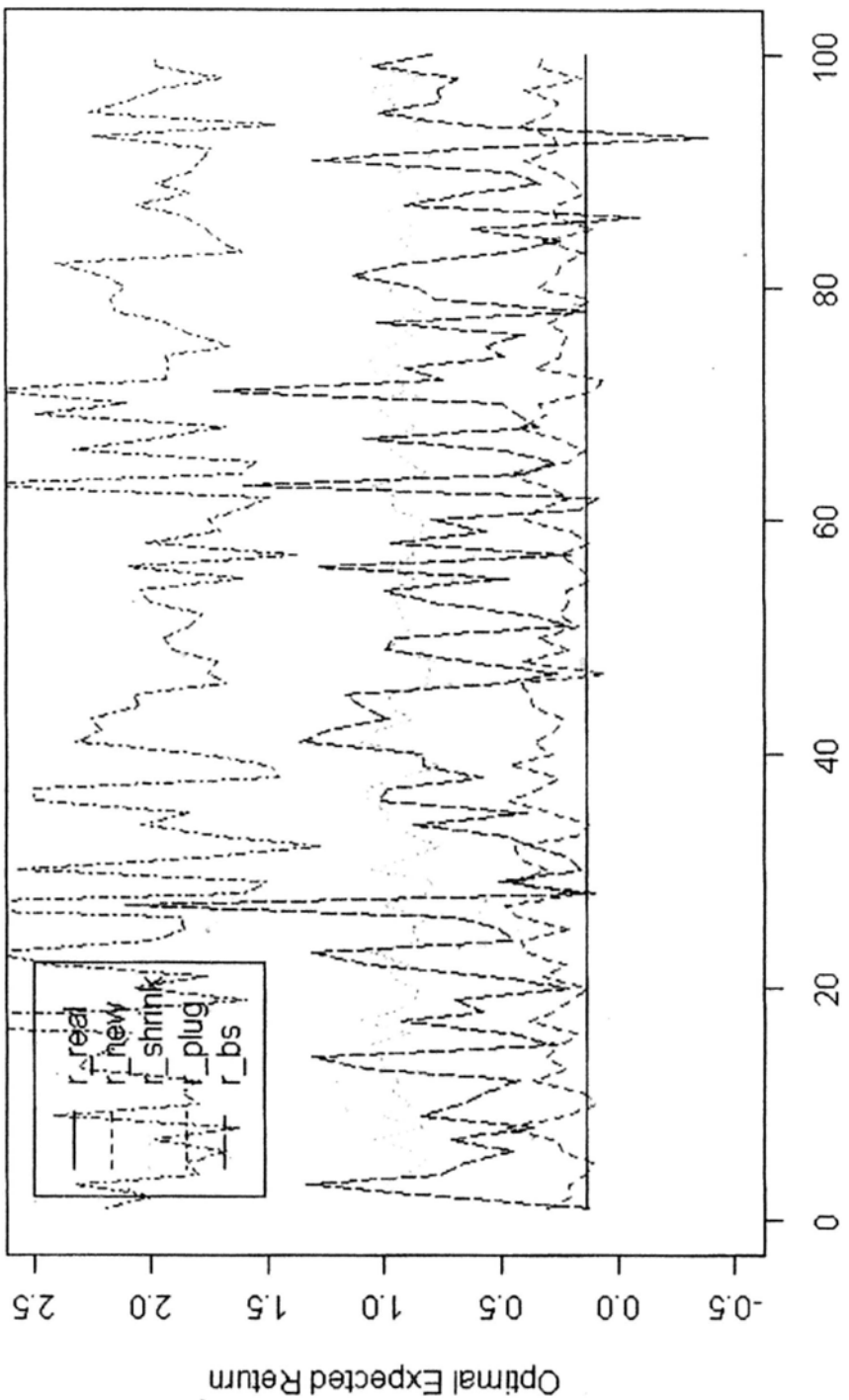

Figure 3.12: Model Comparison $p = 80$, $n = 100$.

□ **End of chapter.**

Chapter 4

Discussion and Further Research

4.1 Introduction

In this chapter, we will first give a discussion of the work of Bai, Liu and Wong (2009a, 2009b) [2, 3]. The impact of the sample mean vector is taken into consideration and then an experiment is conducted to show the importance of it. Since the estimator of Bai et. al (2009a, 2009b) can only be applied when the sample covariance matrix is invertible, we introduce the shrinkage covariance matrix into the framework of their work. However, such an extension is only fulfilled under a strict assumption.

4.2 Discussion

First consider a lemma of Bai, Liu and Wong (2009a) [2].

Lemma 4.2.1. (*see Lemma 3.1 of Bai, Liu and Wong, 2009*). Assume that $\boldsymbol{y_1}, \ldots, \boldsymbol{y_n}$ are n independent random p-vectors of *iid* entries with mean 0 and variance 1. $\boldsymbol{x_k} = \boldsymbol{\mu} + \boldsymbol{z_k}$ with $\boldsymbol{z_k} = \Sigma^{1/2}\boldsymbol{y_k}$, 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 $\boldsymbol{y_k}$'s have finite fourth moments and $p/n \to y \in (0,1)$. Suppose that

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

Then we have

(a)

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

(b)

$$\frac{\mathbf{1}^{\mathrm{T}}S^{-1}\mathbf{1}}{n} \xrightarrow{a.s.} a_2\gamma,$$

(c)

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

The proof of Lemma 4.2.1 is derived from Lemma 2.2.1, Lemma 2.2.2 and Lemma 2.2.3. The details are referred to Bai, Liu and Wong (2009a) [2].

Therefore, we conclude that under the assumptions of Lemma 4.2.1, from Lemma 2.2.1, Lemma 2.2.2 and Lemma 2.2.3, we have the result that for a known mean vector $\boldsymbol{\mu}$,

$$\boldsymbol{\mu}^{\mathrm{T}}h(S^{-1})\boldsymbol{\mu} \xrightarrow{a.s.} \gamma\boldsymbol{\mu}^{\mathrm{T}}h(\Sigma^{-1})\boldsymbol{\mu}. \tag{4.1}$$

From Lemma 4.2.1, we have the result that

$$\hat{\boldsymbol{\mu}}^{\mathrm{T}}h(S^{-1})\hat{\boldsymbol{\mu}} \xrightarrow{a.s.} \gamma\boldsymbol{\mu}^{\mathrm{T}}h(\Sigma^{-1})\boldsymbol{\mu}. \tag{4.2}$$

The above two results are both correct in a mathematical context, and the difference lies in that whether we plug in the real mean vector $\boldsymbol{\mu}$ or we plug in the sample mean vector $\hat{\boldsymbol{\mu}}$. Can the impact of the sample mean vector be ignored using limit theorems? Now, we use an experiment to clarify this problem in the framework of the portfolio theory.

4.2.1 An Experiment

In this experiment, we consider the difference between $\mathrm{E}(r_1^2)$, $\mathrm{E}(r_2^2)$ and $\mathrm{E}(r_3^2)$, where r_1^2, r_2^2 and r_3^2 are defined as:

$$r_1^2 = \boldsymbol{\mu}^{\mathrm{T}} h(\Sigma^{-1})\boldsymbol{\mu},$$

$$r_2^2 = \frac{\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{-1})\hat{\boldsymbol{\mu}}}{\gamma},$$

$$r_3^2 = \frac{\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{-1})\hat{\boldsymbol{\mu}}}{\gamma} - \frac{p-1}{n}.$$

Herein, r_1 denotes the benchmark value, r_2 denotes the optimal expected return using Lemma 4.2.1 and r_3 not only considers the impact of the sample covariance matrix, but also considers the impact of the sample mean vector.

We design the experiment as follows. Consider two settings. All the data sets are generated from a multivariate normal distribution.

(a) Following Bai, Liu and Wong (2009b) [3], specify the true covariance matrix as identity matrix I and suppose that the true mean vector comes from a standard normal distribution, then fixed as known. Consider two cases: $p = 30,\ n = 100$ and $p = 60,\ n = 100$.

(b) Employ the true covariance matrix and the true mean vector of empirical study 2 in Chapter 3. We use the previous two cases, $p = 30$ and $p = 60$. Then simulate 100 observations respectively.

Herein, we do simulation ten times to provide a brief look at the differences. $\mathrm{E}(r_1^2)$, $\mathrm{E}(r_2^2)$ and $\mathrm{E}(r_3^2)$ are approximated by taking all the means of the ten results. See the following two tables, Table 4.1 and Table 4.2.

- (a)

Table 4.1: Comparison of $E(r_1^2)$, $E(r_2^2)$ and $E(r_3^2)$.

	$E(r_1^2)$	$E(r_2^2)$	$E(r_3^2)$
$p = 30,\ n = 100$	21.58	23.51	23.22
$p = 60,\ n = 100$	61.62	63.19	62.60

- (b)

Table 4.2: Comparison of $E(r_1^2)$, $E(r_2^2)$ and $E(r_3^2)$.

	$E(r_1^2)$	$E(r_2^2)$	$E(r_3^2)$
$p = 30,\ n = 100$	0.0058	0.3734	0.0834
$p = 60,\ n = 100$	0.0126	0.7259	0.1359

From Table 4.1 and Table 4.2, we find that in case (a), the impact of the sample mean vector is not obvious. $E(r_1^2)$, $E(r_2^2)$ and $E(r_3^2)$ are very close to each other. However, in case (b), where the true covariance matrix and the true mean vector come from real stock returns, the impact of the sample mean vector becomes important. $E(r_3^2)$ deviates farther from $E(r_2^2)$.

The reason for this is the impact of the sample mean is an integration process on $E(r_1^2)$. The error incurred by the sample mean vector is around $(p-1)/n$. When $p < n$, as p approaches to n, the value of $(p-1)/n$ gets close to 1. The importance of this value depends on the order of magnitude of the optimal expected return. The larger the optimal expected return, the less important is $(p-1)/n$. However, in the real stock market, such an optimal

expected return is very small. Therefore, we cannot ignore this impact just from the perspective of mathematics.

4.3 Further Research

Since the new estimator based on random matrix theory cannot be applied when the sample covariance matrix is singular, in this part, we provide an extension of Bai, Liu and Wong (2009b) [3], using the shrinkage method. The impact of the sample mean vector is taken into consideration.

4.3.1 Data Assumptions

Suppose that each element in the data set

$$x_{tj} \sim N(\mu_j, \sigma_j^2)\ i.i.d,$$

herein $j = 1, 2, \ldots, p$, and $t = 1, 2, \ldots, n$. p, $n \to \infty$, $p/n \to y \in (0, \infty)$; μ_j and σ_j^2 are unknown parameters.

For each asset $\boldsymbol{x}_j$, assume that the sample size n is big enough so that the law of large numbers applies. Then $\hat{\mu}_j$ converges in probability to μ_j and $\hat{\sigma}_j^2$ converges in probability to σ_j^2, where

$$\hat{\mu}_j = \frac{1}{n}\sum_{t=1}^{n} x_{tj},$$

and

$$\hat{\sigma}_j^2 = \frac{1}{n-1}\sum_{t=1}^{n}(x_{tj} - \hat{\mu}_j)^2.$$

Thus we have

$$\frac{x_{tj} - \hat{\mu}_j}{\hat{\sigma}_j} \sim N(0,1),\ \text{as } n \to \infty.$$

Therefore, standardization can be used to transform the data to satisfy the assumption of the standard Marčenko-Pastur (MP) law (1967) [31]. Then the

sample covariance matrix of the standardized data becomes the correlation matrix of the original data. To make the value of $\hat{P}^{*2}$ unchanged, the vector of the sample mean becomes

$$\hat{\boldsymbol{\mu}} = (\hat{\mu}_1/\hat{\sigma}_1, \ldots, \hat{\mu}_p/\hat{\sigma}_p)^{\mathrm{T}}.$$

4.3.2 Error Correction Ratio

In this part, the shrinkage target is specified by the identity matrix I and the shrinkage covariance matrix is:

$$S^* = \alpha I + (1-\alpha)S.$$

Since the data is standardized, we only need to consider the standard M-P law. Now, the density function of the limit spectral distribution of the shrinkage covariance matrix can be deduced from $f_y(x)$. Since

$$\begin{aligned} S^* &= \alpha^* I + (1-\alpha^*)S \\ &= Q[\alpha^* I + (1-\alpha^*)\Lambda]Q', \end{aligned}$$

where $QQ' = I$, the empirical spectral distribution of S^* is

$$\begin{aligned} G_{S^*}(x) &= \frac{1}{p}\sum_{i=1}^{p} \mathbf{1}(\alpha^* + (1-\alpha^*)\lambda_i \le x) \\ &= \frac{1}{p}\sum_{i=1}^{p} \mathbf{1}(\lambda_i \le \frac{x-\alpha^*}{1-\alpha^*}) \\ &= F_S(\frac{x-\alpha^*}{1-\alpha^*}). \end{aligned}$$

Therefore the density function $g_y(x)$ of the limit spectral distribution of S^* is

$$\begin{aligned} g_y(x) &= \frac{1}{1-\alpha^*} f_y(\frac{x-\alpha^*}{1-\alpha^*}) \\ &= \begin{cases} \frac{1}{2\pi(x-\alpha^*)y}\sqrt{(b-\frac{x-\alpha^*}{1-\alpha^*})(\frac{x-\alpha^*}{1-\alpha^*}-a)}, & a \le \frac{x-\alpha^*}{1-\alpha^*} \le b, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

If $y > 1$, $g_y(x)$ has a point mass $1 - 1/y$ at $x = \alpha^*$. Based on Lemma 2.2.1, Lemma 2.2.2 and Lemma 2.2.3, we have the following two propositions about the error correction ratio γ for the shrinkage covariance matrix.

Proposition 4.3.1. For $p < n$ and for a nonrandom p–vector $\boldsymbol{\xi}$,

$$\boldsymbol{\xi}^{\mathrm{T}} h(S^{*-1})\boldsymbol{\xi} \to \gamma_1 \boldsymbol{\xi}^{\mathrm{T}} h(\Sigma^{-1})\boldsymbol{\xi} \quad a.s.,$$

where

$$\gamma_1 = \int_{(1-\alpha^*)a+\alpha^*}^{(1-\alpha^*)b+\alpha^*} \frac{1}{x} g_y(x)\, dx.$$

Proposition 4.3.2. For $p > n$ and for a fixed $\boldsymbol{\xi}$,

$$\boldsymbol{\xi}^{\mathrm{T}} h(S^{*-1})\boldsymbol{\xi} \to \gamma_2 \boldsymbol{\xi}^{\mathrm{T}} h(\Sigma^{-1})\boldsymbol{\xi} \quad a.s.,$$

where

$$\gamma_2 = \frac{1}{\alpha^*}(1 - \frac{1}{y}) + \int_{(1-\alpha^*)a+\alpha^*}^{(1-\alpha^*)b+\alpha^*} \frac{1}{x} g_y(x)\, dx.$$

For notational convenience, define γ as

$$\gamma = \begin{cases} \gamma_1, & p < n, \\ \gamma_2, & p > n. \end{cases}$$

The quantity γ is calculated by numerical methods. Proposition 4.3.1 and Proposition 4.3.2 are demonstrated by simulation results.

Suppose that the true covariance matrix Σ is I and the true mean vector μ is simulated from $N(0, 1)$. The data set is simulated from a p–dimensional multivariate normal distribution using Σ and μ.

To verify Proposition 4.3.1 and Proposition 4.3.2, consider the value of $\mathrm{E}(\boldsymbol{\mu}^{\mathrm{T}} h(S^{*-1})\boldsymbol{\mu})/\gamma$. It is approximated by simulating $\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu}/\gamma$ ten times

and taking the sample average of these values as an estimate. The simulation results are as follows.

Table 4.3: $p = 25,\ n = 50$.

	$\mathrm{E}(\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu})/\gamma$	P^{*2}
$\alpha = 0.2$	25.76	24.44
$\alpha = 0.5$	21.44	21.80
$\alpha = 0.8$	25.58	25.41

Table 4.4: $p = 50,\ n = 100$.

	$\mathrm{E}(\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu})/\gamma$	P^{*2}
$\alpha = 0.2$	48.22	49.50
$\alpha = 0.5$	43.15	44.28
$\alpha = 0.8$	50.39	49.86

Table 4.5: $p = 50,\ n = 25.$

	$\mathrm{E}(\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu})/\gamma$	P^{*2}
$\alpha = 0.2$	49.25	47.25
$\alpha = 0.5$	50.02	49.15
$\alpha = 0.8$	54.22	53.26

Table 4.6: $p = 100,\ n = 50.$

	$\mathrm{E}(\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu})/\gamma$	P^{*2}
$\alpha = 0.2$	95.35	92.20
$\alpha = 0.5$	97.97	98.82
$\alpha = 0.8$	100.73	99.58

4.3.3 New Estimator

Theorem 4.3.1. As $p,\ n \to \infty,\ p/n \to y \in (0, \infty)$,

$$\mathrm{E}(\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}) \xrightarrow{a.s.} \gamma[\frac{p-1}{n} + P^{*2}]$$

Proof. Using double expectation,

$$\begin{aligned} \mathrm{E}(\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}) &= \mathrm{E}[\mathrm{E}(\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^{*-1})\hat{\boldsymbol{\mu}}|\hat{\boldsymbol{\mu}})] \\ &\xrightarrow{a.s.} \gamma \mathrm{E}[\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(\Sigma^{-1})\hat{\boldsymbol{\mu}}]\ as\ \frac{p}{n} \to y \in (0, \infty) \\ &= \gamma[\frac{p-1}{n} + P^{*2}]. \end{aligned}$$

The proof completes. Therefore, the new estimator of the optimal expected return can be specified as

$$\hat{P}^{*2}_{new} = \frac{\hat{\boldsymbol{\mu}}^{\mathrm{T}} h(S^*)\hat{\boldsymbol{\mu}}}{\gamma} - \frac{p-1}{n}.$$

To assess the advantage of the new estimator, some simulation studies are conducted. To be consistent, assume that the real mean $\boldsymbol{\mu}$ comes from a standard normal distribution and the real covariance matrix is I.

The simulation results are summarized in the following four tables. We see that the results from the new estimator are very close to the real values.

Table 4.7: $p = 25,\ n = 50$.

	$\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu}/\gamma$	P^{*2}
$\alpha = 0.2$	15.21	13.29
$\alpha = 0.5$	17.10	16.61
$\alpha = 0.8$	16.87	15.80

Table 4.8: $p = 50,\ n = 100$.

	$\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu}/\gamma$	P^{*2}
$\alpha = 0.2$	51.96	49.74
$\alpha = 0.5$	39.54	42.45
$\alpha = 0.8$	29.94	31.42

Table 4.9: $p = 50$, $n = 25$.

	$\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu}/\gamma$	P^{*2}
$\alpha = 0.2$	50.83	49.05
$\alpha = 0.5$	37.05	44.47
$\alpha = 0.8$	40.07	39.19

Table 4.10: $p = 100$, $n = 50$.

	$\boldsymbol{\mu}^{\mathrm{T}} h(S^*)\boldsymbol{\mu}/\gamma$	P^{*2}
$\alpha = 0.2$	112.94	109.08
$\alpha = 0.5$	98.27	90.22
$\alpha = 0.8$	93.94	88.49

4.3.4 Limitations and Further Work

Although the simulation results of this method are close to the benchmark value, the assumptions are not realistic since the stock returns cannot be *i.i.d.* random variables. Therefore, the further work is to relax the data assumptions to the general case, such as with a covariance structure. Then compute the error correction ratio. Taylor expansions could be used to approximate the limit spectral distribution of the shrinkage covariance matrix.

□ **End of chapter.**

Chapter 5

Conclusion

In this study, we proposed a new estimator for evaluating the optimal expected return for a large dimensional self-financing portfolio.

In the mean-variance portfolio optimization procedure, it is well known that the plug-in optimal expected return is not a good estimator since using the sample mean and the sample covariance matrix of the historical data incurs substantial errors. Rather than constructing new estimators of the mean and the covariance matrix that have less estimation errors, this thesis incorporates the interaction effect of these two quantities and explores how the sample mean and the sample covariance matrix behave based on the idea of conditional expectation.

It is found that the effect of the sample mean is an additive process while the effect of the sample covariance matrix is a multiplicative process. Both of them over-predict the optimal expected return. When the covariance matrix is known, the overestimating value incurred by the sample mean is $(p-1)/n$. It cannot be ignored since the smaller is the return, the more it contributes to the total errors.

In the financial market, the number of stocks can be very large while the sample size should not be too large since the stock market changes significantly across time. Therefore, p/n is substantial and the sample covariance matrix tends to be singular. This thesis used the shrinkage methods to construct a stable covariance matrix which was invertible for both $p < n$ and $p \geq n$. Partial ordering was employed to prove that the shrinkage covariance matrix led to an estimate of the optimal expected return smaller than the plug-in estimate and was closer to the benchmark value.

The new estimator is an improvement of the previous methods. It is demonstrated not only by theory, but also by simulation results.

We also extended our research in the framework of random matrix theory. However, the data assumptions are restricted. The density function of the limit spectral distribution of the shrinkage covariance matrix was deduced, which was used to calculate the error correction ratio incurred by the shrinkage covariance matrix. The impact of the sample mean vector was considered. Further research needs to be pursued under more general assumptions.

□ **End of chapter.**

Bibliography

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis.* Wiley, New York, 3rd edition, 2003.

[2] Z. Bai, H. Liu, and W. K. Wong. Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Mathematical Finance*, 19(4):639–667, 2009a.

[3] Z. Bai, H. Liu, and W. K. Wong. On the Markowitz mean-variance analysis of self-financing portfolios. *Risk and Decision Analysis*, 1:35–42, 2009b.

[4] Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices.* Springer, New York, 2nd edition, 2010.

[5] Z. D. Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, 16(1):611–677, 1999.

[6] Z. D. Bai, B. Q. Miao, and G. M. Pan. On asymptotics of eigenvectors of large sample covariance matrix. *Annals of Probability*, 35(4):1532–1572, 2007.

[7] C. B. Barry. Portfolio analysis under uncertain means, variances and covariances. *Journal of Finance*, 29(2):515–522, 1974.

[8] Z. Bodie, A. Kane, and A. J. Marcus. *Essentials of Investment.* McGraw-Hill Irwin, Boston, 8th edition, 2010.

[9] S. J. Brown. *Optimal Portfolio Choice under Uncertainty: A Bayesian Approach.* PhD thesis, University of Chicago, 1976.

[10] N. H. Chan. *Time Series: Applications to Finance with R and S-Plus(R).* Wiley, New York, 2nd edition, 2010.

[11] K. J. Cohen and J. A. Pogue. An empirical evaluation of alternative portfolio-selection models. *Journal of Business*, 40(2):166–193, 1967.

[12] B. Efron and C. Morris. Limiting the risk of Bayes and empirical Bayes estimators, Part I: the Bayes case. *Journal of American Statistical Association*, 66:807–815, 1971.

[13] B. Efron and C. Morris. Limiting the risk of Bayes and empirical Bayes estimators, Part II: the empirical Bayes case. *Journal of American Statistical Association*, 67:130–139, 1972a.

[14] B. Efron and C. Morris. Empirical Bayes on vector observations: an extension of Stein's method. *Biometrica*, 59:335–347, 1972b.

[15] B. Efron and C. Morris. Stein's estimation rule and its competitors–an empirical Bayes approach. *Journal of American Statistical Association*, 68:117–130, 1973.

[16] B. Efron and C. Morris. Data analysis using stein's estimator and its generalizations. *Journal of American Statistical Association*, 70:311–319, 1975.

[17] B. Efron and C. Morris. Families of minimax estimators of the mean of a multivariate normal distribution. *Annals of Statistics*, 4:11–21, 1976.

[18] B. Efron and C. Morris. Multivariate empirical Bayes and estimation of covariance matrices. *Annals of Statistics*, 4(1):22–32, 1976.

[19] B. F. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap.* Chapman and Hall, London, 1st edition, 1993.

[20] E. J. Elton and M. J. Gruber. Estimating the dependence structure of share prices–implications for portfolio selection. *Journal of Finance*, 28(5):1203–1232, 1973.

[21] E. J. Elton and M. J. Gruber. *Portfolio theory, 25 years after : essays in honor of Harry Markowitz.* North-Holland, Amsterdam, 1979.

[22] E. J. Elton and M. J. Gruber. Modern portfolio theory, 1950 to date. *Journal of Banking & Finance*, 21:1743–1759, 1997.

[23] G. M. Frankfurter, H. E. Phillips, and J. P. Seagle. The effects of uncertain means, variances, and covariances. *Journal of Financial and Quantitative Analysis*, 6(5):1251–1262, 1971.

[24] P. A. Frost and J. E. Savarino. An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305, 1986.

[25] P. Jorion. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3):279–292, 1986.

[26] R. W. Klein and V. S. Bawa. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3:215–231, 1976.

[27] B. Korkie and H. J. Turtle. A mean-variance analysis of self-financing portfolios. *Management Science*, 48(3):427–443, 2002.

[28] O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003.

[29] O. Ledoit and M. Wolf. Honey, I shrunk the sample covariance matrix. *Portfolio Management*, 31(4):110–119, 2004a.

[30] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411, 2004b.

[31] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, 1967.

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

[33] H. Markowitz. *Portfolio Selection.* John Wiley and Sons, New York, 1st edition, 1959.

[34] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York, 1st edition, 1979.

[35] R. C. Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8:323–361, 1980.

[36] R. Michaud. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization.* Oxford University Press, New York, 1st edition, 1998.

[37] R. O. Michaud. The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

[38] R. Roll and S. A. Ross. An empirical investigation of the arbitrage pricing theory. *Journal of Finance*, 35(5):1073–1103, 1980.

[39] S. A. Ross. Mutual fund separation in financial theory - the separation distributions. *Journal of Economic Theory*, 17:254–286, 1978.

[40] W. F. Sharpe. A simplified model for portfolio analysis. *Management Science*, 9(2):277–293, 1963.

[41] C. M. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1:197–206, 1956.

[42] J. W. and C. M. Stein. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1:361–380, 1961.

[43] M. Wolf. Resampling vs. shrinkage for benchmarked managers. *Working Paper. Institute for Empirical Research in Economics*, 2006.