Improvement of Portfolio Selection:

A Comparison of Different Methods

by
Fang Yiyun

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Master of Science in Economics

Thesis Committee:

Jun Yu (Supervisor/Chair)
Professor of Economics and Finance
Singapore Management University

Anthony S Tay
Associate Professor of Economics
Singapore Management University

Aurobindo Ghosh
Assistant Professor of Economics
Singapore Management University

Singapore Management University
2011

Abstract

As one of the most important models in finance, Efficient Portfolio Theory pioneered by Markowitz (1952)has been developed since 1950s. Although it has been widely used in practice, Markowitz's mean-variance model has been questioned about its validity because of its bad estimation performance especially in small samples due to the parameter uncertainty problem. Many strategies have been proposed for the purpose of lower the estimation error of mean-variance model. This paper gives a review of the existing literature with the goal of improving the performance of the Markowitz mean-variance model. We evaluate across five empirical data sets of 11 estimation methods. Among these methods, the combination rules by Tu and Zhou (2010) are practicable in terms of Sharpe ratio, and optimal two-fund rule and shrinkage on the covariance rule are practicable in terms of CEQ return. However, in comparison with the in-sample performance, these models surely still have room to improve.

Keywords: Markowitz mean-variance portfolio, Parameter uncertainty, Estimation error

Contents

Contents **i**

Acknowledgement **iii**

1 Introduction **1**

2 Markowitz Paradigm **4**
- 2.1 Case 1: Without risk-free asset 4
- 2.2 Case 2: With risk-free asset 5
- 2.3 Case 1': Without risk-free asset 6
- 2.4 Case 2': With risk-free asset 8
- 2.5 Utility maximization 10

3 Econometric Approaches **11**
- 3.1 Simple plug-in estimation 12
- 3.2 Naive portfolio 13
- 3.3 Minimum-variance portfolio 14
- 3.4 Bayesian portfolio 14
- 3.5 Optimal two-fund portfolio 15
- 3.6 Three-fund portfolio 15
- 3.7 Bayes-Stein portfolio 16
- 3.8 A combination of sophisticated and naive diversification strategies 17
- 3.9 Shrinkage on the covariance 19

4 Empirical Study **21**
- 4.1 Data Sets 21

4.2 Evaluating Performance 23
4.3 Results for the Data Set Considered 25
4.3.1 Sharpe ratios 25
4.3.2 Certainty equivalent return 29
4.3.3 Summary of finding from the empirical data sets 32

5 Conclusions **33**

References **35**

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Jun YU for the continuous support of my Master study, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Anthony TAY and Prof. Aurobindo GHOSH, for their encouragement, insightful comments, and hard questions.

My sincere thanks also goes to Prof. TSE Yiu Kuen, Prof. Jun YU and Ms Lilian Seah, for giving me warn concern when I suffered from illness during my study journey.

I thank my classmates and seniors in SMU: Jia TANG, Tao ZENG, Yonghui ZHANG, Xiaohu WANG, Ye Chen, Qihui CHEN, and Huaxia ZENG, for being good friends during my Master study in SMU.

Last but not the least, I would like to thank my parents, for supporting me spiritually throughout my life.

I would like to dedicate this thesis to my loving parents who have supported me all the way since the beginning of my studies.

Chapter 1

Introduction

Mean-variance framework about portfolio choice theory pioneered by Markowitz (1952) has been frequently used in practice nowadays in asset allocation problem. It has become a standard platform for portfolio construction in financial investment; see Brandt (2005) for an excellent overview of the literature. According to Markowitz (1952), the mean-variance optimal portfolio is the one minimizes the variance of the portfolio return given a certain level of portfolio return. Whether for academic researchers or practitioners, one very important step in implementing mean-variance method is to relate the theoretical formulation of the problem to the real data. Since the true parameters in the model are unknown and have to be estimated from the real data, there exists a parameter uncertainty problem due to the random errors of the estimated parameters. In this case, the true parameters are the mean of the n-vector of asset returns and the covariance matrix of asset returns. Under the assumption that the n-vector of asset returns follows an independently and identically multi-normal distribution, the most widely used estimates of the mean and covariance matrix of return vector are the maximum likelihood estimates (MLE). Treating the MLEs as true parameters and plugging them into the theoretical formulation of optimal weights give the maximum likelihood estimates of optimal weights.

Although MLE of portfolio weights has good asymptotic property, it is not necessarily the best estimation in finite samples. Since it is difficult to estimate the mean and covariance matrix of the return vector, adding to the linearity problem, the plug-in estimates are seriously biased especially in finite samples.

Kan and Zhou (2007) derive the theoretical formula of the expected loss of out-of-sample performance using MLEs when the true parameters are unknown, showing that the expected loss of out-of-sample performance depends on the number of assets, the time periods of the data sample, the investor's tolerance of risk and the squared Sharpe ratio of the ex ante tangency portfolio of the risky assets.

There are a lot of literature which give a view of the parameter uncertainty problem. Frankfurter, Phillips, and Seagle (1971) state that, since sampling error is so large, portfolios selected according to the Markowitz criterion are likely not more efficient than an equally weighted portfolio for three stocks case. Jobson and Korkie (1980) state that naive formation rules such as the equal weight rule can outperform the Markowitz rule. Kan and Zhou (2007) show that "the standard plugging-in approach that replaces the population parameters by their sample estimates can lead to very poor out-of-sample performance; with parameter uncertainty, holding the sample tangency portfolio and the risk-less asset is never optimal". DeMiguel, Garlappi, and Uppal (2009) compare the sample-based mean-variance model and sophisticated rules designed to reduce estimation error, relative to the naive $1/N$ portfolio. In their study, none of the other rules is consistently better than the $1/N$ rule in terms of Sharpe ratio, certainty-equivalent return, or turnover. They note that "the estimation window needed for the sample-based mean-variance strategy and its extensions to outperform the $1/N$ benchmark is around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets".

In order to reduce estimation error, different econometric approaches have been proposed. Bayesian approach based on the predictive distributions is one general framework. Bayesian approach assumes that the investor cares about the expected utility under predictive distribution which is determined by the prior and the historical data. Brown (1976), Klein and Bawa (1976) give Bayesian optimal portfolio weights under the standard prior on mean and covariance matrix of the return vector. Their bayesian solution suggests taking smaller positions in the risky assets compared to the case when the true parameters are known. In a Bayesian framework, the choice of prior matters and it is not easy to construct useful informative prior in practice. Later, Jorion (1986) develops a Bayes-Stein approach. He uses the average excess return on the sample global minimum-variance portfolio as the shrinkage target, and then gets a Bayes-Stein estimator of the expected mean return vector. Kan and Zhou (2007) propose a three-fund rule that optimally combines the risk free asset, the sample tangency portfolio and the sample global minimum-variance portfolio, which dominates a portfolio combines just the risk free asset and the sample tangency portfolio. Tu and Zhou (2010) study the optimal combinations of the $1/N$ rule with MLE or Kan and Zhou's three-fund rule as a way to improve the performance. Overall, their combinations improves over both the $1/N$ rule and the existing rules substantially in most scenarios. There are also methods that focus on reducing the estimation error of covariance matrix such as Ledoit and Wolf (2004), and Ren and Shimotsu (2009). Also, portfolio rules that exploit moment restrictions imposed by factor models are proposed, like Craig and Pastor (2000). Besides, there are portfolio rules that impose short selling constraints, like Frost and Savarino (1988) and Jagannathan and Ma (2003).

The remainder of the paper is organized as follows. Chapter 2 reviews the theory of portfolio choice in single period case. Chapter 3 presents the different econometric approaches to the portfolio choice problem. In Chapter 4, we then do an empirical study to compare these different econometric approaches. Chapter 5 concludes.

Chapter 2

Markowitz Paradigm

Markowitz Efficient Theory of Markowitz (1952) is by far the most commonly used formulation of portfolio choice problem. Suppose, there are N risky assets with a random return vector R_{t+1}, and a risk-free asset with a known return R_t^f. Then the excess returns are defined as $r_{t+1} = R_{t+1} - R_t^f \mathbf{1}$, and its conditional mean (or risk premia) and covariance matrix are defined as μ_t and Σ_t, respectively. And R_t has a conditional mean $e_t = \mu_t + R_t^f \mathbf{1}$.

2.1 Case 1: Without risk-free asset

Suppose the investor can only allocate all his wealth to N risky assets. In the absence of a risk-free asset, mean-variance criteria basically chooses the vector of portfolio weights, w, which represents the relative allocations of wealth to each of N risky assets, to maximize the expected return of the portfolio under a given level of risk (measured by variance of the total return of the portfolio), or to minimize the risk of the resulting portfolio return $R_{p,t+1} = w' R_{t+1}$, subject to a certain value of the expected portfolio return $\bar{\mu}$. Suppose μ_t and Σ_t do not change across time, then the optimization problem will be

$$\min_w \ w' \Sigma w \tag{2.1}$$

subject to

$$w' e = \bar{\mu} \quad and \quad w' \mathbf{1} = 1. \tag{2.2}$$

The first constraint fixes the return of the portfolio to a certain target, and the second constraint makes sure all the wealth invested in the risky assets. The Lagragian function with Lagragian multipliers λ and γ is given by

$$L(w, \lambda, \gamma) = \frac{1}{2} w' \Sigma w + \lambda(\bar{\mu} - w'e) + \gamma(1 - w'\mathbf{1}). \tag{2.3}$$

Solving the first-order conditions, the optimal portfolio weights are:

$$w^{\star}(\bar{\mu}) = g + \bar{\mu} h \tag{2.4}$$

with

$$g = \frac{1}{D}(B\Sigma^{-1}\mathbf{1} - A\Sigma^{-1}e) \quad \textit{and} \quad h = \frac{1}{D}(C\Sigma^{-1}e - A\Sigma^{-1}\mathbf{1}), \tag{2.5}$$

where $\mathbf{1}$ denotes an $N \times 1$ vector of ones and where $A = \mathbf{1}'\Sigma^{-1}e$, $B = e'\Sigma^{-1}e$, $C = \mathbf{1}'\Sigma^{-1}\mathbf{1}$, and $D = BC - A^2$.

Plugging (2.4) into (2.1), we have the minimized portfolio variance equal to

$$\sigma^2(\bar{\mu}) = w^{\star}(\bar{\mu})'\Sigma w^{\star}(\bar{\mu}) = \frac{C}{D}(\bar{\mu} - \frac{A}{C})^2 + \frac{1}{C}, \tag{2.6}$$

which, after rearranging, is a hyperbola in the $\sigma(\bar{\mu}) - \bar{\mu}$ plane with center $(0, A/C)$ and asymptotes of slopes $\pm\sqrt{D/C}$,

$$\frac{\sigma^2(\bar{\mu})}{\frac{1}{C}} - \frac{(\bar{\mu} - \frac{A}{C})^2}{\frac{D}{C^2}} = 1. \tag{2.7}$$

2.2 Case 2: With risk-free asset

Suppose the investor can allocate his wealth to a risk-free asset with return R_t^f without any borrowing and lending limit. Assume the investor allocate w on N risky assets, then the weight he allocates on the risk-free asset will be $1 - w'\mathbf{1}$. In this case, the the optimization problem becomes:

$$\min_{w} \; w'\Sigma w \tag{2.8}$$

subject to

$$w'e + (1 - w'\mathbf{1})R_t^f = \bar{\mu}. \tag{2.9}$$

If $\bar{\mu} = R_t^f$, the solution is trivial: $w^\star(\bar{\mu}) = \mathbf{0}$. Now we assume $\bar{\mu} > R_t^f$. Using the method of Lagrangian multipliers, we solve

$$\min_{w,\lambda} \; L(w, \lambda) = \frac{1}{2} w'\Sigma w + \lambda(\bar{\mu} - w'e - (1 - w'\mathbf{1})R_t^f). \tag{2.10}$$

By solving the first order conditions, we get

$$w^\star(\bar{\mu}) = \Sigma^{-1}(e - R_t^f \mathbf{1})\frac{\bar{\mu} - R_t^f}{H}, \tag{2.11}$$

where $H = (e - R_t^f\mathbf{1})'\Sigma^{-1}(e - R_t^f\mathbf{1}) = B - 2R_t^f A + (R_t^f)^2 C$. Plugging (2.11) into (2.8), we can obtain

$$\sigma^2(\bar{\mu}) = w^\star(\bar{\mu})'\Sigma w^\star(\bar{\mu}) = \frac{(\bar{\mu} - R_t^f)^2}{H}, \tag{2.12}$$

where $\sigma^2(\bar{\mu})$ denotes the variance of the portfolio return $R(w^\star(\bar{\mu}))$ on $w^\star(\bar{\mu})$. By (2.12) we get

$$\sigma(\bar{\mu}) = \frac{1}{\sqrt{H}}|\bar{\mu} - R_t^f|. \tag{2.13}$$

Therefore, the graph $(\sigma(\bar{\mu}), \bar{\mu})$ forms two half-lines emanating from $(0, R_t^f)$ in the $\sigma(\bar{\mu}) - \bar{\mu}$ plane with slopes $\pm\sqrt{H}$.

Case 1 and Case 2 are expressed in terms of asset return, not asset excess return. People also use excess return to describe mean-variance problem in the literatures.

2.3 Case 1': Without risk-free asset

Suppose the investor can only allocate all his wealth to the N risky assets. Let $\mu_t = e_t - R_t^f\mathbf{1}$ be the conditional mean of $r_{t+1} = R_{t+1} - R_t^f\mathbf{1}$, which is the conditional mean of the vector of the excess returns of the N risky assets. Then the optimization problem becomes:

$$\min_{w} \; w'\Sigma w \tag{2.14}$$

subject to

$$w'\mu = \tilde{\mu} \quad and \quad w'\mathbf{1} = 1, \tag{2.15}$$

where $\tilde{\mu}$ is the desired excess return of the portfolio. In this case, we can have a similar solution to the one of Case **1**, by replacing e and $\bar{\mu}$ with μ and $\tilde{\mu}$, respectively. Thus, the optimal weight can be written as

$$w^{\star}(\tilde{\mu}) = \Lambda_1 + \tilde{\mu}\Lambda_2 \tag{2.16}$$

with

$$\Lambda_1 = \frac{1}{\bar{D}}(\bar{B}\Sigma^{-1}\mathbf{1} - \bar{A}\Sigma^{-1}\mu) \quad and \quad \Lambda_2 = \frac{1}{\bar{D}}(C\Sigma^{-1}\mu - \bar{A}\Sigma^{-1}\mathbf{1}), \tag{2.17}$$

where $\bar{A} = \mathbf{1}'\Sigma^{-1}\mu$, $\bar{B} = \mu'\Sigma^{-1}\mu$, $C = \mathbf{1}'\Sigma^{-1}\mathbf{1}$, and $\bar{D} = \bar{B}C - \bar{A}^2$.

Plugging (2.16) into (2.14), we can get

$$\frac{\sigma^2(\tilde{\mu})}{\frac{1}{C}} - \frac{(\tilde{\mu} - \frac{\bar{A}}{C})^2}{\frac{\bar{D}}{C^2}} = 1, \tag{2.18}$$

which is a hyperbola in the $\sigma(\tilde{\mu}) - \tilde{\mu}$ plane with center $(0, \bar{A}/C)$ and asymptotes of slopes $\pm\sqrt{\bar{D}/C}$. Since $\bar{A} = A - R_t^f$, and

$$\begin{aligned}
\bar{D} &= \bar{B}C - \bar{A}^2 \\
&= D + (2\mu + R_t^f)'\Sigma^{-1}\mathbf{1}(R_t^f)'\Sigma^{-1}\mathbf{1} - (2\mu + R_t^f)'\Sigma^{-1}R_t^f\mathbf{1}'\Sigma^{-1}\mathbf{1} \\
&= D + (2\mu + R_t^f)'\Sigma^{-1}(\mathbf{1}(R_t^f)' - R_t^f\mathbf{1}')\Sigma^{-1}\mathbf{1} \qquad (2.19) \\
&= D + 0 \\
&= D,
\end{aligned}$$

let $\tilde{\mu} + R_t^f = \bar{\mu}$, then (2.18) can be rewritten into

$$\frac{\sigma^2(\bar{\mu})}{\frac{1}{C}} - \frac{(\bar{\mu} - \frac{A}{C})^2}{\frac{D}{C^2}} = 1, \tag{2.20}$$

which is exactly the same formula as (2.7). And the graph of formula (2.18) has the same shape as the graph of formula (2.7), only by moving the center $(0, A/C)$ down to $(0, \bar{A}/C)$.

2.4 Case 2': With risk-free asset

The case that allows the investor to allocate his wealth to a risk-free asset with return R_t^f without any borrowing and lending limit can also be expressed in terms of excess return. Since equation (2.9) is equivalent to

$$w'\mu = w'(e - R_t^f \mathbf{1}) = \bar{\mu} - R_t^f =: \tilde{\mu}, \tag{2.21}$$

the optimization problem can be written as:

$$\min_w \; w'\Sigma w \tag{2.22}$$

subject to

$$w'\mu = \tilde{\mu}. \tag{2.23}$$

This leads to

$$w^\star(\tilde{\mu}) = \frac{\tilde{\mu}}{\mu'\Sigma^{-1}\mu}\Sigma^{-1}\mu = \lambda\Sigma^{-1}\mu, \tag{2.24}$$

where λ is a constant that scales proportionately all elements of Σ^{-1} to achieve the desired portfolio risk premium $\tilde{\mu}$. Thus the variance of the portfolio return is

$$\sigma^2(\tilde{\mu}) = \frac{\tilde{\mu}^2}{\mu'\Sigma^{-1}\mu}, \tag{2.25}$$

and the standard error of the portfolio return is

$$\sigma(\tilde{\mu}) = \frac{1}{\sqrt{\mu'\Sigma^{-1}\mu}}|\tilde{\mu}| = \frac{1}{\sqrt{\mu'\Sigma^{-1}\mu}}|\bar{\mu} - R_t^f| = \frac{1}{\sqrt{H}}|\bar{\mu} - R_t^f|, \tag{2.26}$$

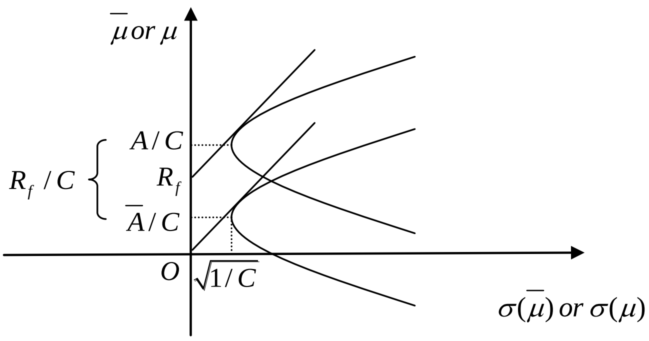

Figure 2.1: Mean-variance frontiers with and without risk-free asset

where $H = (e - R_t^f \mathbf{1})'\Sigma^{-1}(e - R_t^f \mathbf{1}) = \mu'\Sigma^{-1}\mu$. Therefore, the graph $(\sigma(\tilde{\mu}), \tilde{\mu})$ forms two half-lines emanating from $(0, 0)$ in the $\sigma(\tilde{\mu})-\tilde{\mu}$ plane with slopes $\pm\sqrt{H}$.

From expression (2.24), the weights of the tangency portfolio can be found by noting that the weights of the tangency portfolio must sum to one because it lies on the mean-variance frontier of the risky assets. Therefore, for the tangency portfolio:

$$\lambda_{tgc} = \frac{1}{\mathbf{1}'\Sigma^{-1}\mu}, \quad \tilde{\mu}_{tgc} = \frac{\mu'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mu} \quad and \quad w^{\star}_{tgc} = \frac{\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mu}. \tag{2.27}$$

Figure 2.1 illustrates the efficient frontier generated by mean-variance criteria. The two hyperbolas in the figure illustrate the relationship between portfolio return (or portfolio excess return) and standard deviation of the portfolio return of the efficient portfolio in the absence of a risk-free asset. The upper hyperbola represents the efficient frontier in "return-standard deviation" plane while the lower hyperbola represents the efficient frontier in "excess return-standard deviation" plane. The two radials illustrate the cases in the present of a risk-free asset. Similarly, the upper radial is drawn in the "return-standard deviation" plane while the lower one is drawn in the "excess return-standard deviation" plane. Each dot on the frontier gives lowest risk for a given level of expected return (or expected excess return). This figure illustrates the fundamental trade-off between expected return and risk. In the presence of a risk-free asset, the investor allocates fraction w of his wealth to the risky assets and the rest to the risk-free asset.

2.5 Utility maximization

The formulations (2.1)-(2.2), (2.8)-(2.9), (2.14)-(2.15) or (2.22)-(2.23) of the mean-variance problem generate a mapping from a desired portfolio risk premium $\tilde{\mu}$ to the minimum-variance portfolio weights $w^\star$ and resulting portfolio return volatility $\sqrt{w^{\star\prime}\Sigma w^\star}$. However, the choice of the desired risk premium depends on the investor's tolerance for risk. Taking investor's tolerance for risk into consideration, mean-variance problem can be formulated alternatively as an utility maximization problem as following, where the utility function takes a quadratic form:

$$\max_{w} \; w'\mu - \frac{\gamma}{2} w'\Sigma w, \tag{2.28}$$

where γ represents the investor's level of relative risk aversion. The solution to (2.28) is

$$w^\star = \frac{1}{\gamma}\Sigma^{-1}\mu, \tag{2.29}$$

which is exactly the solution to Case 2' when $\lambda = 1/\gamma$, and it links the optimal allocation of the tangency portfolio to the investor's tolerance of the risk.

Chapter 3

Econometric Approaches

From now on, we consider the standard portfolio selection problem that an investor chooses a portfolio which consists a risk-free asset and N risky assets. Using the notation in the last chapter, we denotes $r_{t+1} = R_{t+1} - R_t^f \mathbf{1}$ as the excess rate of return on the N assets at time $t+1$. Assume that r_{t+1} is independent and identically distributed over time and follows a multivariate normal distribution with mean μ and covariance matrix Σ.

Suppose w is the portfolio weights invested on the N risky assets, then $1 - w'\mathbf{1}$ is the weight invested on the risk-free asset. Thus, the excess return on the portfolio at time $t+1$ is $r_{p,t+1} = w' r_{t+1}$, with mean $\mu_p = w'\mu$ and variance $\sigma_p^2 = w'\Sigma w$ respectively. Assume the investor maximizes the quadratic utility function as follows:

$$\max_w \; U(w) = w'\mu - \frac{\gamma}{2} w'\Sigma w, \tag{3.1}$$

where γ measures the investor's level of relative risk aversion. The higher the γ is, the more the investor dislikes risk. The solution to the utility maximization problem is given by

$$w^\star(\mu, \Sigma) = \frac{1}{\gamma} \Sigma^{-1} \mu, \tag{3.2}$$

where μ and Σ are both known, resulting the expected utility as

$$U(w^\star) = \frac{1}{2\gamma} \mu' \Sigma^{-1} \mu. \tag{3.3}$$

However, in practice, we don't know the true values of the parameters μ and

Σ usually. Therefore we cannot compute the true optimal weight $w^\star$ in practice. What we normally do, when implementing the mean-variance portfolio, is to estimate the mean and covariance matrix of the vector of asset excess returns from the previous $T-$period observed data $\Phi(T) = \{r_{t+1}\}_{t=1}^{T}$ and then construct a portfolio for period $T+1$. Therefore, for investors, what matters is finding proper estimate for optimal weight w, which means finding proper estimates for μ and Σ. The following part is focused on the econometric approaches to estimating μ and Σ.

3.1 Simple plug-in estimation

Under the standard i.i.d normal assumption, the maximum likelihood estimators of μ and Σ given the observed sample return data $\Phi(T) = \{r_{t+1}\}_{t=1}^{T}$ are

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_{t+1} \tag{3.4}$$

and

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(r_{t+1}-\hat{\mu})(r_{t+1}-\hat{\mu})'. \tag{3.5}$$

Thus the maximum likelihood estimator of $w^\star$ is $\hat{w}^{ML} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}$. Although that maximum likelihood estimator $\hat{w}^{ML}$ has good asymptotic property, its small sample performance is not good.

Kan and Zhou (2007) state that $\hat{w}^{ML}$ is not optimal in terms of maximizing the expected out-of-sample performance. They provide an analytical expression of the expected out-of-sample performance or the risk function of the plug-in portfolio rule:

$$\rho(w^\star, \hat{w}^{ML}) = (1-k_1)\frac{\theta^2}{2\gamma} + \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)}, \tag{3.6}$$

where $k_1 = (\frac{T}{T-N-2})[2-\frac{T(T-2)}{(T-N-1)(T-N-4)}]$, $\theta^2 = \mu'\Sigma^{-1}\mu$ is the squared Sharpe ratio of the ex ante tangency portfolio of the risky assets and $\rho(w^\star, \hat{w})$ is the expected loss function defined by $\rho(w^\star, \hat{w}) = U(w^\star) - E[\tilde{U}(\hat{w})]$ where $\tilde{U}(\hat{w}) = \hat{w}'\mu - \frac{\gamma}{2}\hat{w}'\Sigma\hat{w}$.

Marx and Hocking (1977) show that $\frac{T-N-2}{T}\hat{\Sigma}^{-1}$ is an unbiased estimator of Σ^{-1} under the i.i.d. normality assumption. Hence, it is suggested to use

$$\tilde{\Sigma} = \frac{1}{T-N-2}\sum_{t=1}^{T}(r_{t+1}-\hat{\mu})(r_{t+1}-\hat{\mu})' = \frac{T}{T-N-2}\hat{\Sigma} \tag{3.7}$$

as an estimator of Σ. Then the new plug-in estimator for the optimal portfolio weights is given by

$$\tilde{w} = \frac{1}{\gamma}\tilde{\Sigma}^{-1}\hat{\mu}, \tag{3.8}$$

which is an unbiased estimator of $w^{\star}$. Kan and Zhou (2007) also prove that $\tilde{w}$ has better expected out-of-sample performance than maximum likelihood estimate $\hat{w}^{ML}$.

3.2 Naive portfolio

The naive strategy is basically to hold a portfolio with equal weight in each of the N risky assets. In that case,

$$w_e = \frac{1}{N}\mathbf{1}. \tag{3.9}$$

That is, the investor allocates all his wealth equally to the risky assets rather than staying cash. This strategy completely ignores the data and does not consider optimization or estimation. The rule is known for a long time and has been used in practice oftenly. Jobson and Korkie (1980) state that "naive formation rules such as equal weight rule can outperform the Markowitz rule." Michaud (1998) notes that "an equally weighted portfolio may often be substantially closer to the true MV optimality than an optimized portfolio." DeMiguel, Garlappi, and Uppal (2009) state that "various extensions to the sample-based mean-variance model that have been proposed in the literature to deal with the problem of estimation error typically do not outperform the $1/N$ benchmark for the seven empirical datasets." Tu and Zhou (2010) find that "for some of the real data sets examined by DeMiguel, Garlappi, and Uppal (2009) and our new data sets here, all of the existing sophisticated strategies, which are the best ones we choose for our study, not only underperform the $1/N$, but also have negative risk-adjusted returns!"

3.3 Minimum-variance portfolio

Under the minimum variance strategy, we choose a portfolio of risky assets that minimize the variance of the return of the portfolio without any consideration of portfolio return. That is,

$$\min_{w} \; w'\Sigma w \quad s.t. \quad w'\mathbf{1} = 1. \tag{3.10}$$

The solution to (3.10) is $w_{min} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$. To implement minimum-variance strategy, we use the estimate of the covariance matrix of the asset returns, $\tilde{\Sigma}$. The estimate of the minimum-variance portfolio weight will be

$$\hat{w}_{min} = \frac{\tilde{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\tilde{\Sigma}^{-1}\mathbf{1}} \tag{3.11}$$

3.4 Bayesian portfolio

Brown (1976) shows that, under the standard diffuse prior on μ and Σ, Bayesian optimal portfolio weights perform better than usual plug-in estimation. Suppose the prior distribution on μ and Σ is

$$p_0(\mu, \Sigma) \propto |\Sigma|^{-\frac{N+1}{2}}, \tag{3.12}$$

then the Bayesian optimal portfolio weights will be

$$w^{Bayes} = \frac{1}{\gamma}(\frac{T-N-2}{T+1})\hat{\Sigma}^{-1}\hat{\mu} \tag{3.13}$$

Kan and Zhou (2007) note that Bayesian rule strictly outperforms simple plug-in estimation by showing analytically that Bayesian rule results in higher expected out-of-sample performance.

3.5 Optimal two-fund portfolio

Kan and Zhou (2007) proposed an optimal two-fund rule by assuming the optimal weights have the form

$$\hat{w} = \frac{c}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}. \tag{3.14}$$

They show that the optimal c under the criteria that it maximizes the expected out-of-sample performance will be

$$c^\star = \frac{(T-N-1)(T-N-4)}{T(T-2)}\frac{\theta^2}{\theta^2+\frac{N}{T}}. \tag{3.15}$$

Since θ^2 is unknown in practice, it needs to be estimated too. To avoid the heavy estimation bias when T is small, they use an adjusted estimator of θ^2, resulting an adjusted estimator of $c^\star$ as

$$\hat{c}^\star = \frac{(T-N-1)(T-N-4)}{T(T-2)}\frac{\hat{\theta}_a^2}{\hat{\theta}_a^2+\frac{N}{T}}, \tag{3.16}$$

where $\hat{\theta}_a^2 = \frac{(T-N-2)\hat{\theta}^2-N}{T} + \frac{2(\hat{\theta}^2)^{N/2}(1+\hat{\theta}^2)^{-(T-2)/2}}{TB_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2,(T-N)/2)}$, $\hat{\theta}^2 = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$, and $B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}dy$. The resulting optimal two-fund portfolio weights are

$$\hat{w}^{two} = \frac{1}{\gamma}\hat{c}^\star\hat{\Sigma}^{-1}\mu. \tag{3.17}$$

3.6 Three-fund portfolio

Kan and Zhou (2007) propose another method called three-fund separation that holding a combination of a risk-free asset, the sample tangency portfolio and a third risky portfolio. They believe that "if there is parameter uncertainty, the use of another risky portfolio can help to diversify estimation risk of the sample tangency portfolio". In their construction, they choose the sample global minimum-variance portfolio as the third risky portfolio. They consider the portfolio rule of the form

$$\hat{w} = \frac{1}{\gamma}(c\hat{\Sigma}^{-1}\hat{\mu} + d\hat{\Sigma}^{-1}\mathbf{1}) \tag{3.18}$$

By maximizing the expected out-of-sample performance of this class of portfolio rules, they obtain the optimal c and d as

$$c^{\star\star} = \frac{(T-N-2)(T-N-4)}{T(T-2)}\left(\frac{\psi^2}{\psi^2 + \frac{N}{T}}\right), \tag{3.19}$$

$$d^{\star\star} = \frac{(T-N-2)(T-N-4)}{T(T-2)}\left(\frac{\frac{N}{T}}{\psi^2 + \frac{N}{T}}\right)\mu_g, \tag{3.20}$$

where $\psi^2 = \mu'\Sigma^{-1}\mu - \frac{(\mu'\Sigma^{-1}\mathbf{1})^2}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = (\mu - \mu_g\mathbf{1})'\Sigma^{-1}(\mu - \mu_g\mathbf{1})$ and $\mu_g = \frac{\mathbf{1}'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$. Therefore, the optimal three-fund portfolio weights are

$$\hat{w}^{\star\star} = \frac{(T-N-2)(T-N-4)}{\gamma T(T-2)}\left[\left(\frac{\psi^2}{\psi^2 + \frac{N}{T}}\right)\hat{\Sigma}^{-1}\hat{\mu} + \left(\frac{\frac{N}{T}}{\psi^2 + \frac{N}{T}}\right)\mu_g\hat{\Sigma}^{-1}\mathbf{1}\right] \tag{3.21}$$

The usual estimates of μ_g and ψ^2 are

$$\hat{\mu}_g = \frac{\mathbf{1}'\hat{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}}, \tag{3.22}$$

$$\hat{\psi}^2 = (\hat{\mu} - \hat{\mu}_g\mathbf{1})'\hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g\mathbf{1}). \tag{3.23}$$

To avoid the problem that $\hat{\psi}^2$ being a heavily biased estimator when T is small, they use the adjusted estimator of ψ^2:

$$\hat{\psi}_a^2 = \frac{(T-N-1)\hat{\psi}^2 - (N-1)}{T} + \frac{2(\hat{\psi}^2)^{(N-1)/2}(1+\hat{\psi}^2)^{-(T-2)/2}}{TB_{\hat{\psi}^2/(1+\hat{\psi}^2)}((N-1)/2, (T-N+1)/2)} \tag{3.24}$$

Thus, there three-fund portfolio weights estimators are

$$\hat{w}^{three} = \frac{(T-N-2)(T-N-4)}{\gamma T(T-2)}\left[\left(\frac{\hat{\psi}_a^2}{\hat{\psi}_a^2 + \frac{N}{T}}\right)\hat{\Sigma}^{-1}\hat{\mu} + \left(\frac{\frac{N}{T}}{\hat{\psi}_a^2 + \frac{N}{T}}\right)\hat{\mu}_g\hat{\Sigma}^{-1}\mathbf{1}\right] \tag{3.25}$$

3.7 Bayes-Stein portfolio

The Bayes-Stein portfolio is actually an application of the idea of shrinkage estimation. Basically, the new estimator shrinks the sample mean toward a common "grand mean". Jorion (1986) takes the grand mean to be the mean of the minimum variance portfolio, μ^{min}. Assuming that the informative prior of μ is

$$p_0(\mu|\Sigma, \mu_g, \lambda) \propto exp[-\frac{1}{2}(\mu - \mathbf{1}\mu_g)'(\lambda\Sigma^{-1})(\mu - \mathbf{1}\mu_g)] \tag{3.26}$$

the Bayes-Stein estimation for μ will be

$$\hat{\mu}^{bs} = (1 - \hat{\phi})\hat{\mu} + \hat{\phi}\hat{\mu}_g\mathbf{1}, \tag{3.27}$$

where

$$\hat{\phi} = \frac{N+2}{(N+2) + T(\hat{\mu} - \hat{\mu}_g\mathbf{1})'\tilde{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g\mathbf{1})}. \tag{3.28}$$

Hence, the Bayes-Stein estimators of portfolio weights are

$$\hat{w}^{bs} = \frac{1}{\gamma}\tilde{\Sigma}^{-1}\hat{\mu}^{bs}. \tag{3.29}$$

3.8 A combination of sophisticated and naive diversification strategies

Tu and Zhou (2010) propose some new methods that combining existing mean-variance portfolio rules with naive portfolio rule. They find that "the combinations improve over both the $1/N$ and the existing rules substantially in most scenarios". First, they propose a combination rule which combines the naive rule and unbiased rule together defined as

$$w^{CML} = (1 - \delta)w_e + \delta\tilde{w} \tag{3.30}$$

They minimize the expected loss function under the w^{CML} rule,

$$\min_{\delta} \; L(w^{\star}, w^{CML}) = \frac{\gamma}{2}[(1 - \delta)^2\pi_1 + \delta^2\pi_2], \tag{3.31}$$

where

$$\pi_1 = (w_e - w^{\star})'\Sigma(w_e - w^{\star}), \tag{3.32}$$

$$\pi_2 = E[(\tilde{w} - w^{\star})'\Sigma(\tilde{w} - w^{\star})], \tag{3.33}$$

getting

$$\delta^{\star} = \frac{\pi_1}{\pi_1 + \pi_2}. \tag{3.34}$$

They obtain the estimator of π_1 being

$$\hat{\pi}_1 = w_e' \hat{\Sigma} w_e - \frac{2}{\gamma} w_e' \hat{\mu} + \frac{1}{\gamma^2} \hat{\theta}_a^2, \tag{3.35}$$

where $\hat{\theta}_a^2$ is defined in section 3.5, and the estimator of π_2 being

$$\hat{\pi}_2 = \frac{1}{\gamma^2}(c_1 - 1)\hat{\theta}_a^2 + \frac{c_1}{\gamma^2}\frac{N}{T}, \tag{3.36}$$

where $c_1 = \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}$. Thus, the estimator of δ is

$$\hat{\delta} = \frac{\hat{\pi}_1}{\hat{\pi}_1 + \hat{\pi}_2}, \tag{3.37}$$

and the resulting estimated optimal rule is

$$\hat{w}^{CML} = (1 - \hat{\delta}) w_e + \hat{\delta} \tilde{w}. \tag{3.38}$$

Tu and Zhou (2010) also try to combine the three-fund rule by Kan and Zhou (2007) with naive portfolio rule together. They consider the combination rule as follows:

$$w^{CKZ} = (1 - \delta_k) w_e + \delta_k \hat{w}^{three}. \tag{3.39}$$

The obtained estimate of δ_k is

$$\hat{\delta}_k = \frac{\hat{\pi}_1 - \hat{\pi}_{13}}{\hat{\pi}_1 - 2\hat{\pi}_{13} + \hat{\pi}_3}, \tag{3.40}$$

where

$$\hat{\pi}_{13} = \frac{1}{\gamma^2}\hat{\theta}_a^2 - \frac{1}{\gamma} w_e' \hat{\mu} + \frac{1}{\gamma c_1}([(\frac{\hat{\psi}_a^2}{\hat{\psi}_a^2 + \frac{N}{T}}) w_e' \hat{\mu} + (\frac{\frac{N}{T}}{\hat{\psi}_a^2 + \frac{N}{T}}) \hat{\mu}_g w_e' \mathbf{1}] \tag{3.41}$$

$$-\frac{1}{\gamma}[(\frac{\hat{\psi}_a^2}{\hat{\psi}_a^2 + \frac{N}{T}}) \hat{\mu}' \tilde{\Sigma}^{-1} \hat{\mu} + (\frac{\frac{N}{T}}{\hat{\psi}_a^2 + \frac{N}{T}}) \hat{\mu}_g \hat{\mu}' \tilde{\Sigma}^{-1} \mathbf{1}]) \tag{3.42}$$

where $\hat{\psi}_a^2$ is defined in section 3.6, and

$$\hat{\pi}_3 = \frac{1}{\gamma^2}\hat{\theta}_a^2 - \frac{1}{\gamma^2 c_1}(\hat{\theta}_a^2 - \frac{N}{T}\frac{\hat{\psi}_a^2}{\hat{\psi}_a^2 + \frac{N}{T}}). \tag{3.43}$$

3.9 Shrinkage on the covariance

The shrinkage estimation on the covariance basically employs the idea of getting a more accurate estimator of covariance matrix in small sample by Ren and Shimotsu (2009),

$$\hat{\Sigma}^{bs} = (1-\hat{\alpha})\hat{\Sigma} + \hat{\alpha}F, \tag{3.44}$$

where F is an estimated covariance matrix of the asset returns implied by a factor model.

Assume the asset returns follow a factor model,

$$r_{it} = \mu_i + X_t\beta_i + \epsilon_{it}, \quad t = 1, ..., T \tag{3.45}$$

where X_t is a vector of factors, and ϵ_{it} is a mean-zero idiosyncratic error for asset i in period t. ϵ_{it} has constant variance δ_{ii} across time, and is uncorrelated to ϵ_{jt} and to the factors. Denote $\mu = [\mu_1, ..., \mu_N]$, $\beta = [\beta_1, ..., \beta_N]$, $\epsilon = [\epsilon_1, ..., \epsilon_N]$, then the factor model is written as

$$r_t = \mu + X_t\beta + \epsilon_t, \quad t = 1, ..., T \tag{3.46}$$

The covariance matrix of r_t implied by the factor model is:

$$\Phi = \beta' Var(X_t)\beta + \Delta, \tag{3.47}$$

where $\Delta = diag(\delta_{ii})$ is the covariance of ϵ_t. Thus, the estimate of Φ can be obtained as:

$$F = b' Var(X_t)b + D, \tag{3.48}$$

where $b = [b_1, ..., b_N]$, $D = diag(d_{ii})$. b_i and d_{ii} denote the OLS estimates of β_i and δ_{ii} by regressing the $i-$th asset returns on an intercept and the factors.

The optimized α is complicated and thus omitted here. Please refer to Ren and Shimotsu (2009) for the specific calculation.

Chapter 4

Empirical Study

In this chapter, we try to understand which econometric approach of mean-variance model performs better by evaluating the performance of the sophisticated strategies discussed in the last chapter across five different empirical data sets of monthly returns, using two performance criteria suggested by DeMiguel, Garlappi, and Uppal (2009):
(1) the out-of-sample Sharpe ratio;
(2) the certainty-equivalent (CEQ) return for the expected utility of a mean-variance investor.

There are 11 portfolio rules we are going to evaluate, including the sample-based simple plugging-in portfolio rule, its different extensions designed to reduce the estimation error and the naive rule which allocates $1/N$ wealth to each of the N assets.

4.1 Data Sets

The real data sets used in our study include Fama-French's size- and book-to-market-sorted portfolio value-weighted monthly return data and industry portfolio value-weighted monthly return data. For the size- and book-to-market-sorted portfolio, we consider two cases: 6 portfolios and 25 portfolios. And for the industry portfolio, we consider three cases, including 10 industry portfolios, 30 industry portfolios, and 49 industry portfolios. Different types of data sets and data sets with different numbers of assets help to figure out how the portfolio rules work under different kinds of data.

(1) 6 portfolios formed on size and book-to-market:
The data is obtained from Kenneth R. French's web site and spanned from July 1963 to December 2010. The portfolios are the intersections of 2 portfolios formed on the size (market equity, ME) and 3 portfolios formed on the ratio of book equity to market equity (BE/ME). The size breakpoint for year t is the median NYSE market equity at the end of June of year t. BE/ME for June of year t is the book equity for the last fiscal year end in t-1 divided by ME for December of t-1. The BE/ME breakpoints are the 30th and 70th NYSE percentiles.

(2) 25 portfolios formed on size and book-to-market:
The data is obtained from Kenneth R. French's web site and spanned from July 1963 to December 2010. The portfolios are the intersections of 5 portfolios formed on size (market equity, ME) and 5 portfolios formed on the ratio of book equity to market equity (BE/ME). The size breakpoints for year t are the NYSE market equity quintiles at the end of June of t. BE/ME for June of year t is the book equity for the last fiscal year end in t-1 divided by ME for December of t-1. The BE/ME breakpoints are NYSE quantiles.

(3) 10 industry portfolios:
The data is obtained from Kenneth R. French's web site and spanned from July 1963 to December 2010. Each NYSE, AMEX, and NASDAQ stock is assigned to an industry portfolio at the end of June of year t based on its four-digit SIC code at that time. (The codes used here are the Compustat SIC codes for the fiscal year ending in calender year t-1. Whenever Compustat SIC codes are not available, CRSP SIC codes are used.) Then the returns from July of t to June of t+1 are computed. The industry categories include (1)consumer nondurables (food, tobacco, textiles, apparel, leather, toys), (2)consumer durables (cars, TV's, furniture, household appliances), (3)manufacturing (machinery, trucks, planes, chemicals, off furn, paper, com printing), (4)energy (oil, gas, and coal extraction and products), (5)business equipment (computers, software, and electronic equipment), (6)telephone and television transmission, (7)wholesale, retail, (8)and some services, healthcare, medical equipment, and drugs, (9)utilities, (10)other.

Table 4.1: List of data sets

#	Data set and source	N	time period
1	Six (2×3) size- and book-to-market portfolio Source: Ken French's web site	6	07/1963-12/2010
2	Twenty-five (5×5) size- and book-to-market portfolio Source: Ken French's web site	25	07/1963-12/2010
3	10 industry portfolio Source: Ken French's web site	10	07/1963-12/2010
4	30 industry portfolio Source: Ken French's web site	30	07/1963-12/2010
5	49 industry portfolio Source: Ken French's web site	49	07/1963-12/2010

(4) 30 industry portfolios:
The data is obtained from Kenneth R. French's web site and spanned from July 1963 to December 2010. The portfolios are constructed in a similar way to the one 10 industry portfolios constructed only with 30 different (or more detailed) categories.

(5) 49 industry portfolios:
The data is obtained from Kenneth R. French's web site and spanned from July 1963 to December 2010. The portfolios are constructed in a similar way to the one 10 industry portfolios constructed only with 49 different (or more detailed) categories.

Table 4.1 has listed the five data sets we used in the empirical study.

4.2 Evaluating Performance

Following DeMiguel, Garlappi, and Uppal (2009), we use a "rolling-sample" approach in our analysis. For a given T-month-long data set of monthly returns, we choose an estimation window of length $M = 120$ or $M = 240$ months. In each month t, starting from $t = M$, we use the data in the previous M months up to month t to implement various portfolio rules. The resulting rules are used to determine the investment decisions for the next month $t + 1$. That is, for example, suppose $w_{k,t}$ are the estimated portfolio weights in month t under a portfolio rule k and suppose the realized excess asset returns in month $t+1$ are r_{t+1}, then the realized excess return in month $t+1$ under the portfolio rule k is $r_{k,t+1} = w'_{k,t} r_{t+1}$. Re-do this process by adding the asset returns of the next period in the data set and dropping the earliest ones, until it reaches the end of the data set. By doing this "rolling-sample" approach, we can get a series of $T - M$ monthly *out-of-sample* returns generated by the 11 different portfolio rules.

Given the time series of *out-of-sample* returns generated by each portfolio rule, we then compute the *certainty-equivalent(CEQ) return* as

$$C\hat{E}Q_k = \hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2, \tag{4.1}$$

in which $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the mean and standard deviation of out-of-sample excess portfolio return. The results we report are for the cases when $\gamma = 1$ and $\gamma = 3$. CEQ return can be regarded as the risk-free rate an investor is willing to accept rather than adopting a particular risky portfolio rule. The CEQ, is the guaranteed amount of money that an individual would view as equally desirable as a risky asset. It is obvious that the higher the CEQ, the better the rule k for investors.

We also calculate the *out-of-sample Sharpe ratio* associated with rule k, defined as the sample mean of out-of-sample excess portfolio return, $\hat{\mu}_k$, divided by the sample standard deviation of out-of-sample excess portfolio return, $\hat{\sigma}_k$:

$$S\hat{R}_k = \frac{\hat{\mu}_k}{\hat{\sigma}_k}. \tag{4.2}$$

Although the true parameters are unknown in the real data, we still calculate the *in-sample Sharpe ratio* and *in-sample certainty-equivalent return* on the purpose of comparing the estimation errors under different rules. They are computed by using the whole sample of excess returns, that is, when $M = T$. The *in-sample Sharpe ratio* is

$$S\hat{R}_k^{is} = \frac{\hat{\mu}_{k'}^{is}\hat{w}_k}{\sqrt{\hat{w}'_k\hat{\Sigma}_k^{is}\hat{w}_k}}, \tag{4.3}$$

and the *in-sample CEQ* is

$$C\hat{E}Q = \hat{\mu}_k^{is\prime}\hat{w}_k - \frac{\gamma}{2}\hat{w}_k^{\prime}\hat{\Sigma}_k^{is}\hat{w}_k. \tag{4.4}$$

4.3 Results for the Data Set Considered

For each data set, we calculate the Sharpe ratio and the CEQ return under different portfolio rules. We set the risk aversion coefficient γ to be 1 or 3. The results are listed from Table 4.2 to Table 4.9.

4.3.1 Sharpe ratios

Table 4.2 gives the Sharpe ratio across all the data sets listed in Table 4.1 for each strategy when $\gamma = 1$ and $M = 120$. The first row of Table 4.2, "in-sample", gives the Sharpe ratio of the Markowitz mean-variance strategy *in-sample*, that is, we assume the whole sample estimates of the parameters to be the true values of the parameters, hence there is no estimation error. The second row of the table gives the Sharpe ratio of the simple plugging-in maximum likelihood estimation. The third row uses the unbiased estimation of portfolio weights. The fourth row gives the Sharpe ratio under the naive $1/N$ strategy. The fifth row uses the minimum variance portfolio rule. The sixth row uses the bayesian rule under a diffuse prior. The seventh row uses the optimal two-fund rule by Kan and Zhou (2007). The eighth row uses the optimal three-fund rule which is also proposed by Kan and Zhou (2007). The ninth row uses the Bayesian-Stein rule. The tenth row gives the Sharpe ratio under the rule which combines the naive rule and maximum likelihood rule together suggested by Tu and Zhou (2010). The eleventh row gives the Sharpe ratio of the combination rule which combines the naive rule and the Kan-Zhou's optimal three-fund rule together which is also suggested by Tu and Zhou (2010). The last one uses the shrinkage-on-covariance method by Fang and Ren.

For example, for the "10 industry portfolios" data set, the in-sample mean-variance portfolio has a Sharpe ratio of 0.1803, while the Sharpe ratio of the simple plugging-in MLE is much less, only 0.0278. Similarly, for the "49 industry

Table 4.2: Sharpe ratio under different portfolio rules when $M = 120$ and $\gamma = 1$

Strategy	F-F portfolios N=6	F-F portfolios N=25	Industry portfolios N=10	Industry portfolios N=30	Industry portfolios N=49
in-sample	0.4115	0.4548	0.1803	0.2836	0.5111
mle	0.3775	0.3246	0.0278	0.1801	0.0254
unbiased	0.3775	0.3246	0.0278	0.1801	0.0254
$1/N$	0.1581	0.1486	0.1288	0.1502	0.1257
min	0.2870	0.2572	0.1555	0.0851	0.0528
bayesian	0.3775	0.3246	0.0278	0.1801	0.0254
two-fund	0.3776	0.3284	0.0294	0.1646	-0.0209
three-fund	0.3794	0.3356	0.0647	0.1748	0.0159
b-s	0.3795	0.3368	0.0717	0.1726	0.0104
cml	0.3795	0.3296	0.0344	0.1882	0.0423
ckz	0.3837	0.3366	0.0715	0.1870	0.0566
shrinkage	0.3477	0.2858	0.0407	0.1992	0.0362

Table 4.3: Sharpe ratio under different portfolio rules when $M = 240$ and $\gamma = 1$

Strategy	FF N=6	FF N=25	Industry portfolios N=10	Industry portfolios N=30	Industry portfolios N=49
in-sample	0.4115	0.4548	0.1803	0.2836	0.5111
mle	0.3870	0.3698	0.0757	0.1063	0.0597
unbiased	0.3870	0.3698	0.0757	0.1063	0.0597
$1/N$	0.1388	0.1369	0.1462	0.1162	0.1281
min	0.3357	0.3407	0.1995	0.1521	0.1052
bayesian	0.3870	0.3698	0.0757	0.1063	0.0597
two-fund	0.3735	0.3351	0.0388	0.0813	0.0701
three-fund	0.3786	0.3687	0.1117	0.1211	0.0664
b-s	0.3760	0.3684	0.1160	0.1237	0.0679
cml	0.3881	0.3717	0.0794	0.1111	0.0687
ckz	0.3790	0.3705	0.1173	0.1286	0.0920
shrinkage	0.3794	0.3766	0.0788	0.1094	0.0541

Table 4.4: Sharpe ratio under different portfolio rules when $M = 120$ and $\gamma = 3$

Strategy	FF N=6	FF N=25	Industry portfolios N=10	Industry portfolios N=30	Industry portfolios N=49
in-sample	0.4115	0.4548	0.1803	0.2801	0.5111
mle	0.3775	0.3246	0.0278	0.1801	0.0254
unbiased	0.3775	0.3246	0.0278	0.1801	0.0254
$1/N$	0.1581	0.1486	0.1288	0.1502	0.1257
min	0.2870	0.2572	0.1555	0.0851	0.0528
bayesian	0.3775	0.3246	0.0278	0.1801	0.0254
two-fund	0.3776	0.3284	0.0294	0.1646	-0.0209
three-fund	0.3794	0.3356	0.0647	0.1748	0.0159
b-s	0.3790	0.3368	0.0717	0.1726	0.0104
cml	0.3782	0.3266	0.0300	0.1830	0.0310
ckz	0.3818	0.3366	0.0684	0.1812	0.0394
shrinkage	0.3470	0.2852	0.0408	0.1989	0.0354

Table 4.5: Sharpe ratio under different portfolio rules when $M = 240$ and $\gamma = 3$

Strategy	FF N=6	FF N=25	Industry portfolios N=10	Industry portfolios N=30	Industry portfolios N=49
in-sample	0.4115	0.4548	0.1803	0.2836	0.5111
mle	0.3870	0.3698	0.0757	0.1063	0.0597
unbiased	0.3870	0.3698	0.0757	0.1063	0.0597
$1/N$	0.1388	0.1369	0.1462	0.1162	0.1281
min	0.3357	0.3407	0.1995	0.1521	0.1052
bayesian	0.3870	0.3698	0.0757	0.1063	0.0597
two-fund	0.3735	0.3351	0.0388	0.0813	0.0701
three-fund	0.3786	0.3687	0.1117	0.1211	0.0664
b-s	0.3760	0.3684	0.1160	0.1237	0.0679
cml	0.3873	0.3704	0.0769	0.1079	0.0626
ckz	0.3781	0.3704	0.1162	0.1237	0.0834
shrinkage	0.3797	0.3764	0.0786	0.1094	0.0541

portfolios" data set, the in-sample Sharpe ratio for the mean-variance portfolio strategy is 0.5111, while the Sharpe ratio of MLE is only 0.0254. Although, for the two "size- and book-to-market-sorted portfolios" data sets, the simple plugging-in MLE seems to give a reasonable out-of-sample Sharpe ratio, this strategy doesn't work for the "10 industry portfolios", "30 industry portfolios", and "49 industry portfolios" data sets. It confirms the perils of using traditional sample-based MLE to implement Markowitz's mean-variance portfolios.

Let's look at the results given by the sophisticated strategies designed to reduce the estimation errors. Of course, if all the parameters are known to the investors, then all these designed strategies will outperform the simple plugging-in MLE. However, due to the parameter uncertainty problem, whether these strategies can outperform MLE or not leaves to be a question.

In Table 4.2, the Sharpe ratios for the unbiased strategy and Bayesian strategy are the same as the ones under simple plugging-in MLE. The Bayes-Stein strategy has a higher out-of-sample Sharpe ratio than MLE for all the data sets except for "30 industry portfolios" and "49 industry portfolios". The same situation happens to optimal two-fund rule and optimal three-fund rule.

As for the combination rules, the rule that combines naive rule and unbiased estimation performs well in terms of higher Sharpe ratio than MLE across all these data sets. Moreover, the rule that combines naive rule and Kan-Zhou's optimal three-fund rule generates even higher Sharpe ratios across most of the data sets.

Also, we find that shrinkage method has a higher out-of-sample Sharpe ratio for all the industry data sets than simple plugging-in MLE.

When we increase the rolling window M to 240, things are slightly different but not too much. But we find that most of the rules fail for the 2×3 size- and book-to-market-sorted portfolios. The reason for that might be the improving of estimation accuracy when the length of rolling window increases. If we change the risk aversion rate γ to 3, there are no big differences.

Among all the above case, only the combination rule (cml) always outperforms simple plugging-in MLE in terms of out-of-sample Sharpe ratio. Other portfolio rules such as optimal three-fund rule and shrinkage on the covariance rule perform better in many cases but not all cases.

Table 4.6: CEQ return under different portfolio rules when $M = 120$ and $\gamma = 1$

Strategy	FF N=6	FF N=25	Industry portfolios N=10	Industry portfolios N=30	Industry portfolios N=49
in-sample	0.0847	0.1034	0.0163	0.0402	0.1306
mle	-0.0065	-0.3167	-0.0658	-0.6047	-1.5482
unbiased	0.0125	-0.1232	-0.0524	-0.2796	-0.5008
$1/N$	-14.6305	-12.8966	-9.2684	-16.1834	-12.2495
min	-9.0708	-6.7345	-6.1585	-9.3280	-9.1599
bayesian	0.0145	-0.1187	-0.0514	-0.2736	-0.4923
two-fund	0.0462	0.0434	-0.0026	0.0024	-0.0269
three-fund	0.0423	0.0255	-0.0087	-0.0375	-0.0763
b-s	0.0376	-0.0117	-0.0084	-0.0783	-0.1638
cml	0.0462	-0.1179	-0.0532	-0.2684	-0.4624
ckz	0.0462	0.0211	-0.0147	-0.0423	-0.0809
shrinkage	0.0585	-0.0113	-0.0473	-0.2068	-0.5903

Moreover, in the case when $M = 240$ and $\gamma = 3$, the out-of-sample Sharpe ratio for MLE is less than that for the naive $1/N$ rule for three of five data sets. That means the effect of estimation error sometimes is so large that the optimal diversification fails. For instance, for the data set "10 industry portfolio", MLE results in a Sharpe ratio of 0.0597 compared to its in-sample value of 0.5111, and 0.1281 for the $1/N$ strategy.

4.3.2 Certainty equivalent return

Since the CEQ return can be interpreted as the risk-free rate that an investor is willing to accept in stead of adopting a given risky portfolio, the higher the CEQ return, the better the optimal rule.

When $M = 120$ and $\gamma = 1$, the in-sample MLE always has a positive CEQ return across all these data sets, while the simple plugging-in MLE gives a negative CEQ return for all the data sets. Surprisingly, bayesian rule, optimal two fund rule, optimal three fund rule, Bayes-Stein rule, combination rules, and shrinkage rule all generate higher CEQ return than simple plugging-in MLE for all the data sets. Furthermore, shrinkage rule has a decent value 0.0585, which is the high-

Table 4.7: CEQ return under different portfolio rules when $M = 240$ and $\gamma = 1$

Strategy	FF N=6	FF N=25	Industry portfolios N=10	Industry portfolios N=30	Industry portfolios N=49
in-sample	0.0847	0.1034	0.0163	0.0402	0.1306
mle	0.0405	-0.0640	-0.0203	-0.2228	-0.3223
unbiased	0.0460	-0.0177	-0.0173	-0.1577	-0.1913
$1/N$	-13.3065	-11.6986	-8.5390	-14.7439	-10.8981
min	-8.5516	-5.5325	-6.2033	-6.7251	-6.7873
bayesian	0.0466	-0.0164	-0.0171	-0.1561	-0.1895
two-fund	0.0506	0.0340	-0.0014	-0.0123	0.0006
three-fund	0.0519	0.0437	0.0023	-0.0279	-0.0328
b-s	0.0493	0.0306	0.0029	-0.0348	-0.0455
cml	0.0467	-0.0168	-0.0173	-0.1521	-0.1843
ckz	0.0527	0.0439	0.0015	-0.0276	-0.0325
shrinkage	0.0704	0.0543	-0.0149	-0.1399	-0.2276

Table 4.8: CEQ return under different portfolio rules when $M = 120$ and $\gamma = 3$

Strategy	FF N=6	FF N=25	Industry portfolios N=10	Industry portfolios N=30	Industry portfolios N=49
in-sample	0.0282	0.0345	0.0054	0.0134	0.0435
mle	-0.0022	-0.1056	-0.0219	-0.2016	-0.5161
unbiased	0.0042	-0.0411	-0.0175	-0.0932	-0.1669
$1/N$	-45.6529	-40.2445	-28.9479	-50.3046	-38.0250
min	-29.8272	-22.2287	-19.6166	-28.7339	-27.9376
bayesian	0.0048	-0.0396	-0.0171	-0.0912	-0.1641
two-fund	0.0154	0.0145	-0.0009	0.0008	-0.0090
three-fund	0.0141	0.0085	-0.0029	-0.0125	-0.0254
b-s	0.0125	-0.0039	-0.0028	-0.0261	-0.0546
cml	0.0043	-0.0410	-0.0176	-0.0925	-0.1639
ckz	0.0151	0.0080	-0.0045	-0.0140	-0.0274
shrinkage	0.0194	-0.0039	-0.0158	-0.0688	-0.1963

Table 4.9: CEQ return under different portfolio rules when $M = 240$ and $\gamma = 3$

Strategy	FF N=6	FF N=25	Industry portfolios N=10	Industry portfolios N=30	Industry portfolios N=49
in-sample	0.0282	0.0345	0.0045	0.0134	0.0435
mle	0.0135	-0.0213	-0.0068	-0.0743	-0.1074
unbiased	0.0153	-0.0059	-0.0058	-0.0526	-0.0638
$1/N$	-41.3904	-36.4577	-26.8686	-45.5205	-33.9235
min	-28.6656	-19.1085	-20.0975	-21.3378	-21.1591
bayesian	0.0155	-0.0055	-0.0057	-0.0520	-0.0632
two-fund	0.0169	0.0113	-0.0005	-0.0041	-0.0002
three-fund	0.0173	0.0146	0.0008	-0.0093	-0.0109
b-s	0.0164	0.0102	0.0010	-0.0116	-0.0152
cml	0.0154	-0.0060	-0.0058	-0.0521	-0.0632
ckz	0.0175	0.0150	0.0006	-0.0098	-0.0109
shrinkage	0.0235	0.0180	-0.0050	-0.0466	-0.0759

est CEQ return for the “6 size- and book-to-market-sorted portfolios” data set. Optimal two fund rule performs best for the rest of the data sets, among which the CEQ returns for “25 size- and book-to-market-sorted portfolios” and “30 industry portfolios” data sets are positive and the ones for “10 industry portfolios” and “49 industry portfolios” are negative.

When M is increased to 240, shrinkage rule has the best performance for the “6 size- and book-to-market-sorted portfolios” and “25 size- and book-to-market-sorted portfolios” data sets. Bayes-Stein rule has the best performance for the “10 industry portfolios” data set. And the optimal two fund rule performs best for the “30 industry portfolios” and “49 industry portfolios”data sets.

When $M = 120$ and $\gamma = 3$, bayesian rule, optimal two fund rule, optimal three fund rule, Bayes-Stein rule, combination rules, and shrinkage rule still generate higher CEQ return than simple plugging-in MLE for all the data sets. Shrinkage rule gives the highest CEQ return for “6 size- and book-to-market-sorted portfolios” data set. Optimal two fund rule gives the highest CEQ return for the rest four data sets. When M is increased to 240, the best rule for “25 size- and book-to-market-sorted portfolios” data set becomes shrinkage on the covariance rule and the best rule for “10 industry portfolios” data set becomes Bayes-Stein rule.

4.3.3 Summary of finding from the empirical data sets

From the above discussion, we find that, in terms of Sharpe ratio, the combination rule that combines naive rule and unbiased estimation together always outperforms simple plugging-in MLE. The combination rule that combines naive rule and three-fund rule together outperforms simple plugging-in MLE for most of the cases. The reason why it fails when rolling window is 240 might be the improving of estimation accuracy of MLE when the sample size increases. Hence, combination rules are highly recommended if the investor takes the out-of-sample Sharpe ratio as the most important measure of the performance for a portfolio rule.

In terms of CEQ return, optimal two-fund rule performs best most often for all these data sets. Shrinkage on the covariance rule has the best performs for "size- and book-to-market-sorted portfolios" data sets. Since shrinkage on the covariance rule takes factor model into consideration, that's maybe the reason why it performs better for "size- and book-to-market-sorted portfolios" data sets. Overall, the optimal two-fund rule and shrinkage on the covariance rule are recommended if the investor takes the out-of-sample CEQ return as the most important measure of the performance for a portfolio rule.

It all depends on the investor's preference which criteria he would like to take as a measure of performance or both.

Moreover, none of the strategies discussed above can beat the *in-sample* Sharpe ratio and *in-sample* CEQ return. It means that there still exists quite much room for the improvement of the estimation approach of the mean-variance problem.

Chapter 5

Conclusions

Markowitz's mean-variance portfolio theory has been widely used in practice since it was born. However, a lot of researchers have doubts on its usefulness since it faces severely estimation error. Quite many new approaches have been proposed such as Bayesian rule, optimal two-fund rule, optimal three-fund rule, Bayes-Stein rule, combination rules, shrinkage rule and so on, on the purpose of reducing its estimation error.

In this paper, we give a review of the existing literature about the portfolio rules designed to improve the estimation of mean-variance problem. We provide an empirical study using five real data sets. We find that, the optimal combination of the naive rule with unbiased estimation and the optimal combination of the naive rule with the three-fund rule can perform well most of the time across different data sets in terms of out-of-sample Sharpe ratio. Shrinkage on the covariance rule performs well in terms of CEQ return for "size- and book-to-market-sorted portfolios" data sets and optimal two-fund rule performs well in terms of CEQ return for the rest of the data sets. Overall, in comparison with all these existing rules, the combination rules proposed by Tu and Zhou (2010) are the most recommended rules if the investor views the out-of-sample Sharpe ratio as the most important measure of portfolio performance, and the optimal two-fund rule and shrinkage on the covariance rule are the most recommended ones if the investor views the out-of-sample CEQ return as the most important measure.

However, since parameter uncertainty problem exists anyway, there's still a lot miles to go. How to make an investment decision concerns about many aspects, not only in finding out more efficient estimation rule, but also in choosing what assets to invest in and how many assets to invest in. We believe there's still a lot of work to do in this area.

References

M.W. Brandt. Portfolio choice problems. *Handbook of Financial Econometrics, forthcoming*, 2005. 1

S.J. Brown. Optimal portfolio choice under uncertainty: A bayesian approach. 1976. 2

M.K.A. Craig and L. Pastor. Asset pricing models: implications for expected returns and portfolio selection. *Review of Financial Studies*, 13(4):883–916, 2000. 3

V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22 (5):1915, 2009. ISSN 0893-9454. 2, 13, 21, 23

G.M. Frankfurter, H.E. Phillips, and J.P. Seagle. Portfolio selection: The effects of uncertain means, variances, and covariances. *Journal of Financial and Quantitative Analysis*, 6(5):1251–1262, 1971. 2

P.A. Frost and J.E. Savarino. For better performance. *The Journal of Portfolio Management*, 15(1):29–34, 1988. 3

R. Jagannathan and T. Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1684, 2003. 3

JD Jobson and B. Korkie. Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554, 1980. 2, 13

P. Jorion. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(03):279–292, 1986. ISSN 0022-1090. 2, 16

R. Kan and G. Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(03):621–656, 2007. 2, 3, 12, 13, 14, 15, 18, 25

R.W. Klein and V.S. Bawa. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231, 1976. 2

O. Ledoit and M. Wolf. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004. 3

H.M. Markowitz. Portfolio Selection, 1952. *Journal of Finance*, 7:77–91, 1952. i, 1, 4

DL Marx and RR Hocking. Moments of Certain Functions of Elements in the Inverse Wishart Matrix. In *Annual Meeting of the American Statistical Association, Chicago*, 1977. 12

R.O. Michaud. Efficient asset management: a practical guide to stock portfolio optimization and asset allocation. 1998. 13

Y. Ren and K. Shimotsu. Improvement in finite sample properties of the Hansen-Jagannathan distance test. *Journal of Empirical Finance*, 16(3):483–506, 2009. 3, 19

J. Tu and G. Zhou. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 2010. ISSN 0304-405X. i, 3, 13, 17, 18, 25, 33