Optimal estimation for economic gains: portfolio choice with parameter uncertainty

RAYMOND KAN and GUOFU ZHOU*

First draft: May 2003
This version: January 2005

*Kan is from the University of Toronto and Zhou is from Washington University in St. Louis. We thank Devraj Basu, Stephen Brown, Alex David, Phil Dybvig, Heber Farnsworth, Lorenzo Garlappi, Campbell Harvey, Philippe Jorion, Ambrus Kecskés, Marcin Kacperczyk, John Liechty, Merrill Liechty, Hong Liu, Peter Müller, Ľuboš Pástor, Tim Simin, Marcel Rindisbacher, Robert Stambaugh, Tan Wang, Chu Zhang and seminar participants at McMaster University, Penn State University, Sichuan University, Southwest University of Finance and Economics, Tsinghua University, University of California at Irvine, Washington University in St. Louis, University of Texas at Dallas, York University and participants at the 2004 Northern Finance Meetings for helpful discussions and comments. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.

Optimal estimation for economic gains: portfolio choice with parameter uncertainty

Abstract

In this paper, we advocate incorporating the economic objective function into parameter estimation by analyzing the optimal portfolio choice problem of a mean-variance investor facing parameter uncertainty. We show that, in estimating the optimal portfolio weights, the standard plug-in approach that replaces the population parameters by their sample estimates can lead to significant utility losses. While Bayesian approach accounts for parameter uncertainty by using predictive densities, its portfolio rule under diffused prior makes little improvements. In contrast, our proposed new rule provides a significant improvement of utility over the plug-in approaches. We further show that with parameter uncertainty, holding the sample tangency portfolio and the riskfree asset is never optimal. An investor can benefit by holding some other risky portfolios that help reduce the estimation risk, suggesting that the presence of estimation risk completely alters the theoretical recommendation of a two-fund portfolio.

1. Introduction

Theoretical models often assume that an economic agent making an optimal financial decision knows the true parameters of the model. But the true parameters are rarely if ever known to the decision-maker. In reality, model parameters have to be estimated, and hence the model's usefulness depends partly on how good the estimates are. This gives rise to an estimation risk in virtually all financial models. At present, the estimation risk is commonly minimized based on statistical criteria such as minimum variance and asymptotic efficiency. Can the parameters be estimated such that expected utility is maximized? This paper provides some answers.

A leading example of parameter uncertainty arises from the classic portfolio choice problem of investors. Markowitz's (1952) seminal work provides a mean-variance framework for investors to obtain the optimal portfolio as a combination of the tangency portfolio and a riskless asset (two-fund separation). Despite its limitation as a single-period model, the mean-variance framework is one of the most important benchmark models used in practice today (see, e.g., Litterman (2003)). However, the framework requires knowledge of both the mean and covariance matrix of the asset returns, which in practice are unknown and have to be estimated from the data. The standard approach, ignoring the estimation risk, simply treats the estimates as the true parameters and plugs them into the optimal portfolio formula. Using predictive distributions pioneered by Zellner and Chetty (1965), Brown (1976) shows that the plug-in method is generally outperformed by the Bayesian decision rule under a diffuse prior (Bawa, Brown, and Klein (1979) provide an extensive survey of the early work). In fact, as our analytical derivation later will show, the Bayesian decision rule is uniformly better than the plug-in method in that it always yields higher expected utility no matter what the true parameter values are. This provides both direct and indirect theoretical support for a number of recent studies, such as Kandel and Stambaugh (1996), Barberis (2000), Pástor (2000), Pástor and Stambaugh (2000), Xia (2001), Tu and Zhou (2004), and Kacperczyk (2004) that use the Bayesian predictive approach to account for parameter uncertainty. Nevertheless, it is possible to estimate the parameters optimally in such a way as to yield a decision rule that is uniformly better than the Bayesian approach (under a diffuse prior), as we will show.

While there exist alternative ways in dealing with parameter uncertainty,[1] our study focuses on the well-defined and yet unsolved problem in the classical mean-variance framework: how should a mean-variance investor optimally estimate the portfolio weights given the return data? Although the mean-variance framework is a simple model, it allows us to obtain analytical results that provide insights into solving portfolio choice problems in more general models.

In this paper, we first study, in the presence of parameter uncertainty, how to optimally estimate the portfolio weights if the investor invests only in the usual two-funds: the riskless asset and the sample tangency portfolio. Although TerHorst, DeRoon, and Werkerzx (2002) analyze a similar problem, their study assumes a known covariance matrix. In contrast, we allow for the more realistic case in which both asset mean and covariance are unknown. Under the standard assumption of normality on the asset returns, we obtain a simple closed-form formula for the optimal weights in the two-fund universe. Because both the plug-in and Bayesian approaches are two-fund portfolios, they are dominated by our proposed optimal portfolio. In particular, we find that a simplified form of the optimal portfolio, whose construction does not rely on any unknown parameters, always yields greater expected utility than both the plug-in and Bayesian approaches (under a diffuse prior) no matter what the true parameter values.

The second problem we study is whether a three-fund portfolio can improve the expected utility even further. If the true parameters are known, as assumed in theory, then the two-fund separation holds and there is no point of analyzing a three-fund portfolio. However, when the parameters are unknown, the tangency portfolio is obtained with estimation error. Intuitively, additional portfolios could be useful if they provide diversification of the estimation risk. Indeed, we show that the optimal portfolio weights can be solved analytically in a three-fund universe that consists of the riskless asset, the sample tangency portfolio, and the sample global minimum-variance portfolio. Therefore, a three-fund portfolio rule can dominate all the previous two-fund rules. This finding has powerful implications. It

[1]For examples, Garlappi, Uppal, and Wang (2004) and Lutgens (2004) study robust portfolio rule that maximizes the worst case performance for a model with parameters that fall within a particular confidence interval. Harvey, Liechty, Liechty, and Müller (2004) study optimal portfolio problem under a Bayesian setting when the returns follow a skew-normal distribution.

says that the recommendation of a theoretical result, like holding a two-fund portfolio here, can be altered completely in the presence of parameter uncertainty, to holding a three-fund (perhaps even more) portfolio.

To better estimate the expected returns, Jorion (1986) provides an interesting Bayes-Stein shrinkage estimator, and shows by simulation that the resulting portfolio rule can often generate higher expected utility in repeated samples than the Bayesian approach (under a diffuse prior). We provide a comparison of Jorion's rule with our optimal three-fund rule and show that Jorion's rule is effectively also a three-fund portfolio rule. As both Jorion's rule and our optimal three-fund rule are not analytically tractable, we use simulations to compare their performance. We find that Jorion's rule is better than the Bayesian approach under a diffuse prior, and that our optimal three-fund rule outperforms even Jorion's rule.

In another attempt to provide better estimates of expected returns, MacKinlay and Pástor (2000) makes an assumption on the covariance matrix based on the link between expected return and covariance matrix under factor based models. With the use of this special covariance matrix, they show that the resulting estimated tangency portfolio can perform much better than the traditional sample tangency portfolio that are based on the unrestricted estimates. However, the link imposes non-trivial nonlinear restrictions in estimating the expected returns, and hence the associated estimation problem has to be solved numerically. In this paper, we provide a simple analytical solution to the maximum likelihood estimation problem, which allows us to perform a large scale simulation study on the MacKinlay and Pástor portfolio rule. We show that their portfolio rule has excellent performance when the one-factor model holds reasonably well for the expected returns of the assets. However, when the one-factor model does not provide a good approximation of the expected returns, then our optimal three-fund rule significantly outperforms MacKinlay and Pástor (2000) rule.

The remainder of the paper is organized as follows. Section 2 provides the optimal decision rule when the investment universe is only the riskless asset and the sample tangency portfolio. Section 3 solves the optimal portfolio rule when the investment universe is enlarged by adding the sample global minimum-variance portfolio. Section 4 analyzes the shrinkage estimators. Section 5 analyzes the portfolio rule of MacKinlay and Pástor (2000) and Section 6 compares the performance of all the portfolio rules with parameters calibrated from real data, and Section 7 concludes.

2. Two-fund portfolio rules

In this section, we first discuss the mean-variance portfolio problem in the presence of estimation risk. Then, we analyze the classical plug-in methods for estimating the optimal portfolio weights of the mean-variance theory, review the Bayesian predictive solution and compare it with the classical plug-in estimates. Finally, we provide our optimal portfolio rule when the investor is concerned with investing in the universe of the riskless asset and the sample tangency portfolio.

2.1. The problem

Consider the standard portfolio choice problem of an investor who chooses a portfolio in the universe of N risky assets and a riskless asset. Denote by r_{ft} and r_t the rates of returns on the riskless asset and N risky assets at time t, respectively. We define excess returns as $R_t \equiv r_t - r_{ft} 1_N$, where 1_N is an N-vector of ones. The standard assumption on the probability distribution of R_t is that R_t is independent and identically distributed over time. In addition, we assume R_t follows a multivariate normal distribution with mean μ and covariance matrix Σ.

Given portfolio weights w, an $N \times 1$ vector, on the risky assets, the excess return on the portfolio at time t is $R_{pt} = w' R_t$. The investor is assumed to choose w so as to maximize the mean-variance objective function

$$U(w) = E[R_{pt}] - \frac{\tau}{2} \operatorname{Var}[R_{pt}] = w'\mu - \frac{\tau}{2} w'\Sigma w, \tag{1}$$

where τ is the coefficient of relative risk aversion. When μ and Σ are known, the solution to the investor's optimal portfolio choice problem is

$$w^* = \frac{1}{\tau} \Sigma^{-1} \mu, \tag{2}$$

and the resulting utility is

$$U(w^*) = \frac{1}{2\tau} \mu' \Sigma^{-1} \mu = \frac{\theta^2}{2\tau}, \tag{3}$$

where $\theta^2 = \mu'\Sigma^{-1}\mu$ is the squared Sharpe ratio of the *ex ante* tangency portfolio of the risky assets. Given the relative risk aversion parameter τ, this is the maximum utility that the investor can obtain when the portfolio weights w^* are computed based on the true parameters.

In practice, w^* is not computable because μ and Σ are unknown. To implement the mean-variance theory of Markowitz (1952), the optimal portfolio weights are usually chosen by a two-step procedure. Suppose an investor has T periods of observed returns data $\mathbf{\Phi}_T = \{R_1,\ R_2, \cdots, R_T\}$ and would like to form a portfolio for period $T+1$. First, the mean and covariance matrix of the asset returns are estimated based on the observed data. Second, these sample estimates are then treated as if they were the true parameters, and are simply plugged into (2) to compute the optimal portfolio weights. We call such a portfolio rule the plug-in rule. More generally, a portfolio rule is defined as a function of the historical returns data Φ_T,

$$\hat{w} = f(R_1, R_2, \cdots, R_T). \tag{4}$$

Intuitively, the investor who uses $\hat{w}$ should be worse off than the investor who knows the true parameters. How do we assess the utility loss from using a given portfolio rule? Based on standard statistical decision theory, we define the loss function of using $\hat{w}$ as

$$L(w^*, \hat{w}) = U(w^*) - U(\hat{w}). \tag{5}$$

As $\hat{w}$ is not equal to w^* in general, the utility loss is strictly positive. However, $\hat{w}$ is a function of $\mathbf{\Phi}_T$, so the loss depends on the realizations of the historical returns data. It is important for decision purposes to consider the average losses involving actions taken under the various outcomes of $\mathbf{\Phi}_T$. The expected loss function is called the risk function and it is defined as

$$\rho(w^*, \hat{w}) = E[L(w^*, \hat{w})] = U(w^*) - E[U(\hat{w})], \tag{6}$$

where the expectation is taken with respect to the true distribution of $\mathbf{\Phi}_T$. Thus, for a given μ and Σ (or a given w^*), $\rho(w^*, \hat{w})$ represents the expected loss over $\mathbf{\Phi}_T$ that is incurred in using the portfolio rule $\hat{w}$.

The risk function provides a criterion for ranking various portfolio rules and the rule that has the lowest risk is the most preferred. Brown (1976), Jorion (1986), Frost and Savarino

(1986), Stambaugh (1997), and TerHorst, DeRoon, and Werkerzx (2002) are examples of using $\rho(w^*, \hat{w})$ to evaluate portfolio rules. Instead of ranking portfolio rules using the risk function $\rho(w^*, \hat{w})$, we can equivalently rank them by their expected utility $E[U(\hat{w})]$. Note that $E[U(\hat{w})]$ is the expected utility under the true distribution of returns across repeated random samples of Φ_T. It should not be confused with the expected utility based on the true distribution or the predictive distribution. $E[U(\hat{w})]$ is in fact the average of the utility where the average is taken over all possible realizations of Φ_T. So, $E[U(\hat{w})]$ is the utility an investor can achieve on average under parameter uncertainty when he follows the portfolio rule $\hat{w}$. This is an objective criterion for evaluating two competing portfolio choice rules. In general, one portfolio rule will generate higher expected utility than another over certain parameter values of (μ, Σ), but lower over some other values. In this case, the two portfolio rules do not uniformly dominate each other, and which one is preferable depends on the actual values of μ and Σ. However, some portfolio rules are inadmissible in the sense that there exists another portfolio rule that generates higher expected utility for every possible choice of (μ, Σ). Clearly, inadmissible portfolio rules should be eliminated from consideration.

In the jargon of statistical decision theory (see, e.g., Berger (1985) and Lehmann and Casella (1998)), our paper takes a frequentist approach, not a Bayesian one, in evaluating portfolio rules. The standard Bayesian optimal portfolio is constructed to maximize the expected utility based on the predictive distribution of R_{T+1} conditional on the historical returns $\mathbf{\Phi}_T$. By design, any other rules, including the rule based on the true parameters of the model, are suboptimal in maximizing the utility based on the predictive distribution obtained from a given prior. Therefore, to compare Bayesian portfolio rules under various priors and to compare them with the rules obtained under the classical framework, the risk function is a useful objective criterion. It answers the question of how various rules perform in repeated samples as compared with the case when the true parameters are assumed to be known.

2.2. Understanding estimation risk

Under the assumption that R_t is i.i.d. normal, the sample mean and covariance matrix

$\hat{\mu}$ and $\hat{\Sigma}$, defined as

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \tag{7}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})', \tag{8}$$

are the sufficient statistics of the historical returns data $\mathbf{\Phi}_T$. Therefore, we only need to consider portfolio rules that are functions of $\hat{\mu}$ and $\hat{\Sigma}$.

The standard plug-in portfolio rule is to replace μ and Σ in (2) by $\hat{\mu}$ and $\hat{\Sigma}$ and the estimated portfolio weights using the plug-in rule are

$$\hat{w} = \frac{1}{\tau}\hat{\Sigma}^{-1}\hat{\mu}. \tag{9}$$

Statistically, $\hat{\mu}$ and $\hat{\Sigma}$ are the maximum likelihood estimators of μ and Σ, so $\hat{w}$ is also a maximum likelihood estimator of $w^* = \Sigma^{-1}\mu/\tau$. Therefore, asymptotically, $\hat{w}$ is the most efficient estimator of the unknown parameter vector w^*. In statistics, the maximum likelihood estimator is usually regarded as a very good estimator. However, as will be shown below, this estimator of w^* is not optimal in terms of maximizing the expected utility.

It is interesting to compare the standard plug-in estimator $\hat{w}$ given by (9) with the unknown but true optimal weights w^*. Under the normality assumption, it is well-known that $\hat{\mu}$ and $\hat{\Sigma}$ are independent of each other and they have the following exact distributions

$$\hat{\mu} \sim N(\mu, \Sigma/T), \tag{10}$$

$$\hat{\Sigma} \sim W_N(T-1, \Sigma)/T, \tag{11}$$

where $W_N(T-1, \Sigma)$ denotes a Wishart distribution with $T-1$ degrees of freedom and covariance matrix Σ. Since $E[\hat{\Sigma}^{-1}] = T\Sigma^{-1}/(T-N-2)$ (see, e.g., Muirhead, 1982, p. 97), we have

$$E[\hat{w}] = \frac{T}{T-N-2}w^*, \tag{12}$$

when $T > N+2$. This implies that $|\hat{w}_i| > |w_i^*|$, so investors who do not know the true parameters and estimate them by using (9) tend to take bigger positions in the risky assets than those who know the true parameters.

To understand the estimation risk from parameter uncertainties in μ and Σ, we analyze the use of $\hat{w}$ in three cases. The first case is a hypothetical one in which Σ is known and μ is estimated. By fixing a value of Σ, this allows us to understand the cost from estimating μ alone. The second case is also a hypothetical one in which μ is known and Σ is estimated, allowing us to understand the cost from estimating Σ alone. The third case is the realistic one in which neither Σ nor μ is known.

The first case is analytically the easiest among the three. When Σ is known, the portfolio rule is $\hat{w} = \Sigma^{-1}\hat{\mu}/\tau$, so the estimation error in $\hat{w}$ is only due to using $\hat{\mu}$ instead of μ. Since $\hat{\mu}'\Sigma^{-1}\hat{\mu} \sim \chi^2_N(T\mu'\Sigma^{-1}\mu)/T$, we have

$$\begin{aligned} E[U(\hat{w})\,|\,\Sigma] &= E[\hat{w}]'\mu - \frac{\tau}{2}E[\hat{w}'\Sigma\hat{w}] \\ &= \frac{1}{\tau}\mu'\Sigma^{-1}\mu - \frac{1}{2\tau}E[\hat{\mu}'\Sigma^{-1}\hat{\mu}] \\ &= \frac{1}{\tau}\mu'\Sigma^{-1}\mu - \frac{1}{2\tau}\left(\frac{N + T\mu'\Sigma^{-1}\mu}{T}\right) \\ &= \frac{\theta^2}{2\tau} - \frac{N}{2\tau T}. \end{aligned} \tag{13}$$

As a result, the expected utility loss from using $\hat{w}$ rather than w^* is

$$\rho(w^*, \hat{w}|\Sigma) \equiv U(w^*) - E[U(\hat{w})|\Sigma] = \frac{N}{2\tau T}, \tag{14}$$

which means that the investor expects to lose a utility of $N/(2\tau T)$ on average. Intuitively, as the sample size increases, $\hat{\mu}$ becomes a more accurate estimator of μ, so the loss decreases. In the extreme case where $T \to \infty$, the true parameters are learned, so the loss is zero. On the other hand, the greater the number of assets, the greater the number of elements of μ that must be estimated, the more the errors in estimating the tangency portfolio, so the greater the loss. Finally, the more risk averse the investor (the higher τ), the less he invests in the risky assets, so the smaller the impact of the estimation risk.

There are two points worth noting. First, $E[U(\hat{w})|\Sigma]$ can be negative when $\theta^2 < N/T$. Because non-participation in the risky assets yields zero utility, the negative value of $E[U(\hat{w})|\Sigma]$ suggests that sometimes the investor is better off not investing in the risky assets. Intuitively, when θ^2 is small or N/T is large, the risk in estimating the parameters outweighs the gain from investing in the risky assets. However, it should be emphasized that the non-participation result assumes the investor uses the above plug-in method to estimate the optimal portfolio weights. There may exist an alternative portfolio rule that always yields positive expected utility with participation in the risky assets. Second, the case of known Σ is similar to a continuous-time set-up, such as Xia's (2001), where the variance is known because it can be learned without error from continuous observations. However, the drift of a diffusion process depends only on the initial and ending observations and is estimated with error. Equation (14) highlights analytically the impact of the number of assets relative to the length of estimation period on the expected utility loss in discrete time.

To see how uncertainty about Σ alone affects expected utility, consider now the second case where μ is known while Σ is estimated. The optimal weights are now $\hat{w} = \hat{\Sigma}^{-1}\mu/\tau$. Let $W = \Sigma^{-\frac{1}{2}}\hat{\Sigma}\Sigma^{-\frac{1}{2}} \sim W_N(T-1, I_N)/T$. The inverse moments of W are (see, e.g., Haff (1979))

$$E[W^{-1}] = \left(\frac{T}{T-N-2}\right) I_N, \tag{15}$$

$$E[W^{-2}] = \left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\right] I_N. \tag{16}$$

Using these results and assuming $T > N+4$, the expected utility is

$$\begin{aligned} E[U(\hat{w})|\mu] &= \frac{1}{\tau}E[\mu'\hat{\Sigma}^{-1}\mu] - \frac{1}{2\tau}E[\mu'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mu] \\ &= \frac{1}{\tau}E[\mu'\Sigma^{-\frac{1}{2}}W^{-1}\Sigma^{-\frac{1}{2}}\mu] - \frac{1}{2\tau}E[\mu'\Sigma^{-\frac{1}{2}}W^{-2}\Sigma^{-\frac{1}{2}}\mu] \\ &= k_1\frac{\theta^2}{2\tau}, \end{aligned} \tag{17}$$

where

$$k_1 = \left(\frac{T}{T-N-2}\right)\left[2 - \frac{T(T-2)}{(T-N-1)(T-N-4)}\right]. \tag{18}$$

Note that $1-k_1$ is the percentage loss of the expected utility due to the estimation error of $\hat{\Sigma}$. It is straightforward to verify that $k_1 < 1$ and it is a decreasing function of N and an increasing function of T. Therefore, similarly to the earlier case where only μ was unknown, the estimation error of $\hat{\Sigma}$ (and hence the expected utility loss) also increases with the number of assets and decreases with the length of the time series.

Compared with the previous case, the investor will still sometimes avoid investing in the risky assets if he uses the portfolio rule $\hat{w} = \hat{\Sigma}^{-1}\mu/\tau$ because k_1 can be negative for N

large relative to T. However, the cost of not knowing μ (assuming Σ is known) affects the expected utility only by a fixed amount of $N/(2\tau T)$, irrespectively of the magnitude of the true parameters. In contrast, not knowing Σ (assuming μ is known) reduces the expected utility by a constant proportional amount that depends on the squared Sharpe ratio of the tangency portfolio.

Finally, consider the realistic case where both μ and Σ are unknown and have to be estimated from the data. Suppose the estimated optimal weights, $\hat{w}$, are now given by (9). Using the inverse moment properties of the Wishart distribution and the fact that $\hat{\mu}$ and $\hat{\Sigma}$ are independent, we have

$$\begin{aligned} E[U(\hat{w})] &= \frac{1}{\tau}E[\hat{\mu}'\hat{\Sigma}^{-1}\mu] - \frac{1}{2\tau}E[\hat{\mu}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}] \\ &= \frac{1}{\tau}E[\hat{\mu}'\Sigma^{-\frac{1}{2}}W^{-1}\Sigma^{-\frac{1}{2}}\mu] - \frac{1}{2\tau}E[\hat{\mu}'\Sigma^{-\frac{1}{2}}W^{-2}\Sigma^{-\frac{1}{2}}\hat{\mu}] \\ &= k_1\frac{\theta^2}{2\tau} - \frac{NT(T-2)}{2\tau(T-N-1)(T-N-2)(T-N-4)}, \end{aligned} \tag{19}$$

assuming $T > N + 4$. Hence, the expected utility loss is

$$\rho(w^*, \hat{w}) = (1-k_1)\frac{\theta^2}{2\tau} + \frac{NT(T-2)}{2\tau(T-N-1)(T-N-2)(T-N-4)}. \tag{20}$$

This formula explicitly relates the expected utility loss to N, T, τ, and θ^2. The qualitative properties are the same as before. As N or θ^2 increases, the loss increases, and as T or τ increases, the loss decreases. Note that the second term of $\rho(w^*, \hat{w})$ is always greater than $\rho(w^*, \hat{w}|\Sigma)$, so the effects of estimation errors of $\hat{\mu}$ and $\hat{\Sigma}$ on the utility loss are not additive because $\hat{w}$ is a multiplicative function of $\hat{\Sigma}^{-1}$ and $\hat{\mu}$. When $\hat{\Sigma}^{-1}$ is used instead of Σ^{-1} in constructing $\hat{w}$, the estimation error of $\hat{\mu}$ is further magnified, which results in the investor taking larger positions in the risky assets.

Note that in past studies of portfolio rules under estimation risk, the expected utility or the risk function of the plug-in portfolio rule is obtained by simulation.[2] In contrast, we provide here an analytical expression. The advantage of the analytical solution is that it allows us not only to provide insights about how to obtain better portfolio rules, but also to

[2]One exception is Brown (1978) who provides an infinite series summation formula for the expected utility in the one risky asset case.

address a number of important issues such as the impact of the error from estimating the covariance matrix of the returns on the expected utility.

There is a general perception that estimation error in expected returns is far more costly than estimation error in the covariance matrix. Indeed, many existing studies of portfolio rules under estimation risk treat the estimation error in the covariance matrix as a second-order effect and focus exclusively on the impact of the estimation error in the expected returns by taking covariance matrix as known. Some simulation studies appear to provide evidence to justify this perception. For example, Chopra and Ziemba (1993) estimate the loss of expected utility from the estimation error of the means and find that it is much higher than the loss that is due to estimation error of the covariances. However, with the aid of our analytical formula for the expected utility, we show that the general perception can be incorrect.

Table 1 reports the expected percentage loss of utility due to estimation errors in $\hat{\mu}$, in $\hat{\Sigma}$, and in both $\hat{\mu}$ and $\hat{\Sigma}$, for various values of N and T. Panel A presents the results for $\theta = 0.2$ and Panel B presents the results for $\theta = 0.4$. The expected percentage loss is not a function of the risk aversion coefficient, so the results in Table 1 are applicable for all values of τ. The first column presents the percentage loss of expected utility due to estimation error in $\hat{\mu}$ alone, i.e., $100(1 - E[U(\hat{w})|\Sigma]/U(w^*))$. The second column presents the percentage loss of expected utility due to estimation error in $\hat{\Sigma}$ alone, i.e., $100(1 - E[U(\hat{w})|\mu]/U(w^*))$. The fourth column presents the percentage loss of expected utility due to estimation errors in both $\hat{\mu}$ and $\hat{\Sigma}$, i.e., $100(1 - E[U(\hat{w})]/U(w^*))$. Since the effects of estimation errors in $\hat{\mu}$ and $\hat{\Sigma}$ are not additive, the third column reports the interactive effect of estimation errors in $\hat{\mu}$ and $\hat{\Sigma}$, whose summation with the first two columns is equal to the fourth column.

$\boxed{\text{Table 1 about here}}$

Assuming $\theta = 0.2$, Panel A shows that when N/T is small, the estimation error in $\hat{\mu}$ indeed accounts for most of the utility loss, often more than ten times the utility loss from the estimation error in $\hat{\Sigma}$. However, when N/T is large, the utility loss due to the estimation error in $\hat{\Sigma}$ is no longer negligible. More importantly, there is a very significant interactive effect between the estimation errors in $\hat{\mu}$ and $\hat{\Sigma}$. For example, when $N = 10$ and $T = 60$, the interactive effect is almost as large as that from estimating μ. Clearly, ignoring the estimation error in $\hat{\Sigma}$ will grossly underestimate the utility loss due to estimation error when N/T is large. Panel B presents the corresponding results for $\theta = 0.4$. With the increase in θ, there are two main differences in the results. First, the percentage loss of expected utility due to the estimation error in $\hat{\mu}$ alone is smaller, while the percentage loss of expected utility due to the estimation error in $\hat{\Sigma}$ alone is independent of θ. As a result, the estimation error in $\hat{\Sigma}$ is relatively more important than before. Second, the percentage loss in expected utility due to the interactive effect also goes down with the increase in θ, so as a whole, the percentage loss in expected utility due to both estimation errors in $\hat{\mu}$ and $\hat{\Sigma}$ is a decreasing function of θ^2. Other than these two differences, the general pattern is the same: when N/T is small, the estimation error in $\hat{\mu}$ is more costly than the estimation error in $\hat{\Sigma}$; however, when N/T is large, the estimation error in $\hat{\Sigma}$ becomes larger and sometimes can be more costly than the estimation error in $\hat{\mu}$. The results in Table 1 suggest that we should not ignore the estimation error in $\hat{\Sigma}$, especially when the ratio N/T is large.

2.3. Three classic plug-in rules

Besides the preceding standard plug-in estimate of the optimal portfolio weights that plugs the maximum likelihood estimator of μ and Σ into the optimal portfolio formula (2) to get the estimated portfolio rule (9), alternative estimates of Σ can be used to obtain different plug-in rules. Two other common estimators of Σ are sometimes used. It is of interest that they in fact can yield higher expected utility than using $\hat{w}$.

The second plug-in approach is to estimate Σ by using an unbiased estimator,

$$\bar{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})' = \frac{T}{T-1}\hat{\Sigma}. \tag{21}$$

Since $\bar{\Sigma}$ is slightly greater than $\hat{\Sigma}$, the resulting optimal portfolio weights invest less aggressively in the risky assets than does $\hat{w}$:

$$\bar{w} \equiv \frac{1}{\gamma}\bar{\Sigma}^{-1}\hat{\mu} = \frac{1}{\gamma}\left(\frac{T-1}{T}\right)\hat{\Sigma}^{-1}\hat{\mu} = \left(\frac{T-1}{T}\right)\hat{w}. \tag{22}$$

However, because $E[\bar{w}] = \frac{T-1}{T-N-2}w^*$, such a portfolio rule still involves taking larger positions in the risky assets relative to the true optimal portfolio. Assuming $T > N + 4$, the expected utility associated with portfolio rule $\bar{w}$ is

$$E[U(\bar{w})] = k_2 \frac{\theta^2}{2\tau} - \frac{N(T-1)^2(T-2)}{2\tau T(T-N-1)(T-N-2)(T-N-4)}, \tag{23}$$

where

$$k_2 = \left(\frac{T-1}{T-N-2}\right)\left[2 - \frac{(T-1)(T-2)}{(T-N-1)(T-N-4)}\right]. \tag{24}$$

Based on this expression, it can then be verified that $E[U(\bar{w})]$ is greater than $E[U(\hat{w})]$, so $\bar{w}$ is a better choice than $\hat{w}$.

The third plug-in approach is to estimate Σ with

$$\tilde{\Sigma} = \frac{1}{T-N-2}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})' = \frac{T}{T-N-2}\hat{\Sigma}. \tag{25}$$

Then, the plug-in estimator for the optimal portfolio weights is

$$\tilde{w} \equiv \frac{1}{\tau}\tilde{\Sigma}^{-1}\hat{\mu} = \frac{T-N-2}{T}\hat{w}. \tag{26}$$

Although $\tilde{\Sigma}$ is not an unbiased estimator of Σ, $\tilde{\Sigma}^{-1}$ is an unbiased estimator of Σ^{-1}, so $\tilde{w}$ is an unbiased estimator of w^*, i.e., $E[\tilde{w}] = w^*$. Hence, over repeated samples, the investor who uses $\tilde{w}$ will on average invest the same amount of money in the risky assets as he would invest in the true optimal portfolio. Assuming $T > N + 4$,

$$E[U(\tilde{w})] = k_3 \frac{\theta^2}{2\tau} - \frac{N(T-2)(T-N-2)}{2\tau T(T-N-1)(T-N-4)}, \tag{27}$$

where

$$k_3 = 2 - \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}. \tag{28}$$

It is straightforward to verify that $E[U(\tilde{w})]$ is greater than $E[U(\bar{w})]$, so the portfolio rule $\tilde{w}$ is better than $\bar{w}$, and hence is also better than $\hat{w}$.

In summary, we have evaluated the expected utilities of three classic plug-in estimators, $\hat{w}, \bar{w}$, and $\tilde{w}$, of the optimal portfolio weights w^*. Interestingly, it is $\tilde{w}$, the unbiased estimator of the unknown optimal portfolio weights, achieves the highest expected utility, while the maximum likelihood estimate yields the lowest.

2.4. Bayesian solution

While the plug-in method ignores the estimation risk, the Bayesian approach based on the predictive distributions pioneered by Zellner and Chetty (1965) provides a general framework that integrates estimation risk into the analysis. Under the classical framework, the utility $U(\hat{w})$ is evaluated conditional on the true parameter w^* being equal to $\hat{w}$ and uncertainty about the goodness of this conditioning is completely ignored. In contrast, w^* is regarded as a random vector in the Bayesian framework. Given a standard diffuse (non-informative) prior on the distribution of μ and Σ, uncertainty about the parameters is summarized by the posterior distribution of the parameters given the data. Integrating out the parameters over this distribution gives the so-called predictive distribution for future asset returns. The optimal portfolio is obtained by maximizing the expected utility under the predictive distribution, i.e.,

$$\begin{aligned}\hat{w}^{\text{Bayes}} &= \operatorname{argmax}_w \int_{R_{T+1}} U(w)p(R_{T+1}|\mathbf{\Phi}_T)\,\mathrm{d}R_{T+1} \\ &= \operatorname{argmax}_w \int_{R_{T+1}}\int_{\mu}\int_{\Sigma} U(w)p(R_{T+1},\mu,\Sigma|\mathbf{\Phi}_T)\,\mathrm{d}\mu\mathrm{d}\Sigma\mathrm{d}R_{T+1}, \qquad (29)\end{aligned}$$

where $U(w)$ is the utility of holding a portfolio w at time $T+1$, $p(R_{T+1}|\mathbf{\Phi}_T)$ is the predictive density, and

$$p(R_{T+1},\mu,\Sigma|\mathbf{\Phi}_T) = p(R_{T+1}|\mu,\Sigma,\mathbf{\Phi}_T)p(\mu,\Sigma|\mathbf{\Phi}_T), \qquad (30)$$

where $p(\mu,\Sigma|\mathbf{\Phi}_T)$ is the posterior density of μ and Σ. Thus the Bayesian approach maximizes the average expected utility over the distribution of the parameters.

Unless otherwise stated, all references to the Bayesian portfolio rule in the rest of the paper assume a diffuse prior. Brown (1976), Klein and Bawa (1976), and Stambaugh (1997) show under the assumption of a standard diffuse prior on μ and Σ,

$$p_0(\mu,\Sigma) \propto |\Sigma|^{-\frac{N+1}{2}}, \qquad (31)$$

the predictive distribution of the asset returns follows a multivariate t-distribution. It can be shown that the Bayesian solution to the optimal portfolio weights has the same formula as for w^* except that now the parameters have to be replaced by their predictive moments, yielding

$$\hat{w}^{\text{Bayes}} = \frac{1}{\tau}\left(\frac{T-N-2}{T+1}\right)\hat{\Sigma}^{-1}\hat{\mu}. \qquad (32)$$

The Bayesian weights differ from the unbiased estimator $\tilde{w}$ only by a factor of $T/(T+1)$. In terms of optimal portfolio, the Bayesian solution also suggests a two-fund separation result: investing only in the riskless asset and the sample tangency portfolio. However, since

$$E[\hat{w}^{\text{Bayes}}] = \left(\frac{T}{T+1}\right) w^*, \tag{33}$$

the Bayesian solution is more conservative than the case where the true parameters are known because it suggests taking smaller positions in the risky assets. Intuitively, the Bayesian approach recognizes the estimation risk explicitly, and hence the risky assets become riskier, while the riskless rate is known for sure. So, all else equal, the riskless asset becomes more attractive and hence the Bayesian investor invests more in it.

Will the Bayesian portfolio rule outperform the classical plug-in methods in repeated samples? Intuitively, this should be the case because the Bayesian portfolio rule incorporates parameter uncertainty into decision-making while the plug-in methods simply ignore it. Simulations by Brown (1976) and Stambaugh (1997) suggest that the Bayesian portfolio rule often outperforms the plug-in method. We provide an analytical proof here to show that the Bayesian portfolio rule always dominates the classical plug-in method. Using the same technique for evaluating $E[U(\hat{w})]$,

$$E[U(\hat{w}^{\text{Bayes}})] \;=\; k_4\frac{\theta^2}{2\tau} - \frac{NT(T-2)(T-N-2)}{2\tau(T+1)^2(T-N-1)(T-N-4)}, \tag{34}$$

where $T > N+4$ and

$$k_4 = \left(\frac{T}{T+1}\right)\left[2 - \frac{T(T-2)(T-N-2)}{(T+1)(T-N-1)(T-N-4)}\right]. \tag{35}$$

Therefore,

$$E[U(\hat{w}^{\text{Bayes}})] - E[U(\tilde{w})] = (k_4 - k_3)\frac{\theta^2}{2\tau} + \frac{N(T-2)(T-N-2)(2T+1)}{2\tau T(T+1)^2(T-N-1)(T-N-4)}. \tag{36}$$

It is easy to see that whenever $T > N+4$,

$$k_4 - k_3 = \frac{(T^2+6T-4) + N[2T(T-N) - 3T - 2(N+4)]}{(T+1)^2(T-N-1)(T-N-4)} > 0 \tag{37}$$

because $2T(T-N) > 8T > 3T + 2(N+4)$. Hence, the explicit expressions for $E[U(\hat{w}^{\text{Bayes}})]$ and $E[U(\tilde{w})]$ show analytically that the Bayesian portfolio rule always strictly outperforms the earlier classical plug-in methods by yielding higher expected utility in repeated samples, regardless of the values of the true parameters. Therefore, the three classical plug-in portfolio rules are inadmissible and they should be replaced by better portfolio rules.

The uniform dominance result suggests that investors are better off using the Bayesian portfolio rule than the classical plug-in rules. It turns out that the Bayesian portfolio rule is still inadmissible because there exists a portfolio rule that uniformly dominates the Bayesian portfolio rule. As shown below, we will obtain such a two-fund rule in closed-form based on our techniques used earlier for evaluating expected utilities.

In a Bayesian framework, informative priors other than the diffuse one may be used. Although there may be countless ways of doing so in principle, it is not an easy matter to construct useful informative priors in practice. For examples, Pástor (2000) and Pástor and Stambaugh (2000) provide interesting priors that incorporate certain beliefs on the usefulness of the CAPM and study their impacts on asset allocation decisions. While understanding how predictive moments are impacted by informative priors is interesting, it is difficult to obtain analytical solution of the risk function for such portfolio rules. To limit the scope of this paper, we will in what follows focus on the diffuse prior only, leaving the study of informative Bayesian portfolio rules for future research.

2.5. Optimal two-fund rule

Theoretically, the estimator of w^* can be any function of the sufficient statistics $\hat{\mu}$ and $\hat{\Sigma}$, i.e.,

$$\hat{w} = f(\hat{\mu}, \hat{\Sigma}). \tag{38}$$

The economic question of interest to the investor is to find a such a function $f(\hat{\mu}, \hat{\Sigma})$ so that the expected utility is maximized. This function can potentially be a very complex nonlinear function of $\hat{\mu}$ and $\hat{\Sigma}$, and there can potentially be infinitely many ways to construct it. However, it is not an easy matter to determine the optimal $f(\hat{\mu}, \hat{\Sigma})$, so we first limit our attention to a special class of portfolio rules that hold just the riskless asset and the sample tangency portfolio.

Both the earlier plug-in and the Bayesian rules suggest holding the riskless asset and the sample tangency portfolio. However, in terms of maximizing expected utility, the weights on the sample tangency portfolio chosen by both the plug-in rule and the Bayesian rule are not necessarily optimal. Indeed, consider the class of two-fund portfolio rules which have weights

$$\hat{w} = \frac{c}{\tau}\hat{\Sigma}^{-1}\hat{\mu}, \tag{39}$$

where c is a constant scalar. For example, the first plug-in and the Bayesian rules specify $c_1 = 1$ and $c_2 = (T-N-2)/(T+1)$, respectively. Using a similar derivation as before, the expected utility of this class of portfolio rules is

$$\begin{aligned} E[U(c\hat{\Sigma}^{-1}\hat{\mu}/\tau)] &= \frac{c\theta^2}{\tau}\left(\frac{T}{T-N-2}\right) \\ &\quad - \frac{c^2}{2\tau}\left(\theta^2 + \frac{N}{T}\right)\left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\right], \end{aligned} \tag{40}$$

assuming $T > N+4$. Differentiating with respect to c, the optimal c is

$$c^* = \left[\frac{(T-N-1)(T-N-4)}{T(T-2)}\right]\left(\frac{\theta^2}{\theta^2 + \frac{N}{T}}\right), \tag{41}$$

which is a product of two terms. If Σ is known, then c^* will consist only of the second term, which thus accounts for the estimation error in $\hat{\mu}$. Similarly, the first term of c^* accounts for the estimation error in $\hat{\Sigma}$. Clearly, both terms are less than one. The value of the second term depends on the relative magnitude of θ^2 and N/T, while the value of the first term depends on the relative magnitude of N and T, but not θ^2.

Expected utility under the optimal choice of $\hat{w}^* = c^*\hat{\Sigma}^{-1}\hat{\mu}/\tau$ is

$$E[U(\hat{w}^*)] = \frac{\theta^2}{2\tau}\left[\frac{(T-N-1)(T-N-4)}{(T-2)(T-N-2)}\right]\left(\frac{\theta^2}{\theta^2 + \frac{N}{T}}\right), \tag{42}$$

which is, of course, higher than the expected utility under both the classical plug-in and the Bayesian rules. Compared to the case of no uncertainty,

$$\frac{E[U(\hat{w}^*)]}{U(w^*)} = \left[\frac{(T-N-1)(T-N-4)}{(T-2)(T-N-2)}\right]\left(\frac{\theta^2}{\theta^2 + \frac{N}{T}}\right) < 1, \tag{43}$$

which is a decreasing function of N and an increasing function of T and θ^2. As a result, the percentage loss of expected utility increases with the number of assets but decreases with both the length of the time series and the squared Sharpe ratio of the tangency portfolio.

Although c^* is optimal, there does not exist a feasible portfolio rule using c^* since θ is unknown in practice. Nevertheless, c^* provides important insights into the optimal decision. In particular, it can yield a simple decision rule that dominates the Bayesian rule. Consider the following rule, which is optimal when $\theta^2 \to \infty$:

$$\hat{w}_* = \frac{c_3}{\tau}\hat{\Sigma}^{-1}\hat{\mu}, \qquad c_3 = \frac{(T-N-1)(T-N-4)}{T(T-2)}. \tag{44}$$

This rule suggests investing $\hat{w}_*$ in the risky assets and $1 - 1_N'\hat{w}_*$ in the riskless asset. Like the Bayesian rule, it is parameter independent (i.e., it only depends on N and T but not on μ and Σ). However, it dominates the Bayesian rule not only when θ approaches infinity, but also for all possible parameter values. The reason is that $f(c) \equiv E[U(c\hat{\Sigma}^{-1}\hat{\mu}/\tau)]$ in (40) is a quadratic function of c, so the expected utility is a decreasing function of c for $c \geq c^*$. Therefore, to show dominance, it suffices to show that $c_2 > c_3 > c^*$. Indeed, when $T > N+4$,

$$c_2 = \frac{T-N-2}{T+1} > \frac{T-N-4}{T} > \left(\frac{T-N-4}{T}\right)\left(\frac{T-N-1}{T-2}\right) = c_3, \tag{45}$$

and $c_3 > c^*$ obviously. Thus, regardless of the value of θ^2, the expected utility is always greater for $\hat{w}_*$ than that under the Bayesian portfolio rule. The expected utility of w^* can be computed explicitly by (40) with $c = c_3$.

The portfolio rule $\hat{w}_*$ can be viewed as a plug-in estimator that estimates Σ by using $\hat{\Sigma}_* \equiv \hat{\Sigma}/c_3$. Incidentally, Haff (1979, Theorem 7) shows that when estimating Σ^{-1}, $\hat{\Sigma}_*^{-1}$ dominates all the estimators that are of the form $c\hat{\Sigma}^{-1}$, when the loss function is defined as $\mathrm{tr}(c\hat{\Sigma}^{-1}\Sigma - I_N)^2$. Although effectively the same estimator of Σ^{-1} is obtained here, our motivation and the loss function are quite different from Haff's.

The optimal scalar c^* provides an additional insight on improving upon using c_3. Without information about the value of θ^2, c_3 represents the best choice of c that maximizes the expected utility. However, if *a priori* $\theta^2 \leq \bar{\theta}^2$, but the exact value of θ^2 is not known, then

$$\bar{c} = c_3\left(\frac{\bar{\theta}^2}{\bar{\theta}^2 + \frac{N}{T}}\right) \tag{46}$$

is a better choice of c because the expected utility $f(c)$ is a decreasing function of c when $c \geq c^*$. Since $c^* < \bar{c} < c_3$, it follows that $f(c^*) > f(\bar{c}) > f(c_3)$. If at the monthly frequency it seems reasonable to believe that $\theta^2 \leq 1$, then $\bar{c} = c_3T/(T+N)$ gives a higher expected utility.

However, this choice requires bounding the Sharpe ratio, so it is not parameter independent, and its performance depends on how the true Sharpe ratio deviates from $\bar{\theta}$. Hence, to avoid ambiguous choices of $\bar{\theta}$, this type of rule will not be studied in the rest of the paper.

To illustrate the magnitude of the expected utility loss due to estimation risk for various two-fund rules, we present two numerical examples. In the first one, we assume an investor with a risk aversion coefficient of $\tau = 3$ chooses a portfolio out of $N = 10$ risky assets and a riskless asset. Assume further that the Sharpe ratio of the *ex ante* tangency portfolio is $\theta = 0.2$. Figure 1 plots the expected utilities (in percentage monthly return) of the investor under various two-fund rules for different lengths of estimation window. If the investor knows μ and Σ, he will hold w^* for the risky assets to achieve a certainty equivalent utility of 0.667%/month (dashed line). If the investor just knows θ, then he will hold the *ex post* tangency portfolio using the optimal weight $\hat{w}^* = c^*\hat{\Sigma}^{-1}\hat{\mu}/\tau$ and his utility is indicated by the solid line. In comparison with using w^*, there is some expected utility loss from using $\hat{w}^*$. Nevertheless, the expected utility is still positive, implying that it makes the investor better off than holding the riskless asset alone. However, this is no longer the case if the investor does not know θ, and if the investor holds the portfolio $\hat{w}_*$ that does not depend on the value of θ. Although this rule is better than the three classic plug-in rules and the Bayesian rule, it results in significant losses in expected utility as indicated by the dotted line, especially when T is small. In fact, an estimation window of at least $T = 250$ months is needed before such a portfolio rule dominates the riskless asset. Finally, the dashed-dotted line shows the expected utility for the standard plug-in portfolio rule $\hat{w} = \hat{\Sigma}^{-1}\hat{\mu}/\tau$. In this case, an estimation window of at least $T = 296$ months is needed before this rule outperforms the riskless asset.

Figure 1 about here

In the second example, we make the same assumptions as in the first example except that there are now $N = 25$ risky assets, and the Sharpe ratio is assumed to be 0.3 instead of 0.2 due to the increase in the number of risky assets. Figure 2 plots the expected utilities of the investor under the four two-fund rules. For w^* and $\hat{w}^*$, the increase in the Sharpe ratio results in higher expected utilities for the investor. However, this is not necessarily true when there is parameter uncertainty and when $\hat{w}_*$ and $\hat{w}$ are used as the estimated portfolio weights. Indeed, by comparing the numbers in Figures 1 and 2, we can see that increasing the number of assets can in fact lead to a decrease in the expected utility, especially when T is small.

Figure 2 about here

These two examples illustrate that while $\hat{w}_*$ improves over $\hat{w}$, it is still a mediocre portfolio rule because it delivers negative expected utility when the parameters are estimated with fewer than 20 years of monthly data. While $\hat{w}^*$ seems a much better rule, it is not feasible as it depends on the unknown parameter θ^2. Therefore, it is important to find a good estimate of θ^2 that will allow the implementation of an approximate optimal two-fund rule. A natural estimator of θ^2 is its sample counterpart,

$$\hat{\theta}^2 = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}. \tag{47}$$

However, $\hat{\theta}^2$ can be a heavily biased estimator of θ^2 when T is small. In the Appendix, we show that $\hat{\theta}^2$ has the following distribution:

$$\hat{\theta}^2 \sim \left(\frac{N}{T-N}\right) F_{N,T-N}(T\theta^2), \tag{48}$$

where $F_{N,T-N}(T\theta^2)$ is a noncentral F distribution with N and $T-N$ degrees of freedom, and a noncentrality parameter of $T\theta^2$. Because of this, the unbiased estimator of θ^2 is

$$\hat{\theta}^2_u = \frac{(T-N-2)\hat{\theta}^2 - N}{T}. \tag{49}$$

However, this estimator can take negative value so it is also undesirable as an estimator of θ^2.

Note that the problem of estimating θ^2 using $\hat{\theta}^2$ is equivalent to the problem of estimating the noncentrality parameter of a noncentral F-distribution using a single observation. This problem has been studied by a number of researchers in statistics. For example, Rukhin (1993) and Kubokawa, Robert, and Saleh (1993) both propose estimators that are superior to the unbiased estimator of θ under the quadratic loss function, whereas Fourdrinier, Philippe, and Robert (2000) and Chen and Kan (2004) provide superior estimators under Stein's type loss function. For our application, we use an adjusted estimator of θ^2 that is due to Kubokawa, Robert, and Saleh (1993). After some simplification as given in the Appendix, this estimator can be written as

$$\hat{\theta}_a^2 = \frac{(T-N-2)\hat{\theta}^2 - N}{T} + \frac{2(\hat{\theta}^2)^{\frac{N}{2}}(1+\hat{\theta}^2)^{-\frac{T-2}{2}}}{TB_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2,(T-N)/2)}, \tag{50}$$

where

$$B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\mathrm{d}y \tag{51}$$

is the incomplete beta function. The first part of this estimator is the unbiased estimator of θ^2 and the second part of the estimator is the adjustment to improve the unbiased estimator when it is too small.

Figure 3 plots $\hat{\theta}_a^2$ and $\hat{\theta}_u^2$ as a function of $\hat{\theta}^2$ for $N = 10$ and $T = 100$. It can be seen that $\hat{\theta}_a^2$ is an increasing and convex function of $\hat{\theta}^2$. When $\hat{\theta}^2$ is equal to zero, $\hat{\theta}_a^2 = 0$. As $\hat{\theta}^2$ gets larger, it becomes more like a linear function of $\hat{\theta}^2$ and behaves almost like the unbiased estimator $\hat{\theta}_u^2$. To understand the intuition of why $\hat{\theta}_a^2$ is a better estimator of θ^2, notice that $(T-N-2)\hat{\theta}^2$ behaves almost like a $\chi^2_N(T\theta^2)$ random variable, and it has an expected value of $T\theta^2 + N$. When $(T-N-2)\hat{\theta}^2$ is large, it is more likely that part of its large value is due to the upward bias of N, so we effectively use the unbiased estimator $\hat{\theta}_u^2$. However, when $(T-N-2)\hat{\theta}^2$ is small, we should not subtract N from $(T-N-2)\hat{\theta}^2$ because a small $(T-N-2)\hat{\theta}^2$ (say less than N) indicates that $(T-N-2)\hat{\theta}^2$ is less than its expected value of N. Therefore, our estimator $\hat{\theta}_a^2$ should be higher than $\hat{\theta}_u^2$ when $\hat{\theta}_u^2$ is small or negative, causing $\hat{\theta}_a^2$ to be a nonlinear function of $\hat{\theta}^2$.

Figure 3 about here

With this estimator of θ^2, the optimal c^* can be estimated using

$$\hat{c}^* = c_3\left(\frac{\hat{\theta}_a^2}{\hat{\theta}_a^2 + \frac{N}{T}}\right), \tag{52}$$

and the associated feasible two-fund optimal portfolio weights are

$$\hat{w}^{\mathrm{II}} = \frac{1}{\tau}\hat{c}^*\hat{\Sigma}^{-1}\hat{\mu}. \tag{53}$$

In comparison with c_3, $\hat{c}^*$ is random and data-dependent, so the expected utility of using $\hat{w}^{\rm II}$ is intractable. Nevertheless, $\hat{w}^{\rm II}$ is expected to outperform $\hat{w}_*$ by design. This must be the case when the estimate of θ^2 is accurate enough. The simulation results reported in Section 6 confirms that this is indeed the case.

Recently, Garlappi, Uppal, and Wang (2004, Proposition 3) propose an interesting two-fund rule that is optimal for an investor who exhibits uncertainty aversion. Their approach is very distinctive in that they directly incorporate parameter uncertainty in the utility function. Their two-fund portfolio rule is given by[3]

$$\hat{w}_{ua} = \frac{c_{ua}}{\tau}\bar{\Sigma}^{-1}\hat{\mu}, \tag{54}$$

where

$$c_{ua} = \begin{cases} 1 - (\varepsilon/\hat{\theta}^2)^{\frac{1}{2}} & \text{if } \hat{\theta}^2 > \varepsilon, \\ 0 & \text{if } \hat{\theta}^2 \leq \varepsilon, \end{cases} \tag{55}$$

with $\varepsilon = NF^{-1}_{N,T-N}(p)/(T-N)$, and $F^{-1}_{N,T-N}(\cdot)$ is the inverse cumulative distribution function of a central F-distribution with N and $T-N$ degrees of freedom and p is a probability. Under the null hypothesis that $\theta = 0$, $\hat{\theta}^2 \sim NF_{N,T-N}/(T-N)$, so using the above portfolio rule, an investor will choose not to invest in the risky assets with probability p if the Sharpe ratio is actually zero. Therefore, p is used to indicate the investor's aversion to uncertainty and an investor with high aversion to uncertainty will choose a higher p. In this paper, we use $p = 0.99$, which is a value that provides good performance based on the empirical results in Garlappi, Uppal, and Wang (2004).[4] This portfolio rule makes intuitive sense. It suggests that when there is uncertainty about θ^2, an investor needs to have enough confidence that $\theta \neq 0$ (i.e., a large enough $\hat{\theta}^2$) before he is willing to invest in the sample tangency portfolio. Otherwise, he will choose to invest in just the riskless asset.

In terms of maximizing the mean-variance expected utility, the uncertainty aversion two-fund rule cannot outperform our theoretical optimal two-fund rule. However, since our optimal two-fund rule has to be estimated, so it is not entirely clear that the uncertainty

[3]Although Garlappi, Uppal, and Wang (2004) do not explicitly state which estimator of Σ they use, it is clear from their context that they use the unbiased estimator of Σ. See also Lutgens (2004, Theorem 1) for a similar portfolio rule.

[4]We also try $p = 0.95$ and the results are qualitatively the same.

aversion two-fund rule is always inferior to our estimated optimal two-fund rule. This issue will be addressed by using simulations in Section 6.

3. Three-fund separation: investing on the ex post frontier

Theoretically, if an investor knows the true parameters, he should only invest in the riskless asset and the tangency portfolio, but the parameters are unknown in practice. A natural approach guided by the standard mean-variance theory is to invest in two funds: the riskless asset and the sample tangency portfolio. This problem was analyzed in detail in the previous section.

However, investing only in the two funds generates a loss in expected utility, as shown below. Intuitively, if there is parameter uncertainty, use of another risky portfolio can help to diversify the estimation risk of the sample tangency portfolio. This is because while both portfolios have estimation errors, their estimation errors are not perfectly correlated. To the extent that the risk-return trade-offs are not constant across the two portfolios, expected utility is higher when the two portfolios are optimally combined. The relative weights in the two portfolios depend on the estimation errors of the two portfolios, their correlation, and their risk-return trade-offs. In addition to the sample tangency portfolio, which risky portfolio should be used? We choose to use the sample global minimum-variance portfolio for two reasons. First, the weights of the global minimum-variance portfolio only depend on $\hat{\Sigma}$ but not $\hat{\mu}$, so the weights can be estimated with higher accuracy. Second, if we limit ourselves to consider just portfolios on the *ex post* minimum-variance frontier, then the sample global minimum-variance portfolio is a natural candidate. Similar to the *ex ante* frontier portfolios, every sample frontier portfolio is a linear combination of two distinct sample frontier portfolios. Hence, it suffices to consider only the sample tangency and global minimum-variance portfolios.[5]

Consider a portfolio rule of the form

$$\hat{w} = \hat{w}(c, d) = \frac{1}{\tau}(c\hat{\Sigma}^{-1}\hat{\mu} + d\hat{\Sigma}^{-1}1_N), \tag{56}$$

[5]It should be emphasized that our method can also be used to analyze other combinations of risky portfolios, and it is possible that other choices of risky portfolios can lead to even higher expected utility than the one that we propose.

where c and d are constants to be chosen optimally. Since the weights of the sample tangency and global minimum-variance portfolios are proportional to $\hat{\Sigma}^{-1}\hat{\mu}$ and $\hat{\Sigma}^{-1}1_N$, respectively, the portfolio rule $\hat{w}(c,d)$ invests in these two sample frontier portfolios and the riskless asset.

Under this class of portfolio rules, the expected utility is

$$
\begin{aligned}
E[U(\hat{w}(c,d))] &= E[\hat{w}(c,d)]'\mu - \frac{\tau}{2}E[\hat{w}(c,d)'\Sigma\hat{w}(c,d)] \\
&= \left(\frac{T}{T-N-2}\right)\frac{1}{2\tau}\left[2(c\mu'\Sigma^{-1}\mu + d\mu'\Sigma^{-1}1_N) - \frac{T(T-2)}{(T-N-1)(T-N-4)}\right. \\
&\quad \times \left.\left(\left(\mu'\Sigma^{-1}\mu + \frac{N}{T}\right)c^2 + 2(\mu'\Sigma^{-1}1_N)cd + (1_N'\Sigma^{-1}1_N)d^2\right)\right], \qquad (57)
\end{aligned}
$$

where $T > N+4$. Differentiating with respect to c and d, the optimal choice of c and d that maximizes the expected utility is

$$
c^{**} = c_3\left(\frac{\psi^2}{\psi^2 + \frac{N}{T}}\right), \qquad (58)
$$

$$
d^{**} = c_3\left(\frac{\frac{N}{T}}{\psi^2 + \frac{N}{T}}\right)\mu_g, \qquad (59)
$$

where

$$
\psi^2 = \mu'\Sigma^{-1}\mu - \frac{(\mu'\Sigma^{-1}1_N)^2}{1_N'\Sigma^{-1}1_N} = (\mu - \mu_g 1_N)'\Sigma^{-1}(\mu - \mu_g 1_N) \qquad (60)
$$

is the squared slope of the asymptote to the *ex ante* minimum-variance frontier, and $\mu_g = (1_N'\Sigma^{-1}\mu)/(1_N'\Sigma^{-1}1_N)$ is the expected excess return of the *ex ante* global minimum-variance portfolio. Therefore, the optimal portfolio weights are

$$
\hat{w}^{**} = \frac{c_3}{\tau}\left[\left(\frac{\psi^2}{\psi^2 + \frac{N}{T}}\right)\hat{\Sigma}^{-1}\hat{\mu} + \left(\frac{\frac{N}{T}}{\psi^2 + \frac{N}{T}}\right)\mu_g\hat{\Sigma}^{-1}1_N\right]. \qquad (61)
$$

Since $d^{**} \neq 0$ unless $\mu_g = 0$, this portfolio rule suggests the use of the sample global minimum-variance portfolio no matter what the true parameters μ and Σ are (except when $\mu_g = 0$). The higher N/T, the greater the investment required in the global minimum-variance portfolio. Intuitively, the greater the number of assets, the greater the difficulty in estimating the weights of the tangency portfolio, and hence the greater the reliance on the optimal portfolio that assumes constant means across assets. This was first pointed out by Jobson, Korkie, and Ratti (1979), who suggest investing only in the sample global minimum-variance portfolio. Since $c^{**} > 0$ whenever $T > N+4$, so investing in just the sample global minimum-variance portfolio is a clearly suboptimal policy under our analysis. The optimal amount to invest in the sample tangency portfolio depends on the relative magnitude of ψ^2 and N/T. The greater the slope of the asymptote to the minimum-variance frontier, the more the investor invests in the sample tangency portfolio because it is potentially more rewarding than investing in the sample global minimum-variance portfolio.

Under the optimal weights $\hat{w}^{**}$, the expected utility is

$$E[U(\hat{w}^{**})] = \frac{\theta^2}{2\tau}\frac{(T-N-1)(T-N-4)}{(T-2)(T-N-2)}\left[1 - \frac{\frac{N}{T}}{\theta^2 + \left(\frac{\theta^2}{\psi^2}\right)\left(\frac{N}{T}\right)}\right], \tag{62}$$

where $T > N+4$. In the presence of parameter uncertainty in both μ and Σ, this is the highest expected utility obtained so far. However, this level of expected utility is unattainable because ψ^2 and μ_g are not known and have to be estimated to implement the above strategy.

To estimate μ_g, we use its sample counterpart

$$\hat{\mu}_g = \frac{\hat{\mu}'\hat{\Sigma}^{-1}1_N}{1_N'\hat{\Sigma}^{-1}1_N}. \tag{63}$$

For ψ^2, we can also use its sample counterpart

$$\hat{\psi}^2 = (\hat{\mu} - \hat{\mu}_g 1_N)'\hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N). \tag{64}$$

In the Appendix, we show that

$$\frac{(T-N+1)\hat{\psi}^2}{N-1} \sim F_{N-1,T-N+1}(T\psi^2), \tag{65}$$

so $\hat{\psi}^2$ shares the same problem with $\hat{\theta}^2$ as being a heavily biased estimator when T is small. Therefore, similarly to $\hat{\theta}^2_a$, we use

$$\hat{\psi}^2_a = \frac{(T-N-1)\hat{\psi}^2 - (N-1)}{T} + \frac{2(\hat{\psi}^2)^{\frac{N-1}{2}}(1+\hat{\psi}^2)^{-\frac{T-2}{2}}}{TB_{\hat{\psi}^2/(1+\hat{\psi}^2)}((N-1)/2, (T-N+1)/2)} \tag{66}$$

to estimate ψ^2. The associated three-fund optimal portfolio weights are then given by

$$\hat{w}^{\mathrm{III}} = \frac{c_3}{\tau}\left[\left(\frac{\hat{\psi}^2_a}{\hat{\psi}^2_a + \frac{N}{T}}\right)\hat{\Sigma}^{-1}\hat{\mu} + \left(\frac{\frac{N}{T}}{\hat{\psi}^2_a + \frac{N}{T}}\right)\hat{\mu}_g\hat{\Sigma}^{-1}1_N\right]. \tag{67}$$

Like $\hat{w}^{\rm II}$, $\hat{w}^{\rm III}$ is random since the weights of the sample tangency and global minimum-variance portfolios depend on the realization of $\hat{\mu}_g$ and $\hat{\psi}^2$. However, since $\hat{w}^{**}$ dominates $\hat{w}^*$, $\hat{w}^{\rm III}$ is expected to dominate $\hat{w}^{\rm II}$. But analytical comparison is difficult because $\hat{\psi}_a^2$ and $\hat{\mu}_g$ are dependent on random data samples. Section 6 provides simulations on $E[U(\hat{w}^{\rm III})]$, and the results indeed show that the expected utilities under this three-fund rule $\hat{w}^{\rm III}$ tend to be higher than those under the two-fund rule $\hat{w}^{\rm II}$.

4. Shrinkage estimators

Since Stein's (1956) seminal work, it is known that when $N > 2$, the sample mean $\hat{\mu}$ is not the best estimator of the population mean μ in terms of the quadratic loss function. This is because Stein's estimator or a shrinkage estimator that shrinks the sample mean appropriately to a constant can have a smaller expected quadratic loss than the sample mean. As a result of Stein's surprising finding, there is a large literature on various shrinkage estimators and the related Bayesian estimators (of which Berger (1985) provides an excellent survey from a Bayesian decision perspective).

In the finance literature, Jorion (1986, 1991) develops an estimator of μ that is motivated by both a shrinkage consideration and a Bayesian analysis (under a suitable informed prior). Jorion's Bayes-Stein estimator of μ is

$$\hat{\mu}^{\rm BS} = (1-v)\hat{\mu} + v\hat{\mu}_g 1_N, \tag{68}$$

where v is the weight given to the shrinkage target $\hat{\mu}_g$, and $1-v$ is the weight on the sample mean. The target used by Jorion (1986) is the average excess return of the sample global minimum-variance portfolio,

$$\hat{\mu}_g = \frac{1_N'\tilde{\Sigma}^{-1}\hat{\mu}}{1_N'\tilde{\Sigma}^{-1}1_N} = \frac{1_N'\hat{\Sigma}^{-1}\hat{\mu}}{1_N'\hat{\Sigma}^{-1}1_N}, \tag{69}$$

and the weight is

$$v = \frac{N+2}{(N+2) + T(\hat{\mu} - \hat{\mu}_g 1_N)'\tilde{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)}, \tag{70}$$

where $\tilde{\Sigma}$ is defined as in (25). From a shrinkage point of view, combining $\hat{\mu}^{\rm BS}$ with $\tilde{\Sigma}$ gives an estimator of the optimal portfolio weights.

Jorion's method, however, is in fact a Bayesian estimation of the optimal portfolio weights, because he replaces Σ in the classic optimal weights formula (equation (2)) with the predictive variance of the asset returns,

$$\mathrm{Var}[R_{T+1}|\Sigma,\lambda,\mathbf{\Phi}_T] = \left(1+\frac{1}{T+\lambda}\right)\Sigma + \frac{\lambda}{T(T+1+\lambda)}\frac{1_N 1_N'}{1_N'\Sigma^{-1}1_N}, \tag{71}$$

where λ is a precision parameter in the informative prior which leads to the shrinkage estimator $\hat{\mu}^{\mathrm{BS}}$

$$p_0(\mu|\Sigma,\mu_g,\lambda) \propto \exp\left[-\frac{1}{2}(\mu - 1_N\mu_g)'(\lambda\Sigma^{-1})(\mu - 1_N\mu_g)\right]. \tag{72}$$

Theoretically, $\mathrm{Var}[R_{T+1}|\mathbf{\Phi}_T]$ should be used after integrating out both Σ and λ from their posterior distributions, but this integration is a formidable task, so the natural approach is to simply use $\mathrm{Var}[R_{T+1}|\Sigma,\lambda,\mathbf{\Phi}_T]$ instead. Although Σ and λ are unknown in $\mathrm{Var}[R_{T+1}|\Sigma,\lambda,\mathbf{\Phi}_T]$, they can be replaced by their sample estimates. In this way, Jorion's empirical Bayes-Stein estimator of the optimal portfolio weights is

$$w^{\mathrm{BS}} = \frac{1}{\tau}(\hat{\Sigma}^{\mathrm{BS}})^{-1}\hat{\mu}^{\mathrm{BS}}, \tag{73}$$

where

$$\hat{\Sigma}^{\mathrm{BS}} = \left(1+\frac{1}{T+\hat{\lambda}}\right)\tilde{\Sigma} + \frac{\hat{\lambda}}{T(T+1+\hat{\lambda})}\frac{1_N 1_N'}{1_N'\tilde{\Sigma}^{-1}1_N} \tag{74}$$

and $\hat{\lambda} = (N+2)/[(\hat{\mu} - \hat{\mu}_g 1_N)'\tilde{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)]$.

Jorion's (1986) approach effectively provides a three-fund rule. Alternatively, our estimated optimal three-fund rule can be thought of as a shrinkage rule with a particular choice of shrinkage estimator of μ and a particular choice of Σ. To see why, rewrite (67) as

$$\begin{aligned} \hat{w}^{\mathrm{III}} &= \frac{c_3}{\tau}\left[\left(\frac{\hat{\psi}_a^2}{\hat{\psi}_a^2 + \frac{N}{T}}\right)\hat{\Sigma}^{-1}\hat{\mu} + \left(\frac{\frac{N}{T}}{\hat{\psi}_a^2 + \frac{N}{T}}\right)\hat{\mu}_g\hat{\Sigma}^{-1}1_N\right] \\ &= \frac{1}{\tau}\hat{\Sigma}_*^{-1}\left[\left(\frac{T\hat{\psi}_a^2}{N+T\hat{\psi}_a^2}\right)\hat{\mu} + \left(\frac{N}{N+T\hat{\psi}_a^2}\right)\hat{\mu}_g 1_N\right]. \end{aligned} \tag{75}$$

With this expression, we can see that the main difference between our estimated optimal three-fund rule and Jorion's shrinkage rule is that our estimated optimal three-fund rule calls for the use of $\hat{\Sigma}_*$ instead of $\hat{\Sigma}^{BS}$ to estimate Σ and the use of the Bayes-Stein shrinkage estimator $\hat{\mu}^{BS}$ with a value of $v = N/(N + T\hat{\psi}_a^2)$ to estimate μ.

Jorion's shrinkage portfolio rule is also a three-fund rule, but it can be suboptimal because it is not constructed for holding optimal proportions in the three funds. Rather, it is motivated by using the standard two-fund optimal portfolio formula with a better estimate for the mean, and this better estimate has the average excess return of the sample global minimum-variance portfolio as the shrinkage target. Since the weights assigned to the sample global minimum-variance and tangency portfolios by Jorion's portfolio rule are not optimal, we expect our optimal three-fund rule to perform better. As the rules are complex functions of $\hat{\mu}$ and $\hat{\Sigma}$, it is difficult to prove it analytically. However, the expected utilities of the two rules can be easily estimated using simulated data sets. In our simulation experiments, the optimal three-fund rule indeed outperforms Jorion's rule.

5. MacKinlay and Pástor portfolio rule

When a factor-based asset pricing model holds, there is a link between expected return and the covariance matrix. MacKinlay and Pástor (2000) exploits this link and imposes an exact one-factor structure to provide a more efficient estimator of the expected returns. Specifically, they assume

$$\Sigma = \sigma^2 I_N + a\mu\mu', \tag{76}$$

where a and σ^2 are positive scalars. Under this assumption, we have

$$\Sigma^{-1} = \frac{1}{\sigma^2}\left(I_N - \frac{a\mu\mu'}{\sigma^2 + a\mu'\mu}\right) \tag{77}$$

and the weights of the tangency portfolio are given by

$$w_T = \frac{\Sigma^{-1}\mu}{1_N'\Sigma^{-1}\mu} = \frac{\mu}{1_N'\mu}. \tag{78}$$

Therefore, one can simply replace Σ by I_N to obtain the weights of the tangency portfolio. While MacKinlay and Pástor (2000) focus their analysis on the tangency portfolio, we extend here their analysis to the case of the optimal portfolio for a given risk aversion coefficient. To obtain the weights of the optimal portfolio, we need to know also a and σ^2 because

$$w^* = \frac{1}{\tau}\Sigma^{-1}\mu = \frac{1}{\tau(\sigma^2 + a\mu'\mu)}\mu. \tag{79}$$

The maximum likelihood estimator of a, σ^2 and μ are obtained by maximizing the log-likelihood function

$$\ln \mathcal{L} = -\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\ln\left(|a\mu\mu' + \sigma^2 I_N|\right) - \frac{1}{2}\sum_{t=1}^{T}(R_t - \mu)'(a\mu\mu' + \sigma^2 I_N)^{-1}(R_t - \mu). \quad (80)$$

As the analytical solution to the maximum likelihood estimator is non-trivial, MacKinlay and Pástor (2000) suggest the use of a constrained quasi-Newton method to obtain the estimates. However, when the number of assets is large, this numerical procedure can be very time-consuming and may also fail to find the global maximum. In the Appendix, we reduce the multiple variable optimization problem to a single variable optimization problem that greatly reduces the computation time and the possibility of finding a local maximum. Let $\tilde{a}$, $\tilde{\sigma}^2$ and $\tilde{\mu}$ be the maximum likelihood estimators, the MacKinlay and Ľuboš portfolio rule is given by

$$\hat{w}_{ML} = \frac{1}{\tau(\tilde{\sigma}^2 + \tilde{a}\tilde{\mu}'\tilde{\mu})}\tilde{\mu}. \quad (81)$$

Instead of using the maximum likelihood estimator, we propose the use of a simplified version of $\hat{w}_{ML}$ that does not require numerical optimization. Let $\hat{U} = \hat{\Sigma} + \hat{\mu}\hat{\mu}'$. Suppose $\hat{\lambda}_1$ is the largest eigenvalue of $\hat{U}$ and $\hat{q}_1$ is the associated eigenvector. We estimate μ using $\tilde{\mu} = (\hat{q}_1'\hat{\mu})\hat{q}_1$ and the weights of the optimal portfolio are estimated by

$$\hat{w}_{ML} = \frac{\tilde{\mu}}{\tau(\hat{\lambda}_1 - \tilde{\mu}'\tilde{\mu})}. \quad (82)$$

In our simulations, we find that our simplified estimator produces numerically identical estimates to the maximum likelihood ones. While the technical details are given in the appendix, we provide the intiution of our simplified estimator here. Note that when $\Sigma = \sigma^2 I_N + a\mu\mu'$, it is easy to verify that the largest eigenvalue of $U = \Sigma + \mu\mu'$ is $\lambda_1 = (a+1)\mu'\mu + \sigma^2$ and the corresponding eigenvector is $q_1 = (\mu'\mu)^{-\frac{1}{2}}\mu$ which is proportional to μ. Therefore, it makes sense to choose $\tilde{\mu}$ to be proportional to $\hat{q}_1$, the sample estimator of q_1. Our simplified estimator makes it clear that under the assumption of the exact one-factor structure, we can use both the information in $\hat{\Sigma}$ and $\hat{\mu}$ to provide a more accurate estimator of μ.

The benefit of using the MacKinlay and Pástor portfolio rule depends on how well (76) approximates the structure of Σ. In other words, it crucially depends on how much information that $\hat{\Sigma}$ contains about μ. When expected returns are well approximated by an exact one-factor structure, we can expect the MacKinlay and Pástor portfolio rule to provide significant benefits in improving expected utility. However, if the exact one-factor structure does not hold, then there could be significant loss. This is because by imposing a misspecifed factor structure, the MacKinlay and Pástor portfolio rule will not converge to the true optimal portfolio as T increases. As analytical expression of the expected utility for the MacKinlay and Pástor portfolio rule is unavailable, we compare its performance together with other portfolio rules in the following section using simulations.

6. Comparison of alternative portfolio rules

In this section, we evaluate the expected utilities of 14 portfolio rules for a mean-variance investor with parameters calibrated from real data. While the rules are developed under the standard normality assumption, we also examine their performance under a more plausible multivariate t-distribution, and find the qualitative results are quite robust to the departure from the normality assumption.

In what follows, we assume that the mean-variance investor has a relative risk aversion of $\tau = 3$. The expected utilities for other values of τ can be obtained by simply rescaling the expected utilities calculated for $\tau = 3$, so the relative rankings of the portfolio rules are independent of the choice of τ. In evaluating these portfolio rules, we consider two scenarios. In the first scenario, we assume there are $N = 10$ risky assets, with their mean and covariance matrix chosen based on the sample estimates from the monthly excess returns on the 10 NYSE size-ranked portfolios from 1926/1–2003/12. For this set of 10 risky assets, our choice of μ and Σ give $\theta = 0.159$, $\psi = 0.130$, and $\mu_g = 0.00444$.[6] In the second scenario, we assume there are $N = 25$ risky assets. Because Fama and French's (1993) 25 portfolios, formed based on size and book-to-market ratio, are the standard test assets in recent empirical asset pricing studies, we assume that the investor invests in these 25 portfolios. The mean and covariance matrix of these 25 portfolios are chosen based on the sample estimates from the monthly excess returns from 1932/1–2003/12.[7] For this set of 25 risky assets, our choice of

[6] The individual elements of μ and Σ are not reported because it can be shown that except foe the MacKinlay and Pástor portfolio rule, the expected utilities of all other portfolio rules are only a function of θ, ψ, and μ_g.

[7] We are grateful to Ken French for making this data available on his website.

μ and Σ give $\theta = 0.344$, $\psi = 0.267$, and $\mu_g = 0.00889$.

Out of the 14 portfolio rules that have been discussed in this paper, the expected utilities can be derived analytically for nine, and for the remaining five, we rely on an efficient simulation method which proceeds as follows. For different lengths of the estimation window T, generate $\hat{\mu}$ and $\hat{\Sigma}$ from (10) and (11). Then, construct the optimal portfolio using the various portfolio rules and compute each corresponding utility. The average utilities across different simulations are then used to approximate the expected utilities. The following simulation results are based on 100,000 simulated samples.

Table 2 reports the results for the 10 asset case. The first three portfolio rules assume that the investor knows some of the parameters. If he knows μ and Σ, the expected utility of his optimal portfolio is given by equation (3), which is 0.419%/month as reported in the first row ("Parameter-certainty Optimal"). If the investor only knows θ, he can invest an optimal amount in the sample tangency portfolio, and the resulting expected utility is reported in the second row ("Theoretical Optimal Two-fund"). Investing in the *ex post* instead of the *ex ante* tangency portfolio generates a substantial loss in expected utility. The loss is a decreasing function of the length of the estimation period, but even for $T = 480$ months, the expected utility from the optimal two-fund rule is only 0.224%/month versus 0.419%/month from holding the *ex ante* tangency portfolio. The third row reports the expected utility of a portfolio that invests optimally in the sample tangency and global minimum-variance portfolios ("Theoretical Optimal Three-Fund"). Implementing this rule requires knowing ψ and μ_g. Compared to the optimal two-fund rule, the gain in expected utility is phenomenal when T is small. When T is large, less weight is put on the sample global minimum-variance portfolio, so the difference between the expected utilities of the optimal two-fund and three-fund rules is not that big. Note that these three rules cannot be implemented in practice, so their expected utilities are only provided as reference points.

Table 2 about here

In the next eight rows, we report the expected utilities of various two-fund rules. The first three rows are for the plug-in methods that estimate μ by $\hat{\mu}$, Σ by $\hat{\Sigma}$, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$, and $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$. The next row is for the Bayesian rule under a diffuse prior which essentially estimates Σ by $(T+1)\hat{\Sigma}/(T-N-2)$. The fifth row reports the expected utility of using the parameter-free optimal two-fund rule, which estimates Σ using Haff's estimator of the covariance matrix $\hat{\Sigma}_* = \hat{\Sigma}/c_3$. All five rules are poor, although the parameter-free optimal two-fund rule dominates all the others.

When the sample size is as large as $T = 360$ months, a history of thirty-years worth of data, one might think it can provide sufficiently accurate estimates of the parameters such that the plug-in methods should work reasonably well. On the contrary, due to the volatilities of the estimates, the expected utilities of using the plug-in methods are in fact negative! This also includes the Bayesian portfolio rule. As the sample decreases from 360 months, the problem is exacerbated. Clearly, sample estimates cannot blindly be substituted for population parameters without significantly decreasing utility.

The next two-fund rule is the "Estimated Optimal Two-fund" rule, which is obtained by replacing the true θ^2 in the optimal two-fund rule by the estimated θ^2. Although the estimated optimal two-fund rule does not deliver the same level of expected utility as the theoretical optimal two-fund rule, it is implementable and performs substantially better than all the plug-in rules. It starts to yield positive expected utility when $T > 120$ months, whereas all the other plug-in rules need $T > 360$ months to yield positive expected utility. Nevertheless, when T is small, the estimate of θ^2 is very volatile, so the estimated optimal two-fund rule still delivers negative expected utility for $T \leq 120$ months.

The second-last two-fund rule is the "Uncertainty Aversion Two-fund" rule of Garlappi, Uppal, and Wang (2004). It is the best rule when $T = 60$ months, even though the expected utility is still negative. But it is dominated by the estimated optimal two-fund rule when $T > 120$ months. The reason is that it invests heavily in the riskless asset. However, it should be pointed out that the uncertainty aversion rule was not designed for maximizing the expected utility of a mean-variance investor, so its under-performance is expected, which does not contradict in any way that it is the best rule under Garlappi, Uppal, and Wang's (2004) uncertainty aversion utility function.

The last two-fund rule is the "Global Minimum-Variance" portfolio rule, which invests

$\Sigma_*^{-1} 1_N \hat{\mu}_g/\tau$ in the risky assets and the rest in the riskless asset.[8] This portfolio rule invests only in the sample global minimum-variance portfolio and the riskless asset, so it is also a two-fund rule. In the Appendix, we show that the expected utility of this global minimum-variance portfolio rule is given by

$$\frac{(T-N-1)(T-N-4)}{(T-2)(T-N-2)}\frac{1}{2\tau}\left(\theta^2-\psi^2+\frac{\frac{(T-N-5)\psi^2}{T-N-1}-\frac{T-4}{T}}{T-N-3}\right). \tag{83}$$

With the parameter specifications here, the simulation results show that this rule is generally dominated by the estimated optimal two-fund rule.

The next two rows in Table 2 report the expected utilities of the two three-fund rules. The first rule is Jorion's shrinkage estimator of the optimal portfolio, which substantially outperforms the plug-in rules and the Bayesian rule. However, it only starts to outperform the estimated optimal two-fund rule when $T > 360$. Therefore, a better estimator of μ alone is not sufficient to beat the estimated optimal two-fund rule. The second three-fund rule is the "Estimated Optimal Three-fund" rule, which is obtained by replacing ψ^2 and μ_g in the theoretical optimal three-fund rule with their estimates. When T is small, the estimated optimal three-fund rule provides an often substantial improvement over Jorion's rule. When T is large, the shrinkage rule and the estimated optimal three-fund rule generate virtually identical expected utilities.

The final row reports the expected utility of the MacKinaly and Pástor portfolio rule. Under this portfolio rule, the expected returns are not estimated using $\hat{\mu}$ but also estimated with the information in $\hat{\Sigma}$. As it turns out under our choices of μ and Σ for the 10 asset case, the one-factor model provides a very good approximation of the expected return. We find that the average absolute expected return error using the best fitted one-factor model is only 0.040%/month. Although when $T \to \infty$, the expected utility under this portfolio rule will converge to only 0.272%/month but not 0.419%/month, the use of the information in $\hat{\Sigma}$ to estimate μ helps tremendously when T is small. When $120 \le T \le 480$, it substantially outperforms all the other rules.

[8] It can be shown that within the class of portfolio rules $d\hat{\Sigma}^{-1}1_N/\tau$, the d that maximizes expected utility is $d^* = c_3\mu_g$, which implies that the optimal weights are $\hat{\Sigma}_*^{-1}1_N\mu_g/\tau$. Our implementable version of the global minimum-variance portfolio rule is obtained by replacing μ_g with $\hat{\mu}_g$.

From the expected utilities of the eleven implementable portfolio rules, an investor is better off by investing just in the riskless asset if $T \leq 60$ because all the portfolio rules that call for investing in the risky assets generate negative expected utility. When $T \geq 120$, the MacKinlay and Pástor is the best rule, with the optimal three-fund rule as the second best overall rule.

From Table 2, the theoretical optimal expected utility is unattainable using existing rules. This is particularly apparent when $T = 60$, but as the sample size increases, the problem diminishes. When $T = 480$, the expected utility of the estimated and theoretical optimal three-fund rules become very close, suggesting that using estimated ψ^2 and μ_g is less of a problem. Nevertheless, the expected utility of the estimated optimal three-fund rule is 0.206%/month, still about 20% less than the expected utility of 0.258%/month from the theoretical optimal three-fund rule. So there is still room for improvement in our estimated optimal three-fund rule, especially when T is small.

Table 3 presents the corresponding results for the 25 asset case. As the number of risky assets increases, two effects occur. The first effect is that there are more parameters to estimate and hence there is more estimation risk, which in turn leads to lower expected utility. The second effect is that with more assets, the Sharpe ratio of the tangency portfolio increases, which in turn leads to higher expected utility in the absence of estimation risk. In our example, the Sharpe ratio of the tangency portfolio in the 25 asset case is about twice as big as it is in the 10 asset case. As a result, the expected utilities for the first three portfolio rules in Table 3 are all higher than their counterparts in Table 2 because the first three portfolio rules assume that some of the parameters are known, so there is little estimation risk. This is not the case for the implementable portfolio rules. For example, when T is small, the plug-in rules in the 25 asset case generate far lower expected utilities than in the 10 asset case. Although the numbers in Tables 2 and 3 are different, the general picture is largely the same. Specifically, the plug-in portfolio rules are all very poor, the estimated optimal two-fund and three-fund rules perform far better than all the plug-in rules and the uncertainty aversion two-fund rules across all T, with the estimated optimal three-fund rule having an edge even over Jorion's shrinkage portfolio rule. The estimated optimal three-fund rule performs particularly well in the 25 asset case, losing out only to the global minimum-variance portfolio rule when $T = 60$, and dominating all the other implementable portfolio rules across T, including the MacKinlay and Pástor portfolio rule. As a result, an investor facing such a portfolio problem is better off using the estimated optimal three-fund rule for $T \geq 120$.

Table 3 about here

One may wonder why the MacKinlay and Pástor portfolio rule does not perform well in the 25 asset case. This has to do with the choice of μ and Σ for the Fama-French 25 size and book-to-market ranked portfolios. It is entirely plausible that we may need more than one factor to explain the expected returns of the Fama-French portfolios. In fact, using the best fitted one-factor model, we find the average absolute expected return error to be 0.155%/month for the 25 portfolios, far larger than the 0.040%/month in the previous 10 asset case. More importantly, even ignoring estimation errors, the expected utility under the MacKinlay and Pástor portfolio rule is still only 0.371%/month, so it is not surprising that this rule does not deliver very good performance for the 25 asset case.

Now let us examine how the performance of the rules may change if the return data follow a multivariate t rather than a multivariate normal distribution. Ever since Fama (1965), many studies find that stock returns have fat-tailed distributions and they are not well approximated by the normal distribution. Blattberg and Gonedes (1974) seems the first in finance to use a t-distribution to model stock returns. In comparison with GARCH models and the like, Kan and Zhou (2004) find that there is little GARCH effect in the monthly stock returns data and a multivariate t-distribution with eight degrees of freedom models the returns of the 25 Fama-French portfolios well. Since the interest here is to examine the robustness of the results to the departure from the normality assumption, we in what follows take a more conservative approach by using a multivariate t-distribution with five degrees of freedom, which represents a significant departure from normality.

Table 4 about here

Table 5 about here

Tables 4 and 5 provide the expected utilities under the same assumptions for Tables 2 and 3 except now the stock returns follow a multivariate t-distribution with five degrees of freedom rather than multivariate normal. Under the multivariate t-distribution, we have no analytical formulas for the expected utilities and hence they are all computed based on 100,000 simulated draws of the data sets. By comparing the results in Tables 2 and 4 with those in Tables 3 and 5, we find that the performance rankings of the portfolio rules remain largely the same under the two different distribution assumptions. However, the expected utilities of the feasible portfolio rules are in general smaller under the multivariate t-distribution assumption than those under the normality assumption. This is intuitive because under the multivariate t-distribution assumption, the estimation errors of $\hat{\mu}$ and $\hat{\Sigma}$ are in general greater than those under the normality assumption (see, e.g., Kan and Zhou (2004)).

In summary, the simulation results suggest that parameter estimates based on statistical criteria alone, such as the maximum likelihood estimator, perform poorly at an economically unacceptable level. Among the 11 implementable rules, the estimated optimal three-fund rule, newly developed based on utility maximization, performs remarkably (when $T > 120$), and offers 65% improvement in the expected utility over the popular maximum likelihood estimator even when T is as large as 480 ($N = 25$). However, when expected returns of the assets can be well approximated by a one-factor model, then it is beneficial to exploit the link between expected returns and covariance matrix. In this case, the MacKinlay and Pástor portfolio rule significantly dominates all the other rules. When a multivariate t-distribution with five degrees of freedom is assumed as the true distribution, the same qualitative conclusions follow, suggesting the results are robust to the departure from the normality assumption.

7. Conclusion

Models for financial decision making often involve unknown parameters that have to be estimated from the data. However, estimation is typically separated from the decision making, and the goodness of the estimates is commonly judged using statistical criteria such as minimum variance and asymptotic efficiency. We argue that it is important to estimate parameters by combining the estimation with the economic objectives at hand. In particular, we show that, in the standard mean-variance framework, the usual maximum likelihood estimate of the optimal portfolio weights is outperformed by alternative sample estimates. These, in turn, are uniformly dominated by the Bayesian approach under a diffuse prior, which accounts for the parameter uncertainty by using predictive densities. The Bayesian solution, in turn, is uniformly dominated by a new two-fund rule that holds the riskless asset and the sample tangency portfolio optimally.

While mean-variance portfolio theory recommends a two-fund solution, we construct an optimal three-fund portfolio rule that improves expected utility over the two-fund rules. Using reasonable choices of parameters, we show that an implementable version of our three-fund rule dominates many of the existing portfolio rules. Nevertheless, we believe there exist even better rules that we encourage future research to discover. Our finding that a three-fund portfolio rule can dominate a two-fund portfolio rule has powerful implications. It says that the recommendation of a theoretical result, like holding a two-fund portfolio here, can be altered completely in the presence of parameter uncertainty, to a holding a three-fund (perhaps even more) portfolio.

Many potential extensions are possible. For example, our analysis can be extended to more complex dynamic portfolio choice problems, such as a set-up like Barberis (2000). In fact, economically better estimates can potentially be sought in any other financial decisions, either in investments or in corporate finance, that involve estimation of unknown parameters with precise economic objectives. Hence, this paper poses many problems for future research. For example, our methodology can be applied to determine the mean-variance optimal hedge ratio in hedging. It can also be used to estimate the discount rate for maximizing the net present value of an investment project. In the asset pricing literature, the market risk premium estimated from sample mean excess returns is generally considered to be too high, but this is in general not the market risk premium an investor use in maximizing the expected utility. Accounting for parameter uncertainty (and perhaps model uncertainty too), what would be the risk premium estimate? These are interesting topics for future research.

Appendix

Proof of (48): Using Theorem 3.2.13 of Muirhead (1982), we have

$$\frac{\hat{\mu}'\Sigma^{-1}\hat{\mu}}{\hat{\mu}(T\hat{\Sigma})^{-1}\hat{\mu}} \sim \chi^2_{T-N}, \tag{A.1}$$

which is independent of $\hat{\mu}$. Therefore, we can write

$$\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} = \frac{T(\hat{\mu}'\Sigma^{-1}\hat{\mu})}{\chi^2_{T-N}}, \tag{A.2}$$

where the numerator and denominator are independent. Since $T\hat{\mu}'\Sigma^{-1}\hat{\mu} \sim \chi^2_N(T\mu'\Sigma^{-1}\mu)$,

$$\hat{\theta}^2 = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} \sim \left(\frac{N}{T-N}\right) F_{N,T-N}(T\theta^2). \tag{A.3}$$

This completes the proof. *Q.E.D.*

Proof of (50): Theorem 3.1 of Kubokawa, Robert, and Saleh (1993) states that if $w \sim \chi^2_p(\delta)/\chi^2_n$, where the numerator and denominator are independent, then the unbiased estimator of δ is $(n-2)w - p$ but under quadratic loss, this unbiased estimator is dominated by

$$\hat{\delta}_a = (n-2)w - \phi_0(w), \tag{A.4}$$

where

$$\phi_0(w) = (n-2)\frac{\int_0^w t^{\frac{p}{2}}(1+t)^{-\frac{n+p}{2}}\mathrm{d}t}{\int_0^w t^{\frac{p}{2}-1}(1+t)^{-\frac{n+p}{2}}\mathrm{d}t}. \tag{A.5}$$

To simplify $\phi_0(w)$, write the integral in the numerator as

$$\int_0^w t^{\frac{p}{2}}(1+t)^{-\frac{n+p}{2}}\mathrm{d}t = \int_0^w \left(\frac{t}{1+t}\right)^{\frac{p}{2}} (1+t)^{-\frac{n}{2}}\mathrm{d}t. \tag{A.6}$$

Using integration by parts on this integral gives

$$\phi_0(w) = p - \frac{2w^{\frac{p}{2}}(1+w)^{-\frac{n+p-2}{2}}}{\int_0^w t^{\frac{p}{2}-1}(1+t)^{-\frac{n+p}{2}}\mathrm{d}t}. \tag{A.7}$$

For the integral in the denominator, we use a change of variables of $y = t/(1+t)$ to obtain

$$\int_0^w t^{\frac{p}{2}-1}(1+t)^{-\frac{n+p}{2}}\mathrm{d}t = \int_0^{\frac{w}{1+w}} y^{\frac{p}{2}-1}(1-y)^{\frac{n}{2}-1}\mathrm{d}y = B_{w/(1+w)}(p/2, n/2). \tag{A.8}$$

Therefore, the adjusted estimator of δ is

$$\hat{\delta}_a = (n-2)w - p + \frac{2w^{\frac{p}{2}}(1+w)^{-\frac{n+p-2}{2}}}{B_{w/(1+w)}(p/2, n/2)}. \tag{A.9}$$

Our adjusted estimator $\hat{\theta}_a^2$ is then obtained by letting $\hat{\delta}_a = T\hat{\theta}_a^2$, $\delta = T\theta^2$, $w = \hat{\theta}^2$, $p = N$, and $n = T - N$ in the equation above. The adjusted estimator $\hat{\psi}_a^2$ is similarly obtained. This completes the proof. *Q.E.D.*

Proof of (65): Let

$$\begin{aligned}
\tilde{A} &= \left([\hat{\mu},\ 1_N]'\Sigma^{-1}[\hat{\mu},\ 1_N]\right)^{-1} = \begin{bmatrix} \hat{\mu}'\Sigma^{-1}\hat{\mu} & \hat{\mu}'\Sigma^{-1}1_N \\ 1_N'\Sigma^{-1}\hat{\mu} & 1_N'\Sigma^{-1}1_N \end{bmatrix}^{-1}, & \text{(A.10)} \\
\hat{A} &= \left([\hat{\mu},\ 1_N]'\hat{\Sigma}^{-1}[\hat{\mu},\ 1_N]\right)^{-1} = \begin{bmatrix} \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} & \hat{\mu}'\hat{\Sigma}^{-1}1_N \\ 1_N'\hat{\Sigma}^{-1}\hat{\mu} & 1_N'\hat{\Sigma}^{-1}1_N \end{bmatrix}^{-1}. & \text{(A.11)}
\end{aligned}$$

From Theorem 3.2.11 of Muirhead (1982), conditional on $\hat{\mu}$,

$$\hat{A} \sim W_2(T - N + 1, \tilde{A}/T). \tag{A.12}$$

Let $\tilde{A}_{ij}$ and $\hat{A}_{ij}$ be the (i,j)th element of $\tilde{A}$ and $\hat{A}$, respectively. It is straightforward to verify that

$$\tilde{A}_{11} = \frac{1}{\tilde{\psi}^2}, \quad \hat{A}_{11} = \frac{1}{\hat{\psi}^2}, \tag{A.13}$$

where $\tilde{\psi}^2 = \hat{\mu}'\Sigma^{-1}\hat{\mu} - (\hat{\mu}'\Sigma^{-1}1_N)^2/(1_N'\Sigma^{-1}1_N)$. From (A.12),

$$\frac{\hat{A}_{11}}{\tilde{A}_{11}/T} = \frac{T\tilde{\psi}^2}{\hat{\psi}^2} \equiv w \sim \chi^2_{T-N+1}, \tag{A.14}$$

and w is independent of $\hat{\mu}$. Since

$$v \equiv T\tilde{\psi}^2 = T\hat{\mu}'\Sigma^{-\frac{1}{2}}\left[I_N - \Sigma^{-\frac{1}{2}}1_N'(1_N'\Sigma^{-1}1_N)^{-1}1_N'\Sigma^{-\frac{1}{2}}\right]\Sigma^{-\frac{1}{2}}\hat{\mu} \sim \chi^2_{N-1}(T\psi^2), \tag{A.15}$$

we have

$$\hat{\psi}^2 = \frac{T\tilde{\psi}^2}{w} = \frac{v}{w} \sim \left(\frac{N-1}{T-N+1}\right) F_{N-1,T-N+1}(T\psi^2). \tag{A.16}$$

This completes the proof. *Q.E.D.*

Almost analytical solution to the maximum likelihood estimation problem in (80):

Under the assumption of Σ in (76), one of the N eigenvalues of Σ is $\sigma^2 + a\mu'\mu$ and the rest are all σ^2. Therefore, we have

$$\ln\left(|a\mu\mu' + \sigma^2 I_N|\right) = (N-1)\ln(\sigma^2) + \ln(\sigma^2 + a\mu'\mu). \tag{A.17}$$

In addition, using the identity

$$\sum_{t=1}^{T}(R_t-\mu)(R_t-\mu)' = T[\hat{\Sigma} + (\hat{\mu}-\mu)(\hat{\mu}-\mu)'], \tag{A.18}$$

we can write

$$\begin{aligned}
&\sum_{t=1}^{T}(R_t-\mu)'(a\mu\mu'+\sigma^2 I_N)^{-1}(R_t-\mu) \\
= \ & \mathrm{tr}\left((a\mu\mu'+\sigma^2 I_N)^{-1}\sum_{t=1}^{T}(R_t-\mu)(R_t-\mu)'\right) \\
= \ & T\left[\mathrm{tr}((a\mu\mu'+\sigma^2 I_N)^{-1}\hat{\Sigma}) + (\hat{\mu}-\mu)'(a\mu\mu'+\sigma^2 I_N)^{-1}(\hat{\mu}-\mu)\right] \\
= \ & \frac{T}{\sigma^2}\left[\mathrm{tr}(\hat{\Sigma}) - \frac{a\mu'\hat{\Sigma}\mu}{\sigma^2+a\mu'\mu} + (\hat{\mu}-\mu)'(\hat{\mu}-\mu) - \frac{a[(\hat{\mu}-\mu)'\mu]^2}{\sigma^2+a\mu'\mu}\right] \\
= \ & \frac{T}{\sigma^2}\left[\mathrm{tr}(\hat{U}) + \frac{\sigma^2(\mu'\mu-2\hat{\mu}'\mu)-a\mu'\hat{U}\mu}{\sigma^2+a\mu'\mu}\right],
\end{aligned} \tag{A.19}$$

where $\hat{U} = \hat{\Sigma} + \hat{\mu}\hat{\mu}'$. Instead of maximizing (80), we can minimize

$$f(\mu, a, \sigma^2) = (N-1)\ln(\sigma^2) + \ln(\sigma^2 + a\mu'\mu) + \frac{1}{\sigma^2}\left[\mathrm{tr}(\hat{U}) + \frac{\sigma^2(\mu'\mu-2\hat{\mu}'\mu)-a\mu'\hat{U}\mu}{\sigma^2+a\mu'\mu}\right] \tag{A.20}$$

to obtain the maximum likelihood estimates.

For a fixed μ, f is only a function of a and σ^2. Setting the derivatives equal to zero, we can find the maximum likelihood estimator of a and σ^2 when conditional on a given value of μ. Conditional on μ, the maximum likelihood estimator of σ^2 and a are given by

$$\tilde{\sigma}^2 = \frac{\mathrm{tr}(\hat{U}) - \frac{\mu'\hat{U}\mu}{\mu'\mu}}{N-1}, \tag{A.21}$$

$$\tilde{a} = 1 - 2\frac{\hat{\mu}'\mu}{\mu'\mu} + \frac{\mu'\hat{U}\mu}{(\mu'\mu)^2} - \frac{\tilde{\sigma}^2}{\mu'\mu}. \tag{A.22}$$

Substituting these values in (A.20), we can write f as a function of μ alone.

$$\begin{aligned} f(\mu) &= N + \ln\left(\mu'\mu - 2\hat{\mu}'\mu + \frac{\mu'\hat{U}\mu}{\mu'\mu}\right) + (N-1)\ln\left(\frac{\mathrm{tr}(\hat{U}) - \frac{\mu'\hat{U}\mu}{\mu'\mu}}{N-1}\right) \\ &= N - (N-1)\ln(N-1) + \ln\left(\mu'\mu - 2\hat{\mu}'\mu + \frac{\mu'\hat{U}\mu}{\mu'\mu}\right) \\ &\quad + (N-1)\ln\left(\mathrm{tr}(\hat{U}) - \frac{\mu'\hat{U}\mu}{\mu'\mu}\right). \end{aligned} \tag{A.23}$$

Let $\hat{Q}\hat{\Lambda}\hat{Q}'$ be the spectral decomposition of $\hat{U}$, where $\hat{\Lambda} = \mathrm{Diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_N)$ are the eigenvalues of $\hat{U}$ in descending order, and the columns of $\hat{Q}$ are the corresponding eigenvectors. By defining $z = \hat{Q}'\mu$ and $\hat{z} = \hat{Q}'\hat{\mu}$, we can find the maximum likelihood estimator by simply minimizing $g(z)$, which is defined as

$$g(z) = \ln\left(z'z - 2\hat{z}'z + \frac{z'\hat{\Lambda}z}{z'z}\right) + (N-1)\ln\left(\sum_{i=1}^{N}\hat{\lambda}_i - \frac{z'\hat{\Lambda}z}{z'z}\right). \tag{A.24}$$

Note that $g(z)$ only depends on the data through $\hat{z}$ and $\hat{\Lambda}$, so the objective function can be evaluated very efficiently. In order to minimize $g(z)$, we first conditional on $z'\hat{\Lambda}z/(z'z) = c$, where $\hat{\lambda}_1 \geq c \geq \hat{\lambda}_N$. Conditional on $z'\hat{\Lambda}z/(z'z) = c$, $g(z)$ can be written as

$$g(z|c) = \ln(z'z - 2\hat{z}'z + c) + (N-1)\ln\left(\sum_{i=1}^{N}\hat{\lambda}_i - c\right), \tag{A.25}$$

so we only need to minimize $z'z - 2\hat{z}'z$ subject to the constraint $z'\hat{\Lambda}z/(z'z) = c$.

Forming the Lagrangian, we have

$$h(z, \phi) = z'z - 2\hat{z}'z - \phi(z'\hat{\Lambda}z - cz'z). \tag{A.26}$$

Taking the partial derivative with respect to z, we obtain the first order condition

$$\frac{\partial h}{\partial z} = 2z - 2\hat{z} - 2\phi(\hat{\Lambda}z - cz) = 0_N. \tag{A.27}$$

Premultiplying the first order condition by z', we note that the optimal z satisfies

$$\tilde{z}'\tilde{z} = \tilde{z}'\hat{z}. \tag{A.28}$$

Solving the first order equation, we have

$$\tilde{z} = [I_N - \phi(\hat{\Lambda} - cI_N)]^{-1}\hat{z}. \tag{A.29}$$

Substituting (A.29) into the constraint $z'(\hat{\Lambda} - cI_N)z = 0$, we obtain the equation

$$p(\phi) = \sum_{i=1}^{N} \frac{(\hat{\lambda}_i - c)\hat{z}_i^2}{[1 - \phi(\hat{\lambda}_i - c)]^2} = 0, \tag{A.30}$$

which can be written as a polynomial equation in ϕ of order $2N - 2$. From the second order equation, we have

$$\frac{\partial h(z, \phi)}{\partial z \partial z'} = 2[(1 + c\phi)I_N - \phi\Lambda]. \tag{A.31}$$

In order for ϕ to be associated with the minimum (i.e., the above matrix to be positive definite), we need $\phi \in (u_N, u_1)$, where $u_i = 1/(\hat{\lambda}_i - c)$. Therefore, although $p(\phi)$ can have as many as $2N - 2$ roots, we only need to search for the root in the interval (u_N, u_1). Note that $\lim_{\phi \to u_N^+} p(\phi) = -\infty$, $\lim_{\phi \to u_1^-}(\phi) = \infty$, and $p(\phi)$ is increasing in ϕ for $u_N < \phi < u_1$, so there is only one root in the interval (u_N, u_1) and we can simply use a line search to find this root. Denote this optimal root as $\tilde{\phi}(c)$, we can then write the objective function $g(z)$ as a function of c alone

$$g(c) = \ln\left(c - \sum_{i=1}^{N} \frac{\hat{z}_i^2}{1 - \tilde{\phi}(c)(\hat{\lambda}_i - c)}\right) + (N-1)\ln\left(\sum_{i=1}^{N} \hat{\lambda}_i - c\right), \tag{A.32}$$

by using the fact that $\tilde{z}'\tilde{z} = \hat{z}'\tilde{z}$ at the optimal solution. Instead of minimizing over $N + 2$ variables, we can now minimize the objective function over only one variable.

In summary, we obtain the maximum likelihood estimator by minimizing $g(c)$ over $\hat{\lambda}_1 \geq c \geq \hat{\lambda}_N$.[9] Let $c^* = \text{argmin}_c g(c)$, the maximum likelihood estimator of $\tilde{\mu}$ is then given by

$$\tilde{\mu} = \hat{Q}\tilde{z} = \hat{Q}[I_N - \tilde{\phi}(c^*)(\hat{\Lambda} - c^* I_N)]^{-1}\hat{z}. \tag{A.33}$$

The maximum likelihood estimator of σ^2 and a are then given by

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{N} \lambda_i - c^*}{N - 1}, \tag{A.34}$$

$$\tilde{a} = \frac{c^* - \tilde{\sigma}^2}{\tilde{\mu}'\tilde{\mu}} - 1. \tag{A.35}$$

[9] For $c = \hat{\lambda}_1$, we have $g(\hat{\lambda}_1) = \ln(\hat{\lambda}_1 - \hat{z}_1^2) + (N-1)\ln(\sum_{i=2}^{N} \hat{\lambda}_i)$. For $c = \hat{\lambda}_N$, we have $g(\hat{\lambda}_N) = \ln(\hat{\lambda}_N - \hat{z}_N^2) + (N-1)\ln(\sum_{i=1}^{N-1} \hat{\lambda}_i)$.

Finally, the weights of the estimated optimal portfolio are given by

$$\hat{w}_{ML} = \frac{\tilde{\mu}}{\tau(\tilde{\sigma}^2 + \tilde{a}\tilde{\mu}'\tilde{\mu})} = \frac{\tilde{\mu}}{\tau(c^* - \tilde{\mu}'\tilde{\mu})}. \tag{A.36}$$

In simulations, we find it is always the case that $c^* = \hat{\lambda}_1$ when the one-factor structure provides a reasonable approximation. In this case, we have $\tilde{\mu} = (\hat{q}_1'\hat{\mu})\hat{q}_1$, where $\hat{q}_1$ is the eigenvector of $\hat{U}$ associated with $\hat{\lambda}_1$, and $\hat{w}_{ML} = \tilde{\mu}/[\tau(\hat{\lambda}_1 - \tilde{\mu}'\tilde{\mu})]$.

Proof of (83): The expected utility of the "global minimum-variance" portfolio rule is given by

$$E\left[U\left(\frac{c_3}{\tau}\hat{\Sigma}^{-1}1_N\hat{\mu}_g\right)\right] = \frac{c_3}{\tau}E[\hat{\mu}_g 1_N'\hat{\Sigma}^{-1}\mu] - \frac{c_3^2}{2\tau}E[\hat{\mu}_g^2(1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N)]. \tag{A.37}$$

Using the fact that $\hat{\mu}$ and $\hat{\Sigma}$ are independent and $E[\hat{\mu}\hat{\mu}'] = \mu\mu' + \Sigma/T$, we can write the two terms in (A.37) as

$$E[\hat{\mu}_g 1_N'\hat{\Sigma}^{-1}\mu] = E\left[\frac{\hat{\mu}'\hat{\Sigma}^{-1}1_N}{1_N'\hat{\Sigma}^{-1}1_N}(1_N'\hat{\Sigma}^{-1}\mu)\right] = E\left[\frac{(1_N'\hat{\Sigma}^{-1}\mu)^2}{1_N'\hat{\Sigma}^{-1}1_N}\right] \tag{A.38}$$

and

$$E[\hat{\mu}_g^2(1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N)] = E\left[\frac{(\mu'\hat{\Sigma}^{-1}1_N)^2}{(1_N'\hat{\Sigma}^{-1}1_N)^2}1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right] + E\left[\frac{(1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N)^2}{T(1_N'\hat{\Sigma}^{-1}1_N)^2}\right]. \tag{A.39}$$

Let

$$\nu = \frac{\Sigma^{-\frac{1}{2}}1_N}{(1_N'\Sigma^{-1}1_N)^{\frac{1}{2}}}, \tag{A.40}$$

$$\eta = \frac{\Sigma^{-\frac{1}{2}}(\mu - \mu_g 1_N)}{\psi}. \tag{A.41}$$

It is easy to verify that ν and η are orthonormal vectors. Denote Q to be an $N \times (N-1)$ orthonormal matrix with its columns orthogonal to ν and its first column is equal to η. Then $[\nu,\ Q]$ form an orthonormal basis of $\mathbb{R}^N$. Let $W = \Sigma^{-\frac{1}{2}}\hat{\Sigma}\Sigma^{-\frac{1}{2}} \sim W_N(T-1, I_N)/T$. We now define an $N \times N$ matrix A as

$$A = \left([\nu,\ Q]'W^{-1}[\nu,\ Q]\right)^{-1}. \tag{A.42}$$

Using Theorem 3.2.11 of Muirhead (1982), we have $A \sim W_N(T-1, I_N)/T$. Partition A into two by two submatrices and denote its (i,j)th block as A_{ij}, with the first element of A

denoted as A_{11}. Using Theorem 3.2.10 of Muirhead (1982), we have

$$\begin{aligned} u \equiv A_{11} - A_{12}A_{22}^{-1}A_{21} &\sim \chi^2_{T-N}/T, && \text{(A.43)} \\ z \equiv -A_{22}^{-\frac{1}{2}}A_{21} &\sim N(0_{N-1}, I_{N-1}/\sqrt{T}), && \text{(A.44)} \\ A_{22} &\sim W_{N-1}(T-1, I_{N-1})/T, && \text{(A.45)} \end{aligned}$$

and they are independent of each other. Let $e_1 = [1,\ 0'_{N-1}]'$ and $c = 1'_N\Sigma^{-1}1_N$. Using the partitioned matrix inverse formula, it can be verified that

$$\begin{aligned} 1'_N\hat{\Sigma}^{-1}1_N &= c\nu'W^{-1}\nu = \frac{c}{u}, && \text{(A.46)} \\ 1'_N\hat{\Sigma}^{-1}\mu &= c\mu_g\nu'W^{-1}\nu + c^{\frac{1}{2}}\psi\eta'W^{-1}\nu = \frac{c\mu_g + c^{\frac{1}{2}}\psi e_1'A_{22}^{-\frac{1}{2}}z}{u}, && \text{(A.47)} \\ 1'_N\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N &= c\nu'W^{-2}\nu = c\nu'W^{-1}[\nu\nu' + QQ']W^{-1}\nu = \frac{c(1+z'A_{22}^{-1}z)}{u^2}. && \text{(A.48)} \end{aligned}$$

Using these expressions, we can express (A.38) as

$$\begin{aligned} E\left[\frac{(1'_N\hat{\Sigma}^{-1}\mu)^2}{1'_N\hat{\Sigma}^{-1}1_N}\right] &= E\left[\frac{\left(c^{\frac{1}{2}}\mu_g + \psi e_1'A_{22}^{-\frac{1}{2}}z\right)^2}{u}\right] && \text{(A.49)} \\ &= \frac{T}{T-N-2}E\left[E\left[\left(c^{\frac{1}{2}}\mu_g + \psi e_1'A_{22}^{-\frac{1}{2}}z\right)^2\middle| A_{22}\right]\right] \\ &= \frac{T}{T-N-2}E\left[c\mu_g^2 + \psi^2 e_1'(TA_{22})^{-1}e_1\middle| A_{22}\right] && \text{(A.50)} \\ &= \frac{T}{T-N-2}\left(\theta^2 - \psi^2 + \frac{\psi^2}{T-N-1}\right). && \text{(A.51)} \end{aligned}$$

The last equality follows from the fact that $\theta^2 - \psi^2 = c\mu_g^2$ and $E[(TA_{22})^{-1}] = I_N/(T-N-1)$.

Similarly, the first term in (A.39) can be expressed as

$$\begin{aligned} & E\left[\frac{(\mu'\hat{\Sigma}^{-1}1_N)^2}{(1'_N\hat{\Sigma}^{-1}1_N)^2}1'_N\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right] \\ =\ & E\left[\frac{(c^{\frac{1}{2}}\mu_g + \psi e_1'A_{22}^{-\frac{1}{2}}z)^2(1+z'A_{22}^{-1}z)}{u^2}\right] \\ =\ & \frac{T^2E[(c\mu_g^2 + 2\psi c^{\frac{1}{2}}\mu_g e_1'A_{22}^{-1}z + \psi^2 z'A_{22}^{-\frac{1}{2}}e_1e_1'A_{22}^{-\frac{1}{2}}z)(1+z'A_{22}^{-1}z)]}{(T-N-2)(T-N-4)} \\ =\ & \frac{T^2E[(c\mu_g^2 + \psi^2 z'A_{22}^{-\frac{1}{2}}e_1e_1'A_{22}^{-\frac{1}{2}}z)(1+z'A_{22}^{-1}z)]}{(T-N-2)(T-N-4)}. && \text{(A.52)} \end{aligned}$$

Using Theorem 3.2.12 of Muirhead (1982), we have $z'A_{22}^{-1}z = u_1/u_2$, where $u_1 \sim \chi^2_{N-1}$ and $u_2 \sim \chi^2_{T-N+1}$, and they are independent of each other. Using this result, the first term in the expectation is

$$c\mu_g^2 E[1 + z'A_{22}^{-1}z] = c\mu_g^2 \left(1 + E\left[\frac{u_1}{u_2}\right]\right) = c\mu_g^2 \left(1 + \frac{N-1}{T-N-1}\right) = \frac{(T-2)(\theta^2 - \psi^2)}{T-N-1}. \tag{A.53}$$

The second term in the expectation is

$$\begin{aligned}
& E[\psi^2 z'A_{22}^{-\frac{1}{2}} e_1 e_1' A_{22}^{-\frac{1}{2}} z(1 + z'A_{22}^{-1}z)] \\
= \; & \psi^2 \left(E[z'A_{22}^{-\frac{1}{2}} e_1 e_1' A_{22}^{-\frac{1}{2}} z] + E[z'A_{22}^{-\frac{1}{2}} e_1 e_1' A_{22}^{-\frac{1}{2}} zz'A_{22}^{-1}z)]\right) \\
= \; & \psi^2 \left(E[e_1'(TA_{22})^{-1}e_1] + E[e_1'(TA_{22})^{-1}e_1 \mathrm{tr}((TA_{22})^{-1})] + 2E[e_1'(TA_{22})^{-2}e_1]\right) \\
= \; & \psi^2 \left[\frac{1}{T-N-1} + \frac{(T-N)(N-1) - 2(N-2)}{(T-N)(T-N-1)(T-N-3)} + \frac{2(T-2)}{(T-N)(T-N-1)(T-N-3)}\right] \\
= \; & \frac{(T-2)\psi^2}{(T-N-1)(T-N-3)},
\end{aligned} \tag{A.54}$$

where the second last equality follows the results in Theorem 3.2 of Haff (1979). Therefore, we have

$$\begin{aligned}
& E\left[\frac{(\mu'\hat{\Sigma}^{-1}1_N)^2}{(1_N'\hat{\Sigma}^{-1}1_N)^2} 1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right] \\
= \; & \frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \left[\theta^2 - \left(\frac{T-N-4}{T-N-3}\right)\psi^2\right].
\end{aligned} \tag{A.55}$$

For the second term in (A.39), it can be expressed as

$$\begin{aligned}
E\left[\frac{(1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N)^2}{T(1_N'\hat{\Sigma}^{-1}1_N)^2}\right] &= E\left[\frac{(1 + z'A_{22}^{-1}z)^2}{Tu^2}\right] \\
&= \frac{TE\left[1 + 2\left(\frac{u_1}{u_2}\right) + \left(\frac{u_1}{u_2}\right)^2\right]}{(T-N-2)(T-N-4)} \\
&= \frac{T\left[1 + \frac{2(N-1)}{T-N-1} + \frac{(N-1)(N+1)}{(T-N-1)(T-N-3)}\right]}{(T-N-2)(T-N-4)} \\
&= \frac{T(T-2)(T-4)}{(T-N-1)(T-N-2)(T-N-3)(T-N-4)}.
\end{aligned} \tag{A.56}$$

Combining all the above results, the expected utility is explicitly evaluated as:

$$E\left[U\left(\frac{c_3}{\tau}\hat{\Sigma}^{-1}1_N\hat{\mu}_g\right)\right]$$

$$
\begin{aligned}
= \quad & \frac{c_3 T}{\tau}\left(\frac{\theta^2}{T-N-2}-\frac{\psi^2}{T-N-1}\right)-\frac{c_3^2 T^2 (T-2)}{2\tau(T-N-1)(T-N-2)} \\
& \times\left[\left(\frac{\theta^2}{T-N-4}-\frac{\psi^2}{T-N-3}\right)+\frac{T-4}{T(T-N-3)(T-N-4)}\right]. \qquad \text{(A.57)}
\end{aligned}
$$

After some simplification, we obtain (83). This completes the proof. *Q.E.D.*

References

Barberis, N., 2000. Investing for the long run when returns are predictable. Journal of Finance 55, 225–264.

Bawa, V. S., Brown, S. J., Klein, R. W., 1979. Estimation Risk And Optimal Portfolio Choice. North-Holland, Amsterdam.

Berger, J. O., 1985. Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

Blattberg, R. C., Gonedes, N. J., 1974. A comparison of the stable and Student distributions as statistical models for stock prices. Journal of Business 47, 244–280.

Brown, S. J., 1976. Optimal portfolio choice under uncertainty. Ph.D. dissertation, University of Chicago.

Brown, S. J., 1978. The portfolio choice problem: comparison of certainty equivalence and optimal Bayes portfolios. Communications in Statistics — Simulation and Computation 7, 321–334.

Chen, R., Kan, R., 2004. Finite sample analysis of two-pass cross-sectional regressions. Working paper, University of Toronto.

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

Fama, E. F., 1965. The behavior of stock market prices. Journal of Business 38, 34–105.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fourdrinier, D., Philippe, A., Robert, C.P., 2000. Estimation of a non-centrality parameter under Stein-type-like losses. Journal of Statistical Planning and Inference 87, 43–54.

Frost, P. A., Savarino, J. E., 1986. An empirical Bayes approach to efficient portfolio selection. Journal of Financial and Quantitative Analysis 21, 293–305.

Garlappi, L., Uppal, R., Wang, T., 2004. Portfolio selection with parameter and model uncertainty: a mult-prior approach. Working paper, University of British Columbia.

Haff, L. R., 1979. An identity for the Wishart distribution with applications. Journal of Multivariate Analysis 9, 531–544.

Harvey, C. R., Liechty, J., Liechty, M., Müller, P., 2003. Portfolio selection with higher moments. Working Paper, Duke University.

Jobson, J. D., Korkie, B., Ratti, V., 1979. Improved estimation for Markowitz portfolios using James-Stein type estimators. Proceedings of the American Statistical Association, Business and Economics Statistics Section 41, 279–284.

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

Jobson, J. D., Korkie, B., 1981. Putting Markowitz theory to work. Journal of Portfolio Management 7, 70–74.

Jorion, P, 1986. Bayes-Stein estimation for portfolio analysis. Journal of Financial and Quantitative Analysis 21, 279–292.

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

Kacperczyk, M., 2004. Asset allocation under distribution uncertainty. Working paper, University of British Columbia.

Kan, R, Zhou, G., 2004. Modeling non-normality using multivariate t: implications for asset pricing. Working paper, Washington University.

Kandel, S., Stambaugh, R. F., 1996. On the predictability of stock returns: An asset-allocation perspective. Journal of Finance 51, 385–424.

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

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

Lutgens, F., 2004. Robust portfolio optimization. Ph.D. dissertation, University of Maastricht.

Litterman, B., 2003. Modern Investment Management. Wiley, New York.

MacKinlay, A. C., Pástor, Ľ., 2000. Asset pricing models: Implications for expected returns and portfolio selection. Review of Financial Studies 13, 883–916.

Markowitz, H. M., 1952. Mean-variance analysis in portfolio choice and capital markets. Journal of Finance 7, 77–91.

Muirhead, R. J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Pástor, Ľ., 2000. Portfolio selection and asset pricing models. Journal of Finance 55, 179–223.

Pástor, Ľ., Stambaugh, R. F., 2000. Comparing asset pricing models: An investment perspective. Journal of Financial Economics 56, 335–381.

Rukhin, A. L., 1993. Estimation of the noncentrality parameter of an F-distribution. Journal of Statistical Planning and Inference 35, 201–211.

Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 197–206 (University of California Press, Berkeley).

Stambaugh, R. F., 1997. Analyzing investments whose histories differ in length. Journal of Financial Economics 45, 285–331.

TerHorst, J. R., DeRoon, F. A., Werkerzx, B. J. M., 2002. Incorporating estimation risk in portfolio choice. Working paper, Tilburg University.

Tu, J., Zhou, G., 2003. Data-generating process uncertainty: What difference does it make in portfolio decisions? Journal of Financial Economics 72, 385–421.

Xia, Y., 2001. Learning about predictability: The effect of parameter uncertainty on dynamic asset allocation. Journal of Finance 56, 205–246.

Zellner, A., Chetty, V. K., 1965. Prediction and decision problems in regression models from the Bayesian point of view. Journal of the American Statistical Association 60, 608–616.

Table 1
Percentage loss of expected utility due to estimation errors in means and covariance matrix of returns

The table presents the percentage loss of expected utility from holding a sample tangency portfolio of N risky assets with the parameters estimated using T periods of historical returns instead of using the true parameters. The first column reports the percentage loss due to the use of the sample average returns $\hat{\mu}$ instead of true expected returns. The second column reports the percentage loss due to the use of the sample covariance matrix $\hat{\Sigma}$ instead of the true covariance matrix. The third column reports the interactive effect from using $\hat{\mu}$ and $\hat{\Sigma}$. The fourth column reports the total percentage loss of expected utility from using $\hat{\mu}$ and $\hat{\Sigma}$. Panel A assumes the Sharpe ratio (θ) of the N risky assets is 0.2 and Panel B assumes $\theta = 0.4$.

Panel A: $\theta = 0.2$

		Percentage Loss of Expected Utility			
N	T	$\hat{\mu}$	$\hat{\Sigma}$	Interaction	$\hat{\mu}$ and $\hat{\Sigma}$
1	60	41.67	4.31	6.18	52.15
	120	20.83	1.90	1.46	24.19
	240	10.42	0.89	0.36	11.66
	360	6.94	0.58	0.16	7.68
	480	5.21	0.43	0.09	5.73
2	60	83.33	6.85	17.61	107.80
	120	41.67	2.93	4.09	48.69
	240	20.83	1.35	0.99	23.17
	360	13.89	0.88	0.43	15.20
	480	10.42	0.65	0.24	11.31
5	60	208.33	16.64	89.69	314.66
	120	104.17	6.44	19.62	130.23
	240	52.08	2.84	4.61	59.53
	360	34.72	1.81	2.01	38.54
	480	26.04	1.33	1.12	28.49
10	60	416.67	42.99	387.46	847.12
	120	208.33	13.95	75.36	297.64
	240	104.17	5.65	16.85	126.67
	360	69.44	3.51	7.23	80.19
	480	52.08	2.54	4.00	58.62
25	60	1041.67	336.67	5211.57	6589.91
	120	520.83	55.53	591.64	1168.01
	240	260.42	17.18	110.77	388.37
	360	173.61	9.81	45.19	228.61
	480	130.21	6.81	24.39	161.42

Table 1
Percentage loss of expected utility due to estimation errors in means and covariance matrix of returns (continued)

Panel B: $\theta = 0.4$

		Percentage Loss in Expected Utility			
N	T	$\hat{\mu}$	$\hat{\Sigma}$	Interaction	$\hat{\mu}$ and $\hat{\Sigma}$
1	60	10.42	4.31	1.55	16.27
	120	5.21	1.90	0.37	7.47
	240	2.60	0.89	0.09	3.58
	360	1.74	0.58	0.04	2.36
	480	1.30	0.43	0.02	1.75
2	60	20.83	6.85	4.40	32.09
	120	10.42	2.93	1.02	14.37
	240	5.21	1.35	0.25	6.81
	360	3.47	0.88	0.11	4.46
	480	2.60	0.65	0.06	3.32
5	60	52.08	16.64	22.42	91.14
	120	26.04	6.44	4.90	37.39
	240	13.02	2.84	1.15	17.01
	360	8.68	1.81	0.50	11.00
	480	6.51	1.33	0.28	8.12
10	60	104.17	42.99	96.86	244.02
	120	52.08	13.95	18.84	84.87
	240	26.04	5.65	4.21	35.91
	360	17.36	3.51	1.81	22.68
	480	13.02	2.54	1.00	16.56
25	60	260.42	336.67	1302.89	1899.98
	120	130.21	55.53	147.91	333.65
	240	65.10	17.18	27.69	109.98
	360	43.40	9.81	11.30	64.51
	480	32.55	6.81	6.10	45.47

Table 2
Expected utilities of various portfolio rules with 10 risky assets when returns follow multivariate normal distribution

The table reports the expected utilities (in percentages per month) of 14 portfolio rules that choose an optimal portfolio of 10 risky assets and a riskless asset for different lengths of the estimation period (T). The excess returns of the 10 risky assets are assumed to be generated from a multivariate normal distribution, with the mean and covariance matrix chosen based on the sample estimates of 10 size-ranked NYSE portfolios. The investor is assumed to have a risk aversion coefficient of three. The expected utilities of the first eight rules and the global minimum-variance rule are obtained analytically. For the other five rules, the expected utilities are approximated using 100,000 simulations.

Portfolio Rule	$T = 60$	$T = 120$	$T = 180$	$T = 240$
Parameter-certainty Optimal	0.419	0.419	0.419	0.419
Theoretical Optimal Two-fund	0.044	0.088	0.122	0.150
Theoretical Optimal Three-fund	0.133	0.168	0.191	0.209
1st Plug-in, $\hat{\Sigma}$	−5.122	−1.531	−0.748	−0.411
2nd Plug-in, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$	−4.936	−1.498	−0.735	−0.404
3rd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−3.110	−1.156	−0.596	−0.329
Bayesian (diffuse prior)	−2.996	−1.130	−0.584	−0.323
Parameter-free Optimal Two-fund	−1.910	−0.879	−0.476	−0.263
Estimated Optimal Two-fund	−0.185	−0.007	0.060	0.102
Uncertainty Aversion Two-fund	**−0.001**	0.004	0.007	0.012
Global Minimum-Variance	−0.152	−0.010	0.040	0.064
Jorion's Shrinkage	−0.899	−0.220	−0.030	0.062
Estimated Optimal Three-fund	−0.343	−0.053	0.051	0.107
MacKinlay-Pástor	−0.060	**0.118**	**0.171**	**0.198**

Portfolio Rule	$T = 300$	$T = 360$	$T = 420$	$T = 480$
Parameter-certainty Optimal	0.419	0.419	0.419	0.419
Theoretical Optimal Two-fund	0.173	0.193	0.210	0.224
Theoretical Optimal Three-fund	0.224	0.237	0.248	0.258
1st Plug-in, $\hat{\Sigma}$	−0.225	−0.107	−0.025	0.034
2nd Plug-in, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$	−0.221	−0.104	−0.023	0.036
3rd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−0.174	−0.072	0.000	0.054
Bayesian (diffuse prior)	−0.170	−0.069	0.002	0.055
Parameter-free Optimal Two-fund	−0.132	−0.043	0.022	0.070
Estimated Optimal Two-fund	0.133	0.157	0.177	0.194
Uncertainty Aversion Two-fund	0.017	0.024	0.032	0.040
Global Minimum-Variance	0.079	0.089	0.096	0.101
Jorion's Shrinkage	0.117	0.155	0.182	0.203
Estimated Optimal Three-fund	0.143	0.169	0.189	0.206
MacKinlay-Pástor	**0.213**	**0.223**	**0.230**	**0.235**

Table 3

Expected utilities of various portfolio rules with 25 risky assets when returns follow multivariate normal distribution

The table reports the expected utilities (in percentages per month) of 14 portfolio rules that choose an optimal portfolio of 25 risky assets and a riskless asset for different lengths of the estimation period (T). The excess returns of the 25 risky assets are assumed to be generated from a multivariate normal distribution, with the mean and covariance matrix chosen based on the sample estimates of Fama and French's 25 size and book-to-market ranked portfolios. The investor is assumed to have a risk aversion coefficient of three. The expected utilities of the first eight rules and the global minimum-variance rule are obtained analytically. For the other five rules, the expected utilities are approximated using 100,000 simulations.

Portfolio Rule	$T = 60$	$T = 120$	$T = 180$	$T = 240$
Parameter-certainty Optimal	1.977	1.977	1.977	1.977
Theoretical Optimal Two-fund	0.241	0.559	0.778	0.937
Theoretical Optimal Three-fund	0.531	0.852	1.019	1.133
1st Plug-in, $\hat{\Sigma}$	−46.367	−6.537	−2.305	−0.837
2nd Plug-in, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$	−44.716	−6.387	−2.254	−0.812
3rd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−12.247	−3.037	−1.072	−0.215
Bayesian (diffuse prior)	−11.785	−2.955	−1.039	−0.197
Parameter-free Optimal Two-fund	−2.736	−1.166	−0.289	0.214
Estimated Optimal Two-fund	−0.047	0.415	0.668	0.851
Uncertainty Aversion Two-fund	−0.038	0.071	0.181	0.320
Global Minimum-Variance	**0.186**	0.490	0.591	0.641
Jorion's Shrinkage	−3.692	−0.201	0.509	0.829
Estimated Optimal Three-fund	−0.022	**0.600**	**0.849**	**1.002**
MacKinlay-Pástor	0.038	0.217	0.269	0.296
Portfolio Rule	$T = 300$	$T = 360$	$T = 420$	$T = 480$
Parameter-certainty Optimal	1.977	1.977	1.977	1.977
Theoretical Optimal Two-fund	1.060	1.156	1.234	1.299
Theoretical Optimal Three-fund	1.221	1.290	1.347	1.395
1st Plug-in, $\hat{\Sigma}$	−0.108	0.324	0.610	0.811
2nd Plug-in, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$	−0.093	0.334	0.617	0.817
3rd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$	0.266	0.574	0.788	0.945
Bayesian (diffuse prior)	0.277	0.582	0.793	0.949
Parameter-free Optimal Two-fund	0.537	0.760	0.924	1.048
Estimated Optimal Two-fund	0.991	1.101	1.190	1.262
Uncertainty Aversion Two-fund	0.466	0.599	0.716	0.816
Global Minimum-Variance	0.671	0.691	0.705	0.716
Jorion's Shrinkage	1.018	1.145	1.238	1.309
Estimated Optimal Three-fund	**1.114**	**1.200**	**1.271**	**1.330**
MacKinlay-Pástor	0.311	0.321	0.329	0.334

Table 4
Expected utilities of various portfolio rules with 10 risky assets when returns follow multivariate t-distribution

The table reports the expected utilities (in percentages per month) of 14 portfolio rules that choose an optimal portfolio of 10 risky assets and a riskless asset for different lengths of the estimation period (T). The excess returns of the 10 risky assets are assumed to be generated from a multivariate t-distribution with five degrees of freedom, with the mean and covariance matrix chosen based on the sample estimates of 10 size-ranked NYSE portfolios. The investor is assumed to have a risk aversion coefficient of three. The expected utilities are approximated using 100,000 simulations.

Portfolio Rule	$T = 60$	$T = 120$	$T = 180$	$T = 240$
Parameter-certainty Optimal	0.419	0.419	0.419	0.419
Theoretical Optimal Two-fund	0.051	0.093	0.127	0.153
Theoretical Optimal Three-fund	0.118	0.159	0.185	0.204
1st Plug-in, $\hat{\Sigma}$	−6.223	−1.796	−0.874	−0.488
2nd Plug-in, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$	−5.997	−1.758	−0.859	−0.480
3rd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−3.786	−1.362	−0.701	−0.396
Bayesian (diffuse prior)	−3.647	−1.332	−0.688	−0.389
Parameter-free Optimal Two-fund	−2.330	−1.042	−0.564	−0.321
Estimated Optimal Two-fund	−0.236	−0.026	0.052	0.097
Uncertainty Aversion Two-fund	**−0.001**	0.004	0.008	0.013
Global Minimum-Variance	−0.238	−0.041	0.022	0.052
Jorion's Shrinkage	−1.163	−0.297	−0.069	0.037
Estimated Optimal Three-fund	−0.468	−0.096	0.027	0.090
MacKinlay-Pástor	−0.095	**0.105**	**0.164**	**0.192**

Portfolio Rule	$T = 300$	$T = 360$	$T = 420$	$T = 480$
Parameter-certainty Optimal	0.419	0.419	0.419	0.419
Theoretical Optimal Two-fund	0.175	0.194	0.210	0.223
Theoretical Optimal Three-fund	0.220	0.233	0.245	0.254
1st Plug-in, $\hat{\Sigma}$	−0.279	−0.148	−0.057	0.006
2nd Plug-in, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$	−0.275	−0.144	−0.055	0.008
3rd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−0.222	−0.109	−0.029	0.027
Bayesian (diffuse prior)	−0.218	−0.106	−0.005	0.029
Parameter-free Optimal Two-fund	−0.175	−0.076	−0.027	0.046
Estimated Optimal Two-fund	0.129	0.154	0.175	0.191
Uncertainty Aversion Two-fund	0.019	0.027	0.035	0.044
Global Minimum-Variance	0.070	0.081	0.090	0.096
Jorion's Shrinkage	0.099	0.141	0.171	0.193
Estimated Optimal Three-fund	0.131	0.159	0.181	0.198
MacKinlay-Pástor	**0.208**	**0.219**	**0.227**	**0.232**

Table 5
Expected utilities of various portfolio rules with 25 risky assets when returns follow multivariate t-distribution

The table reports the expected utilities (in percentages per month) of 14 portfolio rules that choose an optimal portfolio of 25 risky assets and a riskless asset for different lengths of the estimation period (T). The excess returns of the 25 risky assets are assumed to be generated from a multivariate t-distribution with five degrees of freedom, with the mean and covariance matrix chosen based on the sample estimates of Fama and French's 25 size and book-to-market ranked portfolios. The investor is assumed to have a risk aversion coefficient of three. The expected utilities are approximated using 100,000 simulations.

Portfolio Rule	$T = 60$	$T = 120$	$T = 180$	$T = 240$
Parameter-certainty Optimal	1.977	1.977	1.977	1.977
Theoretical Optimal Two-fund	0.287	0.592	0.791	0.938
Theoretical Optimal Three-fund	0.417	0.764	0.949	1.073
1st Plug-in, $\hat{\Sigma}$	−68.535	−9.347	−3.495	−1.532
2nd Plug-in, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$	−66.113	−9.142	−3.427	−1.499
3rd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−18.376	−4.553	−1.849	−0.712
Bayesian (diffuse prior)	−17.694	−4.440	−1.804	−0.687
Parameter-free Optimal Two-fund	−4.278	−1.956	−0.785	−0.136
Estimated Optimal Two-fund	−0.228	0.355	0.633	0.824
Uncertainty Aversion Two-fund	−0.077	0.109	0.254	0.409
Global Minimum-Variance	−0.071	0.338	0.486	0.559
Jorion's Shrinkage	−6.410	−0.871	0.161	0.601
Estimated Optimal Three-fund	−0.426	**0.402**	**0.721**	**0.907**
MacKinlay-Pástor	**0.001**	0.199	0.258	0.288

Portfolio Rule	$T = 300$	$T = 360$	$T = 420$	$T = 480$
Parameter-certainty Optimal	1.977	1.977	1.977	1.977
Theoretical Optimal Two-fund	1.049	1.137	1.212	1.274
Theoretical Optimal Three-fund	1.166	1.237	1.298	1.349
1st Plug-in, $\hat{\Sigma}$	−0.587	−0.038	0.322	0.580
2nd Plug-in, $\bar{\Sigma} = T\hat{\Sigma}/(T-1)$	−0.568	−0.025	0.332	0.586
3rd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−0.097	0.288	0.555	0.753
Bayesian (diffuse prior)	−0.082	0.298	0.563	0.759
Parameter-free Optimal Two-fund	0.264	0.536	0.736	0.890
Estimated Optimal Two-fund	0.962	1.068	1.156	1.229
Uncertainty Aversion Two-fund	0.558	0.686	0.797	0.886
Global Minimum-Variance	0.603	0.633	0.654	0.670
Jorion's Shrinkage	0.847	1.007	1.123	1.213
Estimated Optimal Three-fund	**1.036**	**1.131**	**1.210**	**1.274**
MacKinlay-Pástor	0.305	0.316	0.324	0.330

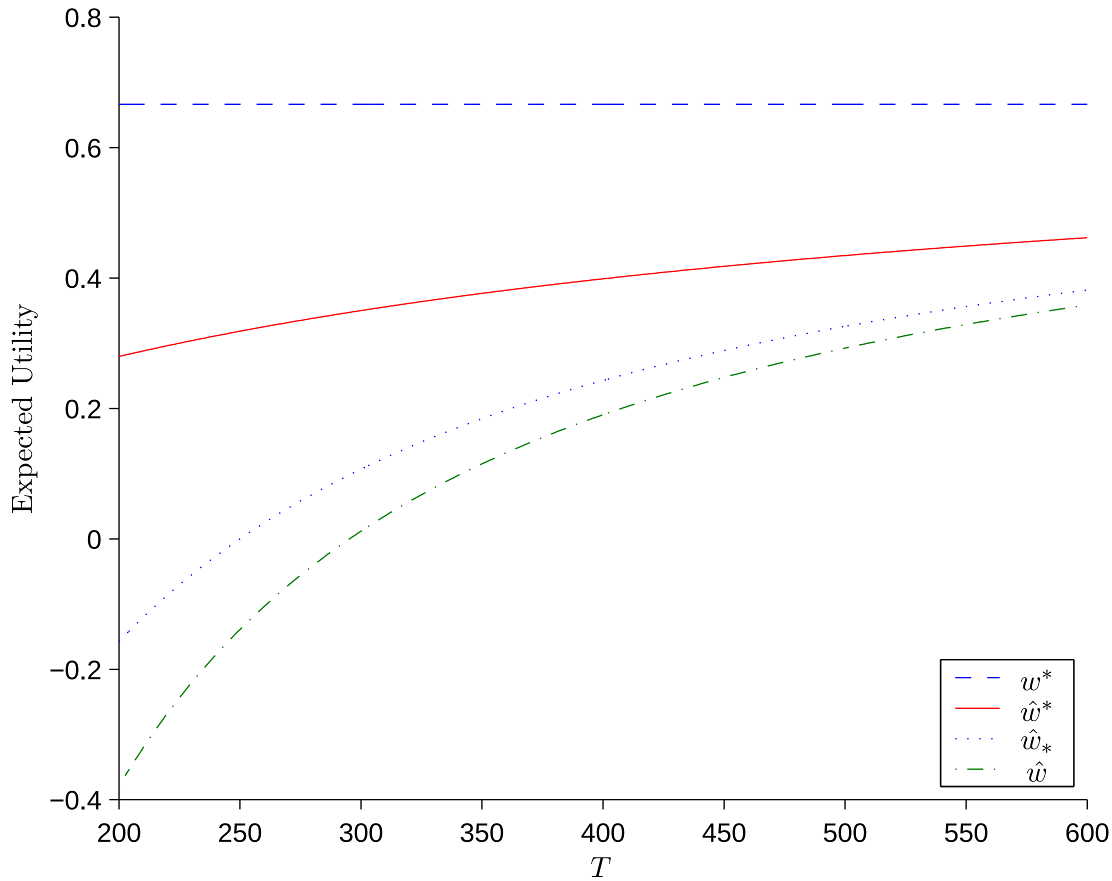

Fig. 1. Expected utility under various two-fund rules with 10 risky assets. The figure plots the expected utility (in percentage monthly returns) of an investor using different two-fund portfolio rules as a function of the length of the estimation period (T). The investor has a relative risk aversion of 3 and chooses an optimal portfolio of 10 risky assets and one riskless asset. The dashed line shows the expected utility of an investor who invests in the *ex ante* tangency portfolio, which has a Sharpe ratio (θ) of 0.2. The solid line shows the expected utility of an investor who knows θ and invests an optimal proportion in the sample tangency portfolio. The dotted line shows the expected utility of an investor who invests an optimal proportion in the sample tangency portfolio, with the weight being only a function of N and T. The dashed-dotted line shows the expected utility of an investor who holds the sample tangency portfolio with a weight determined by plugging the sample means and covariance matrix of the returns into the optimal weight formula.

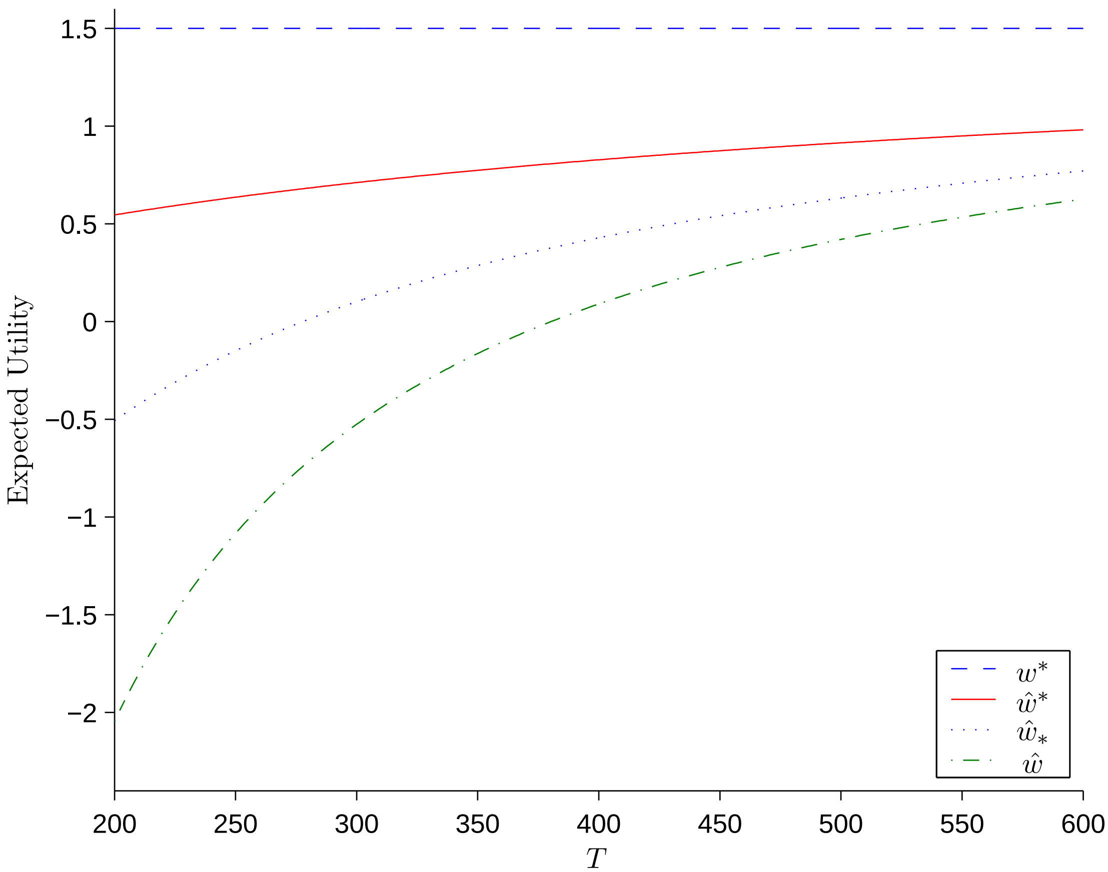

Fig. 2. Expected utility under various two-fund rules with 25 risky assets. The figure plots the expected utility (in percentage monthly return) The figure plots the expected utility (in percentage monthly returns) of an investor using different two-fund portfolio rules as a function of the length of the estimation period (T). The investor has a relative risk aversion of 3 and chooses an optimal portfolio of 25 risky assets and one riskless asset. The dashed line shows the expected utility of an investor who invests in the *ex ante* tangency portfolio, which has a Sharpe ratio (θ) of 0.3. The solid line shows the expected utility of an investor who knows θ and invests an optimal proportion in the sample tangency portfolio. The dotted line shows the expected utility of an investor who invests an optimal proportion in the sample tangency portfolio, with the weight being only a function of N and T. The dashed-dotted line shows the expected utility of an investor who holds the sample tangency portfolio with a weight determined by plugging the sample means and covariance matrix of the returns into the optimal weight formula.

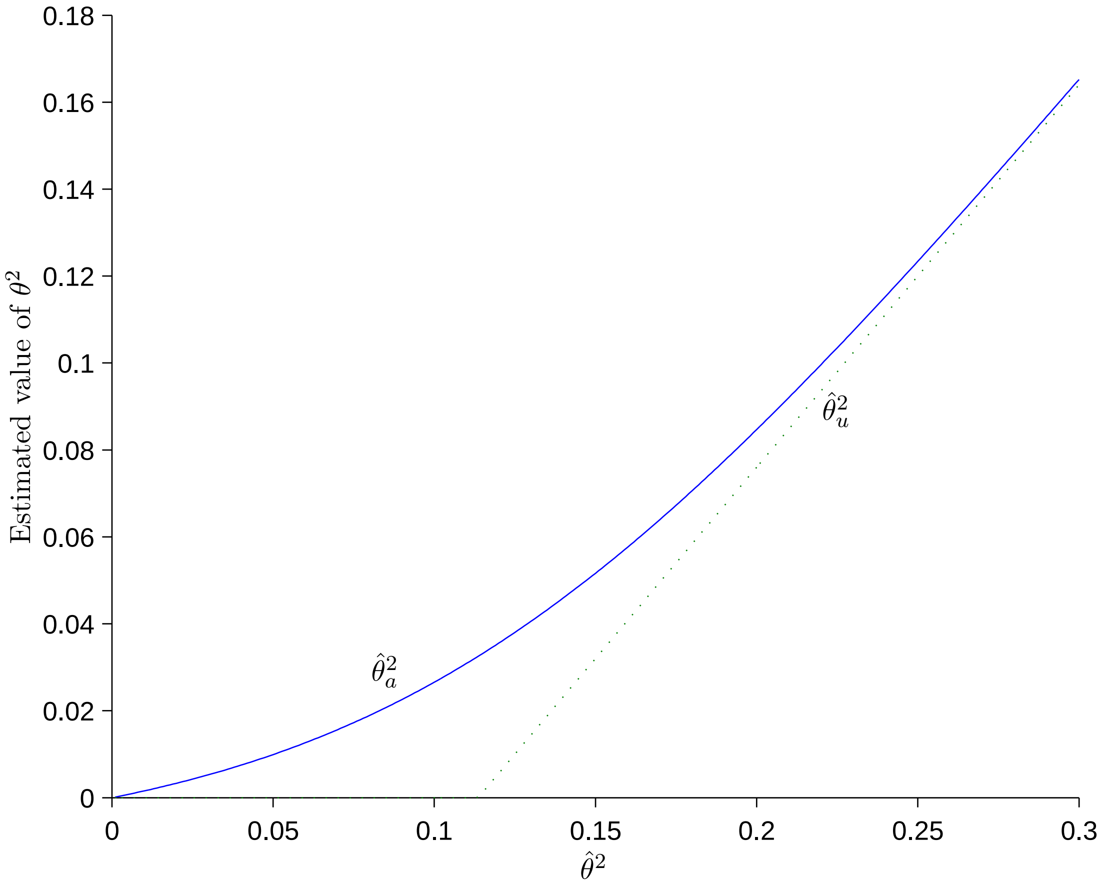

Fig. 3. Representation of two estimators of squared population Sharpe ratio for different values of squared sample Sharpe ratio. The figure plots two estimators of the squared population Sharpe ratio of the tangency portfolio (θ^2) as a function of its sample counterpart $\hat{\theta}^2$ when there are 10 assets and the sample size is 100. The dotted line is for the estimator $\hat{\theta}_u^2$, which is an unbiased estimator of θ^2. The solid line is for the adjusted estimator $\hat{\theta}_a^2$ that is due to Kubokawa, Robert, and Saleh (1993).