JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 42, No. 3, Sept. 2007, pp. 621–656

Optimal Portfolio Choice with Parameter Uncertainty

Raymond Kan and Guofu Zhou*

Abstract

In this paper, we analytically derive the expected loss function associated with using sample means and the covariance matrix of returns to estimate the optimal portfolio. Our analytical results show that the standard plug-in approach that replaces the population parameters by their sample estimates can lead to very poor out-of-sample performance. We further show that with parameter uncertainty, holding the sample tangency portfolio and the riskless asset is never optimal. An investor can benefit by holding some other risky portfolios that help reduce the estimation risk. In particular, we show that a portfolio that optimally combines the riskless asset, the sample tangency portfolio, and the sample global minimum-variance portfolio dominates a portfolio with just the riskless asset and the sample tangency portfolio, suggesting that the presence of estimation risk completely alters the theoretical recommendation of a two-fund portfolio.

I. Introduction

Theoretical models often assume that an economic agent who makes 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 estimation risk in virtually all financial models. At present, estimation risk is commonly minimized based on

*Kan, kan@chass.utoronto.ca, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario M5S 3E6, Canada, and Zhou, zhou@wustl.edu, Olin School of Business, Washington University, St. Louis, MO 63130. We thank Devraj Basu, Phelim Boyle, Stephen Brown (the editor), Alex David, Phil Dybvig, Heber Farnsworth, Lorenzo Garlappi, Campbell Harvey, Philippe Jorion (the referee), Ambrus Kecskés, Marcin Kacperczyk, John Liechty, Merrill Liechty, Hong Liu, Peter Müller, Ľuboš Pástor, Marcel Rindisbacher, Tim Simin, Robert Stambaugh, Raman Uppal, Tan Wang, and Chu Zhang for useful comments. We also thank seminar participants at Chinese University of Hong Kong, McMaster University, Pennsylvania 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, and York University as well as participants at the 2004 Northern Finance Meetings, the Seventh Annual Financial Econometrics Conference at Waterloo, and the 2005 China International Conference in Finance for helpful discussions. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada and the National Bank Financial Professorship in Capital Markets from the University of Toronto. Zhou gratefully acknowledges the China Center for Financial Research (CCFR) of Tsinghua University where part of the paper was written in May 2005.

statistical criteria such as minimum variance and asymptotic efficiency. Can the parameters be estimated in such a way that the out-of-sample performance of the model is maximized? This paper provides some answers.

A leading example of parameter uncertainty arises from the classic portfolio choice problem. Markowitz's (1952) seminal work shows that the optimal portfolio for a mean-variance investor is 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) and Meucci (2005)).[1] 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 estimation risk, simply treats the estimates as the true parameters and plugs them into the optimal portfolio formula derived under the mean-variance framework. 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 out-of-sample performance 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, as we will show, it is possible to estimate the parameters in such a way as to yield a decision rule that is uniformly better than the Bayesian approach (under a diffuse prior).

While there exist alternative ways for dealing with parameter uncertainty,[2] our study focuses on the well-defined and yet unsolved problem in the classic mean-variance framework: how should a mean-variance investor optimally estimate the portfolio weights? 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 settings.

In this paper, the first problem we study is how an investor can optimally estimate the portfolio weights if he invests only in the usual two funds: the riskless asset and the sample tangency portfolio. A similar problem is studied by ter Horst, de Roon, and Werkerzx (2002) assuming a known covariance matrix, but this restrictive assumption is not needed here. When asset returns are normally distributed, we obtain a simple closed-form formula for estimating the optimal weights in the two-fund universe. In particular, we find that a simplified version of the formula, whose construction does not rely on any unknown parameters,

[1]On issues related to this and more general portfolio problems can be found in an excellent survey by Brandt (2005).

[2]For examples, Garlappi, Uppal, and Wang (2007) and Lutgens (2004) study robust portfolio rules that maximize the worst case performance when model parameters fall within a particular confidence interval, and Harvey, Liechty, Liechty, and Müller (2004) study the optimal portfolio problem under a Bayesian setting when the returns follow a skew-normal distribution.

always yields greater expected out-of-sample performance than both the plug-in and Bayesian approaches (under a diffuse prior) no matter what the true parameter values are. A recent paper by Mori (2004) also studies a similar problem under general linear constraints on portfolio weights.[3]

The second problem we study is whether a three-fund portfolio can increase the expected out-of-sample performance even further, i.e., whether a new risky portfolio can be added into the riskless asset and the sample tangency portfolio so as to improve the expected out-of-sample performance. If the true parameters are known, as assumed in theory, then two-fund separation holds and there is no point in 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 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 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 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 out-of-sample performance 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 the Bayesian approach under a diffuse prior is outperformed by Jorion's rule, and that our optimal three-fund rule even outperforms Jorion's rule.

The remainder of the paper is organized as follows. Section II provides the optimal decision rule when the investment universe is only the riskless asset and the sample tangency portfolio. Section III solves the optimal portfolio rule when the investment universe is enlarged by adding the sample global minimum-variance portfolio. Section IV analyzes Jorion's shrinkage portfolio rule. Section V compares the performance of all the portfolio rules with parameters calibrated from real data, and Section VI concludes.

II. 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 classic plug-in methods for estimating the optimal portfolio weights of the mean-variance theory, review the Bayesian predictive solution, and compare it with the classic plug-in estimates.

[3]After completion of this paper, we were alerted that some of our results on the first problem can be found in Mori (2004). Nevertheless, our analysis is intuitive and more relevant to the proposed problem.

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.

A. The Problem

Consider the standard portfolio choice problem of an investor who chooses a portfolio in the universe of a riskless asset and N risky assets. 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 (i.i.d.) 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 $R_{pt} = w'R_t$, so its mean and variance are given by $\mu_p = w'\mu_t$ and $\sigma_p^2 = w'\Sigma w$. The investor is assumed to choose w so as to maximize the mean-variance objective function,

$$U(w) = \mu_p - \frac{\gamma}{2}\sigma_p^2, \tag{1}$$

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

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

and the resulting expected utility is

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

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 γ, 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 $\Phi_T = \{R_1, R_2, \dots, R_T\}$ 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,

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

For an investor who uses a portfolio rule $\hat{w}$, the out-of-sample mean and variance of his portfolio are given by

$$\tilde{\mu}_p = \hat{w}'\mu, \tag{5}$$

$$\tilde{\sigma}_p^2 = \hat{w}'V\hat{w}. \tag{6}$$

Note that as $\hat{w}$ is random rather than fixed, $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ are random variables as functions of the historical returns data.

To establish a comparison of different portfolio rules, one needs to establish an objective function. It is natural to choose an objective function that is based on the average out-of-sample performance of a portfolio rule. The important question is how do we measure the out-of-sample performance of a portfolio rule. A measure that is consistent with the primary objective function is

$$\tilde{U}(\hat{w}) \quad = \quad \tilde{\mu}_p - \frac{\gamma}{2}\tilde{\sigma}_p^2 \quad = \quad \hat{w}'\mu - \frac{\gamma}{2}\hat{w}'\Sigma\hat{w}, \tag{7}$$

which is the expected utility conditional on the weights being chosen as $\hat{w}$. There can be other ways of measuring out-of-sample performance, such as, for example, the out-of-sample Sharpe ratio defined as $\tilde{\mu}_p/\tilde{\sigma}_p$. However, the Sharpe ratio may not be entirely appropriate as a performance measure because it is independent of the leverage of the portfolio, so having a suboptimal weight in the risk-free asset does not affect the Sharpe ratio of a portfolio. Our performance measure $\tilde{U}(\hat{w})$ has an attractive feature because it is a measure of the certainty equivalent of portfolio $\hat{w}$. However, it is important to realize that $\tilde{U}$ is not the expected utility in the usual sense because μ and Σ are unknown to the investor, so $\tilde{U}$ should be interpreted as an out-of-sample performance measure of using portfolio $\hat{w}$.

Note that since $\hat{w}$ is a random variable, the out-of-sample performance $\tilde{U}(\hat{w})$ is also a random variable. It is natural then to evaluate a portfolio rule based on its expected out-of-sample performance $E[\tilde{U}(\hat{w})]$. To justify this measure, we use standard statistical decision theory to define the loss function of using $\hat{w}$ as

$$L(w^*, \hat{w}) \quad = \quad U(w^*) - \tilde{U}(\hat{w}). \tag{8}$$

As $\hat{w}$ is not equal to w^* in general, the loss is strictly positive. However, $\hat{w}$ is a function of Φ_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 Φ_T. The expected loss function is called the risk function and it is defined as

$$\rho(w^*, \hat{w}) \quad = \quad E[L(w^*, \hat{w})] \quad = \quad U(w^*) - E[\tilde{U}(\hat{w})], \tag{9}$$

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

This 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 ter Horst, de Roon, 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 out-of-sample performance $E[\tilde{U}(\hat{w})]$. Note that $E[\tilde{U}(\hat{w})]$ is the expected out-of-sample performance under the true distribution of returns across repeated random samples of Φ_T. So, $E[\tilde{U}(\hat{w})]$ is the out-of-sample performance 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 out-of-sample performance 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 out-of-sample performance for every possible choice of (μ, Σ). Clearly, inadmissible portfolio rules should be eliminated from consideration.

B. 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

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \tag{10}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})', \tag{11}$$

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

We assume $T > N$ so that $\hat{\Sigma}$ is invertible. The standard plug-in portfolio rule is to replace μ and Σ in (2) by $\hat{\mu}$ and $\hat{\Sigma}$. The estimated portfolio weights using the plug-in rule are

$$\hat{w} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}. \tag{12}$$

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 $w^* = \Sigma^{-1}\mu/\gamma$. 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 out-of-sample performance.

It is interesting to compare the standard plug-in estimator $\hat{w}$ given by (12) 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,

$$\hat{\mu} \sim N(\mu, \Sigma/T), \tag{13}$$

$$\hat{\Sigma} \sim W_N(T-1, \Sigma)/T, \tag{14}$$

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

$$E[\hat{w}] = \frac{T}{T-N-2}w^*, \tag{15}$$

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 (12) tend to take bigger positions in the risky assets than those who know the true parameters.

To understand 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. Fixing the value of Σ allows us to understand the cost from estimating μ alone. The second case is also a hypothetical one in which μ is known but Σ is estimated, allowing us to understand the cost from estimating Σ alone. The third case is the more realistic one in which neither Σ nor μ is known and both need to be estimated.

The first case is the easiest one to analyze among the three. When Σ is known, the portfolio rule is $\hat{w} = \Sigma^{-1}\hat{\mu}/\gamma$, so the estimation error in $\hat{w}$ is due only 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

$$
\begin{aligned}
(16) \qquad E[\tilde{U}(\hat{w}) \mid \Sigma] &= E[\hat{w}]'\mu - \frac{\gamma}{2}E[\hat{w}'\Sigma\hat{w}] \\
&= \frac{1}{\gamma}\mu'\Sigma^{-1}\mu - \frac{1}{2\gamma}E[\hat{\mu}'\Sigma^{-1}\hat{\mu}] \\
&= \frac{1}{\gamma}\mu'\Sigma^{-1}\mu - \frac{1}{2\gamma}\left(\frac{N + T\mu'\Sigma^{-1}\mu}{T}\right) \\
&= \frac{\theta^2}{2\gamma} - \frac{N}{2\gamma T}.
\end{aligned}
$$

As a result, the risk function from using $\hat{w}$ rather than w^* is

$$
(17) \qquad \rho(w^*, \hat{w}|\Sigma) \equiv U(w^*) - E[\tilde{U}(\hat{w})|\Sigma] = \frac{N}{2\gamma T},
$$

which means that the investor expects to lose a certainty equivalent return of $N/(2\gamma T)$ 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 γ), the less he invests in the risky assets, so the smaller the impact of estimation risk. Note that the case of known Σ is similar to a continuous-time setup, 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 (17) highlights analytically the impact of the number of assets relative to the length of estimation period on the expected out-of-sample performance in discrete time.[4]

To see how uncertainty about Σ alone affects expected out-of-sample performance, consider now the case where μ is known while Σ has to be estimated. The

[4]When $\theta^2 < N/T$, we have $E[\tilde{U}(\hat{w})|\Sigma] < 0$. Because non-participation in the risky assets yields zero out-of-sample performance, the negative value of $E[\tilde{U}(\hat{w})|\Sigma]$ suggests that the investor is better off not investing in the risky assets when $\theta^2 < N/T$. 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.

optimal weights are now $\hat{w} = \hat{\Sigma}^{-1}\mu/\gamma$. 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))

$$\text{(18)} \qquad E[W^{-1}] = \left(\frac{T}{T-N-2}\right) I_N,$$

$$\text{(19)} \qquad E[W^{-2}] = \left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\right] I_N,$$

where $T > N+4$. Using these results, we know the expected out-of-sample performance is

$$\text{(20)} \quad E[\tilde{U}(\hat{w})|\mu] = \frac{1}{\gamma}E[\mu'\hat{\Sigma}^{-1}\mu] - \frac{1}{2\gamma}E[\mu'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mu]$$
$$= \frac{1}{\gamma}E\left[\mu'\Sigma^{-\frac{1}{2}}W^{-1}\Sigma^{-\frac{1}{2}}\mu\right] - \frac{1}{2\gamma}E\left[\mu'\Sigma^{-\frac{1}{2}}W^{-2}\Sigma^{-\frac{1}{2}}\mu\right]$$
$$= k_1\frac{\theta^2}{2\gamma},$$

where

$$\text{(21)} \qquad k_1 = \left(\frac{T}{T-N-2}\right)\left[2 - \frac{T(T-2)}{(T-N-1)(T-N-4)}\right].$$

Note that $1-k_1$ is the percentage loss of the expected out-of-sample performance 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, similar to the earlier case where only μ was unknown, the estimation error of $\hat{\Sigma}$ (and hence the expected loss in out-of-sample performance) also increases with the number of assets and decreases with the length of the time series.

Compared to the previous case, the investor will still sometimes avoid investing in the risky assets if he uses the portfolio rule $\hat{w} = \hat{\Sigma}^{-1}\mu/\gamma$ 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 out-of-sample performance only by the fixed amount $N/(2\gamma T)$, irrespective of the magnitude of the true parameters. In contrast, not knowing Σ (assuming μ is known) reduces the expected out-of-sample performance by a constant proportional amount that depends on the squared Sharpe ratio of the tangency portfolio.

Finally, consider the 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 (12). Using the inverse moment properties of the Wishart distribution and the fact that $\hat{\mu}$ and $\hat{\Sigma}$ are independent, we have

$$\text{(22)} \quad E[\tilde{U}(\hat{w})] = \frac{1}{\gamma}E[\hat{\mu}'\hat{\Sigma}^{-1}\mu] - \frac{1}{2\gamma}E[\hat{\mu}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}]$$
$$= \frac{1}{\gamma}E[\hat{\mu}'\Sigma^{-\frac{1}{2}}W^{-1}\Sigma^{-\frac{1}{2}}\mu] - \frac{1}{2\gamma}E[\hat{\mu}'\Sigma^{-\frac{1}{2}}W^{-2}\Sigma^{-\frac{1}{2}}\hat{\mu}]$$
$$= k_1\frac{\theta^2}{2\gamma} - \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)},$$

assuming $T > N+4$. Hence, the expected loss in out-of-sample performance is

$$(23) \quad \rho(w^*, \hat{w}) \quad = \quad (1-k_1)\frac{\theta^2}{2\gamma} + \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)}.$$

This formula explicitly relates the expected loss of out-of-sample performance to N, T, γ, and θ^2. The qualitative properties are the same as before. As N or θ^2 increases, the loss increases, and as T or γ 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 out-of-sample performance 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 out-of-sample performance or the risk function of the plug-in portfolio rule is obtained by simulation.[5] 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 address a number of important issues such as the impact of the error from estimating the covariance matrix of the returns on the expected out-of-sample performance.

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 selection in the presence of 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 the 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 out-of-sample performance 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 out-of-sample performance, we show that the general perception can be incorrect.

Table 1 reports the expected percentage loss of out-of-sample performance 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 γ. The first column presents the percentage loss of expected out-of-sample performance due to estimation error in $\hat{\mu}$ alone, i.e., $100(1 - E[\tilde{U}(\hat{w})|\Sigma]/U(w^*))$. The second column presents the percentage loss of expected out-of-sample performance due to estimation error in $\hat{\Sigma}$ alone, i.e., $100(1 - E[\tilde{U}(\hat{w})|\mu]/U(w^*))$. The fourth column presents the percentage loss of expected out-of-sample performance due to estimation errors in both $\hat{\mu}$ and $\hat{\Sigma}$, i.e., $100(1 - E[\tilde{U}(\hat{w})]/U(w^*))$. Since the effects of estimation errors in $\hat{\mu}$ and $\hat{\Sigma}$ are not additive, the third column reports the in-

[5]One exception is Brown (1978) who provides an infinite series summation formula for the expected out-of-sample performance in the one risky asset case. Another exception is Mori (2004) who provides analytical expression of the risk function for the plug-in rule under equality constraints on portfolio weights.

teractive effect of estimation errors in $\hat{\mu}$ and $\hat{\Sigma}$, whose summation with the first two columns is equal to the fourth column.

Assuming $\theta = 0.2$, Panel A of Table 1 shows that when N/T is small, the estimation error in $\hat{\mu}$ indeed accounts for most of the loss of out-of-sample performance, often more than 10 times the loss out-of-sample performance from the estimation error in $\hat{\Sigma}$. However, when N/T is large, the reduction of out-of-sample performance 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 loss of out-of-sample performance 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 out-of-sample performance due to the estimation error in $\hat{\mu}$ alone is smaller, while the percentage loss of out-of-sample performance 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 out-of-sample performance due to the interactive effect also goes down with the increase in θ, so as a whole the percentage loss in expected out-of-sample performance 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.

C. 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 (12), 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 out-of-sample performance than using $\hat{w}$.

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

$$(24) \qquad \bar{\Sigma} \;=\; \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})' \;=\; \frac{T}{T-1}\hat{\Sigma}.$$

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

$$(25) \qquad \bar{w} \;\equiv\; \frac{1}{\gamma}\bar{\Sigma}^{-1}\hat{\mu} \;=\; \frac{1}{\gamma}\left(\frac{T-1}{T}\right)\hat{\Sigma}^{-1}\hat{\mu} \;=\; \left(\frac{T-1}{T}\right)\hat{w}.$$

However, because $E[\bar{w}] = ((T-1)/(T-N-2))w^*$, such a portfolio rule still involves taking larger positions in the risky assets relative to the true optimal port-

TABLE 1

Percentage Loss of Expected Out-of-Sample Performance Due to Estimation Errors in the Means and Covariance Matrix of Returns

Table 1 presents the percentage loss of expected out-of-sample performance 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 out-of-sample performance 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 $\theta = 0.4$.

		Percentage Loss of Expected Out-of-Sample Performance			
N	T	$\hat{\mu}$	$\hat{\Sigma}$	Interaction	$\hat{\mu}$ and $\hat{\Sigma}$
Panel A. $\theta = 0.2$					
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
Panel B. $\theta = 0.4$					
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

folio. Assuming $T > N + 4$, the expected out-of-sample performance associated with portfolio rule $\bar{w}$ is

$$(26) \qquad E[\tilde{U}(\bar{w})] \quad = \quad k_2 \frac{\theta^2}{2\gamma} - \frac{N(T-1)^2(T-2)}{2\gamma T(T-N-1)(T-N-2)(T-N-4)},$$

where

$$(27) \qquad k_2 \quad = \quad \left(\frac{T-1}{T-N-2}\right)\left[2 - \frac{(T-1)(T-2)}{(T-N-1)(T-N-4)}\right].$$

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

The third plug-in approach is to estimate Σ with

$$(28) \qquad \tilde{\Sigma} \quad = \quad \frac{1}{T-N-2}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})' \quad = \quad \frac{T}{T-N-2}\hat{\Sigma}.$$

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

$$(29) \qquad \tilde{w} \quad \equiv \quad \frac{1}{\gamma}\tilde{\Sigma}^{-1}\hat{\mu} \quad = \quad \frac{T-N-2}{T}\hat{w}.$$

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 $T > N + 4$,

$$(30) \qquad E[\tilde{U}(\tilde{w})] \quad = \quad k_3 \frac{\theta^2}{2\gamma} - \frac{N(T-2)(T-N-2)}{2\gamma T(T-N-1)(T-N-4)},$$

where

$$(31) \qquad k_3 \quad = \quad 2 - \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}.$$

It is straightforward to verify that $E[\tilde{U}(\tilde{w})]$ is greater than $E[\tilde{U}(\bar{w})]$, 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 out-of-sample performance 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, that achieves the highest expected out-of-sample performance, while the maximum likelihood estimate yields the lowest.

D. Bayesian Solution

While the plug-in method ignores 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 classic framework, utility is defined with respect to the parameters μ and Σ. The Bayesian approach deals with parameter uncertainty by assuming the investor cares about the expected utility under the predictive distribution $p(R_{T+1}|\Phi_T)$, which is determined by both the historical data and the prior. With a good choice of prior (say highly centered around the true values), there is no doubt that a Bayesian portfolio rule can substantially outperform the classic plug-in rules. However, it is not entirely clear how a good prior can be obtained. For a fair comparison with the classic plug-in rules, we assume a diffuse prior here. Brown (1976), Klein and Bawa (1976), and Stambaugh (1997) show under the standard diffuse prior on μ and Σ,

$$(32) \qquad p_0(\mu, \Sigma) \quad \propto \quad |\Sigma|^{-\frac{N+1}{2}},$$

the Bayesian optimal portfolio weights have the same formula as w^* except for the parameters being replaced by their predictive moments,

$$(33) \qquad \hat{w}^{\text{Bayes}} \quad = \quad \frac{1}{\gamma}\left(\frac{T-N-2}{T+1}\right)\hat{\Sigma}^{-1}\hat{\mu}.$$

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

$$(34) \qquad E[\hat{w}^{\text{Bayes}}] \quad = \quad \left(\frac{T}{T+1}\right)w^*,$$

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 estimation risk explicitly and, hence, the risky assets become riskier, while the riskless rate is known for sure. So, all else being equal, the riskless asset becomes more attractive and, hence, the Bayesian investor invests more in it.

While the Bayesian portfolio rule is optimal by design in maximizing the expected utility based on the predictive distribution of the returns, will it have better out-of-sample performance than the classic plug-in methods? Simulations by Brown (1976) and Stambaugh (1997) suggest that it does outperform the plug-in methods. We provide here an analytical proof for this result.

Using the same technique for evaluating $E[\tilde{U}(\hat{w})]$, we have

$$(35) \qquad E[\tilde{U}(\hat{w}^{\text{Bayes}})] \quad = \quad k_4\frac{\theta^2}{2\gamma} - \frac{NT(T-2)(T-N-2)}{2\gamma(T+1)^2(T-N-1)(T-N-4)},$$

where $T > N+4$ and

$$(36) \qquad k_4 \quad = \quad \left(\frac{T}{T+1}\right)\left[2 - \frac{T(T-2)(T-N-2)}{(T+1)(T-N-1)(T-N-4)}\right].$$

Therefore,

$$(37) \qquad E[\tilde{U}(\hat{w}^{\text{Bayes}})] - E[\tilde{U}(\tilde{w})] \quad = \quad (k_4 - k_3)\frac{\theta^2}{2\gamma} + \frac{N(T-2)(T-N-2)(2T+1)}{2\gamma T(T+1)^2(T-N-1)(T-N-4)}.$$

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

$$(38) \quad k_4 - k_3 \quad = \quad \frac{(T^2 + 6T - 4) + N[2T(T - N) - 3T - 2(N + 4)]}{(T + 1)^2(T - N - 1)(T - N - 4)} \quad > \quad 0$$

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

An intuition for the better out-of-sample performance of the Bayesian portfolio rule is as follows. In the portfolio problem (unlike standard problems where risk functions are used to evaluate parameters), there is a built-in trade-off between mean and variance in the risk function. By not accounting for this trade-off, the plug-in method must fail in a risk function comparison. To some extent, the Bayesian approach exploits some of this trade-off and, hence, leads to a better portfolio rule. Brown (1978) was the first to make such a point. Our procedures in the next section are exact ways to exploit more fully the trade-offs between the mean and variance in the risk function for the portfolio problem.

The uniform dominance result suggests that investors are better off using the Bayesian portfolio rule than the classic plug-in rules. However, it turns out that the Bayesian portfolio rule is still inadmissible because we show later that there exists a portfolio rule that uniformly dominates the Bayesian portfolio rule under the diffuse prior.

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 example, 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 an analytical solution of the risk function for such portfolio rules. To limit the scope of this paper, we will in what follows focus only on the diffuse prior, leaving the study of informative Bayesian portfolio rules for future research.

E. Optimal Two-Fund Rule

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

$$(39) \qquad \hat{w} \quad = \quad f(\hat{\mu}, \hat{\Sigma}).$$

The economic question of interest to the investor is to find a function $f(\hat{\mu}, \hat{\Sigma})$ so that the expected out-of-sample performance is maximized. This function can potentially be a very complex nonlinear function of $\hat{\mu}$ and $\hat{\Sigma}$, and there can 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 class of portfolio rules that hold just the riskless asset and the sample tangency portfolio, and then turn to a more general three-fund portfolio rule in the next section.

Although both the plug-in and the Bayesian rules suggest holding the riskless asset and the sample tangency portfolio, their weights on the sample tangency portfolio are not necessarily optimal in terms of maximizing the expected out-of-sample performance. Indeed, consider the class of two-fund portfolio rules that have weights

$$\hat{w} = \frac{c}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}, \tag{40}$$

where c is a constant scalar. All of the previous rules are special cases of this class. 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, we know that the expected out-of-sample performance of this class of portfolio rules is

$$\begin{aligned} E[\tilde{U}(c\hat{\Sigma}^{-1}\hat{\mu}/\gamma)] &= \frac{c\theta^2}{\gamma}\left(\frac{T}{T-N-2}\right) \\ &\quad - \frac{c^2}{2\gamma}\left(\theta^2 + \frac{N}{T}\right)\left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\right], \end{aligned} \tag{41}$$

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

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

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 out-of-sample performance under the optimal choice of $\hat{w}^* = c^* \hat{\Sigma}^{-1}\hat{\mu}/\gamma$ is

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

which is, of course, higher than the expected out-of-sample performance under both the classic plug-in and Bayesian rules. Compared to the case of no uncertainty,

$$\frac{E[\tilde{U}(\hat{w}^*)]}{U(w^*)} = \left[\frac{(T-N-1)(T-N-4)}{(T-2)(T-N-2)}\right]\left(\frac{\theta^2}{\theta^2 + \frac{N}{T}}\right) < 1, \tag{44}$$

which is a decreasing function of N and an increasing function of T and θ^2. As a result, the percentage loss of expected out-of-sample performance 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, a feasible portfolio rule using c^* does not exist since θ is unknown in practice. Nevertheless, c^* provides important insights into the optimal decision. In particular, c^* can yield a simple decision rule that dominates the Bayesian rule. Consider the following rule, which is optimal when $\theta^2 \to \infty$:

$$(45) \qquad \hat{w}_* = \frac{c_3}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}, \qquad c_3 = \frac{(T-N-1)(T-N-4)}{T(T-2)}.$$

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 $f(c) \equiv E[\tilde{U}(c\hat{\Sigma}^{-1}\hat{\mu}/\gamma)]$ in (41) is a quadratic function of c, so the expected out-of-sample performance 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 $T > N+4$,

$$(46) \qquad c_2 = \frac{T-N-2}{T+1} > \frac{T-N-4}{T} > \left(\frac{T-N-4}{T}\right)\left(\frac{T-N-1}{T-2}\right) = c_3,$$

and obviously $c_3 > c^*$. Thus, regardless of the value of θ^2, the expected out-of-sample performance is always greater for $\hat{w}_*$ than that under the Bayesian portfolio rule. The expected out-of-sample performance of w^* can be computed explicitly by (41) 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 $\text{tr}(c\hat{\Sigma}^{-1}\Sigma - I_N)^2$. 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 to improve upon using c_3. Without information about the value of θ^2, c_3 represents the best choice of c that maximizes the expected out-of-sample performance. However, if a priori $\theta^2 \leq \bar{\theta}^2$, but the exact value of θ^2 is not known, then

$$(47) \qquad \bar{c} = c_3\left(\frac{\bar{\theta}^2}{\bar{\theta}^2 + \frac{N}{T}}\right)$$

is a better choice of c because the expected out-of-sample performance $f(c)$ 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 out-of-sample performance. 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 loss of out-of-sample performance 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 $\gamma = 3$ chooses a portfolio out of $N = 10$ risky assets and a riskless one. Assume further that the Sharpe ratio of the ex ante tangency portfolio is $\theta = 0.2$. Figure 1 plots the expected out-of-sample performance (in percentage monthly return) of the investor under various two-fund rules for different lengths of the estimation window. If the investor knows μ and Σ, he will hold w^* for the risky assets to achieve a certainty equivalent 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 $\hat{w}^* = c^* \hat{\Sigma}^{-1}\hat{\mu}/\gamma$ and his expected out-of-sample performance as indicated by the solid line. In comparison with using w^*, there is some expected loss of out-of-sample performance from using $\hat{w}^*$. Nevertheless, the expected out-of-sample performance 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 out-of-sample performance as indicated by the dotted line in Figure 1, 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 out-of-sample performance for the standard plug-in portfolio rule $\hat{w} = \hat{\Sigma}^{-1}\hat{\mu}/\gamma$. In this case, an estimation window of at least $T = 296$ months is needed before this rule outperforms the riskless asset.

In the second example, we make the same assumptions as in the first one 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 out-of-sample performance of the investor under the four two-fund rules. For w^* and $\hat{w}^*$, the increase in the Sharpe ratio results in higher expected out-of-sample performance 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 out-of-sample performance, especially when T is small.

These two examples illustrate that while $\hat{w}_*$ improves over $\hat{w}$, it is still a mediocre portfolio rule because it delivers negative expected out-of-sample performance 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,

$$\hat{\theta}^2 = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}. \tag{48}$$

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,

$$\hat{\theta}^2 \sim \left(\frac{N}{T-N}\right) F_{N,T-N}(T\theta^2), \tag{49}$$

FIGURE 1

Expected Out-of-Sample Performance under Various Two-Fund Rules with 10 Risky Assets

Figure 1 plots the expected out-of-sample performance (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 three and chooses an optimal portfolio of 10 risky and one riskless asset. The dashed line shows the expected out-of-sample performance 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 out-of-sample performance of an investor who knows θ and invests an optimal proportion in the sample tangency portfolio. The dotted line shows the expected out-of-sample performance 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 out-of-sample performance 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.

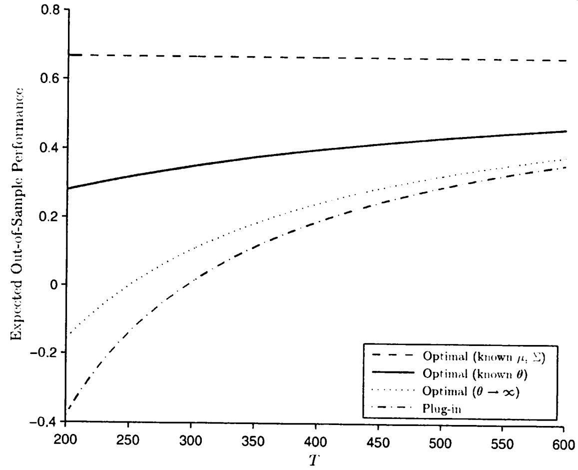
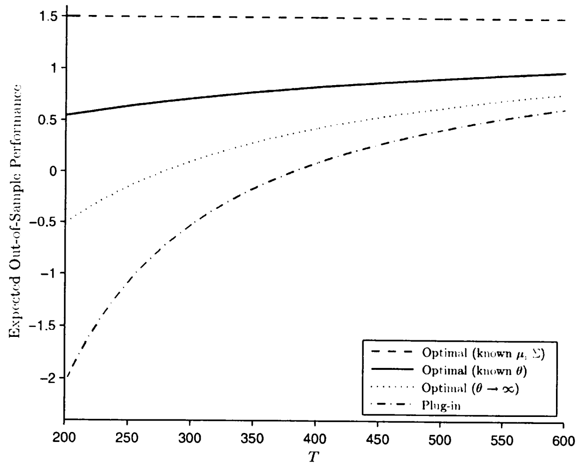


FIGURE 2

Expected Out-of-Sample Performance under Various Two-Fund Rules with 25 Risky Assets

Figure 2 plots the expected out-of-sample performance (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 three and chooses an optimal portfolio of 25 risky and one riskless asset. The dashed line shows the expected out-of-sample performance 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 out-of-sample performance of an investor who knows θ and invests an optimal proportion in the sample tangency portfolio. The dotted line shows the expected out-of-sample performance 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 out-of-sample performance 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.



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

$$(50) \qquad \hat{\theta}^2_u = \frac{(T-N-2)\hat{\theta}^2 - N}{T}.$$

However, this estimator can take negative value so it is also an undesirable 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

$$(51) \qquad \hat{\theta}^2_a = \frac{(T-N-2)\hat{\theta}^2 - N}{T} + \frac{2(\hat{\theta}^2)^{\frac{N}{2}}(1+\hat{\theta}^2)^{-\frac{T-2}{2}}}{TB_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2,(T-N)/2)},$$

where

$$(52) \qquad B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\mathrm{d}y$$

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 the unbiased estimator is too small.

Figure 3 plots $\hat{\theta}^2_a$ and $\hat{\theta}^2_u$ as a function of $\hat{\theta}^2$ for $N=10$ and $T=100$. It can be seen that $\hat{\theta}^2_a$ is an increasing and convex function of $\hat{\theta}^2$. When $\hat{\theta}^2$ is equal to zero, $\hat{\theta}^2_a=0$. As $\hat{\theta}^2$ gets larger, it becomes more like a linear function of $\hat{\theta}^2$ and behaves almost like the unbiased estimator $\hat{\theta}^2_u$. To understand the intuition why $\hat{\theta}^2_a$ 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 $\hat{\theta}^2_u$. 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 $\hat{\theta}^2_a$ should be higher than $\hat{\theta}^2_u$ when $\hat{\theta}^2_u$ is small or negative, causing $\hat{\theta}^2_a$ to be a nonlinear function of $\hat{\theta}^2$.

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

$$(53) \qquad \hat{c}^* = c_3\left(\frac{\hat{\theta}^2_a}{\hat{\theta}^2_a + \frac{N}{T}}\right),$$

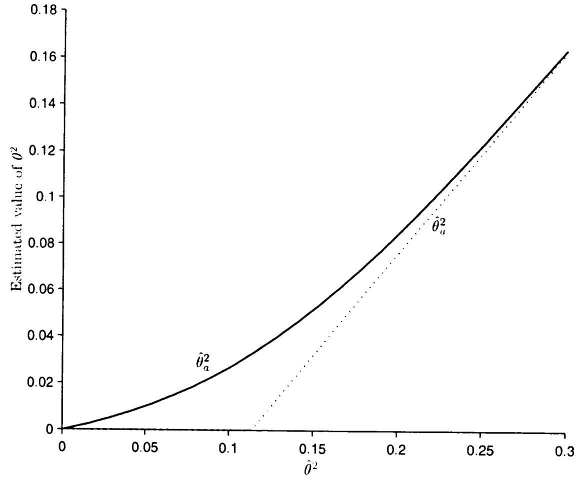
and the associated feasible two-fund optimal portfolio weights are

$$(54) \qquad \hat{w}^{\mathrm{II}} = \frac{1}{\gamma}\hat{c}^*\hat{\Sigma}^{-1}\hat{\mu}.$$

FIGURE 3

Representation of Two Estimators of the Squared Population Sharpe Ratio for Different Values of the Squared Sample Sharpe Ratio

Figure 3 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).



In comparison with c_3, $\hat{c}^*$ is random and data-dependent, so the expected out-of-sample performance of using $\hat{w}^{\text{II}}$ is intractable analytically. Nevertheless, $\hat{w}^{\text{II}}$ 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 V confirm that this is indeed the case.

Recently, Garlappi, Uppal, and Wang (2007), Proposition 3 propose an interesting two-fund rule that is optimal for an investor who exhibits uncertainty aversion. Their approach incorporates parameter uncertainty in the utility function that yields a two-fund portfolio rule,[6]

$$\hat{w}_{ua} = \frac{c_{ua}}{\gamma} \bar{\Sigma}^{-1} \hat{\mu}, \tag{55}$$

where

$$c_{ua} = \begin{cases} 1 - (\varepsilon/\hat{\theta}^2)^{\frac{1}{2}} & \text{if} \quad \hat{\theta}^2 > \varepsilon, \\ 0 & \text{if} \quad \hat{\theta}^2 \leq \varepsilon, \end{cases} \tag{56}$$

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

[6]Although Garlappi, Uppal, and Wang (2007) 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.

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 (2007).[7] 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 out-of-sample performance, the uncertainty aversion two-fund rule cannot outperform our theoretical optimal two-fund rule by design. However, since our optimal two-fund rule has to be estimated, it is not entirely clear whether the uncertainty aversion two-fund rule is always outperformed by our estimated optimal two-fund rule. This issue will be addressed by using simulations in Section V.

III. Three-Fund Separation: Investing on the Ex Post Frontier

Theoretically, if a mean-variance optimizing investor knows the true parameters, he should invest only 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 is analyzed in detail in the previous section.

However, investing in only the two funds generates a loss in expected out-of-sample performance, as shown below. Intuitively, if there is parameter uncertainty, use of another risky portfolio can help to diversify 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 out-of-sample performance 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 depend only 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.[8]

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

[8] 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 out-of-sample performance than the one that we propose.

Consider a portfolio rule of the form,

$$(57)\qquad \hat{w} = \hat{w}(c,d) = \frac{1}{\gamma}\left(c\hat{\Sigma}^{-1}\hat{\mu} + d\hat{\Sigma}^{-1}1_N\right),$$

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 out-of-sample performance is

$$\begin{aligned}(58)\quad & E[\tilde{U}(\hat{w}(c,d))] \\ &= E[\hat{w}(c,d)]'\mu - \frac{\gamma}{2}E[\hat{w}(c,d)'\Sigma\hat{w}(c,d)] \\ &= \left(\frac{T}{T-N-2}\right)\frac{1}{2\gamma}\left[2(c\mu'\Sigma^{-1}\mu + d\mu'\Sigma^{-1}1_N)\right. \\ &\quad - \frac{T(T-2)}{(T-N-1)(T-N-4)} \\ &\quad \left.\times\left(\left(\mu'\Sigma^{-1}\mu + \frac{N}{T}\right)c^2 + 2(\mu'\Sigma^{-1}1_N)cd + (1_N'\Sigma^{-1}1_N)d^2\right)\right],\end{aligned}$$

where $T > N+4$. Differentiating with respect to c and d, we obtain the c and d that maximize the expected out-of-sample performance as

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

$$(60)\qquad d^{**} = c_3\left(\frac{\frac{N}{T}}{\psi^2 + \frac{N}{T}}\right)\mu_g,$$

where

$$(61)\quad \psi^2 = \mu'\Sigma^{-1}\mu - \frac{(\mu'\Sigma^{-1}1_N)^2}{1_N'\Sigma^{-1}1_N} = (\mu - \mu_g 1_N)'\Sigma^{-1}(\mu - \mu_g 1_N)$$

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

$$(62)\qquad \hat{w}^{**} = \frac{c_3}{\gamma}\left[\left(\frac{\psi^2}{\psi^2 + \frac{N}{T}}\right)\hat{\Sigma}^{-1}\hat{\mu} + \left(\frac{\frac{N}{T}}{\psi^2 + \frac{N}{T}}\right)\mu_g\hat{\Sigma}^{-1}1_N\right].$$

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 in only the sample global minimum-variance portfolio. Since $c^{**} > 0$ whenever $T > N + 4$, investing in just the sample global minimum-variance portfolio is clearly suboptimal. 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 out-of-sample performance is

$$(63)\quad E[\tilde{U}(\hat{w}^{**})] = \frac{\theta^2}{2\gamma}\frac{(T-N-1)(T-N-4)}{(T-2)(T-N-2)}\left[1 - \frac{\frac{N}{T}}{\theta^2 + \left(\frac{\theta^2}{\psi^2}\right)\left(\frac{N}{T}\right)}\right],$$

when $T > N + 4$. In the presence of parameter uncertainty in both μ and Σ, this is the highest expected out-of-sample performance obtained so far. However, this level of expected out-of-sample performance 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 and ψ^2, we can use their sample counterparts,

$$(64)\quad \hat{\mu}_g = \frac{\hat{\mu}'\hat{\Sigma}^{-1}1_N}{1_N'\hat{\Sigma}^{-1}1_N},$$

$$(65)\quad \hat{\psi}^2 = (\hat{\mu} - \hat{\mu}_g 1_N)'\hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N).$$

In the Appendix, we show that

$$(66)\quad \frac{(T-N+1)\hat{\psi}^2}{N-1} \sim F_{N-1,T-N+1}(T\psi^2),$$

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 $\hat{\theta}_a^2$, we use

$$(67)\quad \hat{\psi}_a^2 = \frac{(T-N-1)\hat{\psi}^2 - (N-1)}{T} + \frac{2(\hat{\psi}^2)^{\frac{N-1}{2}}(1+\hat{\psi}^2)^{-\frac{T-2}{2}}}{TB_{\hat{\psi}^2/(1+\hat{\psi}^2)}((N-1)/2, (T-N+1)/2)}$$

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

$$(68)\quad \hat{w}^{\mathrm{III}} = \frac{c_3}{\gamma}\left[\left(\frac{\hat{\psi}_a^2}{\hat{\psi}_a^2 + \frac{N}{T}}\right)\hat{\Sigma}^{-1}\hat{\mu} + \left(\frac{\frac{N}{T}}{\hat{\psi}_a^2 + \frac{N}{T}}\right)\hat{\mu}_g\hat{\Sigma}^{-1}1_N\right].$$

Like $\hat{w}^{\mathrm{II}}$, $\hat{w}^{\mathrm{III}}$ 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 $\hat{w}^*$, $\hat{w}^{\mathrm{III}}$ is expected to dominate $\hat{w}^{\mathrm{II}}$. But analytical comparison is difficult because $\hat{\psi}_a^2$ and $\hat{\mu}_g$ are dependent on random data samples. Section V provides simulations on $E[\tilde{U}(\hat{w}^{\mathrm{III}})]$, and the results indeed show that the expected out-of-sample performance under this three-fund rule $\hat{w}^{\mathrm{III}}$ tends to be higher than those under the two-fund rule $\hat{w}^{\mathrm{II}}$.

IV. 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 perspective).

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

$$\hat{\mu}^{\mathrm{BS}} = (1 - v)\hat{\mu} + v\hat{\mu}_g 1_N, \tag{69}$$

where $\hat{\mu}_g$ is the shrinkage target,

$$\hat{\mu}_g = \frac{1_N'\tilde{\Sigma}^{-1}\hat{\mu}}{1_N'\tilde{\Sigma}^{-1}1_N} = \frac{1_N'\hat{\Sigma}^{-1}\hat{\mu}}{1_N'\hat{\Sigma}^{-1}1_N}, \tag{70}$$

which is the average excess return on the sample global minimum-variance portfolio, and v is the weight given to the target,

$$v = \frac{N+2}{(N+2) + T(\hat{\mu} - \hat{\mu}_g 1_N)'\tilde{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)}, \tag{71}$$

where $\tilde{\Sigma}$ is defined as in (28). From a shrinkage point of view, combining $\hat{\mu}^{\mathrm{BS}}$ with $\tilde{\Sigma}$ gives an estimator of the optimal portfolio weights.

Jorion's method is also 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,

$$\mathrm{Var}[R_{T+1} | \Sigma, \lambda, \Phi_T] = \left(1 + \frac{1}{T+\lambda}\right)\Sigma + \frac{\lambda}{T(T+1+\lambda)}\frac{1_N 1_N'}{1_N'\Sigma^{-1}1_N}, \tag{72}$$

where λ is a precision parameter in the following informative prior,

$$p_0(\mu | \Sigma, \mu_g, \lambda) \propto \exp\left[-\frac{1}{2}(\mu - 1_N\mu_g)'(\lambda\Sigma^{-1})(\mu - 1_N\mu_g)\right], \tag{73}$$

which leads to the shrinkage estimator $\hat{\mu}^{\mathrm{BS}}$. Theoretically, $\mathrm{Var}[R_{T+1}|\Phi_T]$ 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 $\mathrm{Var}[R_{T+1}|\Sigma, \lambda, \Phi_T]$ instead. Although Σ and λ are unknown in $\mathrm{Var}[R_{T+1}|\Sigma, \lambda, \Phi_T]$, they can be replaced by their sample estimates. In this way, Jorion's empirical Bayes-Stein estimator of the optimal portfolio weights is

$$w^{\mathrm{BS}} = \frac{1}{\gamma}(\hat{\Sigma}^{\mathrm{BS}})^{-1}\hat{\mu}^{\mathrm{BS}}, \tag{74}$$

where

$$(75) \qquad \hat{\Sigma}^{\mathrm{BS}} = \left(1+\frac{1}{T+\hat{\lambda}}\right)\tilde{\Sigma}+\frac{\hat{\lambda}}{T(T+1+\hat{\lambda})}\frac{1_N1_N'}{1_N'\tilde{\Sigma}^{-1}1_N}$$

and $\hat{\lambda}=(N+2)/[(\hat{\mu}-\hat{\mu}_g1_N)'\tilde{\Sigma}^{-1}(\hat{\mu}-\hat{\mu}_g1_N)]$.

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 (68) as

$$(76) \qquad \begin{aligned} \hat{w}^{\mathrm{III}} &= \frac{c_3}{\gamma}\left[\left(\frac{\hat{\psi}_a^2}{\hat{\psi}_a^2+\frac{N}{T}}\right)\hat{\Sigma}^{-1}\hat{\mu}+\left(\frac{\frac{N}{T}}{\hat{\psi}_a^2+\frac{N}{T}}\right)\hat{\mu}_g\hat{\Sigma}^{-1}1_N\right] \\ &= \frac{1}{\gamma}\hat{\Sigma}_*^{-1}\left[\left(\frac{T\hat{\psi}_a^2}{N+T\hat{\psi}_a^2}\right)\hat{\mu}+\left(\frac{N}{N+T\hat{\psi}_a^2}\right)\hat{\mu}_g1_N\right]. \end{aligned}$$

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 $\hat{\Sigma}^{\mathrm{BS}}$ to estimate Σ and the use of the Bayes-Stein shrinkage estimator $\hat{\mu}^{\mathrm{BS}}$ with a value of $v=N/(N+T\hat{\psi}_a^2)$ to estimate μ.

Although Jorion's shrinkage portfolio rule is a three-fund rule, 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 estimator of 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 out-of-sample performance 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.

V. Comparison of Alternative Portfolio Rules

In this section, we evaluate the expected out-of-sample performance of 13 portfolio rules for a mean-variance investor with parameters calibrated from real data. While the rules are developed under the multivariate 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 normality.

In what follows, we assume that the mean-variance investor has a relative risk aversion of $\gamma=3$. The expected out-of-sample performance for other values of γ can be obtained by simply rescaling the expected out-of-sample performance calculated for $\gamma=3$, so the relative rankings of the portfolio rules are independent of the choice of γ. In evaluating these portfolio rules, we consider two scenarios. In the first one, 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 gives $\theta = 0.159$, $\psi = 0.130$, and $\mu_g = 0.00444$.[9] 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.[10] For this set of 25 risky assets, our choice of μ and Σ gives $\theta = 0.344$, $\psi = 0.267$, and $\mu_g = 0.00889$.

Out of the 13 portfolio rules that have been discussed in this paper, the expected out-of-sample performance can be derived analytically for nine, and for the remaining four we rely on an efficient simulation method that proceeds as follows. For different lengths of the estimation window T, generate a random sample of $\hat{\mu}$ and $\hat{\Sigma}$ from (13) and (14). Then, construct the optimal portfolio using the various portfolio rules and compute each corresponding out-of-sample performance. The average out-of-sample performances across 100,000 simulations are then used to approximate the expected out-of-sample performances.

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 out-of-sample performance 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 out-of-sample performance. The loss is a decreasing function of the length of the estimation period, but even for $T = 480$ months, the expected out-of-sample performance 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 out-of-sample performance 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 out-of-sample performance is significant when T is small. Note that these three rules cannot be implemented in practice, so their expected out-of-sample performances are provided only as reference points.

In the next eight rows of Table 2, we report the expected out-of-sample performances 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 that essentially estimates Σ by $(T+1)\hat{\Sigma}/(T-N-2)$. The fifth row reports the expected out-of-sample per-

[9]The individual elements of μ and Σ are not reported because it can be shown that the expected out-of-sample performance of all our portfolio rules are a function of only θ, ψ, and μ_g.

[10]We are grateful to Ken French for making this data available on his Web site at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

TABLE 2

Expected Out-of-Sample Performance of Various Portfolio Rules with 10 Risky Assets When Returns Follow a Multivariate Normal Distribution

Table 2 reports the expected out-of-sample performance (in percentages per month) of 13 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 out-of-sample performance of the first eight rules and the global minimum-variance rule are obtained analytically. For the other four rules, the expected out-of-sample performances 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, $\tilde{\Sigma} = T\hat{\Sigma}/(T-1)$	−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

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, $\tilde{\Sigma} = T\hat{\Sigma}/(T-1)$	−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

formance 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.[11]

When the sample size is as large as $T = 360$ months, a history of 30-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. Quite the contrary, due to the volatilities of the estimates, the expected out-of-sample performance of using the plug-in methods is in fact negative! This also includes the Bayesian portfolio rule. As the sample size decreases from 360 months, the problem is exacerbated. These results show clearly that blindly substituting sample estimates for population parameters can cause a significant reduction in out-of-sample performance.

[11]One may suspect that the poor performance of the plug-in rules is caused by the sample tangency portfolio falling on the inefficient side of the sample frontier. Modifying them by investing in the sample tangency portfolio only when it is on the efficient side, we find via simulations that the modified rules only provide marginal improvements when T is small, and the improvements are negligible when T is large. More importantly, the modified rules do not outperform our optimal three-fund rule even when T is small.

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 out-of-sample performance 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 out-of-sample performance when $T > 120$ months, whereas all the other plug-in rules need $T > 360$ months to yield positive expected out-of-sample performance. Nevertheless, when T is small, the estimate of θ^2 is very volatile, so the estimated optimal two-fund rule still delivers negative expected out-of-sample performance for $T \leq 120$ months.

The second to last two-fund rule is the uncertainty aversion two-fund rule of Garlappi, Uppal, and Wang (2007). It is the best rule when $T = 60$ months, even though the expected out-of-sample performance is still negative. However, it is dominated by the estimated optimal two-fund rule when $T > 120$ months because it invests too heavily in the riskless asset. Nevertheless, it should be pointed out that the uncertainty aversion rule was not designed for maximizing the expected out-of-sample performance of a mean-variance investor, so its underperformance is expected, which does not contradict in any way that it is the best rule under Garlappi, Uppal, and Wang's (2007) uncertainty aversion utility function.

The last two-fund rule is the global minimum-variance portfolio rule, which invests $\Sigma_*^{-1}1_N\hat{\mu}_g/\gamma$ in the risky assets and the rest in the riskless asset.[12] 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 analytically that the expected out-of-sample performance of this global minimum-variance portfolio rule is given by

$$
\text{(77)} \qquad \frac{(T-N-1)(T-N-4)}{(T-2)(T-N-2)}\frac{1}{2\gamma}\left(\theta^2 - \psi^2 + \frac{\frac{(T-N-5)\psi^2}{T-N-1} - \frac{T-4}{T}}{T-N-3}\right).
$$

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

The last two rows in Table 2 report the expected out-of-sample performance 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 out-of-sample performance.

---

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

From Table 2, the theoretical optimal expected out-of-sample performance 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 out-of-sample performance of the estimated and theoretical optimal three-fund rules becomes very close, suggesting that using estimated ψ^2 and μ_g is less of a problem. Nevertheless, the expected out-of-sample performance of the estimated optimal three-fund rule is 0.206%/month, still about 20% less than the expected out-of-sample performance 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 out-of-sample performance. The second effect is that with more assets, the Sharpe ratio of the tangency portfolio increases, which in turn leads to higher expected out-of-sample performance 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 out-of-sample performance 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 out-of-sample performance 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. As a result, an investor facing such a portfolio problem is better off using the estimated optimal three-fund rule for $T \geq 120$.

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 10 implementable rules, our newly developed estimated optimal three-fund rule performs remarkably well (when $T > 120$) and offers 65% improvement in the expected out-of-sample performance over the popular maximum likelihood estimator even when T is as large as 480 ($N = 25$).[13]

[13]As mentioned earlier, we also performed simulations by assuming the returns follow a multivariate t-distribution (with five degrees of freedom and with the same mean and covariance matrix as in the multivariate normality case). The rankings of the portfolio rules are largely unchanged, suggesting our results are robust to departure from normality. Results are available from the authors.

TABLE 3

Expected Out-of-Sample Performance of Various Portfolio Rules with 25 Risky Assets When Returns Follow a Multivariate Normal Distribution

Table 3 reports the expected out-of-sample performance (in percentages per month) of 13 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 out-of-sample performance of the first eight rules and the global minimum-variance rule are obtained analytically. For the other four rules, the expected out-of-sample performances 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, $\tilde{\Sigma} = T\hat{\Sigma}/(T-1)$	−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
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, $\tilde{\Sigma} = T\hat{\Sigma}/(T-1)$	−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

VI. 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 by 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, however, 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, which is often implemented by holding the sample tangency portfolio and the riskless asset, we show that this is not optimal because a three-fund portfolio rule obtained by combining the usual two funds and the sample global minimum-variance portfolio can improve expected out-of-sample performance substantially. In fact, the three-fund rule dominates many of the existing two-fund portfolio rules. While better rules might be discovered by future research, our finding that a three-fund portfolio rule can dominate the standard sample two-fund portfolio rules 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 holding a three-fund (or perhaps even more) portfolio.

Many potential extensions are possible. For example, it is of interest to extend our analysis to more complex dynamic portfolio choice problems, such as the setups of Barberis (2000) and Aït-Sahalia and Brandt (2001). In fact, economically better estimates can potentially be sought in many financial decisions, either in investments or in corporate finance, that involve estimation of unknown parameters with well-defined economic objectives. Hence, this paper seems to pose many interesting questions for future research. For instance, 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 not necessarily the optimally estimated market risk premium that maximizes investors' economic objective function. Accounting for parameter uncertainty (and perhaps model uncertainty too), what would be the risk premium estimate? This appears to be another interesting topic for future research.

Appendix

Proof of (49). Using Theorem 3.2.13 of Muirhead (1982), we have

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

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

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

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

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

This completes the proof. □

Proof of (51). 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

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

where

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

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

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

Using integration by parts on this integral gives

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

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

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

Therefore, the adjusted estimator of δ is

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

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. □

Proof of (66). Let

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

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

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

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

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

$$\text{(A.13)} \qquad \tilde{A}_{11} \quad = \quad \frac{1}{\tilde{\psi}^2}, \qquad \hat{A}_{11} \quad = \quad \frac{1}{\hat{\psi}^2}.$$

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),

$$\text{(A.14)} \qquad \frac{\hat{A}_{11}}{\tilde{A}_{11}/T} \quad = \quad \frac{T\tilde{\psi}^2}{\hat{\psi}^2} \quad \equiv \quad w \quad \sim \quad \chi^2_{T-N+1};$$

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

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

we have

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

This completes the proof. □

Proof of (77). The expected out-of-sample performance of the "global minimum-variance" portfolio rule is given by

$$\text{(A.17)}\qquad E\left[\tilde{U}'\left(\frac{c_3}{\gamma}\hat{\Sigma}^{-1}1_N\hat{\mu}_g\right)\right] = \frac{c_3}{\gamma}E\left[\hat{\mu}_g 1_N'\hat{\Sigma}^{-1}\mu\right] - \frac{c_3^2}{2\gamma}E\left[\hat{\mu}_g^2\left(1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right)\right].$$

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.17) as

$$\text{(A.18)}\qquad E[\hat{\mu}_g 1_N'\hat{\Sigma}^{-1}\mu] = E\left[\frac{\mu'\hat{\Sigma}^{-1}1_N}{1_N'\hat{\Sigma}^{-1}1_N}(1_N'\hat{\Sigma}^{-1}\mu)\right] = E\left[\frac{(1_N'\hat{\Sigma}^{-1}\mu)^2}{1_N'\hat{\Sigma}^{-1}1_N}\right]$$

and

$$\text{(A.19)}\qquad E\left[\hat{\mu}_g^2\left(1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right)\right] = E\left[\frac{(\mu'\hat{\Sigma}^{-1}1_N)^2}{(1_N'\hat{\Sigma}^{-1}1_N)^2}1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right] + E\left[\frac{(1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N)^2}{T(1_N'\hat{\Sigma}^{-1}1_N)^2}\right].$$

Let $\nu = \Sigma^{-\frac{1}{2}}1_N/(1_N'\Sigma^{-1}1_N)^{\frac{1}{2}}$ and $\eta = \Sigma^{-\frac{1}{2}}(\mu - \mu_g 1_N)/\psi$. 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 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

$$\text{(A.20)}\qquad A = \left([\nu, Q]'W^{-1}[\nu, Q]\right)^{-1}.$$

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

$$\text{(A.21)}\qquad u \equiv A_{11} - A_{12}A_{22}^{-1}A_{21} \sim \chi^2_{T-N}/T,$$

$$\text{(A.22)}\qquad z \equiv -A_{22}^{-\frac{1}{2}}A_{21} \sim N(0_{N-1}, I_{N-1}/\sqrt{T}),$$

$$\text{(A.23)}\qquad A_{22} \sim W_{N-1}(T-1, I_{N-1})/T,$$

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

$$\text{(A.24)}\qquad 1_N'\hat{\Sigma}^{-1}1_N = c\nu'W^{-1}\nu = \frac{c}{u},$$

$$\text{(A.25)}\qquad 1_N'\hat{\Sigma}^{-1}\mu = c\mu_g\nu'W^{-1}\nu + c^{\frac{1}{2}}\psi\eta'W^{-1}\nu = \frac{c\mu_g + c^{\frac{1}{2}}\psi e_1'A_{22}^{-\frac{1}{2}}z}{u};$$

$$\text{(A.26)}\qquad 1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N = c\nu'W^{-2}\nu = c\nu'W^{-1}(\nu\nu' + QQ')W^{-1}\nu = \frac{c(1 + z'A_{22}^{-1}z)}{u^2}.$$

Using these expressions, we can state (A.18) as

$$\text{(A.27)}\quad E\left[\frac{(1_N'\hat{\Sigma}^{-1}\mu)^2}{1_N'\hat{\Sigma}^{-1}1_N}\right] = E\left[\frac{\left(c^{\frac{1}{2}}\mu_g+\psi e_1'A_{22}^{-\frac{1}{2}}z\right)^2}{u}\right]$$
$$= \frac{T}{T-N-2}E\left[E\left[\left(c^{\frac{1}{2}}\mu_g+\psi e_1'A_{22}^{-\frac{1}{2}}z\right)^2\middle|A_{22}\right]\right]$$
$$= \frac{T}{T-N-2}E\left[c\mu_g^2+\psi^2 e_1'(TA_{22})^{-1}e_1\middle|A_{22}\right]$$
$$= \frac{T}{T-N-2}\left(\theta^2-\psi^2+\frac{\psi^2}{T-N-1}\right).$$

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.19) can be expressed as

$$\text{(A.28)}\quad E\left[\frac{(\mu'\hat{\Sigma}^{-1}1_N)^2}{(1_N'\hat{\Sigma}^{-1}1_N)^2}1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right] = E\left[\frac{(c^{\frac{1}{2}}\mu_g+\psi e_1'A_{22}^{-\frac{1}{2}}z)^2(1+z'A_{22}^{-1}z)}{u^2}\right]$$
$$= \frac{T^2E[(c\mu_g^2+2\psi c^{\frac{1}{2}}\mu_g e_1'A_{22}^{-1}z+\psi^2z'A_{22}^{-\frac{1}{2}}e_1e_1'A_{22}^{-\frac{1}{2}}z)(1+z'A_{22}^{-1}z)]}{(T-N-2)(T-N-4)}$$
$$= \frac{T^2E[(c\mu_g^2+\psi^2z'A_{22}^{-\frac{1}{2}}e_1e_1'A_{22}^{-\frac{1}{2}}z)(1+z'A_{22}^{-1}z)]}{(T-N-2)(T-N-4)}.$$

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

$$\text{(A.29)}\quad c\mu_g^2E[1+z'A_{22}^{-1}z] = c\mu_g^2\left(1+E\left[\frac{u_1}{u_2}\right]\right) = c\mu_g^2\left(1+\frac{N-1}{T-N-1}\right)$$
$$= \frac{(T-2)(\theta^2-\psi^2)}{T-N-1}.$$

The second term in the expectation is

$$\text{(A.30)}\quad E[\psi^2z'A_{22}^{-\frac{1}{2}}e_1e_1'A_{22}^{-\frac{1}{2}}z(1+z'A_{22}^{-1}z)]$$
$$= \psi^2\left(E[z'A_{22}^{-\frac{1}{2}}e_1e_1'A_{22}^{-\frac{1}{2}}z]+E[z'A_{22}^{-\frac{1}{2}}e_1e_1'A_{22}^{-\frac{1}{2}}zz'A_{22}^{-1}z)]\right)$$
$$= \psi^2\left(E[e_1'(TA_{22})^{-1}e_1]+E[e_1'(TA_{22})^{-1}e_1\mathrm{tr}((TA_{22})^{-1})]+2E[e_1'(TA_{22})^{-2}e_1]\right)$$
$$= \psi^2\left[\frac{1}{T-N-1}+\frac{(T-N)(N-1)-2(N-2)}{(T-N)(T-N-1)(T-N-3)}+\frac{2(T-2)}{(T-N)(T-N-1)(T-N-3)}\right]$$
$$= \frac{(T-2)\psi^2}{(T-N-1)(T-N-3)}.$$

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

$$\text{(A.31)}\quad E\left[\frac{(\mu'\hat{\Sigma}^{-1}1_N)^2}{(1_N'\hat{\Sigma}^{-1}1_N)^2}1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right] = \frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\left[\theta^2-\left(\frac{T-N-4}{T-N-3}\right)\psi^2\right].$$

For the second term in (A.19), it can be expressed as

$$\begin{aligned}\text{(A.32)}\quad E\left[\frac{(1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N)^2}{T(1_N'\hat{\Sigma}^{-1}1_N)^2}\right] &= E\left[\frac{(1+z'A_{22}^{-1}z)^2}{Tu^2}\right]\\ &= \frac{TE\left[1+2\left(\frac{u_1}{u_2}\right)+\left(\frac{u_1}{u_2}\right)^2\right]}{(T-N-2)(T-N-4)}\\ &= \frac{T\left[1+\frac{2(N-1)}{T-N-1}+\frac{(N-1)(N+1)}{(T-N-1)(T-N-3)}\right]}{(T-N-2)(T-N-4)}\\ &= \frac{T(T-2)(T-4)}{(T-N-1)(T-N-2)(T-N-3)(T-N-4)}.\end{aligned}$$

Combining all the above results, the expected out-of-sample performance is explicitly evaluated as

$$\begin{aligned}\text{(A.33)}\quad & E\left[\tilde{U}'\left(\frac{c_3}{\gamma}\hat{\Sigma}^{-1}1_N\hat{\mu}_g\right)\right]\\ &= \frac{c_3T}{\gamma}\left(\frac{\theta^2}{T-N-2}-\frac{\psi^2}{T-N-1}\right)-\frac{c_3^2T^2(T-2)}{2\gamma(T-N-1)(T-N-2)}\\ &\times\left[\left(\frac{\theta^2}{T-N-4}-\frac{\psi^2}{T-N-3}\right)+\frac{T-4}{T(T-N-3)(T-N-4)}\right].\end{aligned}$$

After some simplification, we obtain (77). This completes the proof. □

References

Aït-Sahalia, Y., and M. Brandt. "Variable Selection for Portfolio Choice." *Journal of Finance*, 56 (2001), 1297–1351.

Barberis, N. "Investing for the Long Run when Returns are Predictable." *Journal of Finance*, 55 (2000), 225–264.

Bawa, V. S.; S. J. Brown; and R. W. Klein. *Estimation Risk And Optimal Portfolio Choice*. Amsterdam: North-Holland (1979).

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

Brandt, M. "Portfolio Choice Problems." In *Handbook of Financial Econometrics*, Y. Ait-Sahalia and L. P. Hansen, eds. Amsterdam: Elsevier (2005).

Brown, S. J. "Optimal Portfolio Choice under Uncertainty." Ph.D. Diss., University of Chicago (1976).

Brown, S. J. "The Portfolio Choice Problem: Comparison of Certainty Equivalence and Optimal Bayes Portfolios." *Communications in Statistics—Simulation and Computation*, 7 (1978), 321–334.

Chen, R., and R. Kan. "Finite Sample Analysis of Two-Pass Cross-Sectional Regressions." Working Paper, University of Toronto (2004).

Chopra, V. K., and W. T. Ziemba. "The Effects of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice." *Journal of Portfolio Management*, 19 (1993), 6–11.

Fama, E. F., and K. R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

Fourdrinier, D.; A. Philippe; and C. P. Robert. "Estimation of a Non-Centrality Parameter under Stein-Type-Like Losses." *Journal of Statistical Planning and Inference*, 87 (2000), 43–54.

Frost, P. A., and J. E. Savarino. "An Empirical Bayes Approach to Efficient Portfolio Selection." *Journal of Financial and Quantitative Analysis*, 21 (1986), 293–305.

Garlappi, L.; R. Uppal; and T. Wang. "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach." *Review of Financial Studies*, 20 (2007), 41–81.

Haff, L. R. "An Identity for the Wishart Distribution with Applications." *Journal of Multivariate Analysis*, 9 (1979), 531–544.

Harvey, C. R.; J. Liechty; M. Liechty; and P. Müller. "Portfolio Selection with Higher Moments." Working Paper, Duke University (2003).

Jobson, J. D.; B. Korkie; and V. Ratti. "Improved Estimation for Markowitz Portfolios Using James-Stein Type Estimators." *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, 41 (1979), 279–284.

Jorion, P. "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis*, 21 (1986), 279–292.

Jorion, P. "Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection." *Journal of Banking and Finance*, 15 (1991), 717–727.

Kacperczyk, M. "Asset Allocation under Distribution Uncertainty." Working Paper, University of British Columbia (2004).

Kandel, S., and R. F. Stambaugh. "On the Predictability of Stock Returns: An Asset-Allocation Perspective." *Journal of Finance*, 51 (1996), 385–424.

Klein, R. W., and V. S. Bawa. "The Effect of Estimation Risk on Optimal Portfolio Choice." *Journal of Financial Economics*, 3 (1976), 215–231.

Kubokawa, T.; C. P. Robert; and A. K. Md. E. Saleh. "Estimation of Noncentrality Parameters." *Canadian Journal of Statistics*, 21 (1993), 45–57.

Litterman, B. *Modern Investment Management: An Equilibrium Approach*. New York: Wiley (2003).

Lutgens, F. "Robust Portfolio Optimization." Ph.D. Diss., University of Maastricht (2004).

Markowitz, H. M. "Mean-Variance Analysis in Portfolio Choice and Capital Markets." *Journal of Finance*, 7 (1952), 77–91.

Meucci, A. *Risk and Asset Allocation*. New York: Springer-Verlag (2005).

Mori, H. "Finite Sample Properties of Estimators for the Optimal Portfolio Weight." *Journal of Japan Statistical Society*, 34 (2004), 27–46.

Muirhead, R. J. *Aspects of Multivariate Statistical Theory*. New York: Wiley (1982).

Pástor, L. "Portfolio Selection and Asset Pricing Models." *Journal of Finance*, 55 (2000), 179–223.

Pástor, L., and R. F. Stambaugh. "Comparing Asset Pricing Models: An Investment Perspective." *Journal of Financial Economics*, 56 (2000), 335–381.

Rukhin, A. L. "Estimation of the Noncentrality Parameter of an F-Distribution." *Journal of Statistical Planning and Inference*, 35 (1993), 201–211.

Stambaugh, R. F. "Analyzing Investments Whose Histories Differ in Length." *Journal of Financial Economics*, 45 (1997), 285–331.

Stein, C. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1 (1956), 197–206.

ter Horst, J. R.; F. A. de Roon; and B. J. M. Werkerzx. "Incorporating Estimation Risk in Portfolio Choice." Working Paper, Tilburg University (2002).

Tu, J., and G. Zhou. "Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions?" *Journal of Financial Economics*, 72 (2003), 385–421.

Xia, Y. "Learning about Predictability: The Effect of Parameter Uncertainty on Dynamic Asset Allocation." *Journal of Finance*, 56 (2001), 205–246.

Zellner, A., and V. K. Chetty. "Prediction and Decision Problems in Regression Models from the Bayesian Point of View." *Journal of the American Statistical Association*, 60 (1965), 608–616.