A Shrinkage Approach to Model Uncertainty and Asset Allocation

Author(s): Zhenyu Wang

Source: *The Review of Financial Studies*, Summer, 2005, Vol. 18, No. 2 (Summer, 2005), pp. 673-705

Published by: Oxford University Press. Sponsor: The Society for Financial Studies.

Stable URL: https://www.jstor.org/stable/3598049

A Shrinkage Approach to Model Uncertainty and Asset Allocation

Zhenyu Wang
University of Texas at Austin

This article takes a shrinkage approach to examine the empirical implications of aversion to model uncertainty. The shrinkage approach explicitly shows how predictive distributions incorporate data and prior beliefs. It enables us to solve the optimal portfolios for uncertainty-averse investors. Aversion to uncertainty about the capital asset pricing model leads investors to hold a portfolio that is not mean-variance efficient for any predictive distribution. However, mean-variance efficient portfolios corresponding to extremely strong beliefs in the Fama–French model are approximately optimal for uncertainty-averse investors. The empirical Bayes approach does not result in optimal portfolios for investors who are averse to model uncertainty.

In standard finance theory, investors optimally allocate their investment funds to assets using a given stochastic model of asset returns. It follows that the optimal asset allocation depends on the choice of the model. Uncertainty about the correct choice of a stochastic model, or simply *model uncertainty*, has recently become a research topic of interest. This issue has been well known as Knightian uncertainty in academic literature. It even appeared recently in the popular press. In the national best seller, *When Genius Failed: The Rise and Fall of Long-Term Capital Management* by Lowenstein (2000), the issue of model uncertainty is described as follows: "There is a key difference between a share of IBM ... and a pair of dice. With dice, there is *risk*—you could, after all, roll snake eyes—but there is no *uncertainty*, because you know (for certain) the chances of getting a 7 and every other result. Investing confronts us with both risk and uncertainty. There is a risk that the price of a share of IBM will fall, and there is uncertainty about how likely it is to do so." The book believes that the partners of Long-Term Capital failed to consider the

For helpful comments and discussions, I thank Torben Andersen, Andrew Ang, Geert Bekaert, Mark Carhart, Fangruo Chen, Wayne Ferson, Larry Epstein, Silverio Foresi, Larry Glosten, Rick Green, Bob Hodrick, Michael Johannes, Bob Litterman, Anthony Lynch, Lubos Pastor, Nick Polson, David Robinson, Tano Santos, Ken Singleton (editor), Rob Stambaugh, an anonymous referee, and seminar participants at Columbia University, Northwestern University, Lehman Brothers, Goldman Sachs, University of Minnesota, Federal Reserve Bank of Atlanta, University of North Carolina at Chapel Hill, University of Texas at Austin, University of Houston, Temple University, University of Southern California, Vanderbilt University, American Finance Association Meeting in 2001, and National Bureau of Economic Research Summer Institute in 2002. I am especially grateful to Gur Huberman and Ravi Jagannathan for their encouragement and suggestions since the beginning of the project. Adam Kolasinski provided excellent research assistance. Some of the earlier versions have been circulated since April 10, 1999 under the title "A shrinkage theory of asset allocation using asset-pricing models." Address correspondence to: Department of Finance, School of Business, University of Texas at Austin, Austin, TX 78712-1179, or e-mail: zw274@mail.utexas.edu.

The Review of Financial Studies Vol. 18, No. 2
doi:10.1093/rfs/hhi014 Advance Access publication February 10, 2005

chances that their models are wrong. When models are likely to be wrong and the correct model is unknown, investors may be averse to the model uncertainty. I examine the empirical implications of investor's aversion to model uncertainty in an asset allocation problem.

Finance professionals often impose restrictions on equilibrium models when specifying the distribution of asset returns. A major reason is that using sample estimates of the mean and variance to obtain optimal allocation over a large number of risky assets is well known to be problematic [Best and Grauer (1991), Britten-Jones (1999), Green and Hollifield (1992), Michaud (1989), Jobson and Korkie (1980), Polson and Tew (2000)]. The resulting portfolio usually contains extremely long and short positions, which are difficult to implement. To solve this problem, portfolio weights are often constrained to be non-negative, but the optimal portfolios are usually corner solutions that assign zero weights to many assets and nonzero weights to a few highly risky assets. Such corner solutions conflict with investors' intuition about the benefit of diversification. More importantly, the optimal portfolio weights are extremely sensitive to variations in the estimated mean and covariance matrix. Another major reason for imposing restrictions on asset-pricing models is that asset prices inconsistent with market equilibrium are believed to be unlikely to persist. Black and Litterman (1991, 1992) discuss reasons for imposing asset-pricing models.

The restriction of an asset-pricing model such as the capital asset pricing model (CAPM) or the Fama–French model reduces the dimension of the estimation problem and leads investors to allocate wealth among the factor portfolios in the model. Although restrictions of asset-pricing models help us obtain portfolios that are more intuitive and easier to implement, we face uncertainty regarding models' pricing ability because all models are rejected in some empirical tests. Shanken (1987) argues that no model is correct and suggests that the empirical Bayesian approach is an appropriate method to apply models.

Bayesian inference, which combines the prior beliefs in models and the information in data, is currently a popular approach in empirical studies of model uncertainty [Frost and Savarino (1986), Pastor and Stambaugh (2000), Polson and Tew (2000)]. Since different prior beliefs lead to different posterior and predictive probability distributions of asset returns, researchers have tried various prior distributions of the pricing errors to see how the priors affect asset allocation. Researchers often take the empirical Bayes approach, in which prior distributions are estimated from the observed samples. In the current literature, Bayesian analysis of asset allocation does not incorporate investors' aversion to uncertainty in prior beliefs.

In the standard Bayesian analysis of asset allocation, although the optimal portfolio is obtained from the predictive distribution implied by a chosen prior, investors may have multiple prior beliefs in models and may be averse to model uncertainty. Their uncertainty aversion should affect their asset allocation. The purpose of this article is to investigate how investors, who are averse to model uncertainty, make asset allocations when they are uncertain about their prior belief in asset-pricing models. This article helps us gain more understanding of the empirically observed asset allocations.

I allow for model uncertainty using the maximin approach, in which investors choose a portfolio that maximizes the minimum expected utility. Investors solve the maxmin problem, in which they choose the portfolio that maximizes the minimum expected utility. The maxmin principle has been a tradition in Bayesian decision theory. Berger (1985) offers a detailed overview on maxmin analysis of Bayesian decisions. Using an axiomatic approach, Gilboa and Schmeidler (1989) further demonstrates that the minimum expected utility represents the preference with aversion to uncertainty about the probability distributions. A solution to the maxmin problem gives the asset allocation optimal for investors who are averse to model uncertainty. Using maxmin analysis, this article examines asset allocations associated with the models and data that attract significant attention in the empirical finance literature. The models are the CAPM related to the mean-variance utility function and Fama–French model derived from empirical experience. The data are the returns on equity portfolios. Many recent articles apply maxmin analysis to finance[1] to examine theoretical issues. I bring uncertainty aversion into empirical studies of asset allocation.

Aversion to uncertainty about an asset-pricing model may lead investors to hold a portfolio that is mean-variance inefficient for any prior belief in the model. Over the portfolios sorted by firm size and book-to-market ratio, there exists no prior belief in the CAPM such that the tangency portfolio with respect to the predictive distribution gives the optimal allocation for uncertainty-averse investors. This can be viewed as a failure of the CAPM in the context of asset allocation. With considerable doubts about both estimates, uncertainty-averse investors do not want to rely on either the estimate restricted by the CAPM or the unrestricted estimate. The optimal allocation by uncertainty-averse investors is not the tangency portfolio for the mean and variance estimated from any prior belief in the model. It is commonly observed that investors do not appear to hold mean-variance efficient portfolios. Aversion to model uncertainty is potentially an important reason.

For some models and data, however, there exist prior beliefs such that the tangency portfolios are optimal for uncertainty-averse investors.

[1] See, for examples, Epstein and Wang (1994), Chamberlain (2000), Hansen and Sargent (2001), Routledge and Zin (2001), Kogan and Wang (2002), Maenhout (2002), and Uppal and Wang (2003).

For example, I show that the tangency portfolio corresponding to the dogmatic prior belief in the Fama–French model is optimal or approximately optimal for uncertainty-averse investors whose priors are centered around the Fama–French model. The reason is that the Fama–French model is a summary of the empirical properties of the portfolios sorted by firm size and book-to-market ratio and gives a more precise estimate of the mean and variance than the unrestricted estimate. The analysis in this article also demonstrates that the tangency portfolio corresponding to the prior belief obtained from the empirical Bayes approach is not optimal for investors with aversion to model uncertainty.

In order to apply maxmin analysis to asset allocation using asset-pricing models, I develop a shrinkage approach for Bayesian analysis of asset allocation, in which the predictive distribution for a given prior is used. I show both the predictive mean and variance of asset returns from the unrestricted sample moments to the estimates restricted by the asset-pricing model. The predictive mean and variance share the same shrinkage factor, which indicates the relation between the mean and variance implied by the asset-pricing model. The shrinkage approach reveals that a Bayesian investor, facing uncertainty about an asset-pricing model, implicitly assigns a weight between the unrestricted estimate and the estimate restricted by the asset-pricing model. The weight is the shrinkage factor. For a given prior distribution, the weight on the estimate restricted by the asset-pricing model is larger if the frontier of the factor portfolios has a higher Sharpe ratio. The weight on the model is large if a long history of stationary data is not available. Investors who take the empirical Bayes approach, in which they use the observed data to estimate the prior, choose the shrinkage factor to be $\frac{1}{2}$, which assigns equal weights to the restricted and unrestricted estimates. The shrinkage approach is useful because it explicitly shows how a prior belief affects the estimates of mean and variance. It allows me to solve the maxmin problem and to obtain the properties of the solutions in the maxmin analysis.

The rest of the article is organized as follows. Section 1 gives an overview of the classic, Bayesian, and maxmin analysis in the context of asset allocation using asset-pricing models. Section 2 develops the shrinkage approach to the Bayesian and maxmin analysis. Section 3 compares the maxmin analysis with the empirical approach. Section 4 examines domestic asset allocations. Conclusions and future research are discussed in Section 5. Mathematical derivations are provided in the appendix.

1. Aversion to Model Uncertainty

In this section, we review the classic, Bayesian, and maxmin analysis in the context of asset allocation using asset-pricing models. There are m risky assets. Let r_{1t} be the $m \times 1$ vector of excess returns over the risk-free rate on the assets during period t. An asset-pricing model is given and there are k factor portfolios in the model. Let r_{2t} be the $k \times 1$ vector of excess returns on the factor portfolios during period t. The time series of T observations, denoted by $R = \{r_t\}_{t=1,\ldots,T} = \{(r'_{1t},\ r'_{2t})'\}_{t=1,\ldots,T}$, are assumed to follow a normal distribution with mean μ and variance Ω, independently across t. The mean and variance are decomposed into the following parts corresponding to the m assets and k factors:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}. \tag{1}$$

The mean and variance can be summarized by the parameters in a regression model:

$$r_{1t} = \alpha + \beta r_{2t} + u_t, \tag{2}$$

where α is the $m \times 1$ vector of Jensen's alpha, β is the $m \times k$ matrix of the betas, and u_t is the $m \times 1$ vector of the residual terms in the regression. The variance of u_t is assumed to be Σ. It follows that the mean and variance of the returns can be expressed as

$$\mu = \begin{pmatrix} \alpha + \beta\mu_2 \\ \mu_2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \beta\Omega_{22}\beta' + \Sigma & \beta\Omega_{22} \\ \Omega_{22}\beta' & \Omega_{22} \end{pmatrix}. \tag{3}$$

The asset-pricing model $\mu_1 = \beta\mu_2$ holds if and only if $\alpha = 0_{m\times 1}$, where $0_{m\times 1}$ is the $m \times 1$ vector of zeros.

In the classic framework of asset allocation using asset-pricing models, investors choose either to believe or not to believe the asset-pricing model. Those who do not believe the asset-pricing model estimate the parameters without restricting α to zero. Denote the maximum likelihood estimates of α, β, and Σ by $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\Sigma}$, respectively. The maximum likelihood estimates of the mean and variance are

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} = \begin{pmatrix} \hat{\alpha} + \hat{\beta}\hat{\mu}_2 \\ \hat{\mu}_2 \end{pmatrix}, \quad \hat{\Omega} = \begin{pmatrix} \hat{\beta}\hat{\Omega}_{22}\hat{\beta}' + \hat{\Sigma} & \hat{\beta}\hat{\Omega}_{22} \\ \hat{\Omega}_{22}\hat{\beta}' & \hat{\Omega}_{22} \end{pmatrix}, \tag{4}$$

where $\hat{\mu}_2$ and $\hat{\Omega}_{22}$ are the sample mean and variance of r_{2t}. The above estimates of $\hat{\mu}$ and $\hat{\Omega}$ are in fact equivalent to the sample mean and variance of r_t. Those investors who do not believe the model make asset allocation based on the unrestricted estimates of the mean and variance in Equation (4). Those who believe the asset-pricing model, however, impose the restriction $\alpha = 0_{m\times 1}$. Let $\bar{\beta}$ and $\bar{\Sigma}$ be the maximum likelihood estimates of β and Σ with the restriction of $\alpha = 0_{m\times 1}$. The restricted maximum likelihood estimates of the mean and variance are

$$\bar{\mu} = \begin{pmatrix} \bar{\beta}\hat{\mu}_2 \\ \hat{\mu}_2 \end{pmatrix}, \quad \bar{\Omega} = \begin{pmatrix} \bar{\beta}\hat{\Omega}_{22}\bar{\beta}' + \bar{\Sigma} & \bar{\beta}\hat{\Omega}_{22} \\ \hat{\Omega}_{22}\bar{\beta}' & \hat{\Omega}_{22} \end{pmatrix}. \tag{5}$$

Those investors who believe the model make asset allocation decisions based on the restricted estimates of the mean and variance.

Discarding the dichotomy between believing and disbelieving the model, the Bayesian framework introduces an informative prior distribution of α to represent an investor's belief in the asset-pricing model. The prior of α, conditional on Σ, is assumed to be a normal distribution with mean $0_{m\times 1}$ and variance $\theta\Sigma$, that is,

$$p(\alpha|\Sigma) = N(0_{m\times 1}, \theta\Sigma). \tag{6}$$

The parameter θ is a positive number that controls the variance of the prior distribution of Jensen's alpha. Since an asset-pricing model does not impose any restrictions on β, Σ, μ_2, or Ω_{22}, prior distributions of these parameters are assumed to be independent and non-informative. Specifically, each of the prior probability density functions of β and μ_2 is proportional to a constant. The prior probability density function of Ω_{22} is proportional to $|\Omega_{22}|^{-(k+1)/2}$, and the prior probability density function of Σ is proportional to $|\Sigma|^{-(m+1)/2}$. These are the standard specifications for non-informative prior distributions [Berger (1985), Bernardo and Smith (1994)].

In Bayesian framework, asset allocation without uncertainty aversion is based on the predictive distribution, which is the distribution of r_{T+1} conditional on the observed data R. Let $E[r_{T+1}|R, \theta]$ and $V[r_{T+1}|R, \theta]$ be the mean and variance of the predictive distribution. They depend on the parameter θ. For a given parameter θ and portfolio x, the mean-variance utility function of an investor is

$$U(x;\theta) = E[r_{T+1}|R,\theta]'x - \tfrac{1}{2}\gamma x' V[r_{T+1}|R,\theta]x, \tag{7}$$

where γ is the degree of risk aversion. Investors choose the following vector of portfolio positions on the risky assets:

$$x(\theta) = \gamma^{-1}(V[r_{T+1}|R,\theta])^{-1}E[r_{T+1}|R,\theta]. \tag{8}$$

Notice that $x(\theta)$ is not a vector of portfolio weights because the elements of the vector do not sum up to 1. In order to examine the return of the optimal portfolio, we need to scale $x(\theta)$ to portfolio weights. That is, the vector of portfolio weights is $\tilde{x}(\theta) = x(\theta)/(1'_{m+k}x(\theta))$, where 1_{m+k} is the $m+k$ dimension vector of ones. It is obvious that the vector of portfolio weights, $\tilde{x}(\theta)$, is independent of the risk aversion parameter γ in Equation (8). It is well known that the risky portfolio $\tilde{x}(\theta)$ maximizes the Sharpe ratio. This portfolio of risky assets is referred to as the tangency portfolio in the finance literature.

Suppose that the investor has uncertainty about the correct model, and is averse to such uncertainty. Such a case can be modeled as the investor having a set of priors parameterized in terms of $\theta \in [0, +\infty)$, and the utility function $\min_{\theta \in [0,+\infty]} U(x, \theta)$ [see Gilboa and Schmeidler (1989)]. The optimal portfolio choice for such an investor is to solve

$$\max_{x} \min_{\theta \in [0,+\infty)} U(x;\theta). \tag{9}$$

In order to show that the solution x^* to this maxmin problem exists and is unique, we employ the shrinkage approach developed in the next section. The solution x^* is sometimes referred to as robust portfolio in the literature. The asset allocation of x^* is optimal for investors who are averse to model uncertainty. Fixing x^*, we can find the solution θ^* to $\min_\theta U(x^*; \theta)$. It is well known that x^* is not necessarily the solution to $\max_x U(x, \theta^*)$. That is, x^* can be different from the asset allocation obtained by setting $\theta = \theta^*$ in Equation (8). The solution x^* and its portfolio weights depend on the risk aversion parameter γ.

The set of predictive distributions obtained by changing θ is not a convex set, that is, a weighted average of two predictive probability distribution functions is not necessarily a predictive distribution. However, Gilboa and Schmeidler (1989) require the set of probability distributions in consideration to be convex in order to show that uncertainty aversion implies a maxmin problem. We can consider the convex hull of the predictive distributions and then apply the implication from uncertainty aversion to maxmin problems established by Gilboa and Schmeidler. Since a maxmin problem implies behavior of uncertain aversion over any set of probability distributions, we can then use the maxmin problem over the set of predictive distributions that is not convex. In fact, the maxmin principle had been used in Bayesian decisions to express aversion to uncertainty [see Berger (1985)] before Gilboa and Schmeidler (1989) established the equivalence axiomatically under the convexity assumption.

The most popular way of solving a maxmin problem is to transform the maxmin problem to a minmax problem. The minmax problem is

$$\min_{\theta \in [0,+\infty)} \max_{x} U(x;\theta). \tag{10}$$

In the minmax problem, investors are conservative and choose the probability distribution that gives the minimum of optimal utility. If $\theta^\dagger$ is a solution to the minmax problem, $x^\dagger$ is the solution to the utility maximization problem for given $\theta^\dagger$. That is, $x^\dagger = \arg\max_x U(x; \theta^\dagger) = x(\theta^\dagger)$. The asset allocation $x^\dagger$ can be understood as a conservative portfolio choice. If we substitute Equation (8) into the utility function in Equation (7), the minmax problem (10) becomes

$$\min_{\theta \in [0,+\infty)} E[r_{T+1}|R,\theta]'(V[r_{T+1}|R,\theta])^{-1}E[r_{T+1}|R,\theta], \tag{11}$$

where, the parameter θ is chosen to minimize the highest Sharpe ratio of the predictive mean and variance. The resulting optimal Sharpe ratio is the greatest lower bound of the Sharpe ratios that an investor can obtain by choosing parameter θ.

Transformation of a maxmin problem to a minmax problem requires some assumptions on the utility function. If the utility function U is (weakly) concave in x and (weakly) convex in θ, the maxmin and minmax problems are equivalent. This is the well-known MiniMax theorem. Sion (1958) and Bazaraa (1993) provide the general form of the theorem. When the MiniMax theorem holds, the maxmin and minmax problems are equivalent and $(x^*; \theta^*) = (x^\dagger; \theta^\dagger)$. It implies $x^* = \arg\max_x U(x; \theta^*) = x(\theta^*)$, that is, the tangency portfolio corresponding to the particular prior belief in the model with $\theta = \theta^*$ will be a portfolio that is optimal for investors averse to model uncertainty. However, when the maxmin and minmax problems are not equivalent, there is no prior such that the tangency portfolio given the prior is optimal for uncertainty-averse investors.

2. The Shrinkage Approach

The shrinkage approach to be developed in this section is new. In the Bayesian approach, different prior beliefs in asset-pricing models lead to different predictive distributions. However, for each given prior belief, it is not clear how much impact it has on the predictive distribution. In other words, we would like to know how a prior belief in an asset-pricing model affects the estimate of the mean and variance. This section develops the shrinkage approach that explicitly shows how a prior belief modifies the sample estimate of the mean and variance toward the estimate restricted by the asset-pricing model. The shrinkage approach is then used to solve the maxmin problem.

The shrinkage formulation is summarized in the following theorem, which relates the predictive mean and variance to the maximum likelihood estimates:

Theorem 1. *Let $\hat{S}$ be the highest Sharpe ratio of the efficient frontier spanned by the sample mean and variance of the factor portfolios, and let*

$$\omega = \frac{1}{1 + T\theta/\left(1 + \hat{S}^2\right)}. \tag{12}$$

Then, the mean and variance of the predictive distribution are

$$E[r_{T+1}|R, \omega] = \omega \begin{pmatrix} \bar{\beta}\hat{\mu}_2 \\ \hat{\mu}_2 \end{pmatrix} + (1 - \omega) \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix}, \tag{13}$$

$$V[r_{T+1}|R, \omega] = \begin{pmatrix} V_{11}(\omega) & V_{12}(\omega) \\ V_{12}(\omega)' & b\hat{\Omega}_{22} \end{pmatrix}, \tag{14}$$

where $V_{11}(\omega)$ and $V_{12}(\omega)$ are given by

$$V_{11}(\omega) = b\big[\omega\bar{\beta} + (1-\omega)\hat{\beta}\big]\hat{\Omega}_{22}\big[\omega\bar{\beta} + (1-\omega)\hat{\beta}\big]' + h\big[\omega\bar{\delta} + (1-\omega)\hat{\delta}\big]\omega\bar{\Sigma} + (1-\omega)\hat{\Sigma}\big], \tag{15}$$

$$V_{12}(\omega) = b\big[\omega\bar{\beta} + (1-\omega)\hat{\beta}\big]\hat{\Omega}_{22}, \tag{16}$$

where $\bar{\delta}$, $\hat{\delta}$, b and h are scalars and defined as follows:

$$\bar{\delta} = \frac{T(T-2)+k}{T(T-k-2)} - \frac{k+3}{T(T-k-2)} \cdot \frac{\hat{S}^2}{1+\hat{S}^2}, \tag{17}$$

$$\hat{\delta} = \frac{(T-2)(T+1)}{T(T-k-2)}, \tag{18}$$

$$b = \frac{T+1}{T-k-2}, \tag{19}$$

$$h = \frac{T}{T-m-k-1}. \tag{20}$$

Equation (13) states that the predictive mean is a weighted average of the estimated means restricted and unrestricted by the asset-pricing model. It is a shrinkage estimator. The shrinkage target is the maximum likelihood estimate of μ under the restriction of the asset-pricing model. According to Efron and Morris (1973), ω is referred to as *the shrinkage factor*. This result is consistent with the well-known observation that posterior mean is a Bayes–Stein estimator. The predictive variance of the assets, $V_{11}(\omega)$, is a quadratic function of the shrinkage factor ω, rather than a linear weighted average. Although this is not a shrinkage estimator in the traditional sense, it can be understood as a nonlinear extension to the classic shrinkage estimator. If the shrinkage factor is one, $V_{11}(\omega)$ is proportional to the estimate restricted by the model. If the shrinkage factor is zero, $V_{11}(\omega)$ is proportional to the unrestricted sample estimate. The theorem also states that the predictive covariance, $V_{12}(\omega)$, between the assets and the factors is proportional to the weighted average of the estimates restricted and unrestricted by the model. The weight is the same shrinkage factor ω.

Theorem 1 establishes the link between the classic and Bayesian frameworks for asset allocation using asset-pricing models. The link is the shrinkage factor ω, which indicates the relative weight between the estimates restricted and unrestricted by the model. More importantly, the predictive mean and variance share exactly the same shrinkage factor. The theorem explicitly shows how a prior belief influences the estimated mean and variance. As expected, the shrinkage factor ω is a decreasing function of θ. When θ approaches 0, the weight assigned to the estimate restricted by the model converges to 1. When θ becomes infinitely large, the weight assigned to the unrestricted estimate converges to 1. Therefore, choosing θ is equivalent to choosing the weight between the estimates restricted and unrestricted by the model.

Equation (12) captures another new feature, stating that the shrinkage factor ω is an increasing function of the Sharpe ratio of the frontier spanned by the factor portfolios. Given T and θ, more weight is assigned to the estimate restricted by the model if the Sharpe ratio of the factor portfolios is higher. This is reasonable. If the asset-pricing model holds, some combination of the factor portfolios should give the same Sharpe ratio as the efficient frontier spanned by all the assets. If the Sharpe ratio of the frontier spanned by the factors is much lower than the Sharpe ratio of the frontier spanned by all the assets, the asset-pricing model is likely to be wrong. If the Sharpe ratio of the factors is higher, it is closer to the highest Sharpe ratio of all the assets, then the model is more likely to be correct. More weight (a higher shrinkage factor) is therefore assigned to the estimate restricted by the model.

Theorem 1 points to a simple way to implement the Bayesian analysis. First, obtain the maximum likelihood estimate subject to the model restriction and the estimate without the restriction; Then, choose ω and use the formula in the theorem to calculate the predictive mean and variance. The shrinkage approach provides a simple and intuitive way to apply Bayesian portfolio selection by choosing ω to reflect the investor's beliefs about the relative importance between the model and data. For example, if the investor thinks that the model and the data are equally important, then they should choose $\omega = \frac{1}{2}$. If the investor thinks that the data are twice as important as the model, they should choose $\omega = \frac{1}{3}$. This is consistent with industry practice, in which people often take average estimates obtained from different models. For example, the asset management group at Goldman Sachs averages the estimate from equilibrium models and that from the views of investors [He and Litterman (1999)].

A major distinction of Theorem 1 from other shrinkage estimation in the literature is that the shrinkage factor for the variance is the same as that for the mean. The work by Polson, Johannes and Stroud (2002), Gron, Jorgensen and Polson (2001), Polson, Jacquier and Rossi (2002), and MacKinlay and Pastor (2000) demonstrate that it is important to estimate variance precisely and consider the relation between mean and variance implied by asset-pricing models. Theorem 1 shows that the predictive mean and variance both depend on the same shrinkage factor. It exhibits the importance of the predictive variance through the relation between the predictive mean and variance implied by an asset-pricing model. In the next section, I show that shrinkage of variance is important for optimal asset allocation among a large number of assets. Most existing shrinkage theorems shrink only the mean of asset returns but assume the variance to be given.[2] Since the main issue in finance is the trade-off between return and risk, the estimation of variance should be at least as important as the mean because variance is closely related to risk. Bayesian inference of variance and its shrinkage formulation are considered in a few studies,[3] but their prior distributions do not consider the restrictions of asset-pricing models. Those studies shrink the predictive mean and variance separately with different shrinkage factors that do not reflect any relation between the mean and variance.

Theorem 1 allows us to obtain useful properties of the maxmin problem and to solve it easily because the predictive mean and variance is a simple function of ω, whose range is from 0 to 1 (rather than from 0 to infinity). The mean-variance utility function of an investor is a function of x and ω, denoted by $U(x;\omega)$. The maxmin problem becomes

$$\max_{x} \min_{\omega\in[0,1]} U(x;\omega). \tag{21}$$

Since $U(x;\omega)$ is continuous in $(x;\omega)$ and the set of possible ω is compact, the function $\tilde{U}(x) \equiv \min_\omega U(x;\omega)$ is continuous in x. Since $U(x;\omega)$ is strictly concave in x, the function $\tilde{U}(x)$ is also strictly concave in x. It follows that there is a unique x^* that solves the maxmin problem (21). With the fixed x^*, solutions ω^* to $\min_\omega U(x^*;\omega)$ always exist but may not be unique.

Theorem 1 also makes it easy to solve for the optimal portfolio x^* in the maxmin problem. It follows that

$$E[r_{T+1}|R,\omega] = \begin{pmatrix} \hat{\mu}_1 + \omega(\bar{\beta}\hat{\mu}_2 - \hat{\mu}_1) \\ \hat{\mu}_2 \end{pmatrix}, \tag{22}$$

$$V[r_{T+1}|R,\omega] = \begin{pmatrix} \Psi_0 + \Psi_1\omega + \Psi_2\omega^2 & b\left[\hat{\beta} + \omega(\bar{\beta} - \hat{\beta})\right]\hat{\Omega}_{22} \\ b\hat{\Omega}_{22}\left[\hat{\beta} + \omega(\bar{\beta} - \hat{\beta})\right]' & b\hat{\Omega}_{22} \end{pmatrix}, \tag{23}$$

where

$$\Psi_0 = b\hat{\beta}\hat{\Omega}_{22}\hat{\beta}' + h\hat{\delta}\hat{\Sigma}, \tag{24}$$

$$\Psi_1 = b(\bar{\beta} - \hat{\beta})\hat{\Omega}_{22}\hat{\beta}' + b\hat{\beta}\hat{\Omega}_{22}(\bar{\beta} - \hat{\beta})' + h\left(\bar{\delta} - \hat{\delta}\right)\hat{\Sigma} + h\hat{\delta}(\bar{\Sigma} - \hat{\Sigma}), \tag{25}$$

$$\Psi_2 = b(\bar{\beta} - \hat{\beta})\hat{\Omega}_{22}(\bar{\beta} - \hat{\beta})' + h\left(\bar{\delta} - \hat{\delta}\right)(\bar{\Sigma} - \hat{\Sigma}). \tag{26}$$

[2] See, for example, Andersen (1971), Black and Litterman (1991), Efron and Morris (1973), Jobson and Korkie, Ratti (1979) and Jorion (1986, 1991). Pastor and Stambaugh (1999) have noticed that if there is one asset and one factor, the posterior mean of Jensen's alpha is approximately a weighted average of the estimated alpha restricted and unrestricted by the model.

[3] See Brown (1976), Frost and Savarino (1986), Jagannathan and Ma (2003), and Ledoit and Wolf (2003).

Then, the predictive mean and variance can be written as

$$E[r_{T+1}|R,\omega] = A_0 + \omega A_1, \tag{27}$$

$$V[r_{T+1}|R,\omega] = B_0 + \omega B_1 + \omega^2 B_2, \tag{28}$$

where

$$A_0 = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix}, \tag{29}$$

$$A_1 = \begin{pmatrix} \bar{\beta}\hat{\mu}_2 - \hat{\mu}_1 \\ 0_{k\times 1} \end{pmatrix}, \tag{30}$$

$$B_0 = \begin{pmatrix} \Psi_0 & b\hat{\beta}\hat{\Omega}_{22} \\ b\hat{\Omega}_{22}\hat{\beta}' & b\hat{\Omega}_{22} \end{pmatrix}, \tag{31}$$

$$B_1 = \begin{pmatrix} \Psi_1 & b(\bar{\beta} - \hat{\beta})\hat{\Omega}_{22} \\ b\hat{\Omega}_{22}(\bar{\beta} - \hat{\beta})' & 0_{k\times k} \end{pmatrix}, \tag{32}$$

$$B_2 = \begin{pmatrix} \Psi_2 & 0_{m\times k} \\ 0_{k\times m} & 0_{k\times k} \end{pmatrix}. \tag{33}$$

Substituting Equations (27) and (28) into the utility function, we obtain

$$U(x,\omega) = U_0(x) + \omega U_1(x) + \omega^2 U_2(x), \tag{34}$$

where

$$U_0(x) = x'A_0 - \tfrac{1}{2}\gamma x' B_0 x, \tag{35}$$

$$U_1(x) = x'A_1 - \tfrac{1}{2}\gamma x' B_1 x, \tag{36}$$

$$U_2(x) = -\tfrac{1}{2}\gamma x' B_2 x. \tag{37}$$

Basic algebra for quadratic functions implies that $\tilde{U}(x) = \min_\omega U(x,\omega)$ and has the following analytical expression:

$$\tilde{U}(x) = \begin{cases} U_0(x) - U_1^2(x)/(4U_2(x)) & \text{if } x \in \mathcal{X}, \\ \min\{U_0(x), U_0(x) + U_1(x) + U_2(x)\}, & \text{if } x \notin \mathcal{X}, \end{cases} \tag{38}$$

where the set $\mathcal{X}$ is defined as

$$\mathcal{X} = \{x : U_2(x) > 0,\ 0 < -0.5U_1(x)/U_2(x) < 1\}. \tag{39}$$

Since $\tilde{U}(x)$ is strictly concave, multivariate optimization programs work well in finding x^* on computer.

The shrinkage is critical for obtaining the properties and solutions of the maxmin problem. It also allows us to check if the utility function $U(x, \omega)$ is convex in ω so that we know if we can apply the MiniMax theorem. The function is convex in ω for any x if and only if $U_2(x)$ is non-negative for any x. This is true if and only if the matrix B_2, or equivalently Ψ_2, is negative semi-definite. Since the nature of Ψ_2 depends on the observed data, convexity of the function $U(x, \omega)$ in ω depends on the observed data. We cannot impose assumptions on the shape of $U(x, \omega)$ because such assumptions can be violated by the observed data. In empirical applications, we should examine eigenvalues of Ψ_2 because Ψ_2 is negative semi-definite if and only if all eigenvalues are non-positive.

It is important to keep in mind that the Minimax theorem is only a sufficient but not necessary condition for the maxmin and minmax problems to be equivalent. Therefore, if the function $U(x, \omega)$ is not convex in ω for every x, we need to solve the minmax problem to find out whether the maxmin and minmax problems are equivalent. We are interested in the equivalent of the two problems because it tells us whether an optimal portfolio for uncertainty-averse investors can be a tangency portfolio given a particular value of θ in the prior. Using the shrinkage factor ω, the minmax problem becomes

$$\min_{\omega \in [0,1]} \max_{x} U(x; \omega). \tag{40}$$

A solution $\omega^\dagger$, referred to as a conservative shrinkage factor in this article, solves

$$\min_{\omega} E[r_{T+1}|R, \omega]' (V[r_{T+1}|R, \omega])^{-1} E[r_{T+1}|R, \omega]. \tag{41}$$

The solution $\omega^\dagger$ exists because the objective function in (41) is continuous and the set of ω is compact. For any given ω, a tangent portfolio determined by the predictive mean and variance is

$$x(\omega) = \gamma^{-1} (V[r_{T+1}|R, \omega])^{-1} E[r_{T+1}|R, \omega]. \tag{42}$$

The portfolio choice corresponding to the minmax solution is $x^\dagger = x(\omega^\dagger)$.

When the maxmin and minmax problems have different solutions, the optimal portfolio for uncertainty-averse investors is not in the set of tangency portfolios, $\mathcal{P}$, which is defined as $\mathcal{P} = \{x : x = x(\omega) \text{ for some } \omega \in [0, 1]\}$. If investors are restricted to the set of tangency portfolios, there will be a loss to them in terms of utility $\min_\omega U(x, \omega)$. The loss can be calculated as

$$L = \max_x \min_\omega U(x, \omega) - \max_{x \in \mathcal{P}} \min_\omega U(x, \omega). \tag{43}$$

3. A Comparison to Empirical Bayes Approach

Bayesian analyses of asset allocation in the current literature do not use maxmin analysis to incorporate aversion to model uncertainty. They often use the empirical Bayes approach to specify prior beliefs. In the standard parametric empirical Bayes approach, researchers specify a functional form of the prior distribution, which sometimes depends on hyperparameters. Researchers estimate the hyperparameters using the marginal distribution of data. Either maximum likelihood or the method of moments can be used to estimate the hyperparameters. This approach is also called the *compound decision problem* in statistics literature. In this approach, the set of data used for estimating hyperparameters is the same set of data used for computing the posterior distribution. Berger (1985) and Bernardo and Smith (1994) provide overviews of the compound decision problems. Since the empirical Bayes approach is often applied in asset allocation, it is important to understand what shrinkage factor it implies and whether the resulting asset allocation is optimal for investors averse to model uncertainty.

For the issues examined in this article, the hyperparameter is θ. In order to determine θ, classic statistical inference is used to estimate the sampling distribution of Jensen's alpha. The estimated sampling distribution is then used as the informative prior distribution of Jensen's alpha. In the appendix, it is shown that, under the null hypothesis of $\alpha = 0_{m\times 1}$, the sampling distribution of the maximum likelihood estimator of α, conditional on the factor and the parameters (α, β, Σ, μ_2, Ω_{22}), is a normal distribution with mean $0_{m\times 1}$ and variance $T^{-1}(1 + \hat{S}^2)\Sigma$. That is,

$$\hat{\alpha}|\Sigma \sim N\left(0_{m\times 1}, \frac{1}{T}(1 + \hat{S}^2)\Sigma\right). \tag{44}$$

This sampling distribution is the multi-factor version of formula (5.3.16) in Campbell, Lo and MacKinley (1997), which has only one factor. It follows from Equation (6) that the parameter in the prior distribution of α should be specified as

$$\theta = T^{-1}\left(1 + \hat{S}^2\right). \tag{45}$$

Substituting this choice of θ into Equation (12) gives $\omega = \frac{1}{2}$. This important result is stated in the following theorem:

Theorem 2. *If the prior distribution of* α *is chosen to satisfy* $V[\alpha|\Sigma] = V[\hat{\alpha}|\Sigma]$, *the shrinkage factor is* $\omega = \frac{1}{2}$.

A researcher taking the empirical Bayes approach, therefore, assigns equal weights to the estimates restricted and unrestricted by the asset-pricing model, implying that the model and data have equal importance. This makes it convenient to compare the asset allocation obtained using the empirical Bayes approach with the allocation obtained using the maxmin approach. Most applications in asset allocation deviate from the standard compound decision problem by choosing θ different from Equation (45). For example, one can estimate a prior variance for alpha using an earlier or different set of data. Alternatively, one can estimate a prior variance of alpha using the cross-sectional variance of estimated alphas in a large cross section of assets. Even when some people choose θ without estimation, they tend to make it more or less consistent with our empirical experience with the data. No matter how a prior variance is inferred from data, it implies a shrinkage factor. If a prior variance of alpha is close to empirical estimate, the implied shrinkage factor is close to $\frac{1}{2}$.

It seems that the empirical Bayes approach is biased toward models — the shrinkage factor is always $\frac{1}{2}$, no matter how many data are observed and used. This is caused by setting the prior mean of Jensen's alpha to zero. One may let $\alpha \mid \Sigma \sim N(\lambda, \theta\Sigma)$, where λ is an $m \times 1$ vector, in the prior distribution and take the empirical Bayes approach to determine both the hyperparameters λ and θ. The resulting predictive mean and variances will be close to the sample estimates.[4] In this case, there is no faith in the asset-pricing model, and the model is in fact not used in asset allocation. Using classical statistical inference to determine both λ and θ can be criticized as non-Bayesian because it relies solely on data. We are only interested in the case of $\lambda = 0$, in which the informative prior distribution of Jensen's alpha imposes a prior belief in the model. Choosing a prior distribution for Jensen's alpha with a zero mean is equivalent to choosing the model as reference. The setup of this kind of prior distribution allows the model to exert influence on the posterior and predictive distributions. The built-in influence of the model is the purpose of this article's exercise — asset allocation *using asset-pricing models.*

In the empirical Bayes approach, researchers sometimes use different sets of data to estimate the prior and posterior distributions. For example, a researcher may use earlier observations to estimate the prior distribution and use later observations to estimate the posterior distribution. Or, a researcher may construct her prior distribution using a different set of assets. Alternatively, a researcher might form her prior opinion by reading other people's research that uses either different set of assets or different period of observations. This kind of empirical Bayes approach can be different from the compound decision problem.

[4] The main reason is that we use data to infer the value of λ. If we set λ to be a known number, a shrinkage result similar to Theorem 1 holds but the shrinkage target is different. In fact, Theorem 1 should be viewed as the case setting $\lambda = 0$.

Suppose the prior distribution is estimated by a set of observed returns $R_o = (Y_o, X_o)$ on some assets and factors. This set of data can be entirely different from, overlap with, or coincide with the data $R = (Y, X)$, which are used for estimating the posterior and predictive distributions. Although the factors in the two sets of data must be the same, the actual observed data X_o and X can be different if the two sets of data cover different periods. Suppose there are T_o observations and m_o assets in the set of data R_o. Let $\hat{\Sigma}_o$ be the estimated variance of the residual in the regression of Y_o on X_o, and $\hat{S}_o$ the highest sample Sharpe ratio of the factor portfolios estimated from X_o. It follows from (44) that

$$\hat{\alpha}|\Sigma \sim N\left(0_{m_o \times 1}, \frac{1}{T_o}(1+\hat{S}_o^2)\Sigma\right). \tag{46}$$

Therefore, we should choose $\theta = T_o^{-1}(1+\hat{S}_o^2)$.

With the above choice of θ, the shrinkage factor will depend on the sample size and the factor Sharpe ratio. Substituting the above θ into Equation (12), we obtain

$$\omega = \left(1 + \frac{T}{T_o} \cdot \frac{1+\hat{S}_o^2}{1+\hat{S}^2}\right)^{-1}. \tag{47}$$

Since the factor Sharpe ratios are usually much smaller than 1, the ratio $(1+\hat{S}_o^2)/(1+\hat{S}^2)$ is not very different from 1. The shrinkage factor is then mainly determined by the relative size of the two sets of samples, that is, $\omega \approx (1+T/T_o)^{-1}$. If more observations are used for estimating the prior variance of Jensen's alpha, more weight will be assigned to the estimate restricted by the asset-pricing model. When researchers arbitrarily determine which data are used for the prior and which are used for the posterior, the resulting shrinkage factor is arbitrary. One should be cautious about the empirical Bayes approach that is not a compound decision problem.

4. Empirical Results

Let us first look at allocations when the factor in the asset-pricing model is the monthly excess return on the value-weighted market index portfolio of NYSE, AMEX, and NASDAQ during the period from July 1963 to December 1998. The excess returns on the assets are the monthly excess returns on the portfolios used by Fama and French (1993) and updated to the end of 1998. In order to compare the CAPM with the Fama–French model, the returns on the two Fama–French factors, SMB and HML, are also included as part of the asset returns. The data are obtained from Kenneth French's Web page. The last five of the Fama–French portfolios that contain the largest firms are excluded because the market, SMB and

Table 1
Asset Allocation Using the CAPM

	$\hat{S} = 0.1219$, $T = 426$							
ω	1.00	0.95	0.90	0.75	0.50	0.30	0.10	0.01
	Optimal portfolio weights							
MKT	1.00	1.19	2.97	0.57	0.69	0.71	0.72	0.73
SMB	0.00	−0.59	−6.03	1.33	0.95	0.88	0.85	0.84
HML	0.00	−0.32	−3.27	0.72	0.52	0.48	0.46	0.45
S-L	0.00	−0.36	−3.65	0.81	0.58	0.53	0.51	0.51
S-2	0.00	0.13	1.32	−0.29	−0.21	−0.19	−0.19	−0.18
S-3	0.00	−0.04	−0.36	0.08	0.06	0.05	0.05	0.05
S-4	0.00	0.35	3.54	−0.78	−0.56	−0.52	−0.50	−0.49
S-H	0.00	0.15	1.57	−0.35	−0.25	−0.23	−0.22	−0.22
2-L	0.00	0.02	0.16	−0.04	−0.03	−0.02	−0.02	−0.02
2-2	0.00	0.07	0.75	−0.17	−0.12	−0.11	−0.11	−0.10
2-3	0.00	0.19	1.94	−0.43	−0.31	−0.28	−0.27	−0.27
2-4	0.00	0.19	1.94	−0.43	−0.31	−0.28	−0.27	−0.27
2-H	0.00	0.08	0.81	−0.18	−0.13	−0.12	−0.11	−0.11
3-L	0.00	−0.21	−2.12	0.47	0.33	0.31	0.30	0.29
3-2	0.00	0.07	0.72	−0.16	−0.11	−0.10	−0.10	−0.10
3-3	0.00	−0.02	−0.21	0.05	0.03	0.03	0.03	0.03
3-4	0.00	0.12	1.26	−0.28	−0.20	−0.18	−0.18	−0.18
3-H	0.00	0.12	1.27	−0.28	−0.20	−0.19	−0.18	−0.18
4-L	0.00	0.30	3.08	−0.68	−0.49	−0.45	−0.43	−0.43
4-2	0.00	−0.35	−3.55	0.78	0.56	0.52	0.50	0.49
4-3	0.00	−0.11	−1.17	0.26	0.18	0.17	0.16	0.16
4-4	0.00	0.05	0.56	−0.12	−0.09	−0.08	−0.08	−0.08
4-H	0.00	−0.05	−0.54	0.12	0.08	0.08	0.08	0.07

This table reports the tangency portfolio weights on assets and factors for various values of the shrinkage factor ω. The assets are the Fama–French portfolios, excluding the five portfolios that contain the largest firms, and the SMB and HML portfolios constructed by Fama and French (1993) according to firms' market capitalization and book-to-market ratio. There is one factor, which is the value-weighted market index return of the NYSE, AMEX, and NASDAQ. The monthly excess returns on these portfolios during the period from July 1963 to December 1998 are used for estimating the mean and variance of the predictive distribution. The sample Sharpe ratio ($\hat{S}$) of the factor portfolios and the number of time-series observations (T) are provided.

HML factors are almost a linear combination of the 25 Fama–French portfolios. Table 1 presents the tangency portfolio weights for various values of the shrinkage factor ω. As expected, most of the weight is on the market portfolio when ω is close to 1.

Since the Fama–French model is empirically successful in fitting the data, it is natural to apply it to asset allocation. The returns on the factors in this model include the returns on the SMB and HML portfolios in addition to the monthly excess returns on the value-weighted market index portfolio. The excess returns on the assets are the monthly excess returns on the Fama–French portfolios considered in Table 1. The tangency portfolio weights for various values of ω are reported in Table 2. When ω is close to 1, most of the portfolio weight is assigned to the three Fama–French factors. When the value of ω is 1 or less, the portfolio weights in Tables 1 and 2 are very similar. In this case, the tangency portfolios are mainly determined by the unrestricted sample estimates.

Table 2
Asset Allocation Using the Fama–French model

	$\hat{S}=0.2551$, $T=426$							
ω	1.00	0.95	0.90	0.75	0.50	0.30	0.10	0.01
	Optimal portfolio weights							
MKT	0.32	0.29	0.24	−0.12	1.46	0.89	0.77	0.74
SMB	0.04	−0.03	−0.13	−0.82	2.25	1.15	0.92	0.86
HML	0.64	0.66	0.68	0.84	0.12	0.38	0.43	0.45
S-L	0.00	−0.04	−0.10	−0.55	1.41	0.71	0.56	0.52
S-2	0.00	0.02	0.04	0.20	−0.51	−0.26	−0.20	−0.19
S-3	0.00	0.00	−0.01	−0.05	0.14	0.07	0.06	0.05
S-4	0.00	0.04	0.10	0.53	−1.36	−0.68	−0.54	−0.51
S-H	0.00	0.02	0.05	0.24	−0.61	−0.30	−0.24	−0.22
2-L	0.00	0.00	0.00	0.02	−0.06	−0.03	−0.02	−0.02
2-2	0.00	0.01	0.02	0.11	−0.29	−0.14	−0.11	−0.11
2-3	0.00	0.02	0.06	0.29	−0.75	−0.38	−0.30	−0.28
2-4	0.00	0.02	0.06	0.29	−0.75	−0.38	−0.30	−0.28
2-H	0.00	0.01	0.02	0.12	−0.31	−0.16	−0.12	−0.12
3-L	0.00	−0.03	−0.06	−0.32	0.82	0.41	0.32	0.30
3-2	0.00	0.01	0.02	0.11	−0.28	−0.14	−0.11	−0.10
3-3	0.00	0.00	−0.01	−0.03	0.08	0.04	0.03	0.03
3-4	0.00	0.02	0.04	0.19	−0.49	−0.24	−0.19	−0.18
3-H	0.00	0.02	0.04	0.19	−0.49	−0.25	−0.19	−0.18
4-L	0.00	0.04	0.09	0.46	−1.19	−0.60	−0.47	−0.44
4-2	0.00	−0.04	−0.10	−0.53	1.37	0.69	0.54	0.51
4-3	0.00	−0.01	−0.03	−0.18	0.45	0.23	0.18	0.17
4-4	0.00	0.01	0.02	0.08	−0.22	−0.11	−0.09	−0.08
4-H	0.00	−0.01	−0.02	−0.08	0.21	0.10	0.08	0.08

This table reports the tangency portfolio weights on assets and factors for various values of the corresponding shrinkage factor ω. The assets are the Fama–French portfolios constructed by Fama and French (1993) according to firms' market capitalization and book-to-market ratio, excluding the five portfolios that contain the largest firms. There are three factors, which are the SMB and HML risk factors constructed by Fama and French (1993) and the value-weighted market index return of the NYSE, AMEX, and NASDAQ. The monthly excess returns on these portfolios during the period from July 1963 to December 1998 are used for estimating the mean and variance of the predicative distribution. The sample Sharpe ratio ($\hat{S}$) of the factor portfolios and the number of time-series observations (T) are provided.

Now, let us examine the optimal allocation with aversion to uncertainty about the CAPM. The solution x^* to the maxmin problem corresponding to each selected degree of risk aversion is scaled into a vector of portfolio weights and reported in columns 2–5 of Table 3. The risk aversion γ is set to be 3, 5, 7, or 9 in this article.[5] It is interesting that the solution to the maxmin problem assigns large positive or negative weights to either SMB or the HML portfolios, even when the CAPM is used to shrink the estimate. The optimal portfolio for an investor with aversion to model uncertainty is very different from the market portfolio prescribed by the CAPM. It is also very different from the portfolio based on the unrestricted sample estimate. It indicates that the utility loss caused by the

[5] Bodie et al. (1999, pp. 191–193) calibrated that $\gamma=2.96$ for U.S. investors. Similarly, Pastor and Stambaugh (2000) calibrated that $\gamma=2.84$.

Table 3
Asset Allocation with Aversion to Model Uncertainty

	The CAPM				Fama–French model			
γ	3.0	5.0	7.0	9.0	3.0	5.0	7.0	9.0
MKT	1.13	0.97	0.98	0.89	0.32	0.32	0.32	0.31
SMB	−0.42	0.20	0.08	0.38	0.04	0.04	0.04	0.03
HML	−0.20	0.05	0.02	0.10	0.64	0.64	0.64	0.64
S-L	0.03	0.00	−0.02	−0.02	0.00	0.00	0.00	0.00
S-2	0.00	−0.01	−0.01	−0.04	0.00	0.00	0.00	0.00
S-3	0.03	−0.07	0.01	0.01	0.00	0.00	0.00	0.00
S-4	0.02	0.05	−0.02	−0.11	0.00	0.00	0.00	0.00
S-H	0.08	−0.04	0.01	0.00	0.00	0.00	0.00	0.00
2-L	0.02	−0.03	0.00	−0.05	0.00	0.00	0.00	0.00
2-2	0.02	−0.02	0.00	−0.01	0.00	0.00	0.00	0.00
2-3	0.02	−0.02	−0.01	0.00	0.00	0.00	0.00	0.00
2-4	0.05	−0.05	−0.04	−0.02	0.00	0.00	0.00	0.00
2-H	0.06	−0.01	0.00	−0.05	0.00	0.00	0.00	0.00
3-L	0.01	0.00	−0.01	−0.01	0.00	0.00	0.00	0.00
3-2	0.02	0.02	0.00	−0.04	0.00	0.00	0.00	0.00
3-3	0.05	0.01	−0.03	−0.06	0.00	0.00	0.00	0.00
3-4	0.03	−0.04	0.00	0.00	0.00	0.00	0.00	0.00
3-H	0.04	0.00	0.01	−0.04	0.00	0.00	0.00	0.00
4-L	0.00	−0.02	0.02	0.02	0.00	0.00	0.00	0.00
4-2	−0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4-3	0.02	0.00	0.02	−0.01	0.00	0.00	0.00	0.00
4-4	0.01	0.01	−0.01	0.03	0.00	0.00	0.00	0.00
4-H	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00

For the assets and data considered in Tables 1 and 2, this table presents the portfolio weights in the solutions to the maxmin problems, for various degree of risk aversion γ, when the CAPM or the Fama–French model is used in asset allocation.

uncertainty about the CAPM is probably as important as the loss caused by the uncertainty about the unrestricted estimate. The optimal allocations by uncertainty-averse investors in no way resemble the tangency portfolios obtained in Table 1. For the CAPM, there is no prior belief that gives the allocation optimal for uncertainty-averse investors. That is, there is no ω such that $x^* = \arg\max_x U(x;\omega)$. Thus, the optimal allocation for uncertainty-averse investors is not a tangency portfolio corresponding to any prior beliefs in the CAPM. In particular, the tangency portfolio obtained using the empirical Bayes approach ($\omega = 0.5$) is not optimal for uncertainty-averse investors.

Naturally, we should also look at the effects of aversion to uncertainty about the Fama–French model. The solutions to the maxmin problems are reported in columns 6–9 of Table 3. When the Fama–French model is used, the portfolio weights in the solution to the maxmin problem is very similar to the tangency portfolio weights implied by the model with ω being close to 1 in Table 2. Therefore, when the Fama–French model is under consideration, an investor with aversion to model uncertainty can approximate his desired asset allocation by setting the shrinkage factor ω

close to 1 or simply assuming that the model holds. It appears puzzling that aversion to uncertainty about the Fama–French model leads to asset allocations similar to the tangency portfolio obtained from the model. In fact, in the set-up of the maxmin problem, the investor is making a choice between two stochastic models — the predictive distribution restricted by the Fama–French model and the unrestricted predictive distribution. The unrestricted predictive distribution may be more likely to be incorrect than the Fama–French model because the estimate depends too much on noisy observations. In this case, aversion to uncertainty about the unrestricted predictive distribution forces the investor to rely more on the Fama–French model. The optimal asset allocation with uncertainty aversion is also very different from the tangency portfolio obtained from the empirical Bayes approach ($\omega = 0.5$) when investors use the Fama–French model.

For uncertainty-averse investors, I calculate the utility loss L, defined in Equation (43), due to the restriction of tangency portfolios. When the CAPM is used for asset allocation, the utility loss L is 1.38%, 2.24%, 0.49%, and 5.88%, respectively, for $\gamma = 3$, 5, 7, and 9. In comparison, when the Fama–French model is used for asset allocation, the utility loss L is zero for $\gamma = 3$ and 5. For $\gamma = 7$ and 9, the utility loss L is not zero but are only 0.02% and 0.05% respectively. This further demonstrates that the tangency portfolio obtained by imposing the Fama–French model is optimal or approximately optimal for uncertainty-averse investors when they consider asset allocation over the portfolios sorted by firm size and book-to-market ratio.

Since the Fama–French model is constructed to fit mainly the data during 1970s and 1980s while the CAPM is motivated from theory, it is natural to examine the models using the data of 1990s. Table 4 presents the results using the data of 1990s. When the CAPM is the reference model, the optimal portfolios for uncertainty-averse investors constructed from the data of 1990s are not drastically different from those constructed from the whole sample. When Fama–French model is used, the optimal portfolios for uncertainty-averse investors using the data of 1990s are very different from the portfolios using the whole sample. The main change is in the weights among the three factor portfolios. However, all the optimal portfolios are heavily concentrated on the factor portfolios if the degree of risk aversion is high. Therefore, the tangency portfolio of Fama–French model still offers a good approximation to the optimal portfolio for uncertainty-averse investors with high degree of risk aversion.

In the above analysis, the maxmin problem is solved directly. The MiniMax Theorem cannot be applied to the problems discussed in this article because $U(x; \omega)$ is not a convex function in ω for the problems and data discussed in this article. As pointed out in Section 2, the function is convex in ω if and only if all the eigenvalues of Ψ_2 are non-positive.

Table 4
Asset Allocation with Uncertainty Aversion Using the 1990s Data

	The CAPM				Fama–French model			
γ	3.0	5.0	7.0	9.0	3.0	5.0	7.0	9.0
MKT	1.09	1.06	1.03	1.05	0.65	0.71	0.73	0.72
SMB	−0.29	−0.28	−0.05	−0.22	−1.30	−0.45	−0.27	−0.29
HML	−0.14	−0.10	−0.06	−0.12	0.87	0.63	0.58	0.59
S-L	0.01	0.01	0.01	0.01	0.04	0.01	0.00	0.00
S-2	0.02	0.03	−0.03	−0.01	0.03	0.01	0.00	0.00
S-3	0.02	0.00	0.01	0.02	0.04	0.01	0.00	0.00
S-4	0.02	0.01	−0.02	0.01	0.02	0.00	0.00	0.01
S-H	0.03	0.04	0.04	0.05	0.09	0.02	−0.01	−0.01
2-L	0.02	0.02	0.01	0.00	0.03	0.00	0.00	0.00
2-2	0.04	0.02	0.01	0.01	0.06	0.01	0.00	0.00
2-3	−0.02	0.00	0.00	0.01	0.01	0.00	0.00	0.00
2-4	0.04	0.05	0.02	0.03	0.13	0.02	−0.01	0.00
2-H	0.06	0.04	0.00	0.03	0.13	0.01	−0.01	0.00
3-L	0.00	0.01	−0.01	0.01	0.00	0.00	0.00	0.00
3-2	0.02	0.02	0.00	0.02	0.04	0.00	0.00	0.00
3-3	0.02	0.00	0.01	0.01	0.02	0.00	0.00	0.00
3-4	0.04	0.04	0.01	0.01	0.09	0.01	0.00	0.00
3-H	0.03	0.02	0.03	0.03	0.07	0.01	0.00	0.00
4-L	−0.02	0.00	−0.01	0.00	−0.01	0.00	0.00	0.00
4-2	0.00	0.02	−0.01	0.01	0.02	0.00	0.00	0.00
4-3	−0.01	0.01	0.00	0.02	0.01	0.00	0.00	0.00
4-4	0.02	0.00	0.01	0.01	0.02	0.00	0.00	0.00
4-H	−0.02	−0.02	−0.01	0.00	−0.04	0.00	0.00	0.00

For the assets considered in Tables 1 and 2 and using the data from January 1990 to December 1998, this table presents the portfolio weights in the solutions to the maxmin problems, for various degree of risk aversion γ, when the CAPM or the Fama–French model is used in asset allocation.

Numerical calculations show that the largest eigenvalue for Ψ_2 is positive for both the CAPM and Fama–French model. Since the conditions in the MiniMax Theorem are sufficient but not necessary, the maxmin and minmax problems may happen to be equivalent. Therefore, we should also solve the minmax problem to compare with the solution to the maxmin problem. In Figure 1, when the CAPM is used, the optimal Sharpe ratio is plotted as a function of the shrinkage factor (the solid curve). Since the curve is strictly decreasing, the conservative shrinkage factor is $\omega^\dagger = 1$. A similar curve is plotted for the Fama–French model (the dotted curve). This curve is also decreasing and thus the conservative shrinkage factor is $\omega^\dagger = 1$. For both the CAPM and Fama–French model, the solutions to the minmax problems assign all the weights to the models, and the corresponding portfolios have positions only on the factor portfolios. Since the asset allocation obtained from the Fama–French model is similar to the allocation obtained from the maxmin problem, the minmax problem and maxmin problem give similar (but not always identical) asset allocations when the Fama–French model is in consideration. This implies that investors who are averse to uncertainty about the Fama–French model can obtain an approximate optimal

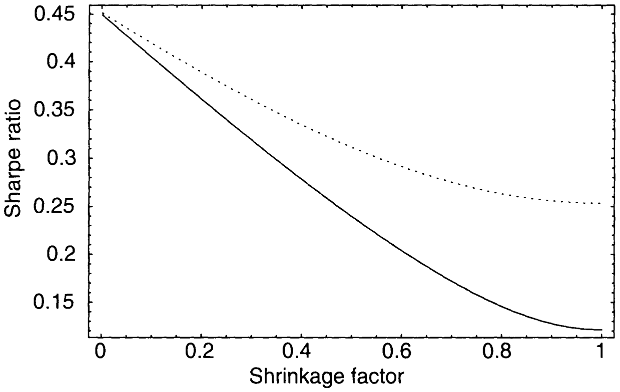

Figure 1
Optimal Sharpe ratio as a function of the shrinkage factor
The highest Sharpe ratio of the efficient frontier spanned by the assets and factors is plotted as a function of the shrinkage factor ω. The solid line is for the CAPM with the assets considered in Table 1. The dotted line is for the Fama–French model with the assets considered in Table 2.

asset allocation by choosing the conservative shrinkage factor. However, this is not true when the CAPM is used.

Results presented in this section show that, when investors are restricted in the set of available predictive distributions, the maxmin and minmax problems are not equivalent in the case of CAPM, and the optimal portfolio solved from the maxmin problem is not a tangency portfolio. This fact is true even if we consider the convex hull of predictive distributions in the maxmin problem. If investors are able to choose probability distributions from the convex hull,[6] the maxmin and minmax problems are still not equivalent in general because the function $U(x, \omega)$ is not always convex in ω. If the maxmin and minmax problems turn out to be equivalent in the convex hull of predictive distributions, the optimal portfolio must be a tangency portfolio for some probability distribution that is not in the set of predictive distributions. This still implies that none of the tangency portfolios obtained from the available predictive distributions are optimal for uncertainty-averse investors.

Besides extending the set of predictive distributions to its convex hull, we can also reduce the set of probability distributions. A smaller set of possible probability distributions represents less uncertainty. Investors may focus on a smaller set of prior beliefs in a model. Particularly interesting sets of available predictive distributions are those obtained by restricting ω between some lower and upper bounds, say $\underline{\omega}$ and $\overline{\omega}$. If we set $\overline{\omega} = 1$, higher $\underline{\omega}$ represents less uncertainty about the belief in a model.

[6] In practice, however, there is no obvious way to determine what prior beliefs generate the additional probability distributions in the convex hull.

Table 5
Uncertainty Averse Asset Allocation with Subsets of Priors

	The CAPM				Fama–French model			
$\overline{\omega}$	1.00	1.00	1.00	0.75	1.00	1.00	1.00	0.75
$\underline{\omega}$	0.25	0.50	0.75	0.25	0.25	0.50	0.75	0.25
MKT	1.07	1.03	0.99	0.57	0.30	0.29	0.32	−0.14
SMB	−0.25	−0.11	0.03	1.34	0.00	−0.01	0.02	−0.86
HML	−0.10	−0.04	0.01	0.72	0.66	0.67	0.65	0.86
S-L	0.02	0.01	0.00	0.81	0.00	0.01	0.00	−0.56
S-2	0.00	0.00	−0.01	−0.29	0.00	0.01	0.00	0.20
S-3	0.01	0.01	0.00	0.08	0.01	0.00	0.00	−0.05
S-4	0.00	0.01	0.01	−0.79	0.00	0.01	0.00	0.54
S-H	0.05	0.01	−0.01	−0.35	0.01	0.00	0.01	0.25
2-L	0.01	0.00	0.00	−0.04	0.01	0.00	0.00	0.03
2-2	0.02	0.01	0.01	−0.16	0.00	0.01	0.00	0.11
2-3	0.01	0.01	0.00	−0.43	−0.01	0.00	0.00	0.30
2-4	0.04	0.02	0.00	−0.43	0.01	0.00	−0.01	0.31
2-H	0.03	0.02	0.00	−0.18	0.01	0.01	0.00	0.13
3-L	0.01	0.01	0.00	0.47	0.01	0.01	0.01	−0.33
3-2	0.00	0.01	−0.01	−0.16	0.01	0.00	0.01	0.11
3-3	0.02	0.01	−0.01	0.04	0.00	0.01	0.00	−0.03
3-4	0.02	0.01	0.00	−0.28	0.01	0.02	0.00	0.19
3-H	0.02	0.01	0.00	−0.28	0.00	0.00	0.00	0.20
4-L	−0.01	0.00	0.00	−0.68	0.00	−0.01	−0.01	0.48
4-2	0.02	0.00	0.00	0.79	0.00	−0.01	−0.01	−0.55
4-3	−0.01	0.00	0.00	0.26	0.00	0.00	0.00	−0.19
4-4	0.00	−0.01	0.00	−0.12	−0.01	0.00	0.00	0.09
4-H	0.00	0.00	−0.01	0.12	0.00	0.00	0.00	−0.08

For the assets and data considered in Tables 1 and 2, this table presents the portfolio weights in the solutions to the maxmin problems, for various bounds ($\underline{\omega}$ and $\overline{\omega}$) of shrinkage factors, when the CAPM or the Fama–French model is used in asset allocation. The degree of risk aversion γ is set to be 3.

Table 5 presents the portfolio weights in the solutions to

$$\max_{x} \min_{\omega \in [\underline{\omega}, \overline{\omega}]} U(x, \omega),$$

where $\overline{\omega} = 1$ and $\underline{\omega} = 0.25$, 0.50, and 0.75. The degree of risk aversion is set to be 3. When the CAPM is used, the change of the set of ω substantially affects uncertainty-averse investor's optimal portfolio. When Fama–French model is used, the optimal portfolio weights are similar to those in Table 3 as long as the set contains $\omega = 1$. Since one might want to center their choice of predictive distributions on the empirical approach, Table 5 also presents the optimal portfolio weights with $\underline{\omega} = 0.25$ and $\overline{\omega} = 0.75$. These portfolios are very different from the optimal portfolios with ω ranging between 0 and 1.

Besides helping to solve the maxmin problem, the shrinkage approach developed in this article offers additional insights into some Bayesian analysis in the literature on asset allocation. To examine the impact of model uncertainty on asset allocation, Pastor and Stambaugh (2000)

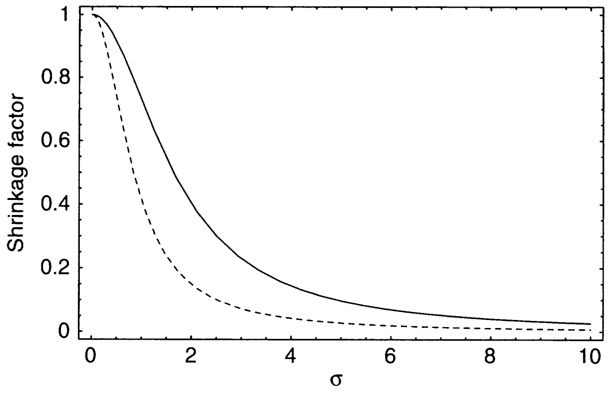

Figure 2
The shrinkage factor as a function of σ
The shrinkage factor ω is plotted as a function of σ (in annual percentage returns) in the prior distribution. The solid line is for the CAPM with the assets considered in Table 1. The dotted line is for the Fama–French model with the assets considered in Table 2.

represent θ in terms of another parameter σ by $\theta = \sigma^2/s^2$, where s^2 is defined as $s^2 = \text{trace}(\hat{\Sigma})/m$. They interpret σ^2 as the average prior variance of Jensen's α and observe that the difference in tangency portfolios using different models "are substantially reduced by modest uncertainty about the models' pricing abilities." In Figure 2, the shrinkage factor ω is plotted as a function of σ. The shrinkage factor drops quickly as σ increases. The sharp reduction of the shrinkage factor explains why modest value of σ substantially reduces models' impact on portfolio choice. The predictive distribution assigns a small weight to the model when σ increases modestly. The reason for the fast drop of the shrinkage factor is the large number of observations. It is easy to see this from Equation (12). Since the factor Sharpe ratio is 0.1219 for the CAPM and 0.2551 for the Fama–French model, the magnitude of the second term in the denominator is in the order of $T\theta$, which equals $T(\sigma^2/s^2)$. When T is large, the shrinkage factor ω drops fast as σ increases. This is reasonable. When more than 30 years of monthly observations are used in the inference of model parameters, it is implicitly assumed that the parameters are constant and the stochastic process of the returns is stationary over 30 years. With such a long period, we should be able to estimate the constant parameters very well with the observed data. Thus, little weight will be assigned to the model when the model's pricing ability is quite uncertain. If the model parameters are varying over time, we are not able to use such a long period of observation in the inference. It will then be more important to apply the restrictions of asset-pricing models.

Theorem 1 can be used to examine whether the shrinkage estimate of the variance is important to optimal portfolio choice. If the predictive variance is not affected by the shrinkage factor, the vector of portfolio weights is the weighted average of the tangency portfolios with and without the restriction of the asset-pricing model. In this case, the tangency portfolio based on the unrestricted sample estimate linearly shrinks to the tangency portfolio of the factors. This observation offers a way to check whether shrinkage of variance is important. The weights in Tables 1 and 2 are, however, nonlinear with respect to the shrinkage factor. In Table 1, the portfolio weight on the market index increases from 1.00 to 2.97 when the shrinkage factor decreases from 1.00 to 0.90. The weight then decreases to 0.57 when the shrinkage factor decreases to 0.75. As the shrinkage factor drops further, the portfolio weight on the market index rises again. In Table 2, we also observe complicated changes of the portfolio weight on the market index. These indicate that the shrinkage estimate of the variance has substantial influence on asset allocation. The shrinkage estimate of the variance could be less important for some small number of assets, but it is more likely to be important when the number of risky assets is large. It is the large number of risky assets that motivates us to impose the restrictions of asset-pricing models, which offers better estimates by reducing the dimension of our problems [Polson and Tew (2000)].

5. Conclusion

Standard investment theory seeks to equate owning a share of stock with rolling a pair of fair dice. In practice, this linkage may be tenuous. Because the true stochastic process of asset returns is uncertain, owning stock is like rolling dice without knowing whether the dice are fair or even how many faces they have. The distinction between owning a stock and participating in a lottery causes investors to allocate their assets differently if investors are averse to model uncertainty. The analysis in this article maintains the distinction as well as the link between investments and lotteries. It is shown that investors with aversion to model uncertainty may choose an asset allocation that is not mean-variance efficient for the probability distribution estimated from any particular prior belief in the model. In most cases, the portfolio obtained from the empirical Bayes approach is suboptimal for uncertainty-averse investors.

This article demonstrates that a Bayesian investor implicitly assigns a weight (shrinkage factor) between the restricted asset-pricing model and the data (or the unrestricted statistical model). For a given prior distribution, the weight on the asset-pricing model is larger if the frontier of the factor portfolios has a higher Sharpe ratio, that is, history lends stronger credibility to the factor-based pricing model. The connection between Bayesian inference and shrinkage estimation is used to understand various issues on asset allocation in the presence of model uncertainty. Not surprisingly, the shrinkage factor derived in this article shows that the weight on the asset-pricing model is large if a long history of stationary data is not available. The shrinkage approach helps us understand the empirical Bayes approach in this context. For example, the standard empirical Bayes approach is shown to assign equal weights to the asset-pricing model and the data.

The methodology developed in the article is useful to gain insight into international asset allocation decisions, in particular the home bias phenomenon. The shrinkage approach can be used to provide an alternative measure of the degree of home bias and examine the effects of world equilibrium models on the estimated international diversification benefits for U.S. investors. Finance professionals sometimes contend that investors' distrust of the world equilibrium model might justify the home bias. When discussing the possible explanations for home bias, French and Poterba (1991) write, "The statistical uncertainty associated with estimating expected returns in equity markets makes it difficult for investors to learn that expected returns in domestic markets are not systematically higher than those abroad. Because it is difficult to estimate ex ante returns, investors may follow their own idiosyncratic investment rules with impunity." The framework developed in this article can be used to study whether aversion to model uncertainty justifies home bias. This line of research has been pursued by Wang (2002).

Although the article focuses on the choice between the estimated probability distributions restricted and unrestricted by an asset-pricing model, economists usually face the choice of several asset-pricing models. It is an interesting but difficult task to develop a Bayesian framework of asset allocation using multiple asset-pricing models. However, it might be possible and useful to implement the shrinkage approach to the choice of multiple asset-pricing models. The shrinkage estimate of the mean can be a weighted average of the estimated means restricted by various models. The shrinkage estimate of the variance might be constructed as a quadratic function of the weights and the estimated variances restricted by the models. The extension of the analysis of model uncertainty to multiple asset-pricing models deserves more research.

In this article, it is assumed that the observations of returns are drawn from identical and independent normal distributions with constant parameters. The asset allocation problem is considered for an investor with mean-variance utility function during a single period. Although this is the simplest problem of asset allocation over multiple risky assets, it offers a starting point for understanding the complicated problems of dynamic asset allocation with more general stochastic processes. Numerous recent articles empirically examine optimal asset allocation problems in multi-period settings when returns are predictable. For example, Ang and Bekaert (2002) and Lynch (2001) examine the dynamic optimal portfolios of multiple risky assets when asset returns are predictable. When returns are not distributed identically and independently over time, it would be useful to incorporate dynamic equilibrium models into asset allocation decisions for multi-period settings. Dynamic optimal allocation over a large number of risky assets is more difficult and sensitive to the estimates of the conditional probability distribution. In this area, it may be even more interesting and fruitful to address the issue of model uncertainty using the shrinkage approach.

Appendix A: Mathematical Derivations

A.1. Two lemmas

It is convenient to prove two lemmas before deriving Theorem 1. The set of unknown parameters is denoted by $\Theta = (\Gamma, \Sigma, \mu_2, \Omega_{22})$, where $\Gamma = (\alpha, \beta)'$. The distribution assumption for asset returns in Section 2 and the prior distribution (6) imply that the posterior distribution of (Γ, Σ) is independent of the posterior distribution of (μ_2, Ω_{22}). Let $X = (r_{21}, \ldots, r_{2T})'$ and $Z = (\iota, X)$, where ι is the $T \times 1$ vector of ones. Also, define $D = \theta^{-1}JJ'$, where J is a $(k+1) \times 1$ vector in which the first element is one and all other elements are zero. The posterior distributions imply the following posterior moments of the parameters:

$$E[\Gamma|\Sigma, R] = (D + Z'Z)^{-1}Z'Z\hat{\Gamma} = \tilde{\Gamma} \equiv (\tilde{\alpha}, \tilde{\beta})', \tag{A.1}$$

$$E[\Sigma|R] = (T\hat{\Sigma} + \hat{\Gamma}'Q\hat{\Gamma})/(T - m - k - 1) \equiv \tilde{\Sigma}, \tag{A.2}$$

$$\mathrm{var}(\mathrm{vec}(\Gamma)|R) = \tilde{\Sigma} \otimes (D + Z'Z)^{-1}, \tag{A.3}$$

$$E[\mu_2|\Omega_{22}, R] = \hat{\mu}_2, \tag{A.4}$$

$$E[\Omega_{22}|R] = \hat{\Omega}_{22}T/(T - k - 2) \equiv \tilde{\Omega}_{22}, \tag{A.5}$$

$$\mathrm{var}(\mu_2|R) = \hat{\Omega}_{22}/(T - k - 2), \tag{A.6}$$

where $Q = Z'(I_T - Z(D + Z'Z)^{-1}Z')Z$, and the notation "$\equiv$" represents the expression "which is denoted by." The above formulae can be obtained by following the standard derivations in the textbooks of Bayesian statistics.

Lemma 1. *The posterior means of the parameters are*

$$E[\Gamma|\Sigma, R] = \omega\bar{\Gamma} + (1 - \omega)\hat{\Gamma} \tag{A.7}$$

$$E[\Sigma|R] = h[\omega\bar{\Sigma} + (1 - \omega)\hat{\Sigma}] \tag{A.8}$$

$$E[\mu_2|\Omega_{22}, R] = \hat{\mu}_2 \tag{A.9}$$

$$E[\Omega_{22}|R] = a\hat{\Omega}_{22}, \tag{A.10}$$

where $a = T/(T - k - 2)$ and h is defined as in Theorem 1.

Proof of Lemma 1. It follows from the properties of the inverse of partitioned matrix that

$$(D + Z'Z)^{-1} = \begin{pmatrix} \theta^{-1} + T & \iota'X \\ X'\iota & X'X \end{pmatrix}^{-1} = \omega G + (1 - \omega)(Z'Z)^{-1}, \tag{A.11}$$

where ω and G are defined as follows:

$$\omega = \frac{\theta^{-1}}{\theta^{-1} + T - \iota' X(X'X)^{-1} X'\iota}, \tag{A.12}$$

$$G = \begin{pmatrix} 0 & 0_{1\times k} \\ 0_{k\times 1} & (X'X)^{-1} \end{pmatrix}. \tag{A.13}$$

It is easy to see that $\iota' X(X'X)^{-1} X'\iota = T\hat{\mu}_2'(\hat{\Omega}_{22} + \hat{\mu}_2\hat{\mu}_2')^{-1}\hat{\mu}_2$. The inverse matrix of $\hat{\Omega}_{22} + \hat{\mu}_2\hat{\mu}_2'$ is $\hat{\Omega}_{22}^{-1} - \hat{\Omega}_{22}^{-1}\hat{\mu}_2\hat{\mu}_2'\hat{\Omega}_{22}^{-1}(1 + \hat{\mu}_2'\hat{\Omega}_{22}^{-1}\hat{\mu}_2)^{-1}$, which the reader may verify by multiplying the two together. It follows that $\iota' X(X'X)^{-1} X'\iota = T(\hat{\mu}_2'\hat{\Omega}_{22}^{-1}\hat{\mu}_2)/(1 + \hat{\mu}_2'\hat{\Omega}_{22}^{-1}\hat{\mu}_2)$. It is well known that $\hat{\mu}_2'\hat{\Omega}_{22}^{-1}\hat{\mu}_2$ is the square of the highest Sharpe ratio of the frontier spanned by the sample mean $\hat{\mu}_2$ and variance $\hat{\Omega}_{22}$. Denoting the Sharpe ratio as $\hat{S}$, we have

$$\iota' X(X'X)^{-1} X'\iota = T\hat{S}^2/(1 + \hat{S}^2), \tag{A.14}$$

and thus $T - \iota' X(X'X)^{-1} X'\iota = T/(1 + \hat{S}^2)$, which gives Equation (12). Equation (A.7) in Lemma 1 follows immediately from substituting Equation (A.11) into Equation (A.1). Since substitution of Equation (A.11) into Q gives

$$Q = \omega(T - \iota' X(X'X)^{-1} X'\iota)JJ', \tag{A.15}$$

we should have

$$\hat{\Gamma}' Q\hat{\Gamma} = \omega(T - \iota' X(X'X)^{-1} X'\iota)\hat{\alpha}\hat{\alpha}'. \tag{A.16}$$

Noticing $T - \iota' X(X'X)^{-1} X'\iota = \iota'[I_T - X(X'X)^{-1}X']\iota$ and letting $M = I_T - \iota\iota'/T$, one can then show that

$$\begin{aligned} \hat{\Gamma}' Q\hat{\Gamma} &= \omega\hat{\alpha}\iota'\left[I_T - X(X'X)^{-1}X'\right]\iota\hat{\alpha}' \\ &= \omega Y'\left\{\left[I - X(X'X)^{-1}X'\right] - \left[M - MX(X'MX)^{-1}X'M\right]\right\}Y \\ &= \omega(T\bar{\Sigma} - T\hat{\Sigma}). \end{aligned} \tag{A.17}$$

Substituting this into Equation (A.2) gives Equation (A.8), which completes the proof of Lemma 1. ■

Lemma 2. *The posterior mean of μ and Ω are*

$$E[\mu|\Omega, R] = \omega\begin{pmatrix} \bar{\beta}\hat{\mu}_2 \\ \hat{\mu}_2 \end{pmatrix} + (1-\omega)\begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix}, \tag{A.18}$$

$$E[\Omega|R] = \begin{pmatrix} \Phi(\omega) & a\left[\omega\bar{\beta} + (1-\omega)\hat{\beta}\right]\hat{\Omega}_{22} \\ a\hat{\Omega}_{22}\left[\omega\bar{\beta} + (1-\omega)\hat{\beta}\right]' & a\hat{\Omega}_{22} \end{pmatrix}, \tag{A.19}$$

where

$$\Phi(\omega) = a\left[\omega\bar{\beta} + (1-\omega)\hat{\beta}\right]\hat{\Omega}_{22}\left[\omega\bar{\beta} + (1-\omega)\hat{\beta}\right]' \tag{A.20}$$

$$+ h[\omega\bar{\lambda} + (1-\omega)\hat{\lambda}]\left[\omega\bar{\Sigma} + (1-\omega)\hat{\Sigma}\right], \tag{A.21}$$

$$\hat{\lambda} = \left[(T-2) - \hat{S}^2/(1+\hat{S}^2)\right]/(T-k-2), \tag{A.22}$$

$$\bar{\lambda} = (T-2)/(T-k-2), \tag{A.23}$$

and ω, $\hat{S}$, and h are defined as in Theorem 1 and a is defined as in Lemma 1.

Proof of Lemma 2. Since the posterior distributions of (Γ, Σ) and (μ_2, Ω_{22}) are independent, we have

$$E[\mu|\Omega, R] = \begin{pmatrix} \tilde{\alpha} + \tilde{\beta}\hat{\mu}_2 \\ \hat{\mu}_2 \end{pmatrix}, \tag{A.24}$$

$$E[\Omega|R] = \begin{pmatrix} E[\beta\Omega_{22}\beta'|R] + \tilde{\Sigma} & \tilde{\beta}\tilde{\Omega}_{22} \\ \tilde{\Omega}_{22}\tilde{\beta}' & \tilde{\Omega}_{22} \end{pmatrix}. \tag{A.25}$$

It follows from Equation (A.3) and the law of iterated expectations that

$$E[\beta\Omega_{22}\beta'|R] = \tilde{\beta}\tilde{\Omega}_{22}\tilde{\beta}' + \mathrm{tr}\left[F\tilde{\Omega}_{22}\right]\tilde{\Sigma}, \tag{A.26}$$

where F is the $k \times k$ submatrix in the lower-right corner of $(D + Z'Z)^{-1}$. Using Equations (A.5), (A.11), (A.14), and $\hat{S}^2 = \hat{\mu}_2'\hat{\Omega}_{22}^{-1}\hat{\mu}_2$, we obtain

$$\mathrm{tr}\left(F\tilde{\Omega}_{22}\right) = \frac{1}{T-k-2}\left[\omega\left(k - \frac{\hat{S}^2}{1+\hat{S}^2}\right) + (1-\omega)k\right]. \tag{A.27}$$

Equation (A.18) in the theorem follows easily from Equations (A.7) and (A.24). Equation (A.19) in the theorem follows from Equations (A.7), (A.8), (A.25), (A.26), and (A.27). This completes the proof of Lemma 2. ■

A.2. Proof of Theorem 1

It follows from the law of iterated expectations that

$$E[r_{T+1}|R] = E[\mu|R], \tag{A.28}$$

$$\mathrm{var}(r_{T+1}|R) = E[\Omega|R] + \mathrm{var}(\mu|R). \tag{A.29}$$

Equation (13) holds because of Equations (A.28) and (A.18). Given Equations (A.29) and (A.19), we only need to figure out $\mathrm{var}(\mu|R)$. This can be written as

$$\begin{aligned} \mathrm{var}(\mu|R) &= \mathrm{var}\left[\begin{pmatrix} \alpha + \beta\mu_2 \\ \mu_2 \end{pmatrix}|R\right] \\ &= \frac{1}{T-k-2}\begin{pmatrix} \tilde{\beta}\hat{\Omega}_{22}\tilde{\beta}' & \tilde{\beta}\hat{\Omega}_{22} \\ \hat{\Omega}_{22}\tilde{\beta}' & \hat{\Omega}_{22} \end{pmatrix} + \begin{pmatrix} E[\mathrm{var}(\alpha + \beta\mu_2|\mu_2, R)|R] & 0_{m\times k} \\ 0_{k\times m} & 0_{k\times k} \end{pmatrix}. \end{aligned} \tag{A.30}$$

The second equality in the above expression follows from

$$\mathrm{var}(\alpha + \beta\mu_2|R) = \mathrm{var}(E[\alpha + \beta\mu_2|\mu_2, R]|R) + E[\mathrm{var}(\alpha + \beta\mu_2|\mu_2, R)|R] \tag{A.31}$$

as well as Equations (A.1), (A.4), (A.5), and (A.6). Since $\alpha + \beta\mu_2 = (I_m \otimes (1\ \mu'_2))\ \mathrm{vec}(\Gamma)$, it follows from Equation (A.3) that

$$\begin{aligned} \mathrm{var}(\alpha + \beta\mu_2|\mu_2, R) &= \left(I_m \otimes (1, \mu_2')\right)\left(\tilde{\Sigma} \otimes (D + Z'Z)^{-1}\right)\left(I_m \otimes (1, \mu_2')\right)' \\ &= \tilde{\Sigma} \otimes \left[(1, \mu_2')(D + Z'Z)^{-1}(1, \mu_2')'\right] \\ &= \rho\tilde{\Sigma}, \end{aligned} \tag{A.32}$$

where $\rho = (1\ \mu_2')\ (D + Z'Z)^{-1}\ (1\ \mu_2')'$. With Equations (A.4) and (A.6), it is straightforward to show that the posterior mean of ρ is $E[\rho|R] = \tilde{\rho}$, where

$$\tilde{\rho} = \mathrm{tr}\left[(D + Z'Z)^{-1}\begin{pmatrix} 1 & \hat{\mu}_2' \\ \hat{\mu}_2 & (T-k-2)^{-1}\hat{\Omega}_{22} + \hat{\mu}_2\hat{\mu}_2' \end{pmatrix}\right]. \tag{A.33}$$

One can use Equations (A.11) and (A.14) to show that

$$\tilde{\rho} = \omega\bar{\rho} + (1-\omega)\hat{\rho}, \tag{A.34}$$

where

$$\bar{\rho} = \frac{k}{T(T-k-2)} + \frac{T-k-3}{T(T-k-2)} \cdot \frac{\hat{S}^2}{1+\hat{S}^2}, \tag{A.35}$$

$$\hat{\rho} = \frac{T-2}{T(T-k-2)}. \tag{A.36}$$

Therefore, we have

$$E[\mathrm{var}(\alpha + \beta\mu_2 | \mu_2, R) | R] = [\omega\bar{\rho} + (1-\omega)\hat{\rho}]\tilde{\Sigma}. \tag{A.37}$$

Substituting this into Equation (A.30), we obtain

$$\begin{aligned} \mathrm{var}(\mu|R) &= \mathrm{var}\left[\begin{pmatrix} \alpha + \beta\mu_2 \\ \mu_2 \end{pmatrix} \Big| R\right] \\ &= \frac{1}{T-k-2}\begin{pmatrix} \tilde{\beta}\hat{\Omega}_{22}\tilde{\beta}' & \tilde{\beta}\hat{\Omega}_{22} \\ \hat{\Omega}_{22}\tilde{\beta}' & \hat{\Omega}_{22} \end{pmatrix} + \begin{pmatrix} [\omega\bar{\rho} + (1-\omega)\hat{\rho}]\tilde{\Sigma} & 0_{m\times k} \\ 0_{k\times m} & 0_{k\times k} \end{pmatrix}. \end{aligned} \tag{A.38}$$

We then combine Equations (A.38), (A.29), and Lemma 1 to get $\mathrm{var}(R_{T+1}|R)$. Finally, Equation (14) is obtained by letting $\bar{\delta} = \bar{\lambda} + \bar{\rho}$ and $\hat{\delta} = \hat{\lambda} + \hat{\rho}$. This completes the proof of Theorem 1.

A.3. Derivation of the distribution in Equation (44)

It follows from the definition of $\hat{\Gamma}$ and Equation (2) that

$$\hat{\Gamma} = \Gamma + (Z'Z)^{-1}Z'U. \tag{A.39}$$

Notice that $\alpha = \Gamma' J$ and $\hat{\alpha} = \hat{\Gamma}' J$. We have

$$\hat{\alpha} = \alpha + U'Z(Z'Z)^{-1}J, \tag{A.40}$$

which implies that $\hat{\alpha}$, conditional on X and the parameters, has a normal distribution with mean α and variance

$$\mathrm{var}(\hat{\alpha}) = E\left[U'Z(Z'Z)^{-1}JJ'(Z'Z)^{-1}Z'U\right]. \tag{A.41}$$

Let U_i be the ith column of U and σ_{ij} be the element of Σ at ith row and jth column. The covariance between $\hat{\alpha}_i$ and $\hat{\alpha}_j$ can be calculated as

$$\begin{aligned} \mathrm{var}(\hat{\alpha}_i, \hat{\alpha}_j) &= E\left[U_i'Z(Z'Z)^{-1}JJ'(Z'Z)^{-1}Z'U_j\right] \\ &= J'(Z'Z)^{-1}J\sigma_{ij}. \end{aligned} \tag{A.42}$$

Using the formula of the inverse of partitioned matrix, one can show that

$$J'(Z'Z)^{-1}J = \left(T - \iota' X(X'X)^{-1}X'\iota\right)^{-1}. \tag{A.43}$$

It then follows from Equation (A.14) that

$$\mathrm{var}(\hat{\alpha}_i, \hat{\alpha}_j) = \frac{1}{T}\left(1 + \hat{S}^2\right)\sigma_{ij}, \tag{A.44}$$

which gives

$$\text{var}(\hat{\alpha}) = \frac{1}{T}\left(1 + \hat{S}^2\right)\Sigma. \tag{A.45}$$

Therefore, under the null hypothesis of $\alpha = 0$, we have

$$\hat{\alpha}|\Sigma \sim N\left(0_{m\times 1}, \frac{1}{T}\left(1 + \hat{S}^2\right)\Sigma\right). \tag{A.46}$$

References

Anderson, T. W., 1984, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New york.

Ang, A., and G. Bekaert, 2002, "International Asset Allocation with Regime Shifts," *Review of Financial Studies*, 15, 1137–1187.

Bazaraa, M. S., 1993, *Nonlinear Programming: Theory and Algorithms* (2nd ed.), John Wiley & Sons, New York.

Berger, J. O., 1985, *Statistical Decision Theory and Bayesian Analysis* (2nd ed.), Springer-Verlag, Berlin.

Bernardo, J. M., and A. F. M. Smith, 1994, *Bayesian Theory*, John Wiley & Sons, New york.

Best, M. J., and R. R. Grauer, 1991, "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results," *Review of Financial Studies*, 4, 315–342.

Black, F., and R. Litterman, 1991, *Global Asset Allocation with Equities, Bonds, and Currencies*, Goldman Sachs Fixed Income Research, Publication, New York.

Black, F., and R. Litterman, 1992, "Global Portfolio Optimization," *Financial Analysts Journal*, September–October, 28–43.

Bodie, Z., K. Alex, and A. J. Marcus, 1999, *Investments* (4th ed.), Irwin McGraw-Hill, New York.

Brown, S. J., 1976, "Optimal Portfolio Choice Under Uncertainty: A Bayesian Approach," Unpublished Ph.D. dissertation, University of Chicago.

Britten-Jones, M., 1999, "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights," *Journal of Finance*, 54, 655–671.

Campbell, J., A. Lo, and A. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Chamberlain, G., 2000, "Econometric Applications of Maxmin Expected Utility," *Journal of Applied Econometrics*, 15, 625–644.

Efron, B., and C. Morris, 1973, "Stein's Estimation Rule and its Competitors—An Empirical Bayes Approach," *Journal of the American Statistical Association*, 68, 117–130.

Epstein, L. G., and T. Wang, 1994, "International Asset Pricing under Knightian Uncertainty," *Econometrica*, 62, 283–322.

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

French, K. R., and J. M. Poterba, 1991, "Investor Diversification and International Equity Markets," *American Economic Review*, 81, 222–226.

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

Gilboa, I., and D. Schmeidler, 1989, "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Green, R. C., and B. Hollifield, 1992, "When will Mean-Variance Efficient Portfolios be well Diversified?" *Journal of Finance*, 47, 1785–1809.

Gron, A., B. Jorgensen, and N. G. Polson, 2001, "Portfolio Choice and Stochastic Volatility," working paper, University of Chicago, Chicago, IL.

Hansen, L. P., and T. J. Sargent, 2001, "Robust Control and Model Uncertainty," *American Economic Review*, 91, 60–66.

He, G., and R. Litterman, 1999, *The Intuition Behind Black–Litterman Model Portfolios*, Goldman Sachs Investment Management Research, Publication, New York.

Jagannathan, R., and T. Ma, 2003, "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58, 1651–1683.

Jobson, J. D., and B. Korkie, 1980, "Estimation for Markowitz Efficient Portfolios," *Journal of the American Statistical Association*, 75, 544–554.

Jobson, J. D., B. Korkie, and V. Ratti, 1979, "Improved Estimation for Markowitz Portfolios using James–Stein type Estimators," in *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, 41, 279–284.

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

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

Kogan, L., and T. Wang, 2002, "A Simple Theory of Asset Pricing under Model Uncertainty," working paper, MIT, Cambridge, MA.

Ledoit, O., and M. Wolf, 2003, "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10, 603–621.

Lowenstein, R., 2000, *When Genius Failed.* Random House, New York.

Lynch, A. W., 2001, "Portfolio Choice and Equity Characteristics: Characterizing the Hedging Demands Induced by Return Predictability," *Journal of Financial Economics*, 62, 67–130.

MacKinlay, A. C., and L. Pastor, 2000, "Asset-Pricing Models: Implications for Expected Returns and Portfolio Selection," *Review of Financial Studies*, 13, 883–916.

Maenhout, P. J., 2002, "Robust Portfolio Rules, Hedging and Asset Pricing," working paper, INSEAD.

Michaud, R., 1989, "The Markowitz Optimization Enigma: Is Optimized Optimal?" *Financial Analyst Journal*, 45, 31–42.

Pastor, L., and R. F. Stambaugh, 1999, "Costs of Equity Capital and Model Mispricing," *Journal of Finance*, 54, 67–121.

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

Polson, N. G., and B. V. Tew, 2000, "Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970–1996," *Journal of Business and Economic Statistics*, 18, 164–173.

Polson, N. G., E. Jacquier, and P. E. Rossi, 2002a, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 20, 69–87.

Polson, N. G., M. Johannes, and J. Stroud, 2002b, "Sequential Optimal Portfolio Allocation: Market and Volatility Timing," working paper, University of Chicago, Chicago, IL.

Routledge, B. R., and S. E. Zin, 2000, "Model uncertainty and liquidity," working paper, Carnegie Mellon University, Pittsburgh, PA.

Shanken, J., 1987, "A Bayesian Approach to Testing Portfolio Efficiency," *Journal of Financial Economics*, 19, 195–215.

Sion, M., 1958, "On General Minimax Theorems," *Pacific Journal of Mathematics*, 8, 171–176.

Uppal, R., and T. Wang 2003, "Model Misspecification and Under Diversification," *Journal of Finance*, 58, 2465–2486.

Wang, Zhenyu 2002, "A Shrinkage Approach to Model Uncertainty and Asset allocation," Working paper, Columbia University, New York.