JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 45, No. 4, Aug. 2010, pp. 959–986

doi:10.1017/S0022109010000335

Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty

Jun Tu and Guofu Zhou*

Abstract

This paper proposes a way to allow Bayesian priors to reflect the objectives of an economic problem. That is, we impose priors on the solution to the problem rather than on the primitive parameters whose implied priors can be backed out from the Euler equation. Using monthly returns on the Fama-French 25 size and book-to-market portfolios and their 3 factors from January 1965 to December 2004, we find that investment performances under the objective-based priors can be significantly different from those under alternative priors, with differences in terms of annual certainty-equivalent returns greater than 10% in many cases. In terms of an out-of-sample loss function measure, portfolio strategies based on the objective-based priors can substantially outperform both strategies under alternative priors and some of the best strategies developed in the classical framework.

I. Introduction

Many finance problems have well-defined economic objectives, but usually no connection has been made between parameter estimation and such objectives. In portfolio choice problems, Zellner and Chetty (1965), Brown (1976), (1978), Klein and Bawa (1976), and Jorion (1986) are earlier Bayesian studies under

*Tu, tujun@smu.edu.sg, Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Rd., Singapore 178899; and Zhou, zhou@wustl.edu, Olin School of Business, Washington University in St. Louis, 1 Brookings Dr., St. Louis, MO 63130. We are grateful to Yacine Aït-Sahalia, Doron Avramov, Anil Bera, Henry Cao, Victor DeMiguel, Lorenzo Garlappi, Eric Ghysels, Bruce Hansen, Yongmiao Hong, Chih-Ying Hsiao, Ravi Jagannathan, Raymond Kan, Hong Liu, and Ľuboš Pástor; seminar participants at Fudan University, Tsinghua University, Washington University in St. Louis, the 2005 Finance Summer Camp of Singapore Management University, the 2006 International Symposium on Financial Engineering and Risk Management at Xiamen University, the 2006 China International Conference in Finance, the 18th Asian Finance Association Annual Meetings, and the 16th Annual Meetings of the Midwest Econometrics Group; and especially to Stephen Brown (the editor) and Martijn Cremers (the referee) for many insightful comments that substantially improved the paper. We also thank Lynnea Brumbaugh-Walter for many helpful editorial comments. Tu acknowledges financial support for this project from Singapore Management University Research Grant C207/MSS6B006.

parameter uncertainty that rely on diffuse and data-based priors.[1] Shanken (1987), Harvey and Zhou (1990), and Kandel, McCulloch, and Stambaugh (1995) use similar priors for asset pricing tests. While Pástor (2000) proposes a new class of priors that incorporates an investor's varying beliefs on an asset pricing model, his study does not address the linkage between priors and the economic objectives at hand, nor do other studies in the economics literature, despite increasing applications of Bayesian decision theory to finance (e.g., Kandel and Stambaugh (1996), Barberis (2000), Brennan and Xia (2001), Avramov (2002), (2004), Cremers (2002), (2006), Cohen, Coval, and Pástor (2005), Tu and Zhou (2004), Wang (2005), Tu (2010), and Pástor and Veronesi (2009)). On the literature about Bayesian portfolio analysis, Bawa, Brown, and Klein (1979) provide an excellent earlier survey, while Fabozzi, Huang, and Zhou (2010) and Avramov and Zhou (2010) review the recent developments.

In this paper, we explore a general approach to forming priors based on economic objectives. To see intuitively how an economic objective function may matter, consider how one may allocate funds between a riskless asset and a risky asset. The optimal portfolio weight w is known to be proportional to μ/σ^2 for a mean-variance investor, where μ and σ^2 are the expected excess mean and the variance of the risky asset, respectively. Even before the investor observes any data, it is likely that he might have some idea of the range for w, say within 0 and 1 with high probability. This implies that μ and σ^2 cannot be arbitrarily assigned, but should be related in such a way that the ratio μ/σ^2 falls mostly into a certain range. This prior on μ and σ^2 is different from other priors, since it links the prior to the economic objective at hand. As it turns out, our applications below show that such objective-based priors can make a substantial difference in portfolio decisions as compared with other priors. For example, using monthly returns on the Fama-French (1993) 25 size and book-to-market (BM) portfolios and their 3 factors from January 1965 to December 2004, we find that investment performances under objective-based priors can be significantly different from those under alternative priors, with differences in terms of annual "certainty-equivalent" returns (CERs) greater than 10% in many cases.

The CER measures the difference in Bayesian utilities had one switched from 1 prior to another, without the ability to decide which of the priors is better. In general, it is difficult to argue that 1 prior is better than another, because what is good or bad has to be defined, and the definition may not be agreeable to all investors. Nevertheless, following the literature on statistical decision (see, e.g., Lehmann and Casella (1998)), we use a loss function approach to distinguish the outcomes of using various priors. The prior that generates the minimum loss is viewed as the best prior. In the portfolio choice problem below, the loss function is well defined. In terms of this loss function, we find that the portfolio strategies based on the objective-based priors significantly outperform the strategies based

[1]In the classical framework, different loss functions might be proposed to account for different objectives (see, e.g., Lehmann and Casella (1998)), but the associated parameter estimates are difficult to obtain. Some of these issues are addressed by Tu and Zhou (2010) and references therein.

on other priors. It is in this sense that the objective-based priors are better than others, and are valuable in the context of making portfolio decisions. Intuitively, the objective-based priors incorporate the economic objective at hand into the prior design, and hence they are likely to be useful, since they place greater emphasis on those parameter values whose implied portfolio weights are more likely to maximize the objective function.

Portfolio weights are the parameters of primary interest in the use of the objective-based priors. The importance of focusing on portfolio weights was recognized at least as early as Brandt (1999) and Britten-Jones (1999). Okhrin and Schmid (2006) provide the distributional properties of portfolio weights. In contrast to these studies in the classical framework (which solve the weights and derive their distribution), we impose priors on the portfolio weights, use the first-order condition (FOC) (the Euler equation) to infer priors on the primitive parameters, and then optimize the utility under the predictive density of the data accounting for parameter estimation errors. Bayesian priors on the portfolio weights have received more attention recently. DeMiguel, Garlappi, Nogales, and Uppal (2009) propose a constrained norm approach for portfolio choice and interpret it as a result of using a suitable prior belief on the portfolio weights. Based on a Markov chain Monte Carlo approach, Chevrier and McCulloch (2008) provide a feasible Bayesian portfolio selection framework that directly translates priors on the portfolio weights into portfolio decisions.

The Bayesian approach under objective-based priors is well suited to address questions related to portfolio weights. In particular, it can be applied to assess the economic importance of asset pricing anomalies[2] (see Schwert (2003) for an excellent survey on anomalies). Following Pástor (2000), we assess the importance of asset pricing anomalies by examining the significance of the CERs when an investor avoids investing in assets associated with anomalies. The investor's degree of belief in the usefulness of anomalies can naturally be represented by the investor's prior weights on assets associated with the anomalies. For instance, if the investor is highly skeptical about the anomalies, he can set his prior weights as 0 on the anomaly assets. This prior can then be updated by data via the Bayesian approach. We find that the CERs can be of significant importance even for an investor with a strong skeptical belief about the profitability of anomalies.

The remainder of the paper is organized as follows. Section II provides the objective-based priors and the associated Bayesian framework. Section III extends the analysis to the case in which asset returns are predictable. Section IV compares various Bayesian portfolio rules based on a Bayesian criterion, and Section V compares these Bayesian rules among themselves and with some classical rules based on an out-of-sample criterion. Section VI analyzes asset pricing anomalies in a Bayesian framework. Section VII concludes.

[2]It can shed light on whether investing in a subset of assets is equivalent to investing in all of them, which is related to the "home bias" puzzle in international finance that investors invest mainly in their own countries. This line of study goes beyond the scope of this paper.

II. The Bayesian Framework

A. The Portfolio Choice Problem

Consider the standard portfolio choice problem in which an investor chooses his optimal portfolio among N risky assets and a riskless asset. Let r_{ft} and r_t be the rates of returns on the riskless asset and N risky assets at time t, respectively. We define $R_t \equiv r_t - r_{ft} 1_N$ as the excess returns (i.e., the returns in excess of the riskless asset), where 1_N is an N-vector of 1s, and we make the standard assumption on the probability distribution of R_t that R_t is independent and identically distributed (i.i.d.) over time and has a multivariate normal distribution with mean μ and covariance matrix V.

To have analytical solutions, we focus our analysis on the standard mean-variance framework, since it is one of the most important models and is widely used in practice.[3] However, our approach can be applied to nonquadratic utilities. This will be discussed briefly below.

In the mean-variance framework, the investor at time T chooses his portfolio weights w so as to maximize the quadratic objective function

$$(1) \qquad U(w) = \mathrm{E}[R_p] - \frac{\gamma}{2} \mathrm{Var}[R_p] = w'\mu - \frac{\gamma}{2} w'Vw,$$

where $R_p = w'R_{T+1}$ is the future uncertain portfolio return and γ is the coefficient of relative risk aversion. It is well known that, when both μ and V are assumed known, the portfolio weights are

$$(2) \qquad w^* = \frac{1}{\gamma} V^{-1}\mu,$$

and the maximized expected utility is

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

where $\theta^2 = \mu' V^{-1}\mu$ is the squared Sharpe ratio of the ex ante tangency portfolio of the risky assets.

However, w^* is not computable in practice because μ and V are unknown. To implement the previously mentioned mean-variance theory of Markowitz (1952), the optimal portfolio weights are usually estimated by using a 2-step procedure. First, the mean and covariance matrix of the asset returns are estimated based on the observed data. Second, these sample estimates are treated as if they were the true parameters and are simply plugged into equation (2) to compute the optimal portfolio weights. This gives rise to a parameter uncertainty problem because the utility associated with the plug-in portfolio weights can be substantially different from $U(w^*)$ due to using the estimated parameters that can be substantially different from the true ones.

[3]See Grinold and Kahn (1999), Litterman (2003), and Meucci (2005) for practical applications of the mean-variance framework; and see Brandt (2009) for an excellent survey of the academic literature.

Like all the studies cited in the introduction, this paper will provide a partial equilibrium analysis of the parameter uncertainty problem. The solutions are derived from the investment perspective of an investor whose trading has no impact on the asset prices. An equilibrium analysis, such as the study of the risk premium on parameter uncertainty in an economy with all Bayesian investors, is an important problem but beyond the scope of this paper.

B. The Standard Bayesian Solution

The Bayesian approach provides a natural solution to the parameter uncertainty problem. Following Zellner and Chetty (1965), the Bayesian optimal portfolio is obtained by maximizing the expected utility under the predictive distribution, that is,

$$\begin{aligned} (4) \qquad \hat{w}^{\text{Bayes}} &= \text{argmax}_w \int_{R_{T+1}} \tilde{U}(w) p(R_{T+1}|\boldsymbol{\Phi}_T)\, \mathrm{d}R_{T+1} \\ &= \text{argmax}_w \int_{R_{T+1}} \int_{\mu} \int_{V} \tilde{U}(w) p(R_{T+1}, \mu, V|\boldsymbol{\Phi}_T)\, \mathrm{d}\mu \mathrm{d}V \mathrm{d}R_{T+1}, \end{aligned}$$

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

$$(5) \qquad p(R_{T+1}, \mu, V|\boldsymbol{\Phi}_T) = p(R_{T+1}|\mu, V, \boldsymbol{\Phi}_T) p(\mu, V|\boldsymbol{\Phi}_T),$$

where $p(\mu, V|\boldsymbol{\Phi}_T)$ is the posterior density of μ and V. In comparing equation (4) with equation (1), the expected utility is maximized in the Bayesian and classical framework under the predictive and true distributions, respectively. However, the evaluation of equation (1) requires treating the 2-step estimates as the true parameters and is hence subject to estimation error, while the Bayesian approach accounts for the estimation error automatically. Brown (1976), Klein and Bawa (1976), and Stambaugh (1997), among others, using the standard diffuse prior on μ and V,

$$(6) \qquad p_0(\mu, V) \propto |V|^{-\frac{N+1}{2}},$$

show that the resulting optimal portfolio weights,

$$(7) \qquad \hat{w}^{\text{Bayes}} = \frac{1}{\gamma}\left(\frac{T-N-2}{T+1}\right) \hat{\Sigma}^{-1}\hat{\mu},$$

are always better than the classical plug-in approach in terms of out-of-sample performance. Kan and Zhou (2007) verify this analytically.

However, neither the classical method nor the diffuse prior approach utilizes any prior information about the parameters. Kan and Zhou (2007) show that the Bayesian solution under a diffuse prior can be dominated by alternative estimators, which indicates clearly that the diffuse prior is not optimal in solving the optimal portfolio problem in the presence of parameter uncertainty. In fact, as shown in Section IV, the diffuse prior implies a strong and unreasonable prior on the cross-sectional variation in the portfolio weights. This seems to be the key reason why the diffuse prior fails to do well. The question then is how to construct useful priors that can improve the investor's expected utility.

C. Priors Based on Asset Pricing Theory

Pástor (2000) and Pástor and Stambaugh (2000) introduce interesting priors that reflect an investor's degree of belief in an asset pricing model. To see how this class of priors is formed, assume $R_t = (y_t, x_t)$, where y_t contains the excess returns of m nonbenchmark positions and x_t contains the excess returns of K $(= N - m)$ benchmark positions. Consider a factor model multivariate regression

$$y_t = \alpha + Bx_t + u_t, \tag{8}$$

where u_t is an $m \times 1$ vector of residuals with 0 means and a nonsingular covariance matrix $\Sigma = V_{11} - BV_{22}B'$, and α and B are related to μ and V through

$$\alpha = \mu_1 - B\mu_2, \qquad B = V_{12}V_{22}^{-1}, \tag{9}$$

where μ_i and V_{ij} $(i,j = 1,2)$ are the corresponding partition of μ and V,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}. \tag{10}$$

For a factor-based asset pricing model, such as the 3-factor model of Fama and French (1993), the restriction is $\alpha = 0$.

To allow for mispricing uncertainty, Pástor (2000) and Pástor and Stambaugh (2000) specify the prior distribution of α as a normal distribution conditional on Σ,

$$\alpha|\Sigma \sim N\left[0, \sigma_\alpha^2\left(\frac{1}{s_\Sigma^2}\Sigma\right)\right], \tag{11}$$

where s_Σ^2 is a suitable prior estimate for the average diagonal elements of Σ. The previous α-Σ link is also explored by MacKinlay and Pástor (2000) in the classical framework. The magnitude of σ_α represents an investor's level of uncertainty about the pricing ability of a given model when $\sigma_\alpha = 0$, the investor believes dogmatically in the model, and there is no mispricing uncertainty. On the other hand, when $\sigma_\alpha = \infty$, the investor believes that the pricing model is entirely useless. Although they provide useful insight, the asset-pricing-theory-based priors are not necessarily connected with the investor's objective function. This issue is addressed later.

D. Priors Incorporating Objectives

Consider now how we construct the objective-based priors formally, the innovation of this paper. The idea is to form an informative prior on model parameters such that the implied optimal portfolio is distributed around some reasonable value. Theoretically, because of certain 1-to-1 mapping, this can also be interpreted as we start from a prior on the optimal portfolio weights first, and then we back out the prior on model parameters.

The idea is analogous to those used by Kandel et al. (1995) and Lamoureux and Zhou (1996), among others. In the context of testing portfolio efficiency, Kandel et al. find that the diffuse prior in fact implies a strong prior on inefficiency of a given portfolio.[4] In the context of market return decomposition, Lamoureux and Zhou find that the diffuse prior implies a concentration on extreme values about predictability. These are examples in which supposedly innocuous diffuse priors on some basic model parameters can actually imply rather strong prior convictions about particular economic dimensions of the problem. That is, diffuse priors can be unreasonable in an economic sense in some applications. As a result, both of the cited studies suggest using informative priors on the model parameters that can imply reasonable priors on functions of interest.

The optimal portfolio weights w are the functions of our interest here and are also the solution to the utility maximization problem. Assume for the moment that no data are available and V is a known matrix. Suppose we have a normal prior on μ,

$$\mu \sim N(\gamma V w_0, V_0), \tag{12}$$

where V_0 is the prior covariance matrix of μ. Both w_0 and V_0 are prior constants to be determined later. Based on the objective function (the quadratic utility here), we know, from the FOC or the Euler equation, that w and μ are related by

$$\mu = \gamma V w, \tag{13}$$

which implies w must have the following prior distribution:

$$w \sim N(w_0, V_0 V^{-1}/\gamma). \tag{14}$$

This says that w has a prior mean of w_0. The magnitude of V_0 determines how close the distribution of the implied portfolio is around w_0. Hence, conditional on V and starting from w_0, we can construct a normal prior on μ such that the implied prior on w is concentrated around w_0. If w_0 is chosen as a desired value, the implied prior distribution on w should be more reasonable than otherwise, as shown in our applications later.

Mathematically, we can also interpret that we start from a prior density on w, equation (14), and then, based on the objective function that provides equation (13), we back out the prior on the primitive parameter μ, equation (12). The mapping is clearly 1-to-1 and unique. When V is treated as unknown, as is the case in general, we can set V as a standard Wishart random variable. Then equation (14) implies some sort of mixture normal (unconditional) distribution for w, but μ is still normal conditional on V. Moreover, (w, V) and (μ, V) still have a 1-to-1 mapping, and a prior on the former uniquely determines a prior on the latter, or vice versa. We make 2 remarks. First, we use a normal prior on μ conditional on V so that it is conjugate. Then, the prior can be easily combined with the likelihood function. Second, the previous procedure works for any utility function. This is because equation (13) is the solution to the Euler equation in the special case of the quadratic utility. For nonquadratic utilities, we can numerically solve

[4]Klein and Brown (1984) provide a generic way to obtain an uninformative prior on nonprimitive parameters, which can potentially be applied to derive an uninformative prior on efficiency.

μ for any given w and V. In this case, if we start from a prior on w, we can always determine the prior on μ. A simple approach for doing so is via simulation. A draw of w determines a draw of μ based on the Euler equation, and this prior in turn can be combined with the likelihood function of the data.

Deferring the choice of V_0, we consider first how to determine a sensible value for w_0. In choosing w_0, without observing any data and without knowing the differences between the risky assets, it is reasonable to treat all the risky assets equally. A diversification consideration would suggest that we assign an equal prior weight across all the risky assets, that is, w_0 is proportional to 1_N, a vector of 1s. In other words, w_0 is proportional to the well-known naive $1/N$ rule that invests equally across all the risky assets, which is the focus of DeMiguel, Garlappi, and Uppal (2009) in their comparison with other rules. The sum of the weights across all the risky assets is the total dollar amount invested in risky assets. To reflect a wide range of this allocation to risky assets, we will consider 2 alternative values, 50% and 100% in later applications.

Another sensible value of w_0 is to take it as the value-weighted market portfolio weights, w_m. Doing so leads to an interesting relation to Black and Litterman's (1992) asset allocation method, which has received considerable attention from many practitioners (see, e.g., Litterman (2003), Meucci (2005)). They argue that, once taking w_0 as the market portfolio weights,

$$\mu_m = \gamma_m V w_0 \tag{15}$$

are the equilibrium expected returns as investors hold the market in equilibrium (with γ_m as the risk aversion parameter of the representative investor). These expected returns, which are used in their asset allocation model, yield more balanced portfolios than the standard solution of the mean-variance framework. Like their model, our approach can also use the equilibrium expected returns as the prior means. However, there are 3 major differences between their approach and ours. First, their prior is formed with a view on the equilibrium returns and is updated by investors' proprietary views. In the absence of the proprietary views, their portfolio decision is based on the equilibrium expected returns, and there is no Bayesian updating. In our case, even if we use the market portfolio weights to determine the equilibrium expected returns, these values will be updated by the data. Second, their procedure ignores uncertainty about the covariance matrix. Third, their procedure does not make use of the predictive distribution.[5]

For the prior specification of V_0, a simple way is to use a value proportional to the identity matrix that implies

$$\mu \sim N(\gamma V w_0, \sigma_\rho^2 I_N), \tag{16}$$

where σ_ρ^2 reflects the degree of uncertainty about μ. A 0 value of σ_ρ^2 implies a dogmatic belief in $\mu_0 = \gamma V w_0$ as the true mean conditional on a given w_0. A value of $\sigma_\rho^2 = \infty$ suggests that μ_0 is not informative at all about the true mean. Other

[5]A formal treatment of their model is beyond the scope of this paper. Zhou (2009) provides a framework for combining their model with the data.

than these 2 extremes, σ_ρ^2 places some modest informative belief on the degree of uncertainty as to how μ is close to μ_0.

However, the identity matrix specification has an undesired property. It measures the difference between μ_d, an alternative value of μ, and μ_0,

$$\mu_d - \mu_0 \neq 0, \tag{17}$$

by placing equal importance on the deviation of each element of μ_d from that of μ_0. While this weighting may be plausible in some applications, it does not measure adequately the investor's assessment of the deviations given his utility function. To see this, let w_d and w_0 be the portfolio weights associated with μ_d and μ_0, respectively, based on the objective function. It is easy to show that (see Appendix)

$$U(w_d) - U(w_0) \approx -\frac{1}{2}[\mu_d - \mu_0]'\Omega^{-1}[\mu_d - \mu_0], \tag{18}$$

where

$$\Omega = -\left\{\left\{\frac{\partial^2 U}{\partial w \partial \mu'}[w_0]\right\}'\left\{\frac{\partial^2 U}{\partial w \partial w'}[w_0]\right\}^{-1}\left\{\frac{\partial^2 U}{\partial w \partial \mu'}[w_0]\right\}\right\}^{-1}. \tag{19}$$

Hence, from the perspective of utility evaluation, the investor weighs the importance of the deviations by Ω^{-1} rather than by the identity matrix. This suggests that a potentially better prior on μ is

$$\mu \sim N\left[\gamma V w_0, \sigma_\rho^2\left(\frac{1}{s_\Omega^2}\Omega\right)\right], \tag{20}$$

where s_Ω^2 is the average of the diagonal elements of Ω. In this way, the investor's objective function, the utility function here, also plays a role in the specification of the prior covariance matrix for μ, in addition to its role in the mean specification based on the FOC. Note that the prior given by expression (20) is invariant to any positive monotonic transformations of the utility function. In the case of the mean-variance utility here, it is easy to verify that $\Omega = \gamma V$. Hence, the previous prior can be written simply as

$$\mu \sim N\left[\gamma V w_0, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right], \tag{21}$$

where V is the covariance matrix of the asset returns and s^2 is the average of the diagonal elements of V. As mentioned earlier, we will use a standard Wishart prior for V. Then, we will have a complete prior specification on all the primitive parameters μ and V.

Consider now the case in which part or all of the data are available for forming priors on the parameters.[6] For simplicity, we assume that 10 years of monthly

[6]Empirical Bayesian analysis allows for such flexible use of data to form priors. See Berger (1985) and references therein. Jorion (1986) seems to be one of the first studies using a Bayesian empirical prior.

data are available. Let $\hat{\mu}_{10}$ and $\hat{V}_{10}$ be the sample mean and covariance matrix, respectively. Then, the standard Bayesian informative prior on μ based on the 10 years' data may be written as

$$\mu \sim N\left[\hat{\mu}_{10}, \sigma_\mu^2\left(\frac{1}{\hat{s}_{10}^2}\hat{V}_{10}\right)\right], \tag{22}$$

where $\hat{s}_{10}^2$ is the average of the diagonal elements of $\hat{V}_{10}$, and σ_μ^2 is a scale parameter that indicates the degree of uncertainty.

Given the data, a Bayesian who uses the objective-based priors can start from the nondata prior (21), update it based on the 10 years' data, and then use this updated prior for his future decision making. The approach is analogous to updating the diffuse prior to get equation (22). The updated prior on μ is given by

$$\mu \sim N\left[\hat{\mu}_{10}^*, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right], \tag{23}$$

where $\hat{\mu}_{10}^* = \gamma V\hat{w}_{10}$, and $\hat{w}_{10}$ is the objective-based Bayesian optimal portfolio weights based on the 10 years' data. It is interesting that the conjugate prior, equation (22), provides a similar covariance structure to that of the objective-based prior. However, their means are entirely different, and they can imply significant differences in portfolio decisions, as shown later.

So far we have assumed the quadratic utility for simplicity because the FOC can be solved analytically in this case. For a more general utility function, however, a numerical approach has to be used to solve it. In this case, one can place a truncated prior around the FOC, rather than a simple normal prior as we did here. Due to its technical nature, we will study these issues elsewhere. In a nutshell, our idea is to use the FOC to generate a prior on the primitive parameters. It is these economics-motivated restrictions that are found helpful in our later applications.

E. Performance Measure

It will be of interest to see what the possible performance differences are when one switches from 1 prior to another. As other cases follow straightforwardly, we illustrate only how to measure the differences in the case when an investor switches from the diffuse prior to the objective-based prior. Following Kandel and Stambaugh (1996) and Pástor and Stambaugh (2000), a plausible measure is the difference in the expected utilities of the 2 priors under the predictive distribution of the latter. Let E^* and V^* be the predictive mean and covariance matrix of the asset returns under the objective-based prior, respectively, and let w_O be the associated optimal portfolio allocation. Then the expected utility of using w_O is given by

$$\mathrm{EU}_O = w_O'\mathrm{E}^* - \frac{\gamma}{2}w_O'V^*w_O, \tag{24}$$

where γ is the degree of the investor's relative risk aversion. The allocation, w_D, which is optimal under the diffuse prior, should have an expected utility of

$$\mathrm{EU}_D = w_D'\mathrm{E}^* - \frac{\gamma}{2}w_D'V^*w_D. \tag{25}$$

Notice that this expected utility is evaluated based on the same E^* and V^* of the objective-based prior. Because of this, the difference

$$\text{CER} = \text{EU}_O - \text{EU}_D \tag{26}$$

is interpreted as the "perceived" CER loss to an investor who is forced to accept the optimal portfolio selection based on the diffuse prior, or the "perceived" CER gain of using the objective-based prior instead of the diffuse prior. Since w_O is optimal under the objective-based prior, the CER is always positive or 0 by construction. The issue is how big this value can be. Generally speaking, values of more than a couple of percentage points per year are deemed as economically significant.[7]

It should be acknowledged that the CER measure tells us only the utility differences from switching one prior into another. It does not say that the prior to be switched from is the better, nor that the one to be switched to is the better. As a result, we will also examine performance differences in terms of an out-of-sample loss function measure in Section V, from the perspective of a frequentist.

III. Objective-Based Priors under Predictability

Kandel and Stambaugh (1996) and Barberis (2000) show that incorporating return predictability plays an important role in portfolio decisions. Avramov (2004) extends this in a multivariate setting. The questions we address here are how the objective-based prior can be constructed and whether it can still make significant differences in portfolio decisions in the presence of predictability.

Following the aforementioned studies, we assume that excess returns are related to M predictive variables by a linear regression[8]

$$R_t = \mu_0 + \mu_1 z_{t-1} + v_t, \tag{27}$$

where z_{t-1} is a vector of M predictive variables, $v_t \sim N(0, \Sigma_{RR})$, and the predictive variables follow a vector autoregression (VAR(1)) process

$$z_t = \psi_0 + \psi_1 z_{t-1} + u_t, \tag{28}$$

with $u_t \sim N(0, \Sigma_{ZZ})$.

In a more compact matrix form, we can write the equations as

$$R = X\Gamma + U_R, \tag{29}$$

$$Z = XA_Z + U_Z, \tag{30}$$

where $R = [R_1, R_2, \ldots, R_T]'$ is a $T \times N$ matrix formed from the returns, $X = [1_T, Z_{-1}]$ is a $T \times (M+1)$ matrix formed from a T-vector of 1s and $Z_{-1} =$

[7]Fleming, Kirby, and Ostdiek (2001) provide a similar measure in the classical framework.

[8]Pástor and Stambaugh (2009) and Wachter and Warusawitharana (2009) are examples of recent studies on predictability, while Rapach, Strauss, and Zhou (2010) find that the predictability even holds up out of sample.

$[z_0, z_1, \ldots, z_{T-1}]'$, $\Gamma = [\mu_0, \mu_1]'$ is a $(M+1) \times N$ matrix of the regression coefficients, $Z = [z_1, z_2, \ldots, z_T]'$, $A_Z = [\psi_0, \psi_1]'$ is a $(M+1) \times M$ matrix of the coefficients in the VAR(1) process, and U_R and U_Z are the corresponding residuals with $\mathrm{vec}(U_R) \sim N(0, \Sigma_{RR} \otimes I_T)$ and $\mathrm{vec}(U_Z) \sim N(0, \Sigma_{ZZ} \otimes I_T)$, respectively.

To highlight the intuition, consider the case of 1 predictive variable with $M = 1$. Assume further that the dividend yield, denoted as DY, is used in the predictive regression such that

$$R_t = \mu_0 + \mu_1 \mathrm{DY}_{t-1} + v_t. \tag{31}$$

To reflect a certain degree of uncertainty about predictability, we use a simple normal prior for μ_1,

$$p_0(\mu_1) \propto N\left[\mu_1^p, \sigma_P^2 \left(\frac{1}{s_{RR}^2} \Sigma_{RR}\right)\right], \tag{32}$$

where μ_1^p is the prior mean on μ_1, σ_P^2 measures the uncertainty about predictability, and s_{RR}^2 is the average of the diagonal elements of Σ_{RR}. Assuming a diffuse prior on all other parameters, we have a complete prior

$$p_0(\Gamma, A_Z, \Sigma_{RR}, \Sigma_{ZZ}) \propto p_0(\mu_1) \times |\Sigma_{RR}|^{-\frac{N+1}{2}} \times |\Sigma_{ZZ}|^{-\frac{M+1}{2}}. \tag{33}$$

This joint prior is informative on predictability, but diffuse otherwise. We henceforth refer to it as the predictability-diffuse prior.

To achieve the goal of utility maximization, the FOC imposes an informative prior on $\mu_0 + \mu_1 \mathrm{DY}_T$, or

$$p_0(\mu_0|\mu_1) \propto N\left[\gamma \Sigma_{RR} w_0 - \mu_1 \mathrm{DY}_T, \sigma_\rho^2 \left(\frac{1}{s_{RR}^2} \Sigma_{RR}\right)\right], \tag{34}$$

where w_0 is the prior portfolio weight, DY_T is the observed DY at time T that is available for portfolio selection at time T, and σ_ρ^2 is the prior scalar of the variance that measures the degree of reliance on the FOC. Hence, we define the objective-based prior as the prior constructed by adding this additional conditional density into the right-hand side of equation (33). In contrast with the predictability-diffuse prior, the objective-based prior reflects not only predictability, but also the economic objective. The marginal prior density of $\Gamma = [\mu_0, \mu_1]'$ can be written succinctly as

$$\begin{aligned} p(\Gamma|\Sigma_{RR}) \propto\ & |\Sigma_{RR}|^{-\frac{1}{2}} \\ & \times \exp\left\{-\frac{1}{2}\mathrm{tr}[\Sigma_{RR}^{-1}(\Gamma - \Gamma_0(\mu_1^p))'\Upsilon(\Gamma - \Gamma_0(\mu_1^p))]\right\}, \end{aligned} \tag{35}$$

where $\Gamma_0'(\mu_1^p) = [\gamma w_0 \Sigma_{RR} - \mu_1^p \mathrm{DY}_T, \mu_1^p]$ is an $N \times 2$ matrix, and $\Upsilon = s^2 \Delta \Psi^{-1} \Delta'$ is a 2×2 matrix with

$$\Delta = \begin{pmatrix} 1 & 0 \\ \mathrm{DY}_T & 1 \end{pmatrix}, \quad \Psi = \begin{pmatrix} \sigma_\rho^2 & 0 \\ 0 & \sigma_P^2 \end{pmatrix}.$$

With this simplification, we can combine the objective prior for all of the parameters with the likelihood function of the data and obtain the posterior densities for Γ and Σ_{RR}:

$$\text{(36)} \qquad \mathrm{vec}(\Gamma)|\Sigma_{RR}, \mathcal{D}_T \sim N[\mathrm{vec}(\widetilde{\Gamma}),\ \Sigma_{RR} \otimes (X'X + \Upsilon)^{-1}],$$

$$\text{(37)} \qquad \Sigma_{RR}|\mathcal{D}_T \sim \mathrm{IW}[S_R, T-1],$$

where

$$\text{(38)} \qquad \widetilde{\Gamma} = (X'X+\Upsilon)^{-1}(X'R+\Upsilon\Gamma_0(\mu_1^p)], \qquad S_R = R'R - \widetilde{\Gamma}'X'X\widetilde{\Gamma}.$$

$\mathcal{D}_T$ denotes the data available at time T, and IW$[\cdot]$ denotes the inverted Wishart distribution. With these results, it is easy to obtain the predictive distribution of the returns for our objective-based prior as well as other functions of interest such as optimal portfolio weights.

IV. A Bayesian Comparison

In this section, we first compare the objective-based priors with their usual alternatives based on the Bayesian criterion of equation (26) under the standard i.i.d. assumption. Then, based on the same criterion, we examine the performances under the various priors when the asset returns are assumed predictable.

The data are monthly returns of the well-known Fama-French (1993) 25 size and BM portfolios and their 3 factors (the market, size, and value factors) from January 1965 to December 2004 plus 10 years of earlier data for forming the data-based priors.[9]

A. CERs under Various Priors

Panel A of Table 1 reports the CERs of switching from the diffuse prior to the objective-based prior in the case in which the sum of the weights is 100% (i.e., $w_0 = 1_N/N$). When we apply the priors to 5 years of monthly data ($T = 60$), the CERs are overwhelmingly large (the reason behind this is analyzed below in detail). They range from an annual rate of 22.66% to 125.47%. However, the greater the σ_p, the smaller the gains. This is because a greater value of σ_p moves the objective-based prior closer to the diffuse prior. In the case in which the sum of the weights is 50%, the results are quite similar. For example, the first entry of 125.47 in Table 1 would become 123.93. We omit those results for brevity.

As the sample size grows, the influence of the priors decreases. This is not surprising because both the posterior and the predictive distributions are completely determined by the data when the sample size is infinity, regardless of the priors. However, with a sample size as large as $T = 480$, Panel A of Table 1 shows

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

TABLE 1

CERs and CSTDs of Switching from Diffuse to Objective-Based Priors

Panel A of Table 1 reports the (annualized) "certainty-equivalent" returns (CERs) of switching from the standard diffuse prior,

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

to the objective-based prior

$$p_O(\mu, V) \quad \propto \quad N\left[\gamma V/N, \sigma_p^2\left(\frac{1}{s^2}V\right)\right] \times |V|^{-\frac{N+1}{2}},$$

where s^2 is the average of the diagonal elements of V, γ is the risk aversion coefficient set to be 3, and σ_p^2 reflects the degree of uncertainty about μ. The data are Fama-French (1993) 25 size and book-to-market portfolios and their 3 factors from January 1965 to December 2004, and T is the sample size starting from January 1965. Panel B reports the prior and posterior means of the cross-section standard deviations (CSTDs) of the optimal portfolio weights implied by the 2 priors. Panel C reports the RATIOs of prior or posterior means of the CSTDs implied by the 2 priors.

	σ_p					
T	1%	2%	3%	4%	5%	∞
Panel A. CERs						
60	125.47	91.36	59.18	36.68	22.66	
120	75.57	31.20	12.33	5.33	2.57	
180	52.54	14.39	4.45	1.71	0.77	
240	38.00	8.27	2.33	0.84	0.37	
360	15.28	2.54	0.65	0.22	0.10	
480	8.70	1.24	0.30	0.10	0.04	
Panel B. Prior and Posterior Means						
Prior	45.46	89.81	137.43	185.74	231.08	457,215.43
60	1.19	2.61	4.07	5.43	6.48	10.56
120	0.95	1.85	2.57	3.02	3.31	4.10
180	0.83	1.55	2.05	2.31	2.46	2.86
240	0.82	1.44	1.82	2.00	2.11	2.35
360	0.78	1.25	1.49	1.59	1.64	1.76
480	0.75	1.15	1.32	1.39	1.43	1.51
Panel C. RATIOs						
Prior	10,058.39	5,090.72	3,326.79	2,461.60	1,978.57	
60	8.87	4.04	2.59	1.95	1.63	
120	4.32	2.22	1.59	1.36	1.24	
180	3.43	1.84	1.39	1.24	1.16	
240	2.87	1.63	1.29	1.17	1.11	
360	2.25	1.40	1.18	1.11	1.07	
480	2.00	1.31	1.14	1.08	1.05	

that the CERs can still be substantial. At $\sigma_p = 1\%$, the CER is greater than 8%, although it eventually decreases to an insignificant amount of 0.04% at $\sigma_p = 5\%$. Overall, it is clear that the objective-based prior, when compared with the diffuse prior, makes a significant difference in portfolio selections.

Now, to understand the large CERs, we want to assess the differences in priors on the implied optimal portfolio weights. Let $w = (w_1, \ldots, w_N)'$ be the portfolio weights. We denote CSTD the cross-section standard deviation,

$$\text{(39)} \qquad \text{CSTD} \quad = \quad \frac{1}{N}\sum_{i=1}^{N}(w_i - \bar{w})^2,$$

where $\bar{w}$ is the cross-section mean. It is clear that CSTD measures the relative holdings across assets. If it is too large, the portfolio weights are obviously unreasonable. Under the objective-based prior, the prior mean of CSTD is straightforward to compute based on random draws of μ and V from their prior distributions.

Under the diffuse prior, however, because of its singularity, its properties can only be examined by using an approximation. We use a normal approximation on μ,

$$\mu \sim N\left[\frac{1}{N}1_N, \lambda I_N\right], \tag{40}$$

where λ is set at 100% to ensure diffuseness. The mean $1_N/N$ is immaterial. Note that 1 key feature of the diffuse prior is that μ and V are independent. The diffuse prior on V is approximated by an inverted Wishart distribution

$$V^{-1} \sim W\left[H^{-1}, \nu\right], \tag{41}$$

with degrees of freedom $\nu = 50$, so that the prior contains only information in a small sample of 50 observations. By the properties of the inverted Wishart distribution, the prior expectation of V equals $H/(\nu - N - 1)$. We specify $H = (\nu - N - 1)\hat{V}_{50}/\hat{s}^2_{50}$, so that $\mathrm{E}(V) = \hat{V}_{50}/\hat{s}^2_{50}$. The value of $\hat{s}^2_{50}$ is set equal to the average of the diagonal elements of the sample covariance matrix $\hat{V}_{50}$. Based on priors (40) and (41), we can make $M = 10{,}000$ draws of μ and V easily and then use them to determine the prior mean of CSTD.

The 1st row of Panel B of Table 1 reports the prior means of CSTD. The last entry, 457,215.43, which is incredibly large, is the prior mean of the CSTD implied by the diffuse prior. Clearly the seemingly diffuse prior on μ and V implies too much cross-section variation in asset positions. In contrast, the prior means of the CSTD implied by the objective-based prior with varying σ_ρ are much smaller. For instance, the first entry, 45.46, implied by the objective-based prior with $\sigma_\rho = 1\%$, though still large, is much smaller and more reasonable.

It is of interest to see how the prior means of CSTDs are updated by the data as more and more data are used, similar to Kandel et al. (1995), Lamoureux and Zhou (1996), and Cremers (2006) in analyzing their functions of interest. Since μ and V can be readily drawn from their posterior distributions, the posterior means of CSTDs are easy to compute. As shown by the rest of the rows of Panel B of Table 1, the posterior means are updated quickly. With a sample size $T = 60$, the posterior means become much smaller than their priors. However, the posterior mean based on the diffuse prior is still large compared with those based on the objective prior with small σ_ρs, despite its sharp decrease relative to the prior mean. As the sample size increases, the posterior means decrease further. In addition, the relative differences among them decrease as well when the sample size increases, as shown more clearly in Panel C using the ratios detailed below.

An alternative way of assessing the difference of a pair of prior means or a pair of posterior means of CSTDs under the 2 priors, namely, the diffuse prior and the objective prior with a given σ_ρ, is to examine the ratio between them, denoted as RATIO in Panel C of Table 1. The 1st row of Panel C reports the ratio of implied prior means of CSTDs. With $\sigma_\rho = 1\%$, the prior means of 457,215.43 and 45.46 under the 2 priors implies a RATIO of 10,058.39, which is incredibly large, indicating the sharp difference between the 2 priors. With $\sigma_\rho = 5\%$, the objective-based prior becomes closer to the diffuse prior, and the RATIO decreases to 1,978.57, still a huge value. When updated by some data, such as with a sample size $T = 60$, as implied by the earlier comparison in prior and posterior means, the RATIOs become much smaller, indicating smaller differences in their portfolio implications. As the sample size increases, the updated RATIOs become even smaller, confirming the earlier increasingly smaller differences in the CERs. In the limit, since the implied optimal portfolio weights should converge under either type of priors, the posterior means of CSTDs should become identical and the RATIOs should approach 1.

Consider now the case in which some of the data, those for the 10 years prior to the estimation window, are used to form informative priors. In this case, the data-based prior, equation (22), plays the role of the earlier diffuse prior, while the corresponding objective-based prior is given by equation (23), which is updated from the previous (no data) prior, equation (21), by the same 10 years' data. For simplicity, we set $\sigma_\mu = \sigma_\rho$ in the comparison. Panel A of Table 2 provides the results. The CERs of switching from the data-based prior to the objective-based prior are substantial when $T \leq 180$ or $\sigma_\mu \leq 2\%$. As in the diffuse prior case in Table 1, the CERs in Table 2 are a decreasing function of σ_ρ. However, unlike the diffuse prior case, they are not necessarily smaller as T increases. For example, quite a few of the CERs when $T = 480$ are even greater than those with fewer samples. There are 2 explanations for this. First, in a given application, the entire sample is only 1 path of all possible realizations of the random asset returns. Since the Bayesian criterion is path dependent, the associated expected utilities will not necessarily be a monotonic function of the sample size.[10] Second, even if they were, their differences, the CERs, may not necessarily be so.

For the same reason as before, the CERs are driven by the prior differences in the optimal portfolio weights. As reported in Panel B of Table 2, the RATIOs are quite large.[11] However, in contrast to the diffuse prior case, they are generally much smaller. This is expected, since the data-based prior already uses part of the data in the prior to reduce its uninformativeness. Qualitatively, though, the results are similar to the earlier case in that they are almost always larger than 1, become smaller, and are approaching 1 as the sample size becomes larger.

Finally, consider the performances of the objective-based prior in comparison to those based on asset pricing models. With x_t as the Fama-French (1993) 3 factors, the degree of belief on the validity of the Fama-French 3-factor model is represented by the alpha prior, equation (11). For simplicity, we assume $\sigma_\alpha = \sigma_\rho$ in the comparison. Panel A of Table 3 provides the results. Similar to the data-based prior case in Table 2, the CERs are economically significant for $T \leq 240$ when $\sigma_\rho \leq 2\%$. However, they are small when $T \geq 360$ and $\sigma_\rho \geq 3\%$. The RATIOs, reported in Panel B of Table 3, explain why there are substantially large CERs, and they also suggest that the objective-based prior implies smaller cross-section variation on the optimal portfolio weights than the asset-pricing-model-based priors. However, the RATIOs do not converge to 1 even when $\sigma_\rho = 5\%$ and $T = 480$. An intuitive explanation is that the validity of asset pricing theory is

[10]For the loss function criterion to be discussed in Section V, the monotonicity holds because all the sample paths are integrated out.

[11]For brevity, we omit results similar to Panel B of Table 1 because there are now 5 cases (of the data-based priors) instead of 1 case (of the diffuse prior) in Table 1.

TABLE 2

CERs and CSTDs of Switching from Data-Based to Objective-Based Priors

Panel A of Table 2 reports the (annualized) "certainty-equivalent" returns (CERs) of switching from the data-based prior

$$p_0(\mu, V) \quad \propto \quad N\left[\hat{\mu}_{10}, \sigma_\mu^2\left(\frac{1}{\hat{s}_0^2}\hat{V}_{10}\right)\right] \times |V|^{-\frac{\nu_V+N+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}HV^{-1}\right\},$$

to the objective-based prior

$$p_0(\mu, V) \quad \propto \quad N\left[\hat{\mu}_{10}^*, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right] \times |V|^{-\frac{\nu_V+N+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}HV^{-1}\right\},$$

where $\hat{\mu}_{10}$ and $\hat{V}_{10}$ are the sample mean and covariance matrix of the prior 10 years' data, $\hat{s}_0^2$ is the average of the diagonal elements of $\hat{V}_{10}$, $H = T_{10}\hat{V}_{10}$, $\nu_V = T_{10}$, $T_{10} = 120$, $\hat{\mu}_{10}^* = \gamma V\hat{w}_{10}$, $\hat{w}_{10}$ is the Bayesian optimal portfolio weights based on the prior 10 years' data, s^2 is the average of the diagonal elements of V, the risk aversion coefficient γ is set to be 3, and σ_ρ^2 is a parameter reflecing the degree of uncertainty about μ. The data are Fama-French (1993) 25 size and book-to-market portfolios and their 3 factors from January 1965 to December 2004, and T is the sample size starting from January 1965. Panel B reports the RATIOs of prior or posterior means of the cross-section standard deviations (CSTDs) of the optimal portfolio weights implied by the 2 priors.

	σ_ρ				
T	1%	2%	3%	4%	5%
Panel A. CERs					
60	53.17	29.57	17.89	12.15	8.66
120	42.72	43.70	31.16	19.41	12.19
180	33.88	14.20	7.49	4.12	2.37
240	17.02	4.25	1.60	0.71	0.36
360	6.37	1.92	0.75	0.34	0.17
480	42.85	8.72	2.37	0.95	0.46
Panel B. RATIOs					
Prior	14.63	6.61	4.32	3.40	2.97
60	5.68	3.34	2.60	2.22	2.01
120	2.79	1.61	1.40	1.35	1.33
180	2.11	1.37	1.29	1.26	1.25
240	1.71	1.25	1.21	1.20	1.20
360	1.48	1.17	1.15	1.14	1.15
480	0.75	0.89	1.00	1.06	1.06

fundamentally different from the other priors, and, therefore, it requires much more data to make the RATIOs converge.

In summary, the economic objective of maximizing a utility function provides useful guidance for choosing priors in Bayesian decision making. Under the Bayesian CER measure, we find that such objective-based priors can make significant differences in portfolio performances compared with both the standard statistical and the asset-pricing-theory-based priors. Even with the sample size as large as $T = 480$, there are still cases where the CERs are economically significant.

B. CERs under Predictability

Consider now what happens to the performances under the various priors when the returns are assumed predictable. For comparison, we allow σ_P, the degree of uncertainty about predictability, to take 2 values, infinity and 50%. When $\sigma_P = \infty$, the investor imposes a no-predictability prior. This is an extreme case, whereas $\sigma_P = 50\%$ may be more reasonable. Table 4 provides the results for $\sigma_P = \infty$ and 50%, respectively. In both cases, the CERs are substantial and more pronounced than in Table 1. For example, with $\sigma_\rho = 1\%$, the gains are 198.32% and 74.72%, compared with 125.47% and 8.70% of the i.i.d. case, when $T = 60$

TABLE 3

CERs and CSTDs of Switching from Fama-French 3-Factor Model-Based to Objective-Based Priors

Panel A of Table 3 reports the (annualized) "certainty-equivalent" returns (CERs) of switching from the Fama-French (1993) 3-factor model-based prior,

$$p_0(\alpha, V) \quad \propto \quad N\left(0, \sigma_\rho^2 \frac{1}{s_\Sigma^2} \Sigma\right) \times |V|^{-\frac{N+1}{2}},$$

to the objective-based prior

$$p_0(\mu, V) \quad \propto \quad N\left[\gamma V/N, \sigma_\rho^2 \left(\frac{1}{s^2} V\right)\right] \times |V|^{-\frac{N+1}{2}},$$

where $\alpha = \mu_1 - B\mu_2$, $\Sigma = V_{11} - V_{12}V_{22}^{-1}V_{21}$, s_Σ^2 is the average of the diagonal elements of Σ. s^2 is the average of the diagonal elements of V, γ is the risk aversion coefficient set to be 3, and σ_ρ^2 reflects the degree of uncertainty about α or μ. The data are Fama-French 25 size and book-to-market portfolios and their 3 factors from January 1965 to December 2004, and T is the sample size starting from January 1965. Panel B reports the RATIOs of prior and posterior means of the cross-section standard deviations (CSTDs) of the optimal portfolio weights implied by the 2 priors.

	σ_ρ				
T	1%	2%	3%	4%	5%
Panel A. CERs					
60	83.19	118.94	123.68	113.31	101.47
120	40.26	38.64	25.49	17.92	13.82
180	32.65	18.51	9.83	6.41	4.85
240	27.68	11.30	5.43	3.39	2.52
360	13.92	3.83	1.65	0.99	0.74
480	8.16	1.81	0.72	0.41	0.30
Panel B. RATIOs					
Prior	7.45	8.05	8.71	9.03	9.57
60	4.76	3.57	2.62	2.13	1.80
120	6.19	4.24	3.23	2.84	2.61
180	6.55	4.26	3.32	2.96	2.79
240	6.07	3.89	3.19	2.90	2.78
360	4.73	3.18	2.71	2.54	2.49
480	3.82	2.62	2.28	2.21	2.14

and 480, respectively. Like the i.i.d. case, the CERs decrease as either σ_ρ or T increases. Overall, the presence of predictability does not weaken the earlier results, but strengthens them.

V. Out-of-Sample Performance

The Bayesian evaluation on the performances of the various priors presented thus far is conditional on the data at hand. The comparison does not speak to the performances of the implied portfolio rules for all possible data sets, which a classical statistician may prefer to see. In this section, based on an out-of-sample criterion, we compare the Bayesian rules among themselves and with some of the classical rules studied by Kan and Zhou (2007).

The new comparison is of interest because the Bayesian CER measure provides only the CER difference had one switched from 1 prior to another, and does not say that 1 prior is better or worse than another. The measure is always positive or 0 by definition. As long as 2 priors (good or bad) are significantly different from each other, the measure will be large and positive. To take a stand, following the statistical decision literature (see, e.g., Lehmann and Casella (1998)), we use a

TABLE 4

CERs of Switching from Predictability-Diffuse to Objective-Based Priors

Table 4 reports the (annualized) "certainty-equivalent" returns (CERs) of switching from the predictability-diffuse prior,

$$p_0(\mu_1) \quad \propto \quad N\left[\hat{\mu}_1^o, \sigma_\beta^2\left(\frac{1}{s_{RR}^2}\Sigma_{RR}\right)\right], \tag{A-7}$$

to the objective-based prior

$$p_0(\mu_0, \mu_1) \quad \propto \quad p_0(\mu_1) \times N\left[\gamma\Sigma_{RR}/N - \mu_1 \mathrm{DY}_T, \sigma_\rho^2\left(\frac{1}{s_{RR}^2}\Sigma_{RR}\right)\right], \tag{A-8}$$

where $\hat{\mu}_1^o$ is the slope of the predictive regression $r_t = \mu_0 + \mu_1 \mathrm{DY}_{t-1} + v_t$, $v_t \sim N(0, \Sigma_{RR})$, based on the previous 10 years' data; s_{RR}^2 is the average of the diagonal elements of Σ_{RR}; σ_β^2 measures the degree of uncertainty about predictability; DY_T is the dividend yield at T; γ is the risk aversion coefficient set to be 3; and σ_ρ^2 reflects the degree of uncertainty in the objective-based prior. The data are Fama-French (1993) 25 size and book-to-market portfolios and their 3 factors from January 1965 to December 2004, and T is the sample size starting from January 1965.

	σ_ρ				
T	1%	2%	3%	4%	5%
Panel A. $\sigma_P = \infty$					
60	198.32	150.32	104.96	71.92	46.82
120	167.51	81.79	37.25	17.73	9.40
180	154.80	63.50	25.73	11.46	5.69
240	140.35	53.62	20.66	9.17	4.57
360	95.36	31.05	10.75	4.50	2.15
480	74.72	22.66	7.48	3.04	1.38
Panel B. $\sigma_P = 50\%$					
60	345.84	256.91	174.46	114.55	73.40
120	157.79	76.50	34.60	16.34	8.65
180	122.65	45.33	17.25	7.36	3.55
240	99.50	33.45	11.82	5.08	2.48
360	100.61	31.55	10.74	4.42	2.14
480	59.16	16.59	5.23	2.10	0.93

loss function approach below to distinguish the outcomes of using various priors. The prior that generates the minimum loss is viewed as the best prior.

Any estimated portfolio strategy is a function of the data. Let w^* and $\tilde{w}$ be the true and estimated optimal portfolios, respectively. The expected utility loss from using $\tilde{w}$ rather than w^* is

$$\rho(w^*, \tilde{w}|\mu, \Sigma) \quad \equiv \quad U(w^*) - \mathrm{E}[U(\tilde{w})|\mu, \Sigma], \tag{42}$$

where the first term on the right-hand side is the true expected utility with the use of the true optimal portfolio. Hence, $\rho(w^*, \tilde{w}|\mu, \Sigma)$ is the utility loss if one plays infinite times the investment game with the estimated rule, whether estimated by a Bayesian approach or a non-Bayesian approach. According to this criterion, the difference in the expected utilities between any 2 estimated rules, $\tilde{w}^1$ and $\tilde{w}^2$, should be

$$\text{Gain} \quad = \quad \mathrm{E}[U(\tilde{w}^1)|\mu, \Sigma] - \mathrm{E}[U(\tilde{w}^2)|\mu, \Sigma]. \tag{43}$$

This is an objective utility gain (loss) of using portfolio strategy $\tilde{w}^1$ versus $\tilde{w}^2$ (if using $\tilde{w}^2$ instead), which is an out-of-sample measure since its value is independent of any single set of observation. If it is 2%, it means that the use of $\tilde{w}^1$ instead of $\tilde{w}^2$ will yield a 2% gain in the expected utility. In this case, if $\tilde{w}^1$ is obtained under prior 1 and $\tilde{w}^2$ is obtained under prior 2, we would say that prior 1 is better than prior 2. This is a criterion widely used in classical statistics to evaluate 2 estimators.[12]

The expected utilities associated with most of the Bayesian portfolio rules are difficult to obtain analytically but can be computed numerically via simulation. To be realistic, we set the true parameter values of the model as the sample mean and covariance matrix of the Fama and French (1993) data used in Section IV. Then, we can simulate a large number of data sets from the assumed normal distribution of asset returns. For any 1 draw of the data set with a sample size T, we conduct a Bayesian analysis for all the Bayesian rules under various priors. Each of the rules provides its estimated optimal portfolio weights. Based on the weights, the expected utility can be computed under the true parameters. Then, the average over all the draws, 10,000 of them, is the expected utility or the out-of-sample performance of the rule (i.e., $\mathrm{E}[U(\tilde{w})|\mu, \Sigma]$). Kan and Zhou (2007) and references therein solve this analytically for some of the popular classical rules. In our comparison below with some classical rules, we use the analytical results whenever available.

Table 5 reports the out-of-sample utility gains if an investor switches from the diffuse prior to the objective-based prior. With the sample size varying from 60 to 480, the objective-based prior outperforms consistently. When $T = 60$, regardless of σ_p, the gains are much greater than other cases when $T \geq 120$, suggesting very poor performance of the diffuse prior with a small sample. However, as the sample size increases, the gains, though economically significant, decrease substantially. Nevertheless, even when the sample size is as large as $T = 480$, the gains can still be greater than 3.5%, certainly of significant economic importance. For the same reason as discussed earlier about the large CERs, the large gains here are also due to the fact that the diffuse prior implies an unreasonable prior on the optimal portfolio weights.

TABLE 5

Out-of-Sample Utility Gains of Switching from Diffuse to Objective-Based Priors

Table 5 reports the out-of-sample utility gains of switching from a diffuse prior to objective-based priors with data sets simulated from a multivariate normal distribution whose mean and covariance matrix are calibrated from the monthly returns of the Fama-French (1993) 25 assets and the associated 3 factors from January 1965 to December 2004. The number of simulated data sets is 1,000. The risk aversion coefficient γ is set to be 3.

	σ_p				
T	1%	2%	3%	4%	5%
60	186.06	185.77	168.78	143.39	118.11
120	43.21	45.25	35.28	25.77	18.89
180	19.55	22.07	15.79	10.84	7.65
240	10.54	13.16	8.92	5.97	4.13
360	3.97	6.25	3.99	2.57	1.75
480	1.56	3.55	2.20	1.39	0.96

[12]The weakness of this criterion is that the gain depends on the true parameters. It is difficult to analytically prove that 1 rule is dominated by another for all possible parameter values or for a set of parameter values of interest. Numerically, we can only claim that 1 rule is better or worse than another for the parameter values under consideration.

When 10 years of monthly data are used to form the priors, Table 6 provides the utility gains of switching from the data-based prior to the objective-based prior. Qualitatively, we reach a similar conclusion as for Table 5. When $T \leq 180$, the gains range from 2.04% to 98.58%. These values are clearly economically significant, but smaller than the diffuse prior case in Table 5. This simply states that the data-based prior provides useful information to portfolio selection, and so it does better than the diffuse prior and has smaller utility differences with the objective-based prior. Moreover, when $T = 480$, some of the gains are no longer economically significant, suggesting that the sample size now becomes large enough to make the data-based prior perform as well as the objective-based prior.

TABLE 6

Out-of-Sample Utility Gains of Switching from Data-Based to Objective-Based Priors

Table 6 reports the out-of-sample utility gains of switching from the data-based to the objective-based priors with data sets simulated from a multivariate normal distribution whose mean and covariance matrix are calibrated from the monthly returns of the Fama-French (1993) 25 assets and the associated 3 factors from January 1965 to December 2004. The number of simulated data sets is 1,000. The risk aversion coefficient γ is set to be 3.

	σ_p				
T	1%	2%	3%	4%	5%
60	72.32	98.58	87.56	68.28	52.60
120	21.97	20.24	13.67	8.96	6.44
180	16.61	9.41	5.10	3.04	2.04
240	12.97	5.21	2.38	1.34	0.85
360	8.11	1.80	0.69	0.41	0.23
480	4.73	0.67	0.28	0.16	0.12

When the objective-based prior is compared with the asset-pricing-model-based prior derived from the Fama-French (1993) 3-factor model, Table 7 provides the results. This prior, like others, underperforms the objective-based prior substantially. However, in comparison with the cases reported earlier in Tables 5 and 6, the asset-pricing-model-based prior does better than the diffuse prior when σ_p is small, but worse than the data-based prior. Since the 3-factor model is not the true data-generating process, it provides less useful information than the data-based prior. On the other hand, since the 3-factor model is still not a bad approximation for the data, it is more useful than the diffuse prior. Overall, we find that the objective-based prior has superior performance and provides a better decision rule than all other priors as judged by the loss function criterion, a widely used approach in the statistical decision literature.

Finally, we compare the Bayesian objective-based prior rule with the classical rules studied by Kan and Zhou (2007). For brevity, we analyze 3 of them here. The first is the maximum likelihood (ML) estimator of the optimal portfolio weights, a popular rule in practice. The other 2 are the shrinkage rule of Jorion (1986) and the 3-fund rule of Kan and Zhou, which are the better performing rules among those compared in Kan and Zhou. Table 8 reports the expected utilities for each of the rules. As is well known, the ML rule performs poorly when the sample is small, say less than 240. Its performance becomes comparable with others only when the sample size is as large as 480. The shrinkage and the 3-fund rules are designed to improve upon the ML and are optimal in certain metrics; hence it is

TABLE 7

Out-of-Sample Utility Gains of Switching from Fama-French 3-Factor Model-Based to Objective-Based Priors

Table 7 reports the out-of-sample utility gains of switching from the Fama-French (1993) 3-factor model-based priors to the objective-based priors with data sets simulated from a multivariate normal distribution whose mean and covariance matrix are calibrated from the monthly returns of the Fama-French 25 assets and the associated 3 factors from January 1965 to December 2004. The number of simulated data sets is 1,000. The risk aversion coefficient γ is set to be 3.

	σ_μ				
T	1%	2%	3%	4%	5%
60	54.39	188.21	237.37	242.78	233.54
120	22.04	55.96	56.07	50.58	45.66
180	10.23	26.14	23.51	19.96	17.44
240	5.68	15.71	13.34	11.10	9.60
360	1.48	6.94	5.43	4.29	3.60
480	0.49	4.28	3.33	2.67	2.30

no surprise that they do much better than the ML rule. However, they depend on a set of estimated parameters that makes their performances still worse than the rule implied by the objective-based prior when $T \leq 120$. But, when $T \geq 240$, they have comparable performances with the latter.

TABLE 8

Out-of-Sample Utilities of Classical Rules and a Bayesian Rule

Table 8 reports the out-of-sample expected utilities of the Bayesian rule under the objective-based prior, the shrinkage rule of Jorion (1986), the 3-fund rule of Kan and Zhou (2007), the maximum likelihood rule ($\hat{V}^{-1}\hat{\mu}/\gamma$), and the 1/N rule, with data sets simulated from a multivariate normal distribution whose mean and covariance matrix are calibrated from the monthly returns of the Fama-French (1993) 25 assets and the associated 3 factors from January 1965 to December 2004. The number of simulated data sets is 1,000. The risk aversion coefficient γ is set to be 3.

	Bayesian σ_μ		Classical Rules			
T	1%	2%	Jorion	Kan-Zhou	$(1/\gamma)\hat{V}^{-1}\hat{\mu}$	$1/N$
60	9.50	9.21	−57.67	1.78	−932.13	4.19
120	17.46	19.50	7.17	16.03	−92.29	4.19
180	23.02	25.53	20.36	23.58	−19.76	4.19
240	27.20	29.82	27.02	28.58	4.99	4.19
360	32.67	34.95	33.79	34.36	24.06	4.19
480	36.22	38.22	37.63	37.89	32.22	4.19

The last column of Table 8 reports yet another comparison with the constant $1/N$ rule. DeMiguel et al. (2009) show that it is difficult for the investment strategies developed thus far to outperform the $1/N$ rule, and they conclude that "there are still many 'miles to go' before the gains promised by optimal portfolio choice can actually be realized out of sample." The results in Table 8 show that the Bayesian objective-based prior rule outperforms not only the 3 classical rules, but also the $1/N$ rule consistently across all sample sizes from $T = 60$ to $T = 480$.[13]

Overall, the proposed objective-based prior rule performs impressively against both other Bayesian rules and the classical rules. The results highlight the importance for investors of basing their priors on the solution to an economic optimization problem. In our study, the objective-based prior essentially says that

[13]Tu and Zhou (2010) propose rules that are based on optimal combinations of the $1/N$ rule with major existing rules and find these new rules outperform the $1/N$ rule in general.

our starting point is a simple approximate solution that diversifies our investments across assets, which imposes suitable constraints on model parameters. Then, we let the data update our prior toward the true but unknown optimal portfolio. Because the prior contains useful information on the whereabouts of the true solution (relative to other priors), it turns out to be very valuable.

VI. Assessing the Importance of Anomalies

In this section, we apply our Bayesian framework to study the importance of Fama and French's (1993) BM portfolio when treated as an anomaly to the capital asset pricing model (CAPM). Since our prior starts from portfolio weights, it is well suited for examining the question of whether or not a given subset of assets is important in the investment decision. In particular, the framework can be used to analyze international diversification and asset pricing anomalies. We focus on anomalies in this paper.

Following Pástor (2000), we assume that the anomalies can be transformed into investable assets and then examine whether including them offers any significant CERs in an asset allocation problem. For simplicity, we consider the case of a single anomaly and assume that the last return, R_{Nt}, is the return associated with the anomaly. If an investor is absolutely skeptical about the anomaly, he could assign a 0 weight to R_{Nt}. While this view is difficult to express by using either the diffuse or the asset pricing theory prior, it fits well into our proposed framework. Let w_1, $(N-1)\times 1$, be his prior portfolio weights on the other assets. The earlier prior,

$$(44) \qquad \mu \sim N\left[\gamma V w_a, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right],$$

then represents the prior centered upon the belief $w_a = (w_1', 0)'$. If the investor is dogmatic about his belief, he will choose his optimal portfolio based on the $N-1$ assets only, and not invest in the anomaly asset at all. The associated optimal portfolio weights for the $N-1$ assets are easily computed based on the predictive moments of those $N-1$ assets, with the weight on R_{Nt} being set at 0. In other words, the investor updates only the first $N-1$ component of w_a in light of the data but does not update his prior weight on the anomaly. Let EU_a be the expected utility associated with this optimal portfolio weight.

Consider now an alternative investment strategy, in which the investor updates w_a as usual, based on the predictive moments of all the N risky assets, despite his prior on R_{Nt} being set at 0. Let EU_b be the expected utility with this updated portfolio. Then the difference between EU_b and EU_a provides the CERs of utilizing the anomaly. This is because, although both EU_a and EU_b are computed under the same skeptical prior, EU_b allows investing in R_{Nt}, while EU_a does not.

While the skeptical prior is reasonable for someone who casts a strong doubt on the anomaly, it does not necessarily reflect well the belief of someone who is open to investing in the anomaly asset even before looking at the data. This means that one may compute EU_b under a more balanced prior. The obvious candidate is the prior that assigns equal weights to all the risky assets. We denote the associated expected utility by EU_c. Then, another measure for the impact of utilizing the anomaly is to compare EU_a with EU_c. Intuitively, the difference between EU_c and EU_a should usually be greater than that between EU_b and EU_a. This is because EU_c and EU_b are computed in the same way except that the former generally uses a better prior than the latter. However, as shown by later applications, the difference between EU_b and EU_c is in fact small. Hence, either $EU_b - EU_a$ or $EU_c - EU_a$ will provide a fairly robust measure for the impact of utilizing the anomaly.

Fama and French's (1993) BM portfolio, HML (high minus low), is a well-known anomaly relative to the CAPM. Zhang (2005) explores, among others, some of the theoretical reasons. Here, following Pástor (2000), we examine the economic importance of the HML portfolio based on the approach outlined in Section II. In this case, we have $N = 2$, since the market index and HML are the only risky assets.

Table 9 reports the CERs, $EU_b - EU_a$, in which EU_a is computed by ignoring the anomaly completely under the skeptical prior. It is seen that, as long as the prior precision is not too tight, with $\sigma_p \geq 2\%$, the gains are over 3.72% across sample sizes. The reason that the CERs are getting greater as σ_p increases is that the prior avoids investing in the HML, and this skeptical prior can be mitigated by a larger value of σ_p. As in the previous section, the risk exposure, either $\sum w_{0i} = 0.5$ or 1, has little to do with the CERs, and we report only the results for the latter case. Overall, the results suggest strongly that the HML portfolio is of great economic significance that makes substantial differences in the asset allocation problem.

TABLE 9

CERs of Utilizing Anomaly under a Skeptical Prior

Based on the market (MKT) and the high minus low (HML) book-to-market portfolios from January 1965 to December 2004, Table 9 reports the (annualized) "certainty-equivalent" returns (CERs) of switching from investing only in the MKT to investing in both the MKT and the HML asset under the skeptical prior,

$$p_0(\mu, V) \quad \propto \quad N\left[\gamma V w_0, \sigma_p^2 \left(\frac{1}{s^2} V\right)\right] \times |V|^{-\frac{N-1}{2}},$$

where s^2 is the average of the diagonal elements of V, γ is the risk aversion coefficient set to be 3, σ_μ^2 reflects the degree of uncertainty about μ, w_0 is set to be $(1\ 0)'$, and T is the sample size starting from January 1965.

	σ_p				
T	1%	2%	3%	4%	5%
60	0.54	3.72	7.72	10.72	12.76
120	1.32	5.18	7.73	9.06	9.68
180	1.97	5.26	6.73	7.35	7.72
240	2.70	6.03	7.27	7.80	8.07
360	4.78	8.60	9.73	10.22	10.47
480	4.55	7.20	8.00	8.27	8.42

Intuitively, an investor who avoids investing entirely in the anomaly under the skeptical prior should do even worse than the one who invests in the anomaly under a more balanced prior that assigns an equal weight to both the market and HML. This is indeed the case, as shown by Table 10. However, the additional impacts are small. Table 11 makes it more apparent. The CERs or the utility differences between the skeptical prior and the balanced prior are less than 1% except in 3 scenarios and are less than 0.46% whenever $\sigma_\rho \geq 3\%$. The results say that even when one starts from such a strong prior that one avoids investing in the HML asset entirely, the impact is less than one would expect. In summary, what drives the CERs here is not the priors about whether or not to invest in the anomaly, but rather whether or not to invest in the anomaly asset at all.

TABLE 10

CERs of Utilizing Anomaly under a More Balanced Prior

Based on the market (MKT) and the high minus low (HML) book-to-market portfolios from January 1965 to December 2004, Table 10 reports the (annualized) "certainty-equivalent" returns (CERs) of switching from investing only in the MKT but not investing in the HML anomaly asset under the skeptical prior

$$p_0(\mu, V) \quad \propto \quad N\left[\gamma V w_a, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right] \times |V|^{-\frac{N+1}{2}},$$

to investing in both the MKT and the HML asset under a more balanced prior

$$p_0(\mu, V) \quad \propto \quad N\left[\gamma V/2, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right] \times |V|^{-\frac{N+1}{2}},$$

where s^2 is the average of the diagonal elements of V, γ is the risk aversion coefficient set to be 3. σ_ρ^2 reflects the degree of uncertainty about μ, $w_a = (1\ 0)'$, and T is the sample size starting from January 1965.

	σ_ρ				
T	1%	2%	3%	4%	5%
60	3.93	6.76	10.07	12.37	13.80
120	3.65	6.49	8.47	9.55	10.13
180	3.50	5.98	7.13	7.64	7.87
240	4.01	6.60	7.68	8.02	8.23
360	5.88	8.99	10.03	10.33	10.57
480	5.40	7.51	8.15	8.33	8.45

TABLE 11

CERs of Switching from a Skeptical Prior to a More Balanced Prior

Based on the market (MKT) and the high minus low (HML) book-to-market portfolios from January 1965 to December 2004, Table 11 reports, while allowing investment in both MKT and HML, the (annualized) "certainty-equivalent" returns (CERs) of switching from a skeptical prior

$$p_0(\mu, V) \quad \propto \quad N\left[\gamma V w_a, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right] \times |V|^{-\frac{N+1}{2}},$$

to a more balanced prior

$$p_0(\mu, V) \quad \propto \quad N\left[\gamma V/2, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right] \times |V|^{-\frac{N+1}{2}},$$

where s^2 is the average of the diagonal elements of V, γ is the risk aversion coefficient set to be 3, σ_ρ^2 reflects the degree of uncertainty about μ, $w_a = (1\ 0)'$, and T is the sample size starting from January 1965.

	σ_ρ				
T	1%	2%	3%	4%	5%
60	2.32	1.07	0.46	0.21	0.09
120	1.22	0.27	0.09	0.03	0.01
180	0.58	0.10	0.02	0.01	0.00
240	0.41	0.06	0.01	0.00	0.00
360	0.22	0.03	0.01	0.00	0.00
480	0.14	0.01	0.00	0.00	0.00

VII. Conclusion

This paper explores the link between Bayesian priors and economic objective functions. Once incorporating the economic objectives into priors to estimate unknown parameters, we find that the performance impacts are economically substantial in a standard portfolio allocation problem, whether the stock returns are predictable or not. Moreover, we find that the objective-based priors offer the superior performance not only when we judge them by using an in-sample Bayesian criterion, but also by using an out-of-sample loss function criterion. In addition, while the shrinkage rule of Jorion (1986) and the 3-fund rule of Kan and Zhou (2007) are excellent rules in the classical framework, we find that the Bayesian rule under the objective-based priors can outperform them substantially, suggesting there is real value in using a prior based on the economic objective at hand. We also apply the methodology to examine asset pricing anomalies, and find that Fama and French's (1993) BM (book-to-market) and HML (high minus low) portfolio factors can make substantial differences in an investor's portfolio decision.

Although our study focuses on a portfolio choice problem, the methodology suggests that economic objective-based priors can be explored in almost any financial decision-making problems with parameter uncertainty. In particular, in cases where a Bayesian framework is deemed appropriate, it is highly likely that the decision maker will have some ideas or a broad range about the optimal solution to a given economic objective even without processing any data for formal Bayesian inference. The point of our paper is that this broad range can be used to form objective-based priors that provide information on the plausible values of model parameters so as to help maximize the objective at hand.

Appendix

Proof of equation (18). Recall that the investor's objective is to maximize his expected utility. If μ_d and μ_0 imply weights of w_d and w_0, respectively, then the utility loss caused by the deviation of w_d from w_0 is

$$
\begin{aligned}
\text{(A-1)}\qquad & U(w_d|\mu_0) - U(w_0|\mu_0) \\
= \;& \frac{\partial U}{\partial w'}[w_0|\mu_0][w_d - w_0] + \frac{1}{2}[w_d - w_0]'\frac{\partial^2 U}{\partial w \partial w'}[w_0|\mu_0][w_d - w_0] \\
& + \frac{1}{6}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial^3 U}{\partial w_i \partial w_j \partial w_k}[w_0|\mu_0][w_{di} - w_{0i}][w_{dj} - w_{0j}][w_{dk} - w_{0k}] + \cdots .
\end{aligned}
$$

Ignoring the higher-order terms and using the first-order condition $(\partial U/\partial w')[w_0|\mu_0] = 0$, we have

$$
\text{(A-2)}\qquad U(w_d|\mu_0) - U(w_0|\mu_0) \approx \frac{1}{2}[w_d - w_0]'\frac{\partial^2 U}{\partial w \partial w'}[w_0|\mu_0][w_d - w_0].
$$

Standard calculus implies

$$
\text{(A-3)}\qquad [w_d - w_0] \approx \left\{\frac{\partial^2 U}{\partial w \partial w'}[w_0|\mu_0]\right\}^{-1}\left\{\frac{\partial U}{\partial w}[w_d|\mu_0] - \frac{\partial U}{\partial w}[w_0|\mu_0]\right\},
$$

and

$$
\text{(A-4)}\qquad \left\{\frac{\partial U}{\partial w}[w_d|\mu_0] - \frac{\partial U}{\partial w}[w_0|\mu_0]\right\} \approx \left\{\frac{\partial^2 U}{\partial w \partial \mu'}[w_0|\mu_0]\right\}[\mu_d - \mu_0].
$$

Therefore, we have equation (18), which says that the utility loss is approximately equal to the weighted average of the deviation of μ_d from μ_0, with the weighting matrix determined by the utility function.

In the case of mean-variance utility, the approximation holds exactly, and it is also easy to verify that

$$\text{(A-5)} \qquad \left\{ \frac{\partial^2 U}{\partial w \partial \mu'} [w_0 | \mu_0] \right\} = I_N,$$

$$\text{(A-6)} \qquad \left\{ \frac{\partial^2 U}{\partial w \partial w'} [w_0 | \mu_0] \right\} = -\gamma V,$$

where V is the covariance matrix of the asset returns. Therefore, in the mean-variance case, $\Omega = \gamma V$. □

References

Avramov, D. "Stock Return Predictability and Model Uncertainty." *Journal of Financial Economics*, 64 (2002), 423–458.

Avramov, D. "Stock Predictability and Asset Pricing Models." *Review of Financial Studies*, 17 (2004), 699–738.

Avramov, D., and G. Zhou. "Bayesian Portfolio Analysis." *Annual Review of Financial Economics*, forthcoming (2010).

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, The Netherlands: North-Holland (1979).

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

Black, F., and R. Litterman. "Global Portfolio Optimization." *Financial Analysts Journal*, 48 (1992), 28–43.

Brandt, M. W. "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach." *Journal of Finance*, 54 (1999), 1609–1645.

Brandt, M. W. "Portfolio Choice Problems." In *Handbook of Financial Econometrics*, Vol. 1, Y. Aït-Sahalia and L. Hansen, eds. Amsterdam, The Netherlands: Elsevier, North-Holland (2009), 269–336.

Brennan, M. J., and Y. Xia. "Assessing Asset Pricing Anomalies." *Review of Financial Studies*, 14 (2001), 905–942.

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

Brown, S. J. "Optimal Portfolio Choice under Uncertainty." Ph.D. Dissertation, 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.

Chevrier, T., and R. E. McCulloch. "Using Economic Theory to Build Optimal Portfolios." Working Paper, University of Chicago (2008).

Cohen, R. B.; J. D. Coval; and L. Pástor. "Judging Fund Managers by the Company They Keep." *Journal of Finance*, 60 (2005), 1057–1096.

Cremers, K. J. M. "Stock Return Predictability: A Bayesian Model Selection Perspective." *Review of Financial Studies*, 15 (2002), 1223–1249.

Cremers, M. "Multifactor Efficiency and Bayesian Inference." *Journal of Business*, 79 (2006), 2951–2998.

DeMiguel, V.; L. Garlappi; N. Nogales; and R. Uppal. "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms." *Management Science*, 55 (2009), 798–812.

DeMiguel, V.; L. Garlappi; and R. Uppal. "Optimal versus Naive Diversification: How Inefficient Is the $1/N$ Portfolio Strategy?" *Review of Financial Studies*, 22 (2009), 1915–1953.

Fabozzi, F.; D. Huang; and G. Zhou. "Robust Portfolios: Contributions from Operations Research and Finance." *Annals of Operations Research*, 176 (2010), 191–220.

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.

Fleming, J.; C. Kirby; and B. Ostdiek. "The Economic Value of Volatility Timing." *Journal of Finance*, 56 (2001), 329–352.

Grinold, R. C., and R. N. Kahn. *Active Portfolio Management: Quantitative Theory and Applications*. New York, NY: McGraw-Hill (1999).

Harvey, C. R., and G. Zhou. "Bayesian Inference in Asset Pricing Tests." *Journal of Financial Economics*, 26 (1990), 221–254.

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

Kan, R., and G. Zhou. "Optimal Portfolio Choice with Parameter Uncertainty." *Journal of Financial and Quantitative Analysis*, 42 (2007), 621–656.

Kandel, S.; R. McCulloch; and R. F. Stambaugh. "Bayesian Inference and Portfolio Efficiency." *Review of Financial Studies*, 8 (1995), 1–53.

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.

Klein, R. W., and S. J. Brown. "Model Selection When There Is 'Minimal' Prior Information." *Econometrica*, 52 (1984), 1291–1312.

Lamoureux, C. G., and G. Zhou. "Temporary Components of Stock Returns: What Do the Data Tell Us?" *Review of Financial Studies*, 9 (1996), 1033–1059.

Lehmann, E. L., and G. Casella. *Theory of Point Estimation*. New York, NY: Springer-Verlag (1998).

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

MacKinlay, A. C., and Ľ. Pástor. "Asset Pricing Models: Implications for Expected Returns and Portfolio Selection." *Review of Financial Studies*, 13 (2000), 883–916.

Markowitz, H. M. "Portfolio Selection." *Journal of Finance*, 7 (1952), 77–91.

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

Okhrin, Y., and W. Schmid. "Distributional Properties of Portfolio Weights." *Journal of Econometrics*, 134 (2006) 235–256.

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

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

Pástor, Ľ., and R. F. Stambaugh. "Predictive Systems: Living with Imperfect Predictors." *Journal of Finance*, 64 (2009), 1583–1628.

Pástor, Ľ., and P. Veronesi. "Learning in Financial Markets." *Annual Review of Financial Economics*, 1 (2009), 361–381.

Rapach, D. E.; J. K. Strauss; and G. Zhou. "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *Review of Financial Studies*, 23 (2010), 821–862.

Schwert, B. "Anomalies and Market Efficiency." In *Handbook of the Economics of Finance*, Vol. 1B, G. M. Constantinides, R. M. Stulz, and M. Harris, eds. Amsterdam, The Netherlands: North-Holland (2003), 937–972.

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

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

Tu, J. "Is Regime Switching in Stock Returns Important in Portfolio Decisions?" *Management Science*, 56 (2010), 1198–1215.

Tu, J., and G. Zhou. "Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions?" *Journal of Financial Economics*, 72 (2004), 385–421.

Tu, J., and G. Zhou. "Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies." *Journal of Financial Economics*, forthcoming (2010).

Wachter, J. A., and M. Warusawitharana. "Predictable Returns and Asset Allocation: Should a Skeptical Investor Time the Market?" *Journal of Econometrics*, 148 (2009), 162–178.

Wang, Z. "A Shrinkage Approach to Model Uncertainty and Asset Allocation." *Review of Financial Studies*, 18 (2005), 673–705.

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.

Zhang, L. "The Value Premium." *Journal of Finance*, 60 (2005), 67–103.

Zhou, G. "Beyond Black-Litterman: Letting the Data Speak." *Journal of Portfolio Management*, 36 (2009), 36–45.