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

Jun Tu and Guofu Zhou*

JEL classification: G11; G12; C11
Keywords: Portfolio choice; Parameter uncertainty; Bayesian priors

Current version: March, 2008

*Singapore Management University and Washington University in St. Louis, respectively. We are grateful to Yacine Aït-Sahalia, Doron Avramov, Anil Bera, Henry Cao, Victor DeMiguel, Lorenzo Garlappi, Eric Ghysels, Bruce Hansen, Yongmiao Hong, Ravi Jagannathan, Raymond Kan, Hong Liu, Ľuboš Pástor, and seminar participants at Fudan University, Tsinghua University, Washington University, 2005 Finance Summer Camp of Singapore Management University, 2006 International Symposium on Financial Engineering and Risk Management at Xiamen University, 2006 China International Conference in Finance, and the 16th Annual Meetings of the Midwest Econometrics Group for many helpful comments. We also thank Lynnea Brumbaugh-Walter for many helpful editorial comments.

Corresponding Author: Guofu Zhou, Olin School of Business, Washington University, St. Louis, MO 63130. Phone: (314) 935-6384 and e-mail: zhou@wustl.edu

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

Economic objectives are often ignored when estimating parameters, though the loss of doing so can be substantial. This paper proposes a way to allow Bayesian priors to reflect the objectives. Using monthly returns of the Fama-French 25 size and book-to-market portfolios and their three factors from January 1965 to December 2004, we find that investment performance under the objective-based priors can be significantly different from that under alternative priors, with differences in terms of annual certainty-equivalent returns greater than 10% in many cases. In terms of out-of-sample performance, the Bayesian rules under the objective-based priors can outperform substantially some of the best rules developed in the classical framework.

I. Introduction

Many finance problems have well-defined economic objectives, but model estimation usually makes no connection to such objectives and is done primarily based on statistical criteria. In the classic framework, different loss functions might be proposed to account for different objectives (Lehmann and Casella, 1998), but the associated parameter estimates are difficult to obtain. In the Bayesian framework, priors are supposed to reflect available information on the problem, but the information on economic objectives is often ignored by diffuse and data-based priors (see, e.g., Shanken, 1987, Harvey and Zhou, 1990, and Kandel, McCulloch, and Stambaugh, 1995). 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), Cohen, Coval and Pástor (2005), Tu and Zhou (2004), and Wang (2005).[1]

This paper is Bayesian. We explore a general approach to form priors based on economic objectives. We focus our analysis on the optimal portfolio selection problem in the standard mean-variance framework due to its simplicity and its wide use in practice. Ever since the publication of Markowitz's (1952) seminal work, extensive research and results have been available in the mean-variance framework that permits analytical insights on the role played by economic objectives. Zellner and Chetty (1965), Brown (1976, 1978), Klein and Bawa (1976), and Jorion (1986) are earlier Bayesian studies on the portfolio selection problem that account for parameter uncertainty. In contrast to their studies and to more recent ones reviewed earlier, we propose in this paper new innovative priors that are closely tied to the first-order conditions (FOCs) of maximizing the economic objectives. We show that such objective-based priors place useful restrictions on model parameters, and these restrictions are fundamentally different from those implied by either diffuse or data-based priors.

To see intuitively how an economic objective function may matter, consider allocating

[1]See Kan and Zhou (2007) and references therein for recent studies on parameter uncertainty in the classical framework.

funds between a riskless asset and a risky one. The optimal portfolio weight is known to be proportional to μ/σ^2 for a mean-variance investor, where μ and σ^2 are the mean and variance of the return on the risky asset in excess of the riskless one. Even before the investor observes any data or does any formal statistical analysis, it is likely that he might have some idea about the range of the optimal portfolio weight, w , should be, say $0 < w < 1$. This implies that μ and σ^2 cannot be arbitrarily assigned, but are related in such a way that the ratio μ/σ^2 falls into the given range. This restriction is a prior originated from choosing w to maximize the objective function. The prior restriction on the optimal solution (between zero and one) imposes a prior restriction on μ and σ^2. Since the latter is based on the FOC of the utility maximization problem, the associated prior is objective-based. Intuitively, the objective-based prior is likely more useful than the diffuse one. Since the prior places greater weight on those parameter values whose implied portfolio weights are likely to maximize the objective function, the resulting expected utility is likely higher. As it turns out, our applications do indeed show that such objective-based restrictions can make a substantial difference in portfolio decisions as compared with other priors. For example, using monthly returns of the popular Fama-French 25 size and book-to-market portfolios and their three factors from January 1965 to December 2004, we find that the difference in investment performance under the objective-based priors versus other priors can be more than 10% in many cases. The objective-based priors can also be formed with data. When some data are available, the researcher can use them to form new informative priors on the portfolio weights, which can in turn be used, based on the FOC relation, to form new objective-based priors on the parameters.

The Bayesian approach under the 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 utility losses when an investor avoids investing in assets associated with anomalies. The investor's degrees of belief on the usefulness of anomalies can naturally

[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 the paper.

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 zeros on the anomaly assets. This prior can then be updated by data via the Bayesian approach. We find that the utility losses 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 with classical rules based on an out-of-sample criterion. Section VI analyzes asset pricing anomalies in a Bayesian framework. Section VII concludes.

II. The Bayesian Framework

In this section, we first review the portfolio choice problem, its standard Bayesian solution and existing prior formulations. Then, we propose our objective-based priors, and discuss the Bayesian criterion for comparing the differences in investment decisions based on various priors.

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 the 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 ones, and make the standard assumption on the probability distribution of R_t that R_t is independent and identically distributed 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, whereas the case of non-quadratic utilities will be discussed briefly later. In the mean-variance framework, the investor at time T chooses his portfolio weights w so as to maximize the quadratic objective function

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

where $R_p = w'R_{T+1}$ is the future uncertain portfolio return and γ is the coefficient of relative risk aversion (which is set at 3 in our empirical applications below). It is well-known that, when both μ and V are assumed known, the portfolio weights are

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

and the maximized expected utility is

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

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 above mean-variance theory of Markowitz (1952), the optimal portfolio weights are usually estimated by using a two-step procedure. First, the mean and covariance matrix of the asset returns are estimated based on the observed data. Second, these sample estimates are then treated as if they were the true parameters, and are simply plugged into (2) to compute the optimal portfolio weights. 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.

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, i.e.,

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

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

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

where $p(\mu, V|\mathbf{\Phi}_T)$ is the posterior density of μ and V. In comparison to Equation (4) with Equation (1), the expected utility is maximized in both the Bayesian and classical framework under the predictive and true distributions, respectively. However, the evaluation of Equation (1) requires treating the two-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,

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

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

However, the diffuse prior improves the plug-in results by only a very small amount since the portfolio weights differ only by $1/(T+1)$ percent. This is not surprising from a statistical point of view. Neither the classical method nor the diffuse prior approach utilizes any prior information about the parameters. Because of diffuse information in both cases, their results should be close. On the other hand, Kan and Zhou (2007) show that the Bayesian solution can be dominated by alternative estimators, which clearly indicates that the diffuse prior is not optimal in solving the optimal portfolio problem in the presence of parameter uncertainty. The question is then how to construct useful priors that can improve the investor's expected utility beyond the use of the diffuse prior.

C. Priors Based on Asset Pricing Theory

Pástor (2000) and Pástor and Stambaugh (2000)) introduce interesting priors that reflect an investors' 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 non-benchmark positions and x_t contains the excess returns of K $(= N - m)$ benchmark positions. Consider a factor model multivariate regression

$$y_t = \alpha + Bx_t + u_t, \tag{7}$$

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

$$\alpha = \mu_1 - B\mu_2, \qquad B = V_{12}V_{22}^{-1}, \tag{8}$$

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

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \; V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}. \tag{9}$$

For a factor-based asset pricing model, such as the three-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,

$$\alpha|\Sigma \sim N\left[0, \sigma_\alpha^2\left(\frac{1}{s_\Sigma^2}\Sigma\right)\right], \tag{10}$$

where s_Σ^2 is a suitable prior estimate for the average diagonal elements of Σ. The above alpha-Sigma 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 is the issue addressed below.

D. The Prior Based on Economic Objectives

Consider now how to form the objective-based priors formally, the innovation of the current paper. Assume for the moment that there is no data available. We would like to illustrate how one can construct informative priors based on the objective function. Suppose that we are interested in forming a normal prior on μ,

$$\mu \sim N(\mu_0, V_0), \tag{11}$$

where μ_0 and V_0 are the prior mean and covariance matrix. To reflect the economic objective, it is natural to link the prior mean to what is implied by the first-order condition (2),

$$\mu \sim N(\gamma V \bar{w}_0, V_0), \tag{12}$$

where $\bar{w}_0$ is our prior mean of the portfolio weights. This says that the prior mean is proportional to both the covariance matrix of the asset returns and $\bar{w}_0$. Conditional on V, the prior expected returns are high on those assets whose risks are high, and the implied mean portfolio weights are $\bar{w}_0$. The magnitude of V_0 determines how close the implied portfolio weights are to $\bar{w}_0$. The remaining question is how to determine a value for $\bar{w}_0$.

In choosing a suitable prior vector $\bar{w}_0$, we assign first a prior value for the sum of its components. This is our prior wealth allocation on the risky assets. To reflect a wide range, we consider two alternative values of 50% and 100%, respectively.[3] Although the total allocation to risky assets is assigned, the portfolio weights across the assets are unknown and should be assigned too. Without any data and without knowing the differences between the assets, it is reasonable to use a *diffuse* or an uninformative value. It treats all assets equally and hence it assigns an equal prior weight across them.

Another sensible prior might be to take w_0 as the value-weighted market portfolio weights, w_m. So doing 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., Grinold and Kahn, 1999, Litterman, 2003, and Meucci, 2005). They argue that, once taking w_0 as the market portfolio weights,

$$\mu_m = \gamma_m V w_m \tag{13}$$

are the equilibrium expected returns as investors hold the market in equilibrium (with γ_m as the risk aversion parameter of the representative investor). It is these expected returns

[3]Notice that 50% and 100% are the implied mean allocations on the risky assets. To impose the condition that the allocation must be in a fixed range, a truncated normal distribution for μ may be used.

that are used in their asset allocation model that yields more balanced portfolios than the standard solution of the mean-variance framework. Like their model, our approach here can also use the equilibrium expected returns as the prior means. However, there are three major differences between their approach and ours. First, their prior is formed with a view on the equilibrium returns and 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, our prior is a prior on the solution, the portfolio weights. Even if we use the market portfolio weights to determine the equilibrium expected returns, these values will be updated by data. Second, their Bayesian procedure is ad hoc in the sense that their Bayesian updating with views does not rely on the posterior that factors into the uncertainty about the covariance matrix. Third, their approach, as seen in their last stage for computing the optimal portfolio weights, ignores the parameter uncertainty problem whereas the predictive distribution used here accounts for it.[4]

Now we need also to have a prior specification for V_0. A simple way of doing so is to use the identity matrix that implies

$$\mu \sim N(\gamma V w_0, \sigma_\rho^2 I_N), \tag{14}$$

where σ_ρ^2 reflects the degree of uncertainty about μ. A zero 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 than these two 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,

$$\mu_d - \mu_0 \neq 0, \tag{15}$$

by placing equal importance on the deviations 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

[4] A formal treatment of their model is beyond the scope of this paper. Zhou (2008) provides a framework for combining the Black and Litterman (1992) model with the data.

be the portfolio weights associated with μ_d and μ_0 based on the objective function. It is easy to show that (see Appendix A)

$$U(w_d) - U(w_0) \approx -\frac{1}{2}[\mu_d - \mu_0]'\Omega^{-1}[\mu_d - \mu_0], \tag{16}$$

where

$$\Omega = -\left\{\left\{\frac{\partial^2 U}{\partial w \partial \mu'}[w_0]\right\}' \left\{\frac{\partial^2 U}{\partial w \partial w'}[w_0]\right\}^{-1} \left\{\frac{\partial^2 U}{\partial w \partial \mu'}[w_0]\right\}\right\}^{-1}. \tag{17}$$

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

$$\mu \sim N\left[\gamma V w_0, \sigma_\rho^2\left(\frac{1}{s_\Omega^2}\Omega\right)\right], \tag{18}$$

where s_Ω^2 is the average of the diagonal elements of Ω. In this way, the investors' 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 by the FOC. Note that the prior given by (18) 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 above prior can be simply written as

$$\mu \sim N\left[\gamma V w_0, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right], \tag{19}$$

where V is the covariance matrix of the asset returns, and s^2 is the average of the diagonal elements of V.

Consider now the case in which part or all of the data are available for forming priors on the parameters.[5] For simplicity, we assume that there are ten years of monthly data 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 ten years data may be written as

$$\mu \sim N\left[\hat{\mu}_{10}, \sigma_\mu^2\left(\frac{1}{\hat{s}_{10}^2}\hat{V}_{10}\right)\right], \tag{20}$$

[5]Empirical Bayesian analysis allows for such flexible use of data to form priors. See Berger (1985) and references therein.

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 non-data prior (19), update it based on the ten years data, and then use this updated prior for his future decision making. The approach is analogous to the way of updating the diffuse prior to get (20). The updated prior on μ is given by

$$\mu \sim N\left[\hat{\mu}_{10}^*, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right], \tag{21}$$

where $\hat{\mu}_{10}^* = \gamma V \hat{w_{10}}$, and $\hat{w_{10}}$ is the objective-based Bayesian optimal portfolio weights based on the ten years data. It is interesting that the conjugate prior, equation (20), provides a similar covariance structure to that of the objective-based prior. However, their means are entirely different, and they can make important differences in portfolio decisions as shown later.

So far we have assumed the quadratic utility for simplicity because the first-order condition 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 first-order condition, 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 of the paper is to use the FOC for the problem at hand to generate a prior on the 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 gains are when one switches from one prior to another. As other cases follow straightforwardly, we illustrate how to measure the gains only from the diffuse prior to the objective-based prior. Following Kandel and Stambaugh (1996) and Pástor and Stambaugh (2000), a plausible measure is the utility gain given by the difference in the expected utilities of the two priors under the predictive distribution of the latter.[6] Let E^* and V^* be the predictive mean and covariance matrix of the asset returns

[6] A classical statistician may prefer to use the true distribution of the data to differentiate the priors. This out-of-sample measure will be analyzed in Section V.

under the objective-based prior, equation (19), and w_O be the associated optimal portfolio allocation. Then the expected utility is given by

$$EU_O = w_O' E^* - \frac{1}{2}\gamma w_O' V^* w_O, \tag{22}$$

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

$$EU_D = w_D' E^* - \frac{1}{2}\gamma w_D' V^* w_D. \tag{23}$$

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

$$CE = EU_O - EU_D \tag{24}$$

is interpreted as the 'perceived' gains of utilizing the objective-based prior, or as the 'perceived' losses in terms of the certainty-equivalent return to an investor who is forced to accept the optimal portfolio selection based on the diffuse prior. Since w_O is optimal under the objective-based prior, CE is always positive or zero by construction. The issue is how big this value can be. Generally speaking, values over a couple of percentage points per year are deemed as economically significant.[7]

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 produce significant economic gains in the presence of predictability.

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

$$R_t = \mu_0 + \mu_1 z_{t-1} + v_t, \tag{25}$$

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

[8] Pástor and Stambaugh (2006), and Wachter and Warusawitharana (2007) are recent Bayesian studies on predictability.

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

$$z_t = \psi_0 + \psi_1 z_{t-1} + u_t, \tag{26}$$

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

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

$$R = X\Gamma + U_R, \tag{27}$$

$$Z = XA_Z + U_Z, \tag{28}$$

where $R = [R_1, R_2, \cdots, R_T]'$ 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 ones and $Z_{-1} = [z_0, z_1, \cdots, z_{T-1}]'$, $\Gamma = [\mu_0, \mu_1]'$ is a $(M+1) \times N$ matrix of the regression coefficients, $Z = [z_1, z_2, \cdots, z_T]'$, $A_Z = [\psi_0, \psi_1]'$ 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 $\text{vec}(U_R) \sim N(0, \Sigma_{RR} \otimes I_T)$ and $\text{vec}(U_Z) \sim N(0, \Sigma_{ZZ} \otimes I_T)$.

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

$$R_t = \mu_0 + \mu_1 DY_{t-1} + v_t. \tag{29}$$

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

$$p_0(\mu_1) \propto N\left[\mu_1^p, \sigma_P^2\left(\frac{1}{s_{RR}^2}\Sigma_{RR}\right)\right], \tag{30}$$

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

$$p_0(\Gamma, A_Z, \Sigma_{RR}, \Sigma_{ZZ}) \propto p_0(\mu_1) \times |\Sigma_{RR}|^{-\frac{N+1}{2}} \times |\Sigma_{ZZ}|^{-\frac{M+1}{2}}. \tag{31}$$

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 first-order condition imposes the following informative prior on $\mu_0 + \mu_1 DY_T$ or

$$p_0(\mu_0|\mu_1) \propto N\left[\gamma\Sigma_{RR}w_0 - \mu_1 DY_T, \sigma_\rho^2\left(\frac{1}{s_{RR}^2}\Sigma_{RR}\right)\right], \tag{32}$$

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 first-order condition. Hence, we define the objective-based prior as the one by adding this additional conditional density into the righthand side of Equation (31). In contrast with the predictability-based prior, the objective-based one reflects not only predictability, but also the economic objective. The marginal prior density of $\Gamma = [\mu_0, \mu_1]'$ can be written succinctly as

$$p(\Gamma|\Sigma_{RR}) \propto |\Sigma_{RR}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}tr[\Sigma_{RR}^{-1}(\Gamma - \Gamma_0(\mu_1^p))'\Upsilon(\Gamma - \Gamma_0(\mu_1^p))]\right\}, \tag{33}$$

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

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

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}:

$$vec(\Gamma)|\Sigma_{RR}, \mathcal{D}_T \sim N[vec(\widetilde{\Gamma}), \ \Sigma_{RR} \otimes (X'X + \Upsilon)^{-1}], \tag{34}$$

$$\Sigma_{RR}|\mathcal{D}_T \sim IW[S_R, T-1], \tag{35}$$

where

$$\widetilde{\Gamma} = (X'X + \Upsilon)^{-1}(X'R + \Upsilon\Gamma_0(\mu_1^p)], \quad S_R = R'R - \widetilde{\Gamma}'X'X\widetilde{\Gamma}, \tag{36}$$

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. Comparison Based on Bayesian Criterion

In this section, we compare first the objective-based priors with their usual alternatives based on the Bayesian criterion of Equation (24). Then, based on the same criterion, we examine the performance under the various priors when the asset returns are assumed predictable.

The data are monthly returns of the well-known Fama-French 25 book-to-market and size portfolios and their three factors from January 1965 to December 2004 plus ten years of earlier data for forming data-based priors.[9]

A. Utility Gains under Various Priors

Table 1 reports the utility gains of switching from the diffuse prior to the objective-based one in the case in which no prior data is used in forming the prior. When we apply the priors to five years worth of data ($T = 60$), the utility gains are overwhelmingly large. They range from an annual rate of 22% to 125%. The large difference is driven by the fact that the predictive moments are very sensitive to prior specifications when $T = 60$. This can also be understood by the simulation results of Kan and Zhou (2007), who show that, with a sample size of $T = 60$, the estimated parameters can be far away from the true ones. There are two additional interesting facts in the table. First, the greater the σ_ρ, the smaller the gains. This is because a higher value of σ_ρ moves the objective prior closer to the diffuse one. Second, the prior exposure to risky assets has little effect on the gains. As the exposure varies from 50% to 100%, the gains change only from 123.93% to 125.47% when $\sigma_\rho = 1\%$, and from 22.36% to 22.66% when $\sigma_\rho = 5\%$. This simply says that the data are informative enough about the risk exposures despite the seemingly large prior differences in them. In other words, the scaling of the total investment is easily updated or corrected by the data, while the relative positions on the assets require much data to estimate correctly their covariance matrix.

As 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 of as

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

large as $T = 480$, Table 1 shows that the utility gains are still substantial. At $\sigma_\rho = 1\%$, the gain is greater than 8%, although it eventually decreases to an insignificant amount of 0.04% at $\sigma_\rho = 5\%$. Overall, it is clear that the objective-based prior makes a significant economic difference in portfolio selections.

Consider now the case in which some of the data, those ten years prior to the estimation window, are used to form informative priors. In this case, the data-based prior, equation (20), plays the role of the earlier diffuse prior, while the corresponding objective-based prior is the one, given by equation (21), that updates the previous (no data) prior, given by equation (19), by the same length of data. For simplicity, we set $\sigma_\mu = \sigma_\rho$ in the comparison. Table 2 provides the results. The objective-based prior outperforms the data-based prior substantially when $T \leq 180$ or $\sigma_\mu \leq 2\%$. Like the diffuse prior case, the gains 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 gains when $T = 480$ are even greater than those with fewer samples. There are two explanations for this. First, in a given application, the entire sample is only one path of all possible realizations of the random asset returns and factors. 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 gains here, may not necessarily be so.

Finally, consider the performance of the objective-based priors in comparison to those based on asset pricing models. With x_t being the Fama-French three-factors, the degree of belief on the validity of the Fama-French three-factor model is represented by the alpha prior, equation (10). For simplicity, we assume $\sigma_\alpha = \sigma_\rho$ in the comparison. Table 3 provides the results. Similar to the data-based prior case, the utility gains are economically significant for all the sample sizes when $\sigma_\rho \leq 2\%$. However, they are small when $T \geq 360$ and $\sigma_\rho \geq 3\%$.

In summary, maximizing a utility function provides important guidance for choosing priors in Bayesian decision making. Using the Fama-French data, we find that such objective-based priors perform differently from both standard statistical and asset-pricing-based priors significantly. Even with sample size as large as $T = 480$, there are cases where the differences

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

in terms of utility gains are still economically significant.

B. Utility Gains under Predictability

Consider now what happens to the performance under the various priors when the returns are assumed predictable. For interest of comparison, we allow σ_P, the degree of uncertainty about predictability, to take two 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 utility gains are substantial and more pronounced than in Table 1. For example, with $\sigma_\rho = 1\%$, the gains are 207.22% and 74.10% compared with 123.93% and 8.87% of the iid case, when $T = 60$ and 480, respectively. Like the iid case, the gains decrease as either σ_ρ or T increases. Overall, the presence of predictability does not weaken the earlier gains, but strengthens them.

V. Out-of-sample Performance

The Bayesian evaluation of the rules presented thus far is conditional on the data at hand, and the Bayesian utility gains measure the economic significance of the differences between investment decisions based on one prior versus another. They do not speak to the performances of the decisions out-of-sample. On the other hand, a classical statistician may prefer to see how the rules perform for all possible data sets. In this section, based on the out-of-sample criterion (detailed below), we compare the Bayesian rules among themselves, and also compare them with some of the classical rules studies by Kan and Zhou (2007).

For the general case in which μ and V are any arbitrarily given parameters, the expected utilities associated with most of the portfolio rules are difficult to obtain analytically. However, they can be easily computed via simulation for any prespecified parameters. To be realistic, we set the true parameters of the model as the sample mean and covariance matrix of the observed monthly data (as used in Section IV). Then, we can simulate a large number of data sets from the assumed normal distribution of asset returns. For any one draw of the data set with a sample size T, each of the rules provides its estimated optimal portfolio weights. Based on the weights, the expected utility can be computed from (1). The average of these expected utility values over all the draws is the out-of-sample performance of the rule (see Kan and Zhou (2007) and references therein for more theoretical discussions). In other words, if one plays the rule over and over again for a large number of times, 1,000 used here, the average is the average utility the rule ensures. In the present mean-variance framework, this is the average risk-adjusted return.

Table 5 reports the differences of the average expected utilities from the objective-based prior to the diffuse prior rules. They can be interpreted as the out-of-sample average utility gains of switching from the diffuse prior to the objective-based one. With sample size varying from 60 to 480, it is seen that the objective-based prior outperform consistently. When $T = 60$, the gains are much greater than the other cases due to the poor performance of the diffuse prior with a small sample size. 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.

When ten years worth of data are used to form the priors, Table 6 provides the the out-of-sample average utility gains of switching from the data-based prior to the objective-based one. Qualitatively, we have a similar conclusion as to Table 5. When $T \leq 180$, the gains range 2.04% to 98.58%. These are clearly economically significant, but smaller than the diffuse prior case. This simply states that the data-based prior provides useful information to portfolio selection, and so it does better than the diffuse prior by having 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 helps the data-based prior to perform as well as the objective-based one.

When the objective-based prior is compared with the asset-pricing prior derived from the Fama-French three-factor model, Table 7 provides the results. This prior, like others, underperforms the objective-based one substantially. In addition, in comparison with others as reported earlier in Table 5 and 6, the asset pricing prior does better than both diffuse and the data-based priors when $\sigma_\rho = 1\%$. Hence, it seems that the three-factor model is not a bad approximation for the true model. Overall, we find that the objective-based prior has superior performances, and is a better decision rule than all other priors as judged by the classical statistical criterion.

Finally, we compare the Bayesian objective-based prior rule with the classical rules studied by Kan and Zhou (2007). For brevity, we analyze three of the classical rules here. The first is the maximum likelihood (ML) estimator of the optimal portfolio weights, a popular rule in practice. The other two are the shrinkage rule of Jorion (1986) and the three-fund rule of Kan and Zhou (2007), which are the better performing rules among those compared in Kan and Zhou (2007). Table 8 reports the average 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 the others only when the sample size is as large as 480. The shrinkage and the three-fund rules are designed to improve upon the ML, and are optimal in certain metrics, and hence it is 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 Bayesian objective-based prior rule when $T \leq 120$. But, when $T \geq 240$, they have comparable performances with the Bayesian rule. It may be noted that, similar to earlier Bayesian comparison, the sum of the prior weights has little impact on performances because the data are quite informative on the total risky position, though much less so on their relative positions.

In summary, the proposed objective-based prior rule performs impressively against both the other Bayesian rules and the classical rules. The results highlight the importance for investors to use their priors on the solution to an optimizing problem. In our case here, the objective-based prior essentially says that our starting point is to diversify our investments so that the prior weights are equal across assets. This implies suitable prior constraints on the parameter values, and then we let the data update our prior toward the true and unknown optimal portfolio. Because the prior contains useful information on the whereabouts of the true solution, it turns out to be very valuable. While beyond the scope of this paper, this seems to have important implications to Bayesian decision making in a number of areas.

VI. Assessing the Importance of Anomalies

In this section, we apply our Bayesian framework to study the importance of Fama and French's (1993) book-to-market portfolio when treated as an anomaly to the 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 gains 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 zero 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,

$$\mu \sim N\left[\gamma V w_a, \sigma_\rho^2\left(\frac{1}{s^2}V\right)\right], \tag{37}$$

then represents the prior centered upon the belief $w_a = (w_1', 0)'$. If the investor is dogmatic about his belief, he will then 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 zero. 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 zero. Let EU_b be the expected utility with this updated portfolio. Then the difference between EU_b and EU_a provides a measure for the utility gains 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 else 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 gains 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 is using a generally better prior than the latter. However, as shown by later applications, the difference between EU_b and EU_c are in fact small. Hence, either $EU_b - EU_a$ or $EU_c - EU_a$ will provide a fairly robust measure for the gains of utilizing the anomaly.

Fama and French's (1993) book-to-market 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 we, following Pástor (2000), examine the economic importance of the HML portfolio based on the approach outlined in Section II.E. In this case, we have $N = 2$ since the market index and HML are the only risky assets.

Table 9 reports the utility gains, $EU_b - EU_a$, 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 $\sigma_\rho \geq 2\%$, the gains are over 3.72% across sample sizes and risk exposures. The reason that the gains are getting greater as σ_ρ increases is that the prior avoids investing in the HML, and this skeptical prior can be mitigated by a larger value of σ_ρ. As in the previous section, the risk exposure, either $\sum w_{0i} = 0.5$ or 1, has little to do with the gains. Overall, the results suggest strongly that the HML portfolio is of great economic significance that yields substantial utility gains in the asset allocation problem.

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 gains are small. Table 11 makes it more apparent. The utility differences between the skeptical prior and the balanced one are less than 1% except in three scenarios, and are less than 0.46% whenever $\sigma_\rho \geq 3\%$. The results say that even when one starts from such a poor prior that one avoids investing in the HML asset entirely, but is willing to let the data to update this prior, then the harm is less than one would have expected. In summary, the priors about the degree of investments in the anomaly asset have little impact, and what drives the utility gains most is the dogmatic belief that the investors will not update their prior investments in the anomaly asset at all after seeing the data.

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 utility gains are 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 an in-sample Bayesian criterion, but also by an out-of-sample classical criterion. In addition, while the shrinkage rule of Jorion (1986) and the three-fund rule of Kan and Zhou (2007) are excellent rules in the classical framework, we find that the Bayesian objective-prior can outperform them substantially, suggesting there is real value-added in using prior solution information for the economic problem at hand. We also apply the methodology to examine asset pricing anomalies and find that Fama and French's (1993) book-to-market portfolio, HML (high minus low), can add substantial value to an investor's allocation 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 as 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 maximize the objective at hand.

Appendix A: Proofs

Proof of Equation (16)

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

$$\begin{aligned}
&U(w_d|\mu_0) - U(w_0|\mu_0) \\
&=\frac{\partial U}{\partial w'}[w_0|\mu_0][w_d - w_0] + \frac{1}{2}[w_d - w_0]'\frac{\partial^2 U}{\partial w \partial w'}[w_0|\mu_0][w_d - w_0] \\
&+\frac{1}{6}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial^3 U}{\partial w_i \partial w_j \partial w_k}[w_0|\mu_0][w_{di} - w_{0i}][w_{dj} - w_{0j}][w_{dk} - w_{0k}] + \cdots .
\end{aligned} \tag{A1}$$

Ignoring the higher order terms and using the first order condition $\frac{\partial U}{\partial w'}[w_0|\mu_0] = 0$, we have

$$U(w_d|\mu_0) - U(w_0|\mu_0) \approx \frac{1}{2}[w_d - w_0]'\frac{\partial^2 U}{\partial w \partial w'}[w_0|\mu_0][w_d - w_0]. \tag{A2}$$

Standard calculus implies

$$[w_d - w_0] \approx \left\{\frac{\partial^2 U}{\partial w \partial w'}[w_0|\mu_0]\right\}^{-1}\left\{\frac{\partial U}{\partial w}[w_d|\mu_0] - \frac{\partial U}{\partial w}[w_0|\mu_0]\right\}, \tag{A3}$$

and

$$\left\{\frac{\partial U}{\partial w}[w_d|\mu_0] - \frac{\partial U}{\partial w}[w_0|\mu_0]\right\} \approx \left\{\frac{\partial^2 U}{\partial w \partial \mu'}[w_0|\mu_0]\right\}[\mu_d - \mu_0]. \tag{A4}$$

Therefore, we have (16), which says that the utility loss is approximately equal to the weighted average of the deviation of μ_d from μ_0, with 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

$$\left\{\frac{\partial^2 U}{\partial w \partial \mu'}[w_0|\mu_0]\right\} = I_N, \tag{A5}$$

$$\left\{\frac{\partial^2 U}{\partial w \partial w'}[w_0|\mu_0]\right\} = -\gamma V, \tag{A6}$$

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

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.

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

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

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

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

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

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

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

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.* McGraw-Hill, New York (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.

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

Litterman, B. *Modern Investment Management: An Equilibrium Approach.* Wiley, New York (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.* Springer-Verlag, New York (2005).

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." CRSP Working Paper No. 617 (2006).

Schwert, B. "Anomalies and Market Efficiency." In *Handbook of the Economics of Finance,* G.M. Constantinides et al., eds. North-Holland, Amsterdam (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., and G. Zhou. "Data-generating Process Uncertainty: What Difference Does it Make in Portfolio Decisions?" *Journal of Financial Economics*, 72 (2004), 385–421.

Wachter, J., and M. Warusawitharana. "Predictable Returns and Asset Allocation: Should a Skeptical Investor Time the Market?" Working Paper, University of Pennsylvania (2007).

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. "An Extension to the Black-Litterman Model: Letting the Data Speak." Working Paper, Washington University in St. Louis (2008).

TABLE 1
Utility Gains of Switching from Diffuse to Objective-based Priors

The table reports the utility gains (annualized) of switching from the standard diffuse prior,

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

to the objective-based prior

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

where σ_ρ^2 reflects the degree of uncertainty about μ and w_0 is proportional to a constant with $\sum w_{0i} = 0.5$ or 1, respectively. The data are Fama-French 25 size and book-to-market portfolios and their three factors from January 1965 to December 2004. T is the sample size from January 1965 to a future month.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	123.93	90.23	58.47	36.17	22.36
1	60	125.47	91.36	59.18	36.68	22.66
0.5	120	76.97	31.80	12.54	5.42	2.61
1	120	75.57	31.20	12.33	5.33	2.57
0.5	180	53.45	14.62	4.54	1.74	0.78
1	180	52.54	14.39	4.45	1.71	0.77
0.5	240	38.72	8.45	2.36	0.85	0.38
1	240	38.00	8.27	2.33	0.84	0.37
0.5	360	15.95	2.66	0.67	0.24	0.10
1	360	15.28	2.54	0.65	0.22	0.10
0.5	480	8.87	1.26	0.31	0.10	0.04
1	480	8.70	1.24	0.30	0.10	0.04

TABLE 2
Utility Gains of Switching from Data-based to Objective-based Priors

The table reports the utility gains (annualized) of switching from the data-based prior

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

to the objective-based prior

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

where $\hat{\mu}_{10}$ is the sample mean of the ten year prior data, and $\hat{\mu}_{10}^* = \gamma V\hat{w_{10}}$, $\hat{w_{10}}$ is the Bayesian optimal portfolio weights based on ten years prior data; w_0 is proportional to a constant with $\sum w_{0i} = 0.5$ or 1, respectively; σ_ρ^2 reflects the degree of uncertainty about μ, $H = T_{10}\hat{V}_{10}$, $\nu_V = T_{10}$, $T_{10} = 120$, and $\hat{V}_{10}$ is the sample covariance of the prior data. The data are Fama-French 25 size and book-to-market portfolios and their three factors from January 1965 to December 2004. T is the sample size from January 1965 to a future month.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	54.25	30.36	18.32	12.28	8.80
1	60	53.17	29.57	17.89	12.15	8.66
0.5	120	44.57	44.56	31.44	19.37	12.16
1	120	42.72	43.70	31.16	19.41	12.19
0.5	180	34.25	14.34	7.50	4.14	2.39
1	180	33.88	14.20	7.49	4.12	2.37
0.5	240	17.76	4.28	1.62	0.73	0.35
1	240	17.02	4.25	1.60	0.71	0.36
0.5	360	6.54	1.97	0.77	0.34	0.17
1	360	6.37	1.92	0.75	0.34	0.17
0.5	480	42.83	8.68	2.37	0.95	0.47
1	480	42.85	8.72	2.37	0.95	0.46

TABLE 3
Utility Gains of Switching from Fama-French Three-factor Model-based to Objective-based Priors

The table reports the utility gains (annualized) of switching from a prior reflecting the degree of belief in the Fama-French three-factor model,

$$p_0(\alpha, B, V) \propto N(B\mu_2, \sigma_\alpha^2 \frac{1}{s_\Sigma^2}\Sigma) \times |V|^{-\frac{N+1}{2}},$$

where $\Sigma = V_{11} - V_{12}V_{22}^{-1}V_{21}$ and s_Σ^2 is the average of the diagonal elements of Σ, to the objective-based prior

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

where σ_ρ^2 reflects the degree of uncertainty about μ and w_0 is proportional to a constant with $\sum w_{0i} = 0.5$ or 1, respectively. The data are Fama-French 25 size and book-to-market portfolios and their three factors from January 1965 to December 2004. T is the sample size from January 1965 to a future month.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	84.32	120.18	125.20	114.53	102.71
1	60	83.19	118.94	123.68	113.31	101.47
0.5	120	42.53	39.94	26.21	18.26	14.11
1	120	40.26	38.64	25.49	17.92	13.82
0.5	180	33.86	18.97	10.05	6.53	4.93
1	180	32.65	18.51	9.83	6.41	4.85
0.5	240	28.62	11.63	5.54	3.45	2.57
1	240	27.68	11.30	5.43	3.39	2.52
0.5	360	14.76	4.04	1.72	1.04	0.76
1	360	13.92	3.83	1.65	0.99	0.74
0.5	480	8.42	1.87	0.74	0.43	0.30
1	480	8.16	1.81	0.72	0.41	0.30

TABLE 4
Utility Gains of Switching from Predictability-diffuse to Objective-based Priors

The table reports the utility gains (annualized) of switching from the predictability-diffuse prior,

(A7) $$p_0(\mu_1) \propto N\left[\hat{\mu_1}^p, \sigma_P^2\left(\frac{1}{s_{RR}^2}\Sigma_{RR}\right)\right],$$

to the objective-based prior

(A8) $$p_0(\mu_0,\mu_1) \propto p_0(\mu_1) \times N\left[\gamma\Sigma_{RR}w_0 - \mu_1 DY_T, \sigma_\rho^2\left(\frac{1}{s_{RR}^2}\Sigma_{RR}\right)\right],$$

where $\hat{\mu_1}^p$ is the slope of the predictive regression $r_t = \mu_0 + \mu_1 DY_{t-1} + v_t$, $v_t \sim N(0,\Sigma_{RR})$, based on previous ten years data, σ_P^2 measures the degree of uncertainty about predictability, DY_T is the dividend yield at T, σ_ρ^2 reflects the degree of uncertainty in the objective-based prior and w_0 is proportional to a constant with $\sum w_{0i} = 0.5$ or 1, respectively. The data are Fama-French 25 size and book-to-market portfolios and their three factors from January 1965 to December 2004. T is the sample size from January 1965 to a future month.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
$\sigma_P = \infty$						
0.5	60	207.22	155.36	109.65	73.29	48.53
1	60	198.32	150.32	104.96	71.92	46.82
0.5	120	168.07	82.86	37.64	17.38	9.22
1	120	167.51	81.79	37.25	17.73	9.40
0.5	180	157.67	65.02	26.73	11.86	5.70
1	180	154.80	63.50	25.73	11.46	5.69
0.5	240	138.21	52.53	20.22	8.70	4.05
1	240	140.35	53.62	20.66	9.17	4.57
0.5	360	96.76	31.76	11.12	4.65	2.26
1	360	95.36	31.05	10.75	4.50	2.15
0.5	480	74.10	22.34	7.49	3.02	1.36
1	480	74.72	22.66	7.48	3.04	1.38

(To be continued)

TABLE 4 (continued)

Utility gains of switching from predictability-diffuse to objective-based Priors

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
$\sigma_P = 50\%$						
0.5	60	357.36	262.51	179.97	114.22	74.64
1	60	345.84	256.91	174.46	114.55	73.40
0.5	120	158.64	77.46	35.00	16.15	8.52
1	120	157.79	76.50	34.60	16.34	8.65
0.5	180	124.64	46.19	17.74	7.54	3.61
1	180	122.65	45.33	17.25	7.36	3.55
0.5	240	97.55	32.42	11.55	4.79	2.14
1	240	99.50	33.45	11.82	5.08	2.48
0.5	360	102.32	32.30	11.01	4.58	2.24
1	360	100.61	31.55	10.74	4.42	2.14
0.5	480	57.75	16.00	5.15	2.02	0.90
1	480	59.16	16.59	5.23	2.10	0.93

TABLE 5
Out-of-sample Utility Gains of Switching from Diffuse to Objective-based Priors

This table reports the average 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 25 assets and the associated three factors from January 1965 to December 2004. For each of the varying sample sizes T below, there are 1000 simulated data sets.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	185.14	185.19	168.42	143.16	118.04
1	60	186.06	185.77	168.78	143.39	118.11
0.5	120	42.50	44.95	35.16	25.68	18.87
1	120	43.21	45.25	35.28	25.77	18.89
0.5	180	18.99	21.88	15.67	10.80	7.66
1	180	19.55	22.07	15.79	10.84	7.65
0.5	240	10.08	13.03	8.91	5.93	4.11
1	240	10.54	13.16	8.92	5.97	4.13
0.5	360	3.64	6.18	3.97	2.56	1.76
1	360	3.97	6.25	3.99	2.57	1.75
0.5	480	1.33	3.52	2.19	1.39	0.95
1	480	1.56	3.55	2.20	1.39	0.96

TABLE 6
Out-of-sample Utility Gains of Switching from Data-based to Objective-based Priors

This table reports the average 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 25 assets and the associated three factors from January 1965 to December 2004. For each of the varying sample sizes T below, there are 1000 simulated data sets.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	71.52	98.15	87.66	68.40	52.92
1	60	72.32	98.58	87.56	68.28	52.60
0.5	120	21.38	20.07	13.53	9.02	6.16
1	120	21.97	20.24	13.67	8.96	6.44
0.5	180	16.38	9.49	5.15	2.94	2.00
1	180	16.61	9.41	5.10	3.04	2.04
0.5	240	12.77	5.16	2.38	1.42	0.89
1	240	12.97	5.21	2.38	1.34	0.85
0.5	360	8.04	1.81	0.67	0.40	0.14
1	360	8.11	1.80	0.69	0.41	0.23
0.5	480	4.70	0.70	0.26	0.16	0.14
1	480	4.73	0.67	0.28	0.16	0.12

TABLE 7
Out-of-sample Utility Gains of Switching from Fama-French Three-factor Model-based to Objective-based Priors

This table reports the average utility gains of switching from the Fama-French three-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 three factors from January 1965 to December 2004. For each of the varying sample sizes T below, there are 1000 simulated data sets.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	53.47	187.62	237.01	242.54	233.47
1	60	54.39	188.21	237.37	242.78	233.54
0.5	120	21.33	55.66	55.95	50.49	45.63
1	120	22.04	55.96	56.07	50.58	45.66
0.5	180	9.67	25.95	23.39	19.92	17.45
1	180	10.23	26.14	23.51	19.96	17.44
0.5	240	5.22	15.58	13.33	11.06	9.58
1	240	5.68	15.71	13.34	11.10	9.60
0.5	360	1.15	6.87	5.41	4.27	3.61
1	360	1.48	6.94	5.43	4.29	3.60
0.5	480	0.26	4.24	3.32	2.67	2.30
1	480	0.49	4.28	3.33	2.67	2.30

TABLE 8
Out-of-sample Utilities of Classical Rules and a Bayesian One

This table reports the average utilities of the Bayesian rule under the objective-based prior, the shrinkage rule of Jorion (1986), the three-fund rule of Kan and Zhou (2007), and the maximum likelihood 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 25 assets and the associated three factors from January 1965 to December 2004. For each of the varying sample sizes T below, there are 1000 simulated data sets.

$\sum w_{0i}$	T	Bayesian σ_ρ		Classical rules		
		1%	2%	Jorion	Kan-Zhou	$\frac{1}{\gamma}\hat{V}^{-1}\hat{\mu}$
0.5	60	8.58	8.63	-57.67	1.78	-932.13
1	60	9.50	9.21	-57.67	1.78	-932.13
0.5	120	16.75	19.20	7.17	16.03	-92.29
1	120	17.46	19.50	7.17	16.03	-92.29
0.5	180	22.45	25.34	20.36	23.58	-19.76
1	180	23.02	25.53	20.36	23.58	-19.76
0.5	240	26.74	29.70	27.02	28.58	4.99
1	240	27.20	29.82	27.02	28.58	4.99
0.5	360	32.34	34.88	33.79	34.36	24.06
1	360	32.67	34.95	33.79	34.36	24.06
0.5	480	36.00	38.18	37.63	37.89	32.22
1	480	36.22	38.22	37.63	37.89	32.22

TABLE 9
Utility Gains of Utilizing Anomaly under a Skeptical Prior

Based on the market (MKT) and the high minus low book-market (HML) portfolios from January 1965 to December 2004, the table reports the utility gains (annualized) of switching from investing only in the MKT to investing in both MKT and HML under the skeptical prior,

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

where σ_ρ^2 reflects the degree of uncertainty about μ and w_0 is proportional to the 2×1 vector $(1\ 0)'$ with $\sum w_{0i} = 0.5$ or 1, respectively. T is the sample size from January 1965 to a future month.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	0.55	3.79	7.73	10.77	12.73
1	60	0.54	3.72	7.72	10.72	12.76
0.5	120	1.35	5.21	7.70	9.04	9.84
1	120	1.32	5.18	7.73	9.06	9.88
0.5	180	1.98	5.26	6.70	7.45	7.70
1	180	1.97	5.26	6.73	7.35	7.72
0.5	240	2.70	6.00	7.27	7.80	8.11
1	240	2.70	6.03	7.27	7.80	8.07
0.5	360	4.82	8.65	9.84	10.25	10.46
1	360	4.78	8.60	9.73	10.22	10.47
0.5	480	4.59	7.28	8.00	8.22	8.36
1	480	4.55	7.20	8.00	8.27	8.42

TABLE 10
Utility Gains of Utilizing Anomaly under a More Balanced Prior

Based on the market (MKT) and the high minus low book-market (HML) portfolios from January 1965 to December 2004, the table reports the utility gains (annualized) of switching from investing only in the MKT but not investing in the HML anomaly asset under the skeptical prior

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

where σ_ρ^2 reflects the degree of uncertainty about μ and w_a is proportional to the 2×1 vector $(1\ 0)'$ with $\sum w_{0i} = 0.5$ or 1, respectively; to investing in both MKT and HML under a more balanced prior

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

where w_0 is proportional to a constant with $\sum w_{0i} = 0.5$ or 1, respectively. T is the sample size from January 1965 to a future month.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	1.66	4.94	8.77	11.41	13.39
1	60	3.93	6.76	10.07	12.37	13.80
0.5	120	2.20	5.78	8.12	9.46	10.04
1	120	3.65	6.49	8.47	9.55	10.13
0.5	180	2.59	5.61	6.93	7.54	7.86
1	180	3.50	5.98	7.13	7.64	7.87
0.5	240	3.24	6.34	7.43	7.91	8.15
1	240	4.01	6.60	7.58	8.02	8.23
0.5	360	5.35	8.83	9.89	10.30	10.54
1	360	5.88	8.99	10.03	10.33	10.57
0.5	480	4.98	7.37	8.04	8.27	8.41
1	480	5.40	7.51	8.15	8.33	8.45

TABLE 11
Utility Gains of Switching from a Skeptical Prior to a More Balanced Prior

Based on the market (MKT) and the high minus low book-market (HML) portfolios from January 1965 to December 2004, the table reports, while allowing to invest in both MKT and HML, the utility gains (annualized) of switching from a skeptical prior

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

where σ_ρ^2 reflects the degree of uncertainty about μ and w_a is proportional to the 2×1 vector $(1\ 0)'$ with $\sum w_{0i} = 0.5$ or 1, respectively; to a more balanced prior

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

where w_0 is proportional to a constant with $\sum w_{0i} = 0.5$ or 1, respectively. T is the sample size from January 1965 to a future month.

$\sum w_{0i}$	T	σ_ρ				
		1%	2%	3%	4%	5%
0.5	60	0.58	0.24	0.12	0.03	0.03
1	60	2.32	1.07	0.46	0.21	0.09
0.5	120	0.30	0.08	0.02	0.01	0.01
1	120	1.22	0.27	0.09	0.03	0.01
0.5	180	0.15	0.02	0.01	0.00	0.00
1	180	0.58	0.10	0.02	0.01	0.00
0.5	240	0.10	0.02	0.00	0.00	0.00
1	240	0.41	0.06	0.01	0.00	0.00
0.5	360	0.06	0.01	0.00	0.00	0.00
1	360	0.22	0.03	0.01	0.00	0.00
0.5	480	0.03	0.00	0.00	0.00	0.00
1	480	0.14	0.01	0.00	0.00	0.00