Bayesian Portfolio Analysis

Doron Avramov*

Finance Department, The Hebrew University of Jerusalem, Mt. Scopus Jerusalem 91905, Israel; email: doron.avramov@huji.ac.il; R.H. Smith School of Business, University of Maryland, College Park, Maryland 20742; email: davramov@rhsmith.umd.edu

Guofu Zhou

Olin Business School, Washington University, St. Louis, Missouri 63130; email: zhou@wustl.edu

Key Words

portfolio choice, parameter uncertainty, informative prior beliefs, return predictability, model uncertainty, learning

Abstract

This paper reviews the literature on Bayesian portfolio analysis. Information about events, macro conditions, asset pricing theories, and security-driving forces can serve as useful priors in selecting optimal portfolios. Moreover, parameter uncertainty and model uncertainty are practical problems encountered by all investors. The Bayesian framework neatly accounts for these uncertainties, whereas standard statistical models often ignore them. We review Bayesian portfolio studies when asset returns are assumed both independently and identically distributed as well as predictable through time. We cover a range of applications, from investing in single assets and equity portfolios to mutual and hedge funds. We also outline challenges for future work.

1. INTRODUCTION

Portfolio selection is one of the most important problems in practical investment management. The first papers in the field go back at least to the mean-variance paradigm of Markowitz (1952), which analytically formalizes the risk-return trade-off in selecting optimal portfolios. Even when the mean variance is a static one-period model, it has widely been accepted by both academics and practitioners. The latter-developed intertemporal capital asset pricing model (ICAPM) of Merton (1973) accounts for the dynamic multiperiod nature of investment-consumption decisions. In an intertemporal economy, the overall demand for risky assets consists of both the mean-variance component as well as a component hedging against unanticipated shocks to time-varying investment opportunities. Empirically, for a wide variety of preferences, hedging demands for risky assets are typically small, even nonexistent (see also Ait-Sahalia & Brandt 2001, Brandt 2009).

We review Bayesian studies of portfolio analysis. The Bayesian approach is potentially attractive. First, it can employ useful prior information about quantities of interest. Second, it accounts for estimation risk and model uncertainty. Third, it facilitates the use of fast, intuitive, and easily implementable numerical algorithms in which to simulate otherwise complex economic quantities. In addition, three building blocks underly Bayesian portfolio analysis: First is the formation of prior beliefs, which are typically represented by a probability density function on the stochastic parameters underlying the stock-return evolution. The prior density can reflect information about events, macroeconomy news, asset pricing theories, as well as any other insights relevant to the dynamics of asset returns. Second is the formulation of the law of motion governing the evolution of asset returns, asset pricing factors, and forecasting variables. Third is the recovery of the predictive distribution of future asset returns, analytically or numerically, incorporating prior information, law of motion, as well as estimation risk and model uncertainty. The predictive distribution, which integrates out the parameter space, characterizes the entire uncertainty about future asset returns. The Bayesian optimal portfolio rule is obtained by maximizing the expected utility with respect to the predictive distribution.

Zellner & Chetty (1965) pioneer the use of predictive distribution in decision making in general. Appearing during the 1970s, the first applications in finance are entirely based on uninformative or data-based priors. Bawa et al. (1979) provide an excellent survey on such applications. Jorion (1986) introduces the hyperparameter prior approach in the spirit of the Bayes-Stein shrinkage prior, whereas Black & Litterman (1992) advocate an informal Bayesian analysis with economic views and equilibrium relations. Recent studies by Pástor (2000) and Pástor & Stambaugh (2000) center prior beliefs around values implied by asset pricing theories. Tu & Zhou (2010) argue that the investment objective provides a useful prior for portfolio selection.

Whereas all the above-noted studies assume that asset returns are identically and independently distributed through time, Kandel & Stamabugh (1996), Barberis (2000), and Avramov (2002) account for the possibility that returns are predictable by macro variables such as the aggregate dividend yield, the default spread, and the term spread. Incorporating predictability provides fresh insights into asset pricing in general and Bayesian portfolio selection in particular.

We review Bayesian portfolio studies when asset returns are assumed to (*a*) be independently and identically distributed (IID), (*b*) be predictable through time by macro conditions, and (*c*) exhibit regime shifts and stochastic volatility. We cover a range of applications, from

investing in the market portfolio, equity portfolios, and single stocks to investing in mutual funds and hedge funds. We also outline existing challenges for future work.

The paper is organized as follows: Section 2 reviews Bayesian portfolio analysis when asset returns are independently and identically distributed through time. Section 3 surveys studies that account for potential predictability in asset returns. Section 4 discusses alternative return-generating processes. Section 5 outlines ideas for future research, and Section 6 concludes.

2. ASSET ALLOCATION WHEN RETURNS ARE INDEPENDENTLY AND IDENTICALLY DISTRIBUTED

Consider $N + 1$ investable assets, one of which is riskless and the other is risky. Risky assets may include stocks, bonds, currencies, mutual funds, and hedge funds. Denote by r_{ft} and r_t the returns on the riskless and risky assets, respectively, at time t. Then, $R_t \equiv r_t - r_{ft} 1_N$ is an N-dimensional vector of time t excess returns on risky assets, where 1_N is an N-vector of ones. The joint distribution of R_t is assumed IID through time with mean μ and covariance matrix V.

For analytical insights, it is useful to review the mean-variance framework pioneered by Markowitz (1952). In particular, consider an optimizing investor who chooses at time T portfolio weights w so as to maximize the quadratic objective function

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

where E and Var denote the mean and variance of the uncertain portfolio rate of return $R_p = w'R_{T+1}$ to be realized at time $T + 1$, and γ is the relative risk-aversion coefficient. When both μ and V are known, the optimal portfolio weights are given by

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

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.

In practice, it is impossible to compute w^* because both μ and V are essentially unknown. One approach is to apply the mean-variance theory in two steps. In the first step, the mean and covariance matrix of asset returns are estimated on the basis of the observed data. Specifically, given a sample of T observations on asset returns, the standard maximum likelihood estimators are

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} R_t \tag{4}$$

and

$$\hat{V} = \frac{1}{T} \sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})'. \tag{5}$$

Then, in the second step, these sample estimates are treated as if they were the true parameters and are simply plugged into Equation 2 to compute the estimated optimal portfolio weights

$$\hat{w}^{\mathrm{ML}} = \frac{1}{\gamma} \hat{V}^{-1} \hat{\mu}. \tag{6}$$

The two-step procedure gives rise to a parameter-uncertainty problem because it is the estimated parameters, not the true ones, that are used to compute optimal portfolio weights. Consequently, the utility associated with the plug-in portfolio weights can be substantially different from the true utility, $U(w^*)$. In particular, denote by θ the vector of the unknown parameters μ and V. Mathematically, the two-step procedure maximizes the expected utility conditional on the estimated parameters, denoted by $\hat{\theta}$, being equal to the true ones

$$\max_{w}[U(w) \, | \, \theta = \hat{\theta}]. \tag{7}$$

Thus, estimation risk is altogether ignored.

2.1. The Bayesian Framework

The Bayesian approach treats θ as a random quantity. One can infer only its probability distribution function. Following Zellner & Chetty (1965), the Bayesian optimal portfolio is obtained by maximizing the expected utility under the predictive distribution. In particular, the utility maximization is formulated as

$$\begin{aligned}
\hat{w}^{\mathrm{Bayes}} &= \mathrm{argmax}_w \int_{R_{T+1}} \tilde{U}(w) p(R_{T+1} \, | \, \boldsymbol{\Phi}_T) \mathrm{d}R_{T+1} \\
&= \mathrm{argmax}_w \int_{R_{T+1}} \int_{\mu} \int_{V} \tilde{U}(w) p(R_{T+1}, \mu, V \, | \, \boldsymbol{\Phi}_T) \mathrm{d}\mu \mathrm{d}V \mathrm{d}R_{T+1},
\end{aligned} \tag{8}$$

where $\tilde{U}(w)$ is the utility of holding a portfolio w at time $T+1$ and $\boldsymbol{\Phi}_T$ is the data available at time T. Moreover, $p(R_{T+1} \, | \, \boldsymbol{\Phi}_T)$ is the predictive density of the time $T+1$ return, which integrates out μ and V from

$$p(R_{T+1}, \mu, V \, | \, \boldsymbol{\Phi}_T) = p(R_{T+1} \, | \, \mu, V, \boldsymbol{\Phi}_T) \, p(\mu, V \, | \, \boldsymbol{\Phi}_T), \tag{9}$$

where $p(\mu, V \, | \, \boldsymbol{\Phi}_T)$ is the posterior density of μ and V. To compare the classical and Bayesian formulations in Equations 7 and 8, notice that expected utility is maximized under the conditional and predictive distributions, respectively. Unlike the conditional distribution, the Bayesian predictive distribution accounts for estimation errors by integrating over the unknown parameter space. The degree of uncertainty about the unknown parameters will thus play a role in the optimal solution.

To gain a better understanding of the Bayesian approach, we consider various specifications for prior beliefs about the unknown parameters. We start with the standard diffuse prior on μ and V. The typical formulation is given by

$$p_0(\mu, V) \propto |V|^{-\frac{N+1}{2}}. \tag{10}$$

Then, assuming that returns on risky assets are jointly normally distributed, the posterior distribution is given by (see, e.g., Zellner 1971),

$$p(\mu, V \mid \mathbf{\Phi}_T) = p(\mu \mid V, \mathbf{\Phi}_T) \times p(V \mid \mathbf{\Phi}_T) \tag{11}$$

with

$$p(\mu \mid V, \mathbf{\Phi}_T) \propto |V|^{-1/2} \exp\left\{-\frac{1}{2}\mathrm{tr}[T(\mu - \hat{\mu})(\mu - \hat{\mu})'V^{-1}]\right\}, \tag{12}$$

$$P(V) \propto |V|^{-\frac{v}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}V^{-1}(T\hat{V})\right\}, \tag{13}$$

where tr denotes the trace of a matrix and $v = T + N$. Moreover, the predictive distribution obeys the expression

$$p(R_{T+1} \mid \mathbf{\Phi}_T) \propto |\hat{V} + (R_{T+1} - \hat{\mu})(R_{T+1} - \hat{\mu})'/(T + 1)|^{-T/2}, \tag{14}$$

which amounts to a multivariate t-distribution with $T - N$ degrees of freedom.

Although recognized by Markowitz (1952), the problem of estimation error did not receive serious attention until the 1970s. Winkler (1973) and Winkler & Barry (1975) are early examples of Bayesian studies on portfolio choice. Brown (1976, 1978) and Klein & Bawa (1976) independently lay out the Bayesian predictive density approach, with Brown (1976) giving an especially thorough explanation of the estimation error problem and the associated Bayesian approach. Bawa et al. (1979) provide an excellent review of the early literature.

Under the diffuse prior in Equation 10, it is known that the Bayesian optimal portfolio weights are

$$\hat{w}^{\mathrm{Bayes}} = \frac{1}{\gamma}\left(\frac{T - N - 2}{T + 1}\right)\hat{V}^{-1}\hat{\mu}. \tag{15}$$

Similar to the classical solution $\hat{w}^{\mathrm{ML}}$, an optimizing Bayesian agent holds the portfolio that is also proportional to $\frac{1}{\gamma}\hat{V}^{-1}\hat{\mu}$, in which the coefficient of proportion is $(T - N - 2)/(T + 1)$. This coefficient can be substantially smaller than one when N is large relative to T. Intuitively, the assets are riskier in a Bayesian framework because parameter uncertainty is an additional source of risk and this risk is accounted for in the portfolio decision. As a result, in the presence of a risk-free security, the overall positions in risky assets are generally smaller in the Bayesian versus classical frameworks.

However, the Bayesian approach based on diffuse prior does not yield significantly different portfolio decisions compared with the classical framework. In particular, $\hat{w}^{\mathrm{ML}}$ is a biased estimator of w^*, whereas the classical unbiased estimator is given by

$$\bar{w} = \frac{1}{\gamma}\frac{T - N - 2}{T}\hat{V}^{-1}\hat{\mu}, \tag{16}$$

which is a scalar adjustment of $\hat{w}^{\mathrm{ML}}$ and differs from the Bayesian counterpart only by a scalar $T/(T+1)$. The difference is independent of N, and is negligible for all practical sample sizes. Moreover, optimal portfolios formed on the basis of both the maximum likelihood and Bayesian procedures imply the same relative proportions among the N risky assets.

One setting in which diffuse priors do yield considerably different results is when some of the N risky assets have longer histories than others (Stambaugh 1997). Otherwise, incorporating parameter uncertainty makes little difference if the diffuse prior is

used. Indeed, to exhibit the decisive advantage of the Bayesian portfolio analysis, it is generally necessary to elicit informative priors that account for events, macro conditions, asset pricing theories, as well as any other insights relevant to the evolution of stock prices.

2.2. Performance Measures

How can one argue that an informative prior is better than the diffuse prior? In general, it is difficult to make a strong case for a prior specification, because what is good or bad has to be defined and the definition may differ among investors. Moreover, ex ante, knowing which prior is closer to the true data-generating process is also difficult.

Following McCulloch & Rossi (1990), Kandel & Stambaugh (1996), and Pástor & Stambaugh (2000), we focus on utility differences for motivating a performance metric. To illustrate, let $\tilde{w}_a$ and $\tilde{w}_b$ be the Bayesian optimal portfolio weights under priors a and b, and let U_a and U_b be the associated expected utilities evaluated by using the predictive density under prior a. Then the difference in the expected utilities,

$$CER = U_a - U_b, \tag{17}$$

is interpreted as the certainty equivalent return (CER) loss perceived by an investor who is forced to accept the portfolio selection $\tilde{w}_b$ even when $\tilde{w}_a$ would be the ultimate choice. The CER is nonnegative by construction. Indeed, the essential question is how big this value is. Generally speaking, values over a couple of percentage points per year are deemed economically significant.

However, the CER does not say prior a is better or worse than prior b. It merely evaluates the expected utility differential if prior b is used instead of prior a, even when prior a is perceived to be the right one. Recall that the true model as well as which one of the priors is more informative about the true data-generating process are unknown.

Following the statistical decision literature (see, e.g., Lehmann & Casella 1998), we can nevertheless 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 one. In the portfolio choice problem here, the loss function is well defined. Because any estimated portfolio strategy, $\tilde{w}$, is a function of the data, the expected utility loss from using $\tilde{w}$ rather than w^* is

$$\rho(w^*, \tilde{w} \,|\, \mu, V) \equiv U(w^*) - E[U(\tilde{w}) \,|\, \mu, V], \tag{18}$$

where the first term on the right-hand side is the true expected utility based on the true optimal portfolio. Hence, $\rho(w^*, \tilde{w} \,|\, \mu, V)$ is the utility loss if one plays infinite times the investment game with $\tilde{w}$, whether estimated via a Bayesian or a non-Bayesian approach. In particular, the difference in expected utilities between any two estimated rules, $\tilde{w}_a$ and $\tilde{w}_b$, should be

$$Gain = E[U(\tilde{w}_a) \,|\, \mu, V] - E[U(\tilde{w}_b) \,|\, \mu, V]. \tag{19}$$

This is an objective utility gain (loss) of using portfolio strategy $\tilde{w}_a$ versus $\tilde{w}_b$. It is considered to be an out-of-sample measure because it is independent of any single set of observations. If the measure is, say, 5%, then using $\tilde{w}_a$ instead of $\tilde{w}_b$ would yield a 5% gain in the expected utility over repeated use of the estimation strategy. In this

case, if $\tilde{w}_a$ is obtained under prior a and $\tilde{w}_b$ is obtained under prior b, one could consider prior a to be superior to prior b. The loss or gain criterion is widely used in the classical statistics to evaluate two estimators. Brown (1976, 1978), Jorion (1986), Frost & Savarino (1986), and Stambaugh (1997), for example, use $\rho(w^*, \hat{w})$ to evaluate portfolio rules.

One cannot compute the loss function exactly because it depends on unknown true parameters. Nevertheless, it is widely used in two major ways. First, alternative estimators can be assessed in simulations with various assumed true parameters. Second, a comparison of alternative estimators can often be made analytically without any knowledge of the true parameters. For example, Kan & Zhou (2007) show that the Bayesian solution $\hat{w}^{\text{Bayes}}$ dominates $\bar{w}$ given in Equation 16, by having positive utility gains regardless of the true parameter values. However, the Bayesian solution is dominated by yet another classical rule:

$$\hat{w}_c = \frac{c}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}, \qquad c = \frac{(T-N-1)(T-N-4)}{T(T-2)}. \tag{20}$$

This again calls for the use of informative priors in Bayesian portfolio analysis.

2.3. Conjugate Prior

The conjugate prior, which retains the same class of distributions, is a natural and common informative prior on any problem in decision making. In our context, the conjugate specification considers a normal prior for μ (conditional on V) and inverted Wishart prior for V. The conjugate prior is given by

$$\mu \,|\, V \sim N(\mu_0, \frac{1}{\tau}V) \tag{21}$$

and

$$V \sim IW(V_0, v_0), \tag{22}$$

where μ_0 is the prior mean, τ is a parameter reflecting the prior precision of μ_0, and v_0 is a similar prior precision parameter on V. Under this prior, the posterior distribution of μ and V obeys the same form as that based on the conjugate prior, except that now the posterior mean of μ is given by a weighted average of the prior and sample means

$$\tilde{\mu} = \frac{\tau}{T+\tau}\mu_0 + \frac{T}{T+\tau}\hat{\mu}. \tag{23}$$

Similarly, V_0 is updated as

$$\tilde{V} = \frac{T+1}{T(v_0+N-1)}\left(V_0 + T\hat{V} + \frac{T\tau}{T+\tau}(\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})'\right), \tag{24}$$

which is a weighted average of the prior variance, sample variance, and deviations of $\hat{\mu}$ from μ_0.

Frost & Savarino (1986) provide an interesting application of the conjugate prior, assuming a priori that all assets exhibit identical means, variances, and patterned covariances. They find that such a prior improves ex post performance. This prior is related the well-known 1/N rule that invests equally across the N assets.

2.4. Hyperparameter Prior

Jorion (1986) introduces hyperparameters η and λ that underlie the prior distribution of μ. In particular, the hyperparameter prior is formulated as

$$p_0(\mu \mid \eta, \lambda) \propto |V|^{-1} \exp\left\{-\frac{1}{2}(\mu - \eta 1_N)'(\lambda V)^{-1}(\mu - \eta 1_N)\right\}. \tag{25}$$

Then, employing diffuse priors on both η and λ and integrating these parameters out from a suitable distribution, the predictive distribution of the future portfolio return can be obtained following Zellner & Chetty (1965). In particular, Jorion's optimal portfolio rule is given by

$$w^{\mathrm{PJ}} = \frac{1}{\gamma}(\hat{V}^{\mathrm{PJ}})^{-1}\hat{\mu}^{\mathrm{PJ}}, \tag{26}$$

where

$$\hat{\mu}^{\mathrm{PJ}} = (1 - \hat{v})\hat{\mu} + \hat{v}\hat{\mu}_g 1_N, \tag{27}$$

$$\hat{V}^{\mathrm{PJ}} = \left(1 + \frac{1}{T + \hat{\lambda}}\right)\bar{V} + \frac{\hat{\lambda}}{T(T + 1 + \hat{\lambda})}\frac{1_N 1_N'}{1_N' \bar{V}^{-1} 1_N}, \tag{28}$$

$$\hat{v} = \frac{N + 2}{(N + 2) + T(\hat{\mu} - \hat{\mu}_g 1_N)' \bar{V}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)}, \tag{29}$$

$$\hat{\lambda} = (N + 2)/[(\hat{\mu} - \hat{\mu}_g 1_N)' \bar{V}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)], \tag{30}$$

$$\bar{V} = T\hat{V}/(T - N - 2), \tag{31}$$

and

$$\hat{\mu}_g = 1_N' \bar{V}^{-1}\hat{\mu}/1_N' \bar{V}^{-1} 1_N. \tag{32}$$

This hyperparameter portfolio rule can be motivated on the basis of the following Bayes-Stein shrinkage estimator (see, e.g., Jobson et al. 1979) of expected return

$$\hat{\mu}^{\mathrm{BS}} = (1 - v)\hat{\mu} + v\mu_g 1_N, \tag{33}$$

where $\mu_g 1_N$ is the shrinkage target, $\mu_g = 1_N' V^{-1}\mu/1_N' V^{-1} 1_N$, and v is the weight given to the target. Jorion (1986) as well as subsequent studies find that w^{PJ} improves w^{ML} substantially, implying that it also outperforms the Bayesian strategy based on the diffuse prior.

2.5. The Black-Litterman Model

Markowitz's portfolio rule $\hat{w}^{\mathrm{ML}}$ typically implies unusually large long and short positions in the absence of portfolio constraints. Moreover, it delivers many zero positions when short sales are not allowed. Black & Litterman (1992) provide a novel solution to this problem. They assume that the investor starts with initial views consistent with market equilibrium and then updates those views with his own views via the Bayesian rule. For instance, if the market equilibrium views are based on the CAPM, the implied optimal

portfolio is the value-weighted index. If the investor has views identical to the market, the market portfolio will be the ultimate choice.

However, what if the investor has different views? Black & Litterman (1992) propose a way to update market views with the investor's own views. Let us formalize the Black-Litterman model. Based on market views, expected excess returns are given by

$$\mu^e = \gamma V w_e, \tag{34}$$

where w_e denotes the value-weighted weights in the stock index and γ is the market risk-aversion coefficient. Assume that the true expected excess return μ is normally distributed with mean μ^e,

$$\mu = \mu^e + \varepsilon^e, \quad \varepsilon^e \sim N(0, \tau V), \tag{35}$$

where ε^e, the deviation of μ from μ^e, is normally distributed with zero mean and covariance matrix τV in which τ is a scalar indicating the degree of belief in how close μ is to the equilibrium value μ^e. In the absence of any views on future stock returns, and in the special case of $\tau = 0$, the investor's portfolio weights must be equal to w_e, the weights of the value-weighted index.

Black & Litterman (1992) consider views about the relative performance of stocks that can be represented mathematically by a single vector equation,

$$P\mu = \mu^v + \varepsilon^v, \quad \varepsilon^v \sim N(0, \Omega), \tag{36}$$

where P is a $K \times N$ matrix summarizing K views, μ^v is a K-vector summarizing the prior means of the view portfolios, and ε^v is the residual vector. The views may be formed on the basis of news, events, or analysis on the economy and investable assets. The covariance matrix of the residuals, Ω, measures the degree of confidence the investor has in his own views. Applying Bayes' rule to the beliefs in the market equilibrium relationship and the investor's own views, as formulated in Equations 35 and 36, Black & Litterman (1992) obtain the Bayesian updated expected returns and risks as

$$\bar{\mu}^{BL} = [(\tau V)^{-1} + P'\Omega^{-1}P]^{-1}[(\tau V)^{-1}\mu^e + P'\Omega^{-1}\mu^v] \tag{37}$$

and

$$\bar{V}^{BL} = V + [(\tau V)^{-1} + P'\Omega^{-1}P]^{-1}. \tag{38}$$

Replacing V by $\hat{V}$ and plugging these two updated estimates into Equation 6, one obtains the Black-Litterman solution to the portfolio choice problem.

Note that the Black-Litterman expected return, $\bar{\mu}^{BL}$, is a weighted average of the equilibrium expected return and the investor's views about expected return. Intuitively, the less confident the investor is about his views, the closer $\bar{\mu}^{BL}$ is to the equilibrium value, and so the closer the Black-Litterman portfolio is to w_e. He & Litterman (1999) mathematically show this is the case. Hence, the Black-Litterman model tilts the investor's optimal portfolio away from the market portfolio according to the strength of the investor's views. Because the market portfolio is a reasonable starting point that takes no extreme positions, any suitably controlled tilt should also yield a portfolio without any extreme positions. This property is one of the major reasons the Black-Litterman model is popular in practice.

Whereas the Black-Litterman model is considered to be a Bayesian approach, it is not entirely Bayesian. For one, the data-generating process is not spelled out explicitly. Moreover, the Bayesian predictive density is not used anywhere. Zhou (2009) treats the

investor's view as yet another layer of priors and combines this and the equilibrium prior with the data-generating process, resulting in a formal Bayesian treatment and an extension of the Black-Litterman model.

2.6. Asset Pricing Prior

Pástor (2000) and Pástor & Stambaugh (2000) introduce interesting priors that reflect an investor's degree of belief in the ability of an asset pricing model to explain the cross sectional dispersion in expected returns. In particular, let $R_t = (y_t, x_t)$, 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

$$y_t = \alpha + Bx_t + u_t, \tag{39}$$

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

$$\alpha = \mu_1 - B\mu_2, \qquad B = V_{12}V_{22}^{-1}, \tag{40}$$

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

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}. \tag{41}$$

A factor-based asset pricing model, such as the three-factor model of Fama & French (1993), implies the restrictions $\alpha = 0$ for all nonbenchmark assets.

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

$$\alpha \mid \Sigma \sim N\left[0, \sigma_\alpha^2\left(\frac{1}{s_\Sigma^2}\Sigma\right)\right], \tag{42}$$

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 & Pástor (2000) in a classical framework. The magnitude of σ_α represents an investor's level of uncertainty about the pricing ability of a given model. On the one hand, 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 disregards the pricing model as entirely useless.

This asset pricing prior also has the Bayes-Stein shrinkage interpretation. In particular, the prior on α implies a prior mean on μ, say μ_0. Accordingly, the predictive mean is

$$\mu_p = \tau\mu_0 + (1 - \tau)\hat{\mu}, \tag{43}$$

where τ inversely depends on the sample size and positively on the level of prior confidence in the pricing model. Similarly, the predictive variance is a mixture of prior and sample variances.

2.7. Objective-Based Prior

Priors are generally placed on the parameters μ and V, not on the optimal portfolio weights. Indeed, a diffuse prior on the parameters may be interpreted as a diffuse prior on the optimal portfolio weights as well. However, in various applications, seemingly innocuous diffuse priors on some basic model parameters can actually imply rather strong

prior convictions about particular economic dimensions of the problem. For example, in the context of testing portfolio efficiency, Kandel et al. (1995) find that a diffuse prior on model parameters implies a rather strong prior on inefficiency of a given portfolio. Klein & 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. In the context of return predictability, Lamoureux & Zhou (1996) find that the diffuse prior implies a prior concentration on either high or low degrees of return predictability. Thus, it is important to form informative priors on the model parameters that can imply reasonable priors on functions of interest.

Tu & Zhou (2010) advocate a method for constructing priors on the unobserved parameters based on a prior on the solution of an economic objective. In maximizing an economic objective, a Bayesian agent may have some idea about the range for the solution even before observing the data. Thus, the aim is to form a prior on the solution, from which the prior on the parameters can be backed out. For instance, the investor may have a prior corresponding to equal or value-weighted portfolio weights. The prior on optimal weights can then be transformed into a prior on μ and V. Such priors on the primitive parameters are called objective-based priors.

Formally, the objective-based prior starts from a prior on w,

$$w \sim N(w_0, V_0 V^{-1}/\gamma), \tag{44}$$

where w_0 and V_0 are suitable prior constants with known values and then back out a prior on μ,

$$\mu \sim N\left[\gamma V w_0, \sigma_\rho^2\left(\frac{1}{s^2} V\right)\right], \tag{45}$$

where s^2 is the average of the diagonal elements of V. The prior on V can be taken as the usual inverted Wishart distribution.

Using monthly returns on the Fama-French 25 size and book-to-market portfolios and three factors from January 1965 to December 2004, Tu & Zhou (2009) find that investment performance under objective-based priors can be significantly different from that under diffuse and asset pricing priors, with differences in annual certainty-equivalent returns greater than 10% in many cases. In terms of the loss function measure, portfolio strategies based on the objective-based priors can substantially outperform both strategies under the alternative priors.

3. PREDICTABLE RETURNS

So far, asset returns are assumed to be IID and thus unpredictable through time. However, Keim & Stambaugh (1986), Campbell & Shiller (1988), and Fama & French (1989) identify business-cycle variables, such as the aggregate dividend yield and the default spread, that predict future stock and bond returns. Such predictive variables, when incorporated in studies that deal with the time-series and cross-sectional properties of expected returns, provide fresh insights into asset pricing and portfolio selection. In asset pricing, Lettau & Ludvigson (2001) and Avramov & Chordia (2006a) show that factor models with time-varying risk premia and/or risk are reasonably successful relative to their unconditional counterparts. Focusing on portfolio selection, Kandel & Stambaugh (1996) analyze investments when returns are potentially predictable.

3.1. One-Period Models

In particular, consider a one-period optimizing investor who must allocate at time T funds between the value-weighted NYSE index and one-month Treasury bills. The investor makes portfolio decisions based on estimating the predictive system

$$r_t = a + b'z_{t-1} + u_t \tag{46}$$

and

$$z_t = \theta + \rho z_{t-1} + v_t, \tag{47}$$

where r_t is the continuously compounded NYSE return in month t in excess of the continuously compounded T-bill rate for that month, z_{t-1} is a vector of M predictive variables observed at the end of month $t-1$, b is a vector of slope coefficients, and u_t is the regression disturbance in month t. The evolution of the predictive variables is essentially stochastic. Typically, a first-order vector autoregression is employed to model that evolution. The residuals in Equations 46 and 47 are assumed to obey the normal distribution. In particular, let $\eta_t = [u_t, v_{t'}]'$ then $\eta_t \sim N(0, \Sigma)$, where

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_v \end{bmatrix}. \tag{48}$$

The distribution of r_{T+1}, the time $T+1$ NYSE excess return, which is conditional on data and model parameters, is $N(a + b'z_T, \sigma_u^2)$. Assuming an inverted Wishart prior distribution for Σ and a multivariate normal prior for the intercept and slope coefficients in the predictive system, the Bayesian predictive distribution $P(r_{T+1} | \Phi_T)$ obeys the student t density. Then, for a power utility investor with a parameter of relative risk aversion denoted by γ, the optimization problem is

$$\omega^* = \arg\max_{\omega} \int_{r_{T+1}} \frac{[(1-\omega)\exp(r_f) + \omega \exp(r_f + r_{T+1})]^{1-\gamma}}{1-\gamma} P(r_{T+1} | \Phi_T) dr_{T+1}, \tag{49}$$

subject to ω being nonnegative. An analytic solution for the optimal portfiolio is unfeasible, but a proper solution can easily be obtained numerically. In particular, given G independent draws for R_{T+1} from the predictive distribution, the optimal portfolio is found by maximizing the quantity

$$\frac{1}{G} \sum_{g=1}^{G} \frac{\left\{(1-\omega)\exp(r_f) + \omega \exp(r_f) + R_{T+1}^{(g)}\right\}^{1-\gamma}}{1-\gamma} \tag{50}$$

subject to ω being nonnegative. Kandel & Stambaugh (1996) show that even when the statistical evidence on predictability is weak, as reflected through the R^2 in the regression in Equation 46, the current values of the predictive variables, z_T, can exert a substantial influence on the optimal portfolio.

3.2. Multiperiod Models

Whereas Kandel & Stambaugh (1996) study asset allocation in a single-period framework, Barberis (2000) analyzes multiperiod investment decisions, considering both a buy-and-hold investor as well as an investor who dynamically rebalances the optimal

stock-bond allocation. Implementing long-horizon asset allocation in a buy-and-hold setup is straightforward. In particular, let K denote the investment horizon; then $R_{T+K} = \sum_{k=1}^{K} r_{T+k}$ is the cumulative (continuously compounded) return over the investment horizon. Avramov (2002) shows that the distribution for R_{T+K} conditional on the data (denoted Φ_T) and set of parameters (denoted Θ) is given by

$$R_{T+K}|, \Theta, \Phi_T \sim N(\lambda, \Upsilon), \tag{51}$$

where

$$\lambda = Ka + b'\left[(\rho^K - I_M)(\rho - I_M)^{-1}\right]z_T + b'\left[\rho(\rho^{K-1} - I_M)(\rho - I_M)^{-1} - (K-1)I_M\right](\rho - I_M)^{-1}\theta, \tag{52}$$

$$\Upsilon = K\sigma_u^2 + \sum_{k=1}^{K}\delta(k)\Sigma_v\delta(k)' + \sum_{k=1}^{K}\sigma_{uv}\delta(k)' + \sum_{k=1}^{K}\delta(k)\sigma_{vu}, \tag{53}$$

and

$$\delta(k) = b'\left[\left(\rho^{k-1} - I_M\right)(\rho - I_M)^{-1}\right]. \tag{54}$$

Drawing from the Bayesian predictive distribution is done in two steps. First, draw the model parameters Θ from their posterior distribution. Second, conditional on model parameters, draw R_{T+K} from the normal distribution formulated in Equations 51–54. The optimal portfolio can then be found using Equation 50 with R_{T+K} replacing R_{T+1} and Kr_f replacing r_f.

Incorporating dynamic rebalancing, intermediate consumption, and learning could establish a nontrivial challenge for recovering the optimal portfolio choice. Brandt et al. (2005) nicely address the challenge using a tractable simulation–based method. Pástor & Veronesi (2009) comprehensively survey the literature on learning in financial markets.

Essentially, the IID setup corresponds to $b = 0$ in the predictive regression (Equation 46), which yields $\lambda_{iid} = Ka$ and $\Upsilon_{iid} = K\sigma_u^2$ in Equations 52 and 53. The conditional mean and variance in an IID world increase linearly with the investment horizon. Thus, there is no horizon effect when (a) returns are IID and (b) estimation risk is not accounted for, as shown by Samuelson (1969) and Merton (1969) in an equilibrium framework. Incorporating estimation risk, Barberis (2000) shows that the allocation to equity diminishes with the investment horizon, as stocks appear to be riskier in longer horizons. Accounting for both return predictability and estimation risk, Barberis (2000) shows that investors allocate more heavily to equity the longer their horizon.

Although estimation risk plays virtually no role in the single-period case, it plays an important role in long-horizon investment decisions. Barberis shows that a long-horizon investor who ignores it may overallocate to stocks by a sizeable amount. Even when the predictors evolve stochastically, both Kandel & Stambaugh (1996) and Barberis (2000) assume that the initial value of the predictive variables z_0 is nonstochastic. With a stochastic initial value, the distribution of future returns conditioned on model parameters no longer obeys a well-known distributional form. Stambaugh (1999) addresses this problem by implementing the Metropolis Hastings algorithm, a Markov chain Monte Carlo procedure introduced by Metropolis et al. (1953) and generalized by Hastings (1970). Several other powerful numerical Bayesian algorithms exist, including the Gibbs sampler and data

augmentation (see a review by Chib & Greenberg 1996), which make the Bayesian approach broadly applicable. Additional advantages of the Bayesian approach are the ability it gives a Bayesian investor to incorporate model uncertainty as well as to consider prior views about the degree of predictability explained by asset pricing models. The latter two advantages of the Bayesian approach are explained below.

3.3. Model Uncertainty

As noted above, financial economists have identified economic variables that appear to predict future asset returns. However, the "correct" predictive regression specification has remained an open issue for several reasons. For one, existing equilibrium pricing theories are not explicit about which variables should enter the predictive regression. This ambiguity is undesirable, as it renders the empirical evidence subject to data-overfitting concerns. Indeed, Bossaerts & Hillion (1999) confirm in-sample return predictability but fail to detect out-of-sample predictability. Moreover, the multiplicity of potential predictors also makes the empirical evidence difficult to interpret. For example, one may find an economic variable statistically significant on the basis of a particular collection of explanatory variables, but often not on the basis of a competing specification. Given that the true set of predictive variables is unknown, the Bayesian methodology of model averaging described below is attractive, as it explicitly incorporates model uncertainty in asset allocation decisions.

Bayesian model averaging has been used to study heart attacks in medicine, traffic congestion in transportation economy, hot hands in basketball, and economic growth in the macroeconomy literature. In finance, Bayesian model averaging facilitates a flexible modeling of investors' uncertainty about potentially relevant predictive variables in forecasting models. In particular, it assigns posterior probabilities to a wide set of competing return-generating models (overall, 2^M models). It then uses the probabilities as weights on the individual models to obtain a composite-weighted model. This optimally weighted model is then employed to investigate asset allocation decisions. Bayesian model averaging contrasts sharply with the traditional classical approach of model selection. In the latter approach, one uses a specific criterion (e.g., adjusted R^2) to select a single model and then operates as if that selected model is correct. Implementing model-selection criteria, the econometrician views the selected model as the true one with a unit probability and discards the other competing models as worthless, thereby ignoring model uncertainty. Accounting for model uncertainty, Avramov (2002) shows that Bayesian model averaging outperforms, ex post out-of-sample, the classical approach of model-selection criteria, generating smaller forecast errors and being more efficient. Ex ante, an investor who ignores model uncertainty suffers considerable utility loses.

The Bayesian weighted predictive distribution of R_{T+K} averages over the model space and integrates over the posterior distribution that summarizes the within-model parameter uncertainty about Θ_j:

$$P(R_{T+K} \,|\, \Phi_T) = \sum_{j=1}^{2^M} P(\mathcal{M}_j \,|\, \Phi_T) \int_{\Theta_j} P(\Theta_j \,|\, \mathcal{M}_j, \Phi_T) P(R_{T+K} \,|\, \mathcal{M}_j, \Theta_j, \Phi_T) d\Theta_j, \qquad (55)$$

where j is the model identifier, $P(\mathcal{M}_j \,|\, \Phi_T)$ is the posterior probability that model $\mathcal{M}_j$ is the correct one, and Θ_j denotes the parameters of model j. Drawing from the weighted

predictive distribution is done in three steps: First, draw the model from the distribution of models. Then, conditional upon the model, implement the two steps noted above for drawing future returns from the model-specific Bayesian predictive distribution.

3.4. Prior About the Extent of Predictability Explained by Asset Pricing Models

As noted above, the Bayesian approach facilitates incorporating economically motivated priors. In the context of return predictability, the classical approach has examined whether predictability is explained by rational pricing or whether it is due to asset pricing misspecification (see, e.g., Campbell 1987, Ferson & Korajczyk 1995, Kirby 1998). Studies such as these approach finance theory by focusing on two polar viewpoints: rejecting or not rejecting a pricing model based on hypothesis tests. The Bayesian approach incorporates pricing restrictions on predictive regression parameters as a reference point for a hypothetical investor's prior belief. The investor uses the sample evidence about the extent of predictability to update various degrees of belief in a pricing model and then allocates funds across cash and stocks. Pricing models are expected to exert stronger influence on asset allocation when the prior confidence in their validity is stronger and when they explain much of the sample evidence on predictability.

Avramov (2004) models excess returns on N investable assets as

$$r_t = \alpha(z_{t-1}) + \beta f_t + u_{rt}, \tag{56}$$

$$\alpha(z_{t-1}) = \alpha_0 + \alpha_1 z_{t-1}, \tag{57}$$

$$f_t = \lambda(z_{t-1}) + u_{ft}, \tag{58}$$

and

$$\lambda(z_{t-1}) = \lambda_0 + \lambda_1 z_{t-1}, \tag{59}$$

where f_t is a set of K monthly excess returns on portfolio-based factors, α_0 stands for an N-vector of the fixed component of asset mispricing, α_1 is an $N \times M$ matrix of the time-varying component, and β is an $N \times K$ matrix of factor loadings. A conditional version of an asset pricing model (with fixed beta) implies the relation

$$\mathbb{E}(r_t \mid z_{t-1}) = \beta\lambda(z_{t-1}) \tag{60}$$

for all t, where $\mathbb{E}$ denotes the expected value operator. Equation 60 imposes restrictions on the parameters and the goodness of fit in the multivariate predictive regression

$$r_t = \mu_0 + \mu_1 z_{t-1} + v_t, \tag{61}$$

where μ_0 is an N-vector and μ_1 is an $N \times M$ matrix of slope coefficients. In particular, note that by adding to the right-hand side of Equation 61 the quantity $\beta(f_t - \lambda_0 - \lambda_1 z_{t-1})$, subtracting the (same) quantity βu_{ft}, and decomposing the residual in Equation 61 into two orthogonal components as $v_t = \beta u_{ft} + u_{rt}$, we reparameterize the return-generating process (Equation 61) as

$$r_t = (\mu_0 - \beta\lambda_0) + (\mu_1 - \beta\lambda_1)z_{t-1} + \beta f_t + u_{rt}. \tag{62}$$

Matching the right-hand-side coefficients in Equation 62 with those in Equation 56 yields

$$\mu_0 = \alpha_0 + \beta\lambda \tag{63}$$

and

$$\mu_1 = \alpha_1 + \beta\lambda_1. \qquad (64)$$

The relation in Equation 64 indicates that return predictability, if it exists, is due to the security-specific model mispricing component ($\alpha_1 \neq 0$) and/or the common component in risk premia that varies ($\lambda_1 \neq 0$). When mispricing is precluded, the regression parameters that conform to the asset pricing model are

$$\mu_0 = \beta\lambda_0, \qquad (65)$$

$$\mu_1 = \beta\lambda_1. \qquad (66)$$

Avramov (2004) shows that asset allocation is extremely sensitive to the imposition of model restrictions on predictive regressions. Indeed, an investor who believes those restrictions are perfectly valid but is forced to allocate funds while disregarding the model's implications faces an enormous utility loss. Furthermore, asset allocations depart considerably from those dictated by the pricing models when the prior allows even minor deviations from the underlying models.

3.5. Time-Varying Beta

Although the above discussion assumes that beta is constant, accounting for time-varying beta is straightforward. Avramov & Chordia (2006b) model the $N \times K$ matrix of factor loadings as

$$\beta(z_t) = \beta_0 + \beta_1(I_K \otimes z_t), \qquad (67)$$

where $\otimes$ denotes the Kronecker product. These authors show that the mean and covariance matrix of asset returns in the presence of time-varying alpha, beta, and risk premia (assuming informative priors) can be expressed as

$$\mu_T = \hat{\alpha}(z_T) + \hat{\beta}(z_T)(\hat{a}_f + \hat{A}_f z_T), \qquad (68)$$

and

$$\Sigma_T = \mathcal{P}_1\hat{\beta}(z_T)\hat{\Sigma}_{ff}\hat{\beta}(z_T)' + \mathcal{P}_2\hat{\Psi}, \qquad (69)$$

where the $\hat{x}$ notation stands for the maximum likelihood estimators, Σ_{ff} is the covariance matrix of u_{ft}, and Ψ is the covariance matrix of u_{rt}, assumed to be diagonal. The predictive variance in Equation 69 is larger than its maximum likelihood analog as it incorporates the factors $\mathcal{P}_1$ and $\mathcal{P}_2$, where $\mathcal{P}_1$ is a scalar greater than one and $\mathcal{P}_2$ is a diagonal matrix such that each diagonal entry is greater than one.

3.6. Out-of-Sample Performance

Stock return predictability continues to be a subject of research controversy. Skepticism exists as a result of concerns relating to data mining, statistical biases, and weak out-of-sample performance of predictive regressions. Foster et al. (1997), Bossaerts & Hillion (1999), and Stambaugh (1999) address such concerns. Moreover, if firm-level predictability indeed exists, it is not clear whether it is driven by time variation in alpha, beta, or the equity premium.

Relative to the IID setup, incorporating predictability does improve performance of investments in equity portfolios, single stocks, mutual funds, and hedge funds. Focusing on equity portfolios, Avramov (2004) shows that optimal portfolios based on dogmatic beliefs in conditional pricing models deliver the lowest Sharpe ratios. In addition, completely disregarding pricing-model implications results in the second lowest Sharpe ratios. Much higher Sharpe ratios are obtained when asset allocations are based on the so-called shrinkage approach, in which inputs for portfolio optimization combine the underlying pricing model and the sample evidence on predictability.

Avramov & Chordia (2006b) show that incorporating business-cycle predictors benefits a real-time optimizing investor who must allocate funds across 3123 NYSE-AMEX stocks and cash. Investment returns are positive when adjusted by the Fama-French and momentum factors as well as by size, book-to-market, and past-return characteristics. The investor optimally holds small-cap, growth, and momentum stocks and loads less (more) heavily on momentum (small-cap) stocks during recessions. Returns on individual stocks are predictable out-of-sample due to alpha variation. In contrast, beta variation plays no role. Whereas Avramov (2004) and Avramov & Chordia (2006b) focus on multisecurity paradigms, Wachter & Warusawitharana (2009) document the superior out-of-sample performance of the Bayesian approach in market timing. That is, the equity premium is also predictable by macro conditions.

3.7. Investing in Mutual and Hedge Funds

In an IID setup, Baks et al. (2001) explore the role of prior information about fund performance in making investment decisions. These authors consider a mean-variance optimizing investor who is skeptical about the ability of a fund manager to pick stocks and time the market. They find that even with a high degree of skepticism about fund performance the investor would allocate considerable amounts to actively managed funds.

Baks et al. (2001) define fund performance as the intercept in the regression of the fund's excess returns on excess return of one or more benchmark assets. Pástor & Stambaugh (2002a, 2002b), however, recognize the possibility that the intercept in such regressions could be a mix of fund performance and model mispricing. In particular, consider the case wherein benchmark assets used to define fund performance are unable to explain the cross-section dispersion of passive assets, that is, the sample alpha in the regression of nonbenchmark passive assets on benchmarks assets is nonzero. Then model mispricing emerges in the performance regression. Thus, Pástor & Stambaugh (2002a, 2002b) formulate prior beliefs on both performance and mispricing.

Geczy et al. (2005) apply the Pástor-Stambaugh methodology to study the cost of investing in socially responsible mutual funds. Comparing portfolios of these funds to those constructed from the broader fund universe reveals the cost of imposing the socially responsible investment constraint on investors seeking the highest Sharpe ratio. This socially responsible investment cost depends crucially on the investor's views about the validity of asset pricing models and managerial skills in stock picking and market timing. Busse & Irvine (2006) also apply the Pástor-Stambaugh methodology to compare the performance of Bayesian estimates of mutual fund performance with standard classical-based measures using daily data. They find that Bayesian alphas based on the CAPM are particularly useful for predicting future standard CAPM alphas.

Baks et al. (2001) and Pástor & Stambaugh (2002a, 2002b) assume that the prior on alpha is independent across funds. However, as shown by Jones & Shanken (2005), under the independence assumption, the maximum posterior mean alpha increases without bound as the number of funds increases and "extremely large" estimates could randomly be generated, even when fund managers have no skill. Instead, Jones & Shanken (2005) propose incorporating prior dependence across funds. Then, investors aggregate information across funds to form a general belief about the potential for abnormal performance. Each fund's alpha estimate is shrunk toward the aggregate estimate, mitigating extreme views.

Avramov & Wermers (2006) and Avramov et al. (2010) extend the Avramov (2004) methodology to study investments in mutual funds and hedge funds, respectively, when fund returns are potentially predictable. Avramov & Wermers (2006) show that long-only strategies that incorporate predictability in managerial skills outperform their Fama-French and momentum benchmarks by 2–4% per year by timing industries over the business cycle, and by an additional 3–6% per year by choosing funds that outperform their industry benchmarks. Similarly, Avramov et al. (2010) show that incorporating predictability substantially improves out-of-sample performance for the entire universe of hedge funds as well as for various investment styles. The major source of investment profitability is again predictability in managerial skills. In particular, long-only strategies that incorporate such predictability outperform their Fung & Hsieh (2004) benchmarks by more than 14% per year. The economic value of predictability emerges for different rebalancing horizons and alternative benchmark models. It is also robust to adjustments for backfill bias, incubation bias, illiquidity, and style composition.

4. ALTERNATIVE DATA-GENERATING PROCESSES

The data-generating processes for asset returns discussed thus far are either IID normal or predictable with IID disturbances. Such specifications facilitate a tractable implementation of Bayesian portfolio analysis. To provide a richer model of the interaction between the stock market and economic fundamentals, Pástor & Stambaugh (2009a) advocate a predictive system allowing aggregate predictors to be imperfectly correlated with the conditional expected return. Subsequently, Pástor & Stambaugh (2009b) find that stocks are substantially more volatile over long horizons from an investor's perspective, which seems to have profound implications for long-term investments.

Incorporating regime shifts in asset returns is also potentially attractive, as stock prices tend to rise or fall persistently during certain periods. Tu (2010) extends the asset pricing framework (Equation 39) to capture economic regimes. In particular, he models benchmark and nonbenchmark assets as

$$y_t = \alpha^{s_t} + B^{s_t} x_t + u_t^{s_t}, \tag{70}$$

where $u_t^{s_t}$ is an $m \times 1$ vector with zero means and a nonsingular covariance matrix, Σ^{s_t}, and s_t is an indicator of the states. Under the usual normal assumption of model residuals, the regime shift formulation is identical to the specification (Equation 39) in each regime. Tu (2010) shows that uncertainty about regime is more important than model mispricing. Hence, the correct identification of the data-generating process can have significant impact on portfolio choice.

To incorporate latent factors and stochastic volatility in the asset pricing formulation (Equation 39), Han (2006) allows x_t in

$$y_t = \alpha + Bx_t + u_t \tag{71}$$

to follow the latent process

$$x_t = c + CX_{t-1} + v_t. \tag{72}$$

In addition, the vector of residuals u_t could display stochastic volatilities. In such a dynamic factor multivariate stochastic volatility model, Han finds that the corresponding dynamic strategies significantly outperform various benchmark strategies out of sample, and the outperformance is robust to different performance measures, investor's objective functions, time periods, and assets. In addition, Nardari & Scruggs (2007) extend Geweke & Zhou (1996) to provide an alternative stochastic volatility model with latent asset pricing factors. In their model, mispricing with respect to the arbitrage pricing theory pioneered by Ross (1976) can be accommodated.

Because the true data-generating process is unknown, there is uncertainty about whether a given process adequately fits the data. For example, previous studies typically assume that stock returns are conditionally normal. However, the normality assumption is strongly rejected by the data. Tu & Zhou (2004) find that the t distribution can better fit the data. Kacperczyk (2008) provides a general framework for treating data-generating-process uncertainty.

5. EXTENSIONS AND FUTURE RESEARCH

Even though Bayesian analysis of portfolio selection has impressively evolved over the past three decades, there is still a host of applications of Bayesian methodologies to be carried out. For one, the Bayesian methodology can be applied to account for estimation risk and model uncertainty in managing long-short portfolios, international asset allocation, hedge fund speculation, defined-benefit pensions, as well as portfolio selection with various risk controls. In addition, there are still virtually untouched asset pricing theories to be accounted for in forming informative prior beliefs.

Mean-variance utility has long been the baseline for asset allocation in practice. (See, for instance, Grinold & Kahn (1999), Litterman (2003), and Meucci (2005), who discuss various applications of the mean-variance framework.) Indeed, controlling for factor exposures and imposing trading constraints, among other real-time trading impediments, can easily be accommodated within the mean-variance framework with either analytical insights or fast numerical solutions. In addition, the intertemporal hedging demand is typically small relative to the mean-variance component. Theoretically, however, it would be of interest to consider alternative sets of preferences.

Employing alternative utility specifications must be done with extra caution. In particular, as emphasized by Geweke (2001), the predictive density under iso-elastic preferences is typically student t. Unrestricted utility maximization under the t predictive density can encounter a divergence problem, but the divergence problem can be addressed by imposing suitable portfolio constraints. Moreover, the divergence problem disappears with a suitable adjustment of the degrees of freedom of the t distribution. Harvey et al. (2004) is an excellent example of portfolio selection with higher moments that has an interpretation well grounded in economic theory. Ang et al. (2005) and Hong et al. (2007)

advocate a Bayesian portfolio analysis that allows the data-generating process to be asymmetric.

A different class of recursive utility functions is found to be useful in accounting for asset pricing patterns unexplained by the CAPM of Sharpe (1964) and Lintner (1965) and the consumption-based CAPM (CCAPM) of Rubinstein (1976), Lucas (1978), Breeden (1979), and Grossman & Shiller (1981). In particular, Bansal & Yaron (2004) utilize Esptein & Zin (1989) preferences to explain asset pricing puzzles at the aggregate level. Avramov et al. (2009) consider Duffie & Epstein (1992) preferences to explain the counterintuitive cross-sectional negative relations between average stock returns and the three apparent risk measures (*a*) credit risk, (*b*) dispersion, and (*c*) idiosyncratic volatility. Recursive preferences are also employed by Zhou & Zhu (2009), who can justify the large negative market-variance risk premium. Indeed, to our knowledge, there are no Bayesian studies utilizing the recursive utility framework, nor are there any Bayesian priors that exploit information on such potentially promising asset pricing models. Future work should form prior beliefs based on long-run risk formulations.

Finally, portfolio analysis based on specifications that depart from IID stock returns (see the multivariate process formulated in Sections 3 and 4) is challenging to solve in multiperiod investment horizons. Much future research in this area is called for.

6. CONCLUSION

In making portfolio decisions, investors often confront parameter estimation errors and possible model uncertainty. In addition, investors may have prior information about the investment problem that can arise from news, events, macroeconomic analysis, and asset pricing theories. The Bayesian approach is well-suited to neatly account for these features, whereas the classical statistical analysis disregards any potentially relevant prior information. Hence, Bayesian portfolio analysis is likely to play an increasing role in making investment decisions in practical investment management.

Although enormous progress has been made in developing various priors and methodologies for applying the Bayesian approach in standard asset-allocation problems, there are still investment problems that are open for future Bayesian studies. Moreover, much more should be done to allow Bayesian portfolio analysis to go beyond popular mean-variance utilities as well as to consider more general and realistic data-generating processes.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

Ait-Sahalia Y, Brandt M. 2001. Variable selection for portfolio choice. *J. Finance* 56:1297–351
Ang A, Bekaert G, Liu J. 2005. Why stocks may disappoint. *J. Financ. Econ.* 76:471–508

Avramov D. 2002. Stock return predictability and model uncertainty. *J. Financ. Econ.* 64:423–58

Avramov D. 2004. Stock return predictability and asset pricing models. *Rev. Financ. Stud.* 17:699–738

Avramov D, Chordia T. 2006a. Asset pricing models and financial market anomalies. *Rev. Financ. Stud.* 19:1001–40

Avramov D, Chordia T. 2006b. Predicting stock returns. *J. Financ. Econ.* 82:387–415

Avramov D, Wermers R. 2006. Investing in mutual funds when returns are predictable. *J. Financ. Econ.* 81:339–77

Avramov D, Cederburg S, Hore S. 2009. *Cross-sectional asset pricing puzzles: an equilibrium perspective.* Work. Pap., Univ. Maryland

Avramov D, Kosowski R, Naik N, Teo M. 2010. Hedge funds, managerial skill, and macroeconomic variables. *J. Financ. Econ.* Forthcoming

Baks K, Metrick A, Wachter J. 2001. Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation. *J. Finance* 56:45–85

Bansal R, Yaron A. 2004. Risks for the long run: a potential resolution of asset pricing puzzles. *J. Finance* 59:1481–509

Barberis N. 2000. Investing for the long run when returns are predictable. *J. Finance* 55:225–64

Bawa VS, Brown SJ, Klein RW. 1979. *Estimation Risk and Optimal Portfolio Choice.* Amsterdam: North-Holland

Black F, Litterman R. 1992. Global portfolio optimization. *Financ. Anal. J.* 48:28–43

Bossaerts P, Hillion P. 1999. Implementing statistical criteria to select return forecasting models: What do we learn? *Rev. Financ. Stud.* 12:405–28

Brandt M. 2009. Portfolio choice problems. In *Handbook of Financial Econometrics*, Vol. 1, ed. Y Ait-Sahalia, LP Hansen. Amsterdam: North-Holland

Brandt M, Goyal A, Santa-Clara P, Stroud JR. 2005. A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *Rev. Financ. Stud.* 18:831–71

Breeden DT. 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *J. Financ. Econ.* 7:265–96

Brown SJ. 1976. *Optimal portfolio choice under uncertainty.* PhD thesis. Univ. Chicago

Brown SJ. 1978. The portfolio choice problem: comparison of certainty equivalence and optimal Bayes portfolios. *Commun. Stat. Simulat. Comput.* 7:321–34

Busse J, Irvine PJ. 2006. Bayesian alphas and mutual fund persistence. *J. Finance* 61:2251–88

Campbell JY. 1987. Stock returns and the term structure. *J. Financ. Econ.* 18:373–99

Campbell JY, Shiller RJ. 1988. *The dividend ratio model and small sample bias: a Monte Carlo study.* Tech. Work. Pap., NBER

Chib S, Greenberg E. 1996. Markov Chain Monte Carlo simulation methods in econometrics. *Econ. Theory* 12:409–31

Duffie D, Epstein LG. 1992. Asset pricing with Stochastic differential utility. *Rev. Financ. Stud.* 5:411–36

Epstein LG, Zin SE. 1989. Substitution, risk aversion, and the temporal behavior of consumption growth and asset returns: a theoretical framework. *Econometrica* 57:937–69

Fama EF, French KR. 1989. Business conditions and expected returns on stocks and bonds. *J. Financ. Econ.* 25:23–49

Fama EF, French KR. 1993. Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* 33:3–56

Ferson WE, Korajczyk RA. 1995. Do arbitrage pricing models explain the predictability of stock returns? *J. Bus.* 68:309–49

Foster FD, Smith T, Whaley RE. 1997. Assessing goodness-of-fit of asset pricing models: the distribution of the maximal R^2. *J. Finance* 52:591–607

Frost PA, Savarino JE. 1986. An empirical Bayes approach to efficient portfolio selection. *J. Financ. Quant. Anal.* 21:293–305

Fung W, Hsieh DA. 2004. Hedge fund benchmarks: a risk-based approach. *Financ. Anal. J.* 60:65–80

Geczy CC, Stambaugh RF, Levin D. 2005. *Investing in socially responsible mutual funds*. Work. Pap.,
   SSRN

Geweke J. 2001. A note on some limitations of CRRA utility. *Econ. Lett.* 71(3):341–45

Geweke J, Zhou G. 1996. Measuring the pricing error of the arbitrage pricing theory. *Rev. Financ.
   Stud.* 9:557–87

Grinold RC, Kahn RN. 1999. *Active Portfolio Management*. New York: McGraw-Hill

Grossman SJ, Shiller RJ. 1981. The determinants of the variability of stock market prices. *Am. Econ.
   Rev.* 71:222–27

Han Y. 2006. Asset allocation with a high dimensional latent factor stochastic volatility model.
   Rev. Financ. Stud. 19:237–71

Harvey CR, Liechty JC, Liechty MW, Müller P. 2004. *Portfolio selection with higher moments*. Work.
   Pap., Duke Univ.

Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications.
   Biometrika 57:97–109

He G, Litterman R. 1999. *The intuition behind Black-Litterman model portfolios*. Work. Pap., Goldman
   Sachs Quant. Resour. Group

Hong Y, Tu J, Zhou G. 2007. Asymmetries in stock returns: statistical tests and economic evaluation.
   Rev. Financ. Stud. 20:1547–81

Jobson JD, Korkie B, Ratti V. 1979. Improved estimation for Markowitz portfolios using James-Stein
   type estimators. *Proc. Am. Stat. Assoc. Bus. Econ. Stat. Sect.* 41:279–84

Jones C, Shanken J. 2005. Mutual fund performance and learning across funds. *J. Financ. Econ.*
   78:507–52

Jorion P. 1986. Bayes-Stein estimation for portfolio analysis. *J. Financ. Quant. Anal.* 21:279–92

Kacperczyk M. 2008. *Asset allocation under distribution uncertainty*. Work. Pap., Univ. Br.
   Columbia

Kan R, Zhou G. 2007. Optimal portfolio choice with parameter uncertainty. *J. Financ. Quant. Anal.*
   42:621–56

Kandel S, McCulloch R, Stambaugh RF. 1995. Bayesian inference and portfolio efficiency. *Rev.
   Financ. Stud.* 9(1):1–53

Kandel S, Stambaugh RF. 1996. On the predictability of stock returns: an asset-allocation perspective.
   J. Finance 51:385–424

Keim DB, Stambaugh RF. 1986. Predicting returns in bond and stock markets. *J. Financ. Econ.*
   12:357–90

Kirby C. 1998. The restrictions on predictability implied by rational asset pricing models. *Rev. Financ.
   Stud.* 11:343–82

Klein RW, Bawa VS. 1976. The effect of estimation risk on optimal portfolio choice. *J. Financ. Econ.*
   3:215–31

Klein RW, Brown SJ. 1984. Model selection when there is 'minimal' prior information. *Econometrica*
   52:1291–312

Lamoureux C, Zhou G. 1996. Temporary components of stock returns: What do the data tell us? *Rev.
   Financ. Stud.* 9:1033–59

Lehmann EL, Casella G. 1998. *Theory of Point Estimation*. New York: Springer-Verlag

Lettau M, Ludvigson S. 2001. Consumption, aggregate wealth, and expected stock returns. *J. Finance*
   56:815–49

Lintner J. 1965. Security prices, risk and maximal gains from diversification. *J. Finance* 20:587–615

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

Lucas RE Jr. 1978. Asset prices in an exchange economy. *Econometrica* 46:1429–45

MacKinlay AC, Pástor L. 2000. Asset pricing models: implications for expected returns and portfolio
   selection. *Rev. Financ. Stud.* 13:883–916

Markowitz HM. 1952. Mean-variance analysis in portfolio choice and capital markets. *J. Finance*
   7:77–91

McCulloch R, Rossi PE. 1990. Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory. *J. Financ. Econ.* 28(1–2):7–38

Merton R. 1969. Lifetime portfolio selection under uncertainty: the continuous time case. *Rev. Econ. Stat.* 51:247–57

Merton RC. 1973. An intertemporal capital asset pricing model. *Econometrica* 41:867–87

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–92

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

Nardari F, Scruggs J. 2007. Bayesian analysis of linear factor models with latent factors, multivariate stochastic volatility, and APT pricing restrictions. *J. Financ. Quant. Anal.* 42:857–92

Pástor L. 2000. Portfolio selection and asset pricing models. *J. Financ.* 55:179–223

Pástor L, Stambaugh RF. 2000. Comparing asset pricing models: an investment perspective. *J. Financ. Econ.* 56:335–81

Pástor L, Stambaugh RF. 2002a. Mutual fund performance and seemingly unrelated assets. *J. Financ. Econ.* 63:315–49

Pástor L, Stambaugh RF. 2002b. Investing in equity mutual funds. *J. Financ. Econ.* 63:351–80

Pástor L, Stambaugh RF. 2009a. Predictive systems: living with imperfect predictors. *J. Finance* 64:1583–628

Pástor L, Stambaugh RF. 2009b. *Are stocks really less volatile in the long run?* Work. Pap., Univ. Chicago/Univ. Penn.

Pástor L, Veronesi P. 2009. Learning in financial markets. *Annu. Rev. Financ. Econ.* 1:361–81

Ross S. 1976. The arbitrage theory of capital asset pricing. *J. Econ. Theory* 13:341–60

Rubinstein M. 1976. The valuation of uncertain income streams and the pricing of options. *Bell J. Econ.* 7:407–25

Samuelson PA. 1969. Lifetime portfolio selection by dynamic stochastic programming. *Rev. Econ. Stat.* 51:239–46

Sharpe WF. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *J. Finance* 19:425–42

Stambaugh RF. 1997. Analyzing investments whose histories differ in length. *J. Financ. Econ.* 45:285–331

Stambaugh RF. 1999. Predictive regressions. *J. Financ. Econ.* 54:375–421

Tu J. 2010. Is regime switching in stock returns important in portfolio decisions? *Manag. Sci.* 56:1198–1215

Tu J, Zhou G. 2004. Data-generating process uncertainty: What difference does it make in portfolio decisions? *J. Financ. Econ.* 72:385–421

Tu J, Zhou G. 2010. Incorporating economic objectives into Bayesian priors: portfolio choice under parameter uncertainty. *J. Financ. Quant. Anal.* doi:10.1017/S0022109010000335 (E-pub ahead of print)

Wachter JA, Warusawitharana M. 2009. Predictable returns and asset allocation: Should a skeptical investor time the market? *J. Econom.* 148:162–78

Winkler RL. 1973. Bayesian models for forecasting future security prices. *J. Financ. Quant. Anal.* 8:387–405

Winkler RL, Barry CB. 1975. A Bayesian model for portfolio selection and revision. *J. Finance* 30:179–92

Zellner A. 1971. *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley

Zellner A, Chetty VK. 1965. Prediction and decision problems in regression models from the Bayesian point of view. *J. Am. Stat. Assoc.* 60:608–16

Zhou G. 2009. Beyond Black-Litterman: letting the data speak. *J. Portfol. Manage.* 36:36–45

Zhou G, Zhu Y. 2009. *A long-run risks model with long- and short-run volatilities: explaining predictability and volatility risk premium.* Work. Pap., SSRN

Contents

Errata

An online log of corrections to *Annual Review of Financial Economics* articles
may be found at http://financial.annualreviews.org