THE JOURNAL OF FINANCE • VOL. LXI, NO. 5 • OCTOBER 2006

Bayesian Alphas and Mutual Fund Persistence

JEFFREY A. BUSSE and PAUL J. IRVINE*

ABSTRACT

We use daily returns to compare the performance predictability of Bayesian estimates of mutual fund performance with standard frequentist measures. When the returns on passive nonbenchmark assets are correlated with fund holdings, incorporating histories of these returns produces a performance measure that predicts future performance better than standard measures do. Bayesian alphas based on the Capital Asset Pricing Model (CAPM) are particularly useful for predicting future standard CAPM alphas. Over our sample period, priors consistent with moderate to diffuse beliefs in managerial skill dominate more skeptical prior beliefs, a result that is consistent with investor cash flows.

SINCE THE 1960S, THE ACADEMIC LITERATURE has analyzed mutual funds with an expanding array of performance measures. The majority of the measures are based on ordinary least squares (OLS) estimation of factor model regressions. In two early studies, Jensen (1968, 1969) uses the single-factor Capital Asset Pricing Model (CAPM). Subsequently, performance measures based on factor models have evolved along with the empirical asset pricing literature to incorporate passive assets such as size, book-to-market (Fama and French (1993)), and momentum (Jegadeesh and Titman (1993)). See, for example, Ippolito (1989), Elton et al. (1993), and Carhart (1997). Besides relying on least squares regression to estimate performance, studies typically use passive asset returns that are contemporaneous with fund returns. However, Pástor and Stambaugh (PS, 2002a) demonstrate that mutual fund performance measures need not be restricted to information on fund and passive assets over the life of the fund. Indeed, they find that incorporating a long time series of passive asset returns using Bayesian methods produces estimates of fund performance which are more precise.

The approach of PS (2002a) for measuring fund performance may be particularly well suited to examining another topic central to the fund literature,

*Busse is from the Goizueta Business School, Emory University, and Irvine is from the Terry College of Business, University of Georgia. We thank an anonymous referee, Stephen Brown, Richard Green (the editor), Lubos Pástor, John Scruggs, Jay Shanken, Chris Stivers, and seminar participants at University of Cincinnati, University of Colorado, Erasmus University, University of Georgia, Simon Fraser University, University of Texas at Dallas, Tilburg University, Vanderbilt University, Yale University, the 2003 American Finance Association meetings in Washington, DC, the Berkeley Program in Finance, and the CIRANO Fund Management Conference for useful comments. We thank Rob Stambaugh for preliminary discussions on this topic. Thanks to Ron Harris for research assistance.

namely, mutual fund performance persistence. Persistence studies examine whether funds sustain relative performance over adjacent periods. Using factor model regressions and passive asset returns that coincide with the fund returns, Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995), Malkiel (1995), Elton, Gruber, and Blake (1996), and Gruber (1996) all find evidence of mutual fund persistence. However, using a four-factor model, Carhart (1997) attributes persistence to fund expenses and momentum security holdings. More recently, Bollen and Busse (2005) find persistence beyond expenses or momentum across shorter, quarterly periods. The persistence of mutual fund performance therefore remains an open question. Since PS (2002a) show that their Bayesian, long-history factor approach estimates performance more precisely, incorporating their measures in an analysis of persistence could uncover evidence that is not apparent, given noisier standard measures.

We use daily fund returns and the framework of PS (2002a) to examine whether Bayesian alphas predict future performance better than standard performance measures. The intuition behind using the PS (2002a) approach to predict future fund performance is as follows. Suppose that a value fund outperformed a growth fund last year simply because the returns of value stocks exceeded the returns of growth stocks last year. To predict the value fund's future performance, the Bayesian approach estimates the abnormal performance of value stocks, relative to a pricing model, using data that include and precede the recent annual measurement period. In this example, the ability of the Bayesian measure to predict future performance more accurately than a standard performance measure depends on which returns better reflect future value stock returns, more recent value stock returns, or value stock returns over a longer time period. The Bayesian approach applies Stambaugh's (1997) insight that truncating a set of returns so that all return series have equal length is inefficient. Thus, in this example, the abnormal return of value stocks is estimated more precisely by using a long time series of historical value stock returns. We find that when the returns on passive nonbenchmark assets are correlated with fund holdings, incorporating histories of these returns in a Bayesian framework produces a performance measure that predicts future performance better than standard measures do.

The Bayesian alpha's advantage over standard measures stems in part from the additional data used in its estimation, including both long-history passive assets and also fund expenses. When we incorporate long-history passive assets in a frequentist measure, the predictability of the frequentist measure improves substantially.

We also use Bayesian measures to shed further light on the dynamics of mutual fund cash flows. An extensive literature examines the relationship between mutual fund performance and subsequent investor cash flows. Several earlier papers, including Ippolito (1989), Gruber (1996), Chevalier and Ellison (1997), and Sirri and Tufano (1998), find strong positive relationships between performance measures, including returns and single- and multifactor alphas, and subsequent cash flow. Given its sensitivity to a set of priors, the Bayesian alpha

is well suited to further study the behavior of investors with respect to fund performance. By varying the priors used to estimate the Bayesian alpha, we infer which set of prior beliefs best reflects investor behavior. Our evidence suggests that investors believe managers can earn abnormal returns: Investors choose funds with the highest estimates of past performance and do not invest simply in the lowest expense funds. Such behavior has been generally rewarded, with year-to-year risk-adjusted fund performance net of expenses persisting over our sample period. Frequentist measures cannot use fund cash flows to infer investor beliefs about managerial skill because frequentist measures are invariant to investor priors.

Finally, we use Bayesian measures to examine whether daily fund returns dominate monthly fund returns for estimating abnormal performance. We find that performance estimates based on daily fund and passive asset returns predict future performance substantially better than estimates based on monthly returns. With daily data, we can precisely estimate frequentist or Bayesian measures of abnormal performance over intervals of 1 year or less. Monthly estimates of alpha, on the other hand, require multiyear measurement intervals—a period of time over which managerial skill and fund factor loadings may not persist.

Although the Bayesian approach in our paper is most closely related to PS (2002a), it is also related to other papers in the growing literature that employs Bayesian techniques to examine mutual funds. Baks, Metrick, and Wachter (2001) find that particular priors can justify investment in active mutual funds, even for investors who have little belief in either managerial skill or the asset pricing model. PS (2002b) use Bayesian methods to form optimal portfolios of mutual funds, and determine that portfolio choice and subsequent performance are sensitive to the investor's prior beliefs in managerial skill and the asset pricing model. Baks (2006) uses a Bayesian approach to distinguish between performance attributable to the fund manager and performance attributable to the fund family. Jones and Shanken (2005) incorporate prior beliefs about the aggregate performance of mutual fund managers to estimate fund performance. Finally, Cohen, Coval, and Pástor (2005) infer the skill of unknown managers through commonality among their holdings and those of managers of known skill. They contrast their skill estimate with a standard estimate and a simple Bayesian estimate.

The article proceeds as follows. Section I outlines the computation of the performance measures used in the empirical analysis. Section II describes the data. Section III presents the empirical analysis. Section IV concludes.

I. Performance Measures

A. Standard Performance Measures

Fund abnormal performance is often measured by alpha, defined as the intercept in a regression of the fund's excess returns on the returns of one or more passive assets, that is,

$$r_{A,t} = \alpha_A + \beta_A' r_{B,t} + \varepsilon_{A,t}, \tag{1}$$

where $r_{A,t}$ is the time-t excess return of fund A, $r_{B,t}$ is the time-t excess return of the passive asset(s), and α_A is the fund's alpha.

The single-factor CAPM consists of one passive asset, the excess return on the market portfolio. However, the ability of the market portfolio to price all assets accurately has been called into question frequently over the last two decades. Passive portfolios composed of small stocks (Banz (1981)), high book-to-market stocks (Fama and French (1992)), and stocks with high past returns (Jegadeesh and Titman (1993)) produce positive returns after controlling for market risk. Consequently, a positive single-factor alpha could represent holdings in these passive assets (see Elton et al. (1993)), rather than skill in selecting specific securities. To address this issue, multifactor specifications include additional passive indices in $r_{B,t}$. Fama and French (1993) augment the market portfolio with size and book-to-market factors, and Carhart (1997) includes momentum as a fourth passive asset.

We refer to the one, three, and four passive assets in the CAPM, Fama–French, and Carhart four-factor models as benchmark assets. Although fund alpha is defined with respect to the benchmark assets, the fund may also be exposed to other nonbenchmark passive assets. For the small, book-to-market, and momentum passive assets, the benchmark or nonbenchmark classification depends on the pricing model. For example, for the CAPM (Fama–French) model, the size and book-to-market passive assets are nonbenchmark (benchmark) assets. We classify four additional passive assets as nonbenchmark assets in all three pricing models. The first captures the differential return between stocks with low betas with respect to a book-to-market factor (i.e., growth stocks) and stocks with high book-to-market betas, similar to the factor used by Daniel and Titman (1997). The remaining three passive assets capture the dynamics of industry-specific returns and are related to the industry portfolios of Moskowitz and Grinblatt (1999).

We estimate the alphas of the nonbenchmark assets by regressing them on the benchmark(s),

$$r_{N,t} = \alpha_N + \beta_N' r_{B,t} + \varepsilon_{N,t}, \tag{2}$$

where $r_{N,t}$ are the nonbenchmark assets, $r_{B,t}$ are the excess returns of the benchmark asset(s), and α_N are the nonbenchmark alphas. The set of nonbenchmark assets varies with the benchmark model. For example, for the CAPM (Fama–French) model, there are seven (five) nonbenchmark assets, consisting of all passive assets that are not benchmark assets.

Regressing excess fund returns on both the benchmark and nonbenchmark assets gives

$$r_{A,t} = \delta_A + c_{A,N}' r_{N,t} + c_{A,B}' r_{B,t} + u_{A,t}, \tag{3}$$

where we refer to the intercept in equation (3), δ_A, as fund A's stock selection skill, and $c_{A,N}$ are the fund's exposures to the nonbenchmark assets. If we

substitute equation (2) into equation (3) and compare the result to equation (1), we find that

$$\alpha_A = \delta_A + c'_{A,N}\alpha_N. \tag{4}$$

For a given benchmark model, equation (4) decomposes the fund's alpha into stock selectivity (δ_A) and fund style ($c'_{A,N}\alpha_N$), the incremental abnormal return attributable to fund exposure to the nonbenchmark assets.

Mutual fund studies typically estimate the standard performance measures of equations (1) to (4) directly from the time series of returns for the fund and for the benchmark and nonbenchmark assets, all over the same time period. Stambaugh (1997), however, documents the advantages of using prior-period passive asset returns, finding that long-horizon returns provide more precise estimates of the moments of correlated short-horizon returns. In our context, the return history of mutual funds is relatively short. According to Morningstar's *Principia Pro*, for example, the median age of U.S. domestic equity funds is 4.8 years as of January 2005. Conversely, we can estimate nonbenchmark alphas, α_N, using data as far back as the 1960s, the starting point for book values on Compustat. Using the long-period estimates of the nonbenchmark α_N in equation (4) should produce more accurate predictions of future performance associated with fund style.

Consequently, in addition to the frequentist alphas calculated from passive asset returns that are contemporaneous to the mutual fund returns (which we refer to as *standard* measures), we also estimate frequentist measures that use long-history passive assets. For the long-history frequentist measures, we estimate equation (2) using the entire time history of benchmark and nonbenchmark asset returns, which begin two decades before our fund returns. We estimate equation (3) as before, using the passive asset returns that coincide with the fund returns. We then use equation (4) to combine the nonbenchmark alphas in equation (2) with the stock selectivity and nonbenchmark asset loadings in equation (3).

B. Bayesian Performance Measures

The Bayesian alpha, introduced by PS (2002a), provides an alternative to the conventional measures in equations (1) to (4). Similar to the long-history frequentist measures described above, the Bayesian measure uses passive asset returns from periods of time that precede the mutual fund returns. In addition, the Bayesian measure incorporates a flexible set of prior beliefs about both managerial skill and the validity of an asset pricing model.

We use the techniques described in PS (2002a, 2002b) to calculate Bayesian alphas. For consistency, we use their notation. We also use the same set of benchmark and nonbenchmark assets introduced in Section I.A. To estimate the posterior distributions of the elements of equation (4), we first specify a conditional prior distribution of α_N, the nonbenchmark abnormal return. Conditional on Σ, the covariance matrix for $\varepsilon_{N,t}$, we specify the prior distribution for α_N as

$$\alpha_N \mid \Sigma \sim N\left(0, \sigma^2_{\alpha_N}\left(\frac{1}{s^2}\Sigma\right)\right). \tag{5}$$

We use $\sigma^2_{\alpha_N}$, the marginal prior variance of each element in α_N, to adjust the prior belief in the ability of the benchmark(s) to explain the returns on the nonbenchmark assets. A prior of $\sigma^2_{\alpha_N} = 0$ is equivalent to setting $\alpha_N = 0$, corresponding to perfect confidence in the ability of the benchmark assets to price the returns of the nonbenchmark assets. Alternatively, $\sigma^2_{\alpha_N} = \infty$ corresponds to a diffuse prior for α_N, which allows the data to dominate the posterior estimate. Prior variances between these two extremes signify that the prior belief in the ability of the benchmark assets to price the nonbenchmark assets is centered at $\alpha_N = 0$, but some belief in model mispricing exists. We use an empirical Bayes approach to set s^2 in equation (5) equal to the average of the diagonal elements of Σ from OLS estimation of equation (2) for each nonbenchmark asset.

We construct the priors for the estimation of the skill and the benchmark and nonbenchmark asset loadings in equation (3) conditional on σ^2_u, the variance of $u_{A,t}$. Conditional on σ^2_u, we specify the priors for δ_A and $c_A = (c'_{AN}\ \ c'_{AB})'$ as independent normal distributions:

$$\delta_A \,|\, \sigma^2_u \sim N\left(\delta_0, \left(\frac{\sigma^2_u}{E\left(\sigma^2_u\right)}\right)\sigma^2_\delta\right) \tag{6}$$

and

$$c_A \,|\, \sigma^2_u \sim N\left(c_0, \left(\frac{\sigma^2_u}{E\left(\sigma^2_u\right)}\right)\Phi_c\right). \tag{7}$$

Following PS (2002b), we set σ^2_δ, the marginal prior variance of δ_A, to finite values and δ_0 equal to -1 multiplied by the fund's annual expense ratio divided by the number of return data points per year. Intuitively, we center the priors at a belief that managers possess no skill and that investing in mutual funds will underperform the benchmark and nonbenchmark assets by the fund's expenses. Setting $\sigma^2_\delta = 0$ implies no belief in managerial skill and is equivalent to setting $\delta_A = \delta_0$. Conversely, a diffuse skill prior ($\sigma^2_\delta = \infty$) allows the data to dominate the posterior skill estimate. Finite prior variances between these extremes allow for varying degrees of belief in the possibility of managerial skill.

The prior variance scale factor, $\sigma^2_u / E(\sigma^2_u)$, in equations (6) and (7) controls for fund idiosyncratic risk. Idiosyncratic risk affects the degree to which the skill and factor loading posteriors deviate from their priors. Following an empirical Bayes approach, we set $E(\sigma^2_u)$ equal to the cross-sectional mean of $\hat{\sigma}^2_u$ from OLS regressions of equation (3) for each fund type, sorted by investment objective. We also follow an empirical Bayes approach to set the parameters c_0 and Φ_c equal to the OLS estimate of the sample cross-sectional moments of $\hat{c}_A$, again estimated separately for each investment objective. Factor loading priors thus reflect the average factor loading of the fund type, c_0, the cross-sectional

deviation of loadings across funds of that type, Φ_c, and the idiosyncratic risk of each particular fund.

We combine the priors specified in equations (5)–(7) with fund, benchmark, and nonbenchmark returns as in PS (2002a) to produce estimates of the posterior distributions of the elements of equation (4). See Appendix for details. Designating the posterior means by tilde, we have

$$\tilde{\alpha}_A = \tilde{\delta}_A + \tilde{c}_{A,N}\tilde{\alpha}_N. \tag{8}$$

We use the posterior mean estimates in equation (8) as the Bayesian estimates of alpha, $\tilde{\alpha}_A$, managerial skill, $\tilde{\delta}_A$, and fund style, $\tilde{c}_{A,N}\tilde{\alpha}_N$. The Bayesian skill estimate, $\tilde{\delta}_A$, more closely approximates the intercept in a regression of excess fund returns on all of the benchmark and nonbenchmark assets as the priors on managerial skill become more diffuse (i.e., as σ_δ^2 grows large). Similarly, the Bayesian alpha, $\tilde{\alpha}_A$, more closely approximates the intercept in a regression of excess fund returns on the benchmark assets (i.e., the standard alpha) as both priors become more diffuse, corresponding to large $\sigma_{\alpha N}^2$ and large σ_δ^2. With diffuse prior beliefs, the Bayesian estimate differs from the standard alpha because we estimate the $\tilde{\alpha}_N$ from passive asset returns that include and precede the standard alpha measurement period.

C. Time-Varying Factor Loadings

Sunder (1980) and Ohlson and Rosenberg (1982) find significant time variation in CAPM factor loading estimates. Recently, Mamaysky, Spiegel, and Zhang (2003) incorporate time-varying factor loadings in the estimation of mutual fund alphas and betas. They note that even if security returns follow a factor model with constant loadings, fund factor loadings will time-vary when their portfolios have a time-varying mix of securities.

Therefore, we also estimate performance in a time-varying parameter model, first in a frequentist setting and then adapted to a Bayesian framework, allowing for flexible prior beliefs. We specify a parsimonious model of parameter variation in which factor coefficients, $\beta_{A,t} = (\delta_{A,t} \;\; c_{A,N,t} \;\; c_{A,B,t})$, follow a random walk:

$$\beta_{A,t} = \beta_{A,t-1} + \eta_{A,t}. \tag{9}$$

Adapting equation (3) to the time-varying parameter model yields

$$r_{A,t} = \delta_{A,t} + c_{A,N,t}r_{N,t} + c_{A,B,t}r_{B,t}, \quad t = 1, \ldots, T, \tag{10}$$

where $\delta_{A,t}, c_{A,N,t}$, and $c_{A,B,t}$ vary over time. Imposing the structure of equation (9) on the coefficients' time-series processes allows for estimation of the coefficients even when there is only one observation per fund-period.[1] We

[1] We follow Sunder (1980) and model the time-varying factor loadings as a random walk. We rely on Ohlson and Rosenberg (1982), who examine various alternative time-varying specifications for the single-factor model. They conclude that the most crucial improvement in the model's goodness of fit is obtained by allowing for variation in beta, whether random or sequential.

estimate our time-varying frequentist model parameters with generalized least squares, adapted for time variation in the parameters. Gibbs sampling techniques yield estimates of the time-varying parameters in the Bayesian model.[2] To test predictability, we use the time-T coefficient estimates to compute time-varying parameter model alphas:

$$\alpha_{A,T} = \delta_{A,T} + c_{A,N,T}\alpha_N. \tag{11}$$

These alphas differ from equation (4) because they incorporate the most recent estimate of the time-series coefficients for skill and nonbenchmark factor loadings. If the time-varying parameter alphas capture fund dynamics, then they could potentially improve predictability relative to constant-parameter model alphas.

II. Data

The mutual fund sample, taken from Busse (1999), consists of daily returns on 230 open-end equity funds. Nonspecialized domestic funds are included, the names of which are taken from the December 1984 edition of Weisenberger's Mutual Funds *Panorama*. To be included in the sample, funds must have at least $15 million in total net assets (as of December 1984) as well as daily net asset values (NAVs) and dividends available from Interactive Data Corporation. The daily NAVs and dividends are combined to form daily total returns. The sample is divided into three investment objectives: maximum capital gains (68 funds), growth (107 funds), and growth and income (55 funds). See Busse (1999) for further details.

Compared to the mutual fund sample available from the Center for Research in Security Prices (CRSP) Survivor-Bias Free U.S. Mutual Fund Database, our sample size is small. The advantage of our sample is that it consists of daily returns, whereas CRSP's mutual fund returns are predominantly monthly.[3] Daily returns are important in the context of analyzing fund persistence because they allow for pricing model estimation over short measurement intervals. For instance, Bollen and Busse (2005) find that persistence attributable to skill disappears for 3-year ranking intervals. With monthly returns, it would not be feasible to estimate alpha over a short measurement interval.

The prior for δ_A in equation (6) incorporates fund expenses. We take annual fund expenses from the 1985 to 1995 editions of Weisenberger. We use quarterly total net assets from Standard & Poor's Micropal to estimate quarterly normalized net cash flow for each fund. We define normalized cash flow as

[2] An appendix detailing the specifics of the time-varying frequentist and Bayesian estimation is available from the authors upon request.

[3] Suppliers such as Standard & Poor's Micropal have daily data on a wider cross-section of mutual funds. However, such data are prone to considerable error, including incorrect dividends, ex-dividend dates, and NAVs. The sample in this study is corrected for most of these errors. See Busse (1999).

$$\mathrm{NCF}_t = \frac{\mathrm{TNA}_t - \mathrm{TNA}_{t-1}(1 + R_t)}{\mathrm{TNA}_{t-1}}, \tag{12}$$

where TNA_t are the fund's total net assets at the end of quarter t, and R_t is the fund's total return during quarter t.

Our models include eight passive assets, including the market portfolio, size, book-to-market, momentum, a stock characteristic-balanced measure, and three industry factors. The factors are daily in frequency. The CRSP NYSE/AMEX/Nasdaq value-weighted market return (MKT) proxies for the market portfolio. For size, we use the Fama and French (1993) value-weighted small market capitalization minus big market capitalization (SMB) factor. For the book-to-market factor, we use the Fama–French value-weighted high book-to-market minus low book-to-market (HML) factor. We take the SMB and HML factors from Ken French's web site. For momentum, we use a value-weighted version of the 1-year high return minus low return (up minus down, UMD) factor. Carhart (1997) uses a variation of this momentum factor at a monthly frequency. We construct our daily series UMD factor following the monthly procedure described on Ken French's web site, except with daily returns.

We construct the characteristic-balanced measure, denoted CMS, as described in Daniel and Titman (1997), Davis, Fama, and French (2000), and PS (2000, 2002a), except at a daily frequency.[4] The factor captures the return difference between stocks with low HML betas and stocks with high HML betas in a multiple regression that also includes MKT and SMB as independent variables.

We use the 20 industry groups described in Moskowitz and Grinblatt (1999) to construct three industry factors, IND1–3, at a daily frequency. We first compute the cross-product matrix from the residuals obtained by regressing the excess returns of the 20 industry portfolios on the other benchmark and nonbenchmark assets. We extract the three eigenvectors corresponding to the three largest eigenvalues from the cross-product matrix to form the principal components matrix. To form the three industry factors, IND1–3, we then normalize the rows of the principal components matrix to a mean square of one. See Connor and Korajczyk (1986).[5]

To compute daily excess returns on the funds and on the CRSP market return, we use the CRSP monthly *T30RET* 30-day T-bill return divided equally over the trading days in the month as the risk-free rate. We use a common starting date of July 1, 1968 for all factors.[6]

[4] The procedures used to create this factor vary slightly across Daniel and Titman (1997), Davis, Fama, and French (2000), and PS (2000, 2002a). We specifically follow PS (2002a).

[5] Connor and Korajczyk (1986) show that, as the number of assets increases, the normalized rows of the principal components matrix converge to the factor realizations.

[6] July 1, 1968 is the earliest date available for the daily CMS factor. CMS uses HML loadings resulting from a 60-month regression of excess stock returns on MKT and fixed-weight SMB and HML factors (see Daniel and Titman (1997)). The fixed-weight HML factor begins in July 1963, 6 months after the earliest book values on Compustat.

III. Empirical Analysis

A. Methodology

The paper's main empirical procedure sorts funds into deciles based on the mean of the Bayesian posterior alpha distribution or frequentist alpha estimated during a ranking period, and then examines standard alphas of the deciles during the following post-ranking period. Two main considerations determine the length of the ranking period. First, since the portfolios of actively managed funds evolve over time, a short measurement interval should lead to more accurate factor loading estimates. Second, since the precision of the factor loading estimates increases as the number of fund returns used in their estimation increases, a long measurement interval should lead to greater precision if the funds' portfolios do not change substantially over time. Also, after controlling for size, book-to-market, and momentum, Carhart (1997) finds no evidence of performance persistence unrelated to fees over 3-year measurement intervals, whereas Bollen and Busse (2005) find persistence in four-factor alphas over quarterly periods. After considering these issues, we use an annual ranking period. Later, we compare our annual ranking-period results to shorter quarterly period results and longer 3-year results.

For the post-ranking period, we use a quarterly interval because Bollen and Busse (2005) find that alpha persistence deteriorates over post-ranking periods greater than a quarter. With a quarterly post-ranking period, we estimate Bayesian alphas over overlapping annual time periods that begin each quarter during the 1985-to-1995 sample period, and we then examine the nonoverlapping quarterly post-ranking performance of the funds.

Over each annual ranking period, we estimate Bayesian alphas using 11 different skill prior variances ranging from 10^{-13} to 10^{-3} and eight model mispricing prior variances ranging from 10^{-11} to 10^{-4} for each benchmark model (CAPM, Fama–French, and four-factor). We find that these variances span the set of prior beliefs that give different posterior mean estimates of the Bayesian alphas. Empirically, $\sigma_\delta^2 \leq 10^{-13}$ ($\sigma_{\alpha N}^2 \leq 10^{-11}$) produces posterior mean estimates equivalent to $\sigma_\delta^2 = 0$ ($\sigma_{\alpha N}^2 = 0$), and $\sigma_\delta^2 \geq 10^{-3}$ ($\sigma_{\alpha N}^2 \geq 10^{-4}$) produces estimates equivalent to $\sigma_\delta^2 = \infty$ ($\sigma_{\alpha N}^2 = \infty$). For each ranking period, we estimate the Bayesian alphas using the daily nonbenchmark asset data from July 1968 through the end of that ranking period. To the extent that a longer history of nonbenchmark assets produces more precise estimates of nonbenchmark abnormal returns (Stambaugh (1997)), predictive ability is likely to increase over the sample for a given fund as more and more nonbenchmark asset data become available. The post-ranking-period standard alphas of the deciles indicate the extent to which the Bayesian alphas predict future performance. By sorting on the mean of the Bayesian posterior alpha distribution and examining the subsequent standard alpha rather than the subsequent Bayesian alpha, we avoid induced correlation between the ranking- and post-ranking-period measures that would arise from using some of the same historical passive asset returns in both Bayesian measures. We use the same benchmark model in both

the ranking-period Bayesian measure and the post-ranking-period standard measure.

In addition to estimating the Bayesian alphas during the annual ranking periods, we also examine two different frequentist measures for each model, namely, the standard measure that uses only benchmark assets over the same time period as the fund returns, and the long-history measure described earlier. The long-history measure uses the same set of nonbenchmark assets as the Bayesian measure. Finally, in one version of our empirical analysis, we incorporate time-varying parameters into our Bayesian and frequentist performance measures. For the Bayesian (frequentist) sorts, we weight the standard alphas in the post-ranking-period deciles either equally or by the inverse of the posterior variance of the Bayesian alpha (variance of the frequentist alpha) from the ranking period.[7]

In addition to examining the post-ranking-period alphas of the Bayesian alpha-sorted deciles, we also examine the post-ranking-period cash flows of the Bayesian-sorted deciles. Cash flows indicate the extent to which investors use the same type of information incorporated in the Bayesian alphas to determine where to invest.

We evaluate the relationship between the ranking-period Bayesian or frequentist alpha and post-ranking-period standard alpha or cash flow via three measures. To examine whether the relationship is statistically significant, we compute Spearman rank correlation coefficients between the ranking-period decile ranking and the post-ranking-period alpha or cash flow decile ranking. To examine economic significance, we compute the difference in post-ranking-period alpha or cash flow between the top and bottom deciles (d1−d10). For certain analyses, we also report the post-ranking-period alpha for the top decile (d1).

B. Main Results

B.1. Bayesian Alpha Predictability

Table I and Figure 1 show statistics that assess the relationship between the mean of the Bayesian posterior alpha distribution and subsequent standard alpha. The table reports average Spearman rank, d1−d10, and d1 statistics for several skill prior variances, σ_δ^2, and model mispricing prior variances, $\sigma_{\alpha N}^2$. The figure shows d1−d10 statistics for each prior variance combination. In both the table and the figure, Panel A uses the CAPM single-factor model, Panel B uses the Fama–French three-factor model, and Panel C uses the Carhart

[7] We estimate frequentist alpha variance as

$$\hat{\sigma}_A^2 = \hat{\sigma}_{\delta_A}^2 + \hat{\sigma}_{\hat{c}'_{A,N}\alpha_N}^2 + 2\text{cov}(\delta_A, \hat{c}'_{A,N}\alpha_N),$$

where we compute $\hat{\sigma}_{\delta_A}$, $\hat{\sigma}_{\hat{c}'_{A,N}\alpha_N}$, and $\text{cov}(\delta_A, \hat{c}'_{A,N}\alpha_N)$ from 1,000 iterations of a bootstrap simulation. In the bootstrap simulation, we randomly order, with replacement, days in the fund and long-history ranking periods and then use OLS to estimate equations (2) and (3) on the reordered data.

Table I

Performance Predictability of Bayesian versus Frequentist Alphas, Including Nonbenchmarks

The table reports statistics that describe the extent to which Bayesian and frequentist alphas predict subsequent standard alphas. In Panels A1, A2, B1, B2, C1, and C2, for each combination of skill prior variance and model mispricing prior variance, we sort funds into deciles based on their Bayesian alpha posterior mean during an annual ranking period. In Panels A3, B3, and C3, we sort funds into deciles based on their frequentist alpha during an annual ranking period. For each decile, we compute the standard alpha over the following quarterly post-ranking period. We estimate the ranking-period Bayesian alphas and long-history frequentist alphas using one benchmark (MKT) and seven nonbenchmarks (SMB, HML, UMD, CMS, and IND1–3) in Panel A, using three benchmarks (MKT, SMB, and HML) and five nonbenchmarks (UMD, CMS, and IND1–3) in Panel B, and using four benchmarks (MKT, SMB, HML, and UMD) and four nonbenchmarks (CMS and IND1–3) in Panel C. We estimate the ranking-period standard alphas and all post-ranking-period standard alphas using the same set of benchmarks that we use in the Bayesian and long-history frequentist alpha estimation, but no nonbenchmarks. We estimate all alphas using daily returns. The Spearman correlation measures the relationship between the ranking-period Bayesian alpha posterior mean or frequentist alpha decile ranking and the post-ranking-period standard alpha decile ranking. d1−d10 is the difference between the post-ranking-period standard alpha (daily percentage) for the top and bottom ranking-period deciles, and d1 is the post-ranking-period standard alpha (daily percentage) for the top ranking-period decile. In Panels A1, A2, B1, B2, C1, and C2 (Panels A3, B3, and C3), we weight the standard alphas in the post-ranking-period deciles either equally or by the inverse of the posterior variance of the Bayesian alpha (variance of the frequentist alpha) from the ranking period ("Precision weighted"). Each statistic in Panels A1, A2, B1, B2, C1, and C2 further represents the mean associated with that parameter (skill prior variance or model mispricing prior variance) over all combinations of the other parameter. A two-tailed Spearman test based on decile rankings is significant at the 10%, 5%, and 1% levels for correlations of 0.564, 0.648, and 0.794, respectively. The sample consists of 230 mutual funds. The mutual fund sample period is from January 2, 1985 to December 29, 1995. The factor sample period is from July 1, 1968 to December 29, 1995.

Prior or Data Series Length	Equal Weighted			Precision Weighted		
	Spearman	d1−d10	d1	Spearman	d1−d10	d1
Panel A: CAPM						
A1. Bayesian by skill prior variance						
1E−13	0.715	0.0042	−0.0014	0.894	0.0065	0.0006
1E−10	0.712	0.0043	−0.0015	0.868	0.0065	0.0005
1E−8	0.879	0.0097	0.0015	0.933	0.0119	0.0040
1E−6	0.982	0.0141	0.0045	0.974	0.0150	0.0069
1E−3	0.961	0.0144	0.0047	0.965	0.0153	0.0071
A2. Bayesian by model mispricing prior variance						
1E−11	0.874	0.0118	0.0027	0.892	0.0113	0.0036
1E−10	0.891	0.0120	0.0026	0.925	0.0117	0.0036
1E−8	0.856	0.0086	0.0015	0.914	0.0109	0.0043
1E−6	0.804	0.0077	0.0009	0.932	0.0101	0.0035
1E−4	0.814	0.0078	0.0009	0.928	0.0100	0.0035
A3. Frequentist						
Standard	0.867	0.0076	0.0008	0.879	0.0084	0.0027
Long-history	0.927	0.0142	0.0061	0.915	0.0162	0.0095

(*continued*)

Table I—*Continued*

Prior or Data Series Length	Equal Weighted			Precision Weighted		
	Spearman	d1−d10	d1	Spearman	d1−d10	d1
Panel B: Fama–French						
B1. Bayesian by skill prior variance						
1E−13	0.794	0.0136	0.0069	0.870	0.0126	0.0058
1E−10	0.792	0.0138	0.0070	0.850	0.0127	0.0058
1E−8	0.977	0.0196	0.0100	0.983	0.0188	0.0099
1E−6	0.897	0.0224	0.0107	0.959	0.0217	0.0111
1E−3	0.930	0.0220	0.0108	0.958	0.0215	0.0112
B2. Bayesian by model mispricing prior variance						
1E−11	0.689	0.0131	0.0052	0.810	0.0128	0.0052
1E−10	0.716	0.0132	0.0050	0.845	0.0133	0.0054
1E−8	0.910	0.0201	0.0110	0.955	0.0191	0.0104
1E−6	0.949	0.0207	0.0110	0.960	0.0196	0.0104
1E−4	0.947	0.0207	0.0110	0.964	0.0197	0.0105
B3. Frequentist						
Standard	0.952	0.0213	0.0109	0.988	0.0226	0.0125
Long-history	0.988	0.0211	0.0105	0.939	0.0208	0.0107
Panel C: Four-Factor						
C1. Bayesian by skill prior variance						
1E−13	0.745	0.0068	−0.0011	0.888	0.0092	0.0003
1E−10	0.703	0.0075	−0.0009	0.865	0.0097	0.0005
1E−8	0.952	0.0208	0.0046	0.965	0.0201	0.0053
1E−6	0.865	0.0222	0.0058	0.964	0.0222	0.0073
1E−3	0.971	0.0219	0.0057	0.961	0.0221	0.0072
C2. Bayesian by model mispricing prior variance						
1E−11	0.904	0.0158	0.0030	0.939	0.0153	0.0038
1E−10	0.918	0.0158	0.0028	0.967	0.0154	0.0038
1E−8	0.893	0.0153	0.0027	0.923	0.0167	0.0042
1E−6	0.833	0.0150	0.0024	0.905	0.0170	0.0041
1E−4	0.823	0.0150	0.0024	0.903	0.0170	0.0041
C3. Frequentist						
Standard	0.879	0.0226	0.0058	0.988	0.0223	0.0075
Long-history	0.976	0.0222	0.0065	0.988	0.0230	0.0078

four-factor model. Skill prior variance ranges from 10^{-13}, which essentially precludes the possibility that managerial skill exists, to the relatively diffuse 10^{-3}. Model mispricing prior variance, which reflects uncertainty over the model's ability to price the nonbenchmark assets, varies from a high level of confidence in the pricing model, 10^{-11}, to a relatively diffuse 10^{-4}.

The table shows results from equally weighting the alphas in post-ranking-period deciles, as well as results from weighting the alphas by the inverse of their posterior variance (labeled *precision weighted*). The figure only shows

equally weighted results. The table also shows the relationships between frequentist alphas—including measures that use and do not use long-history passive assets—and subsequent standard alphas.

The equal-weighted results in Panels A1, A2, B1, B2, C1, and C2 of Table I show that the Spearman rank correlation between the ranking-period Bayesian alpha mean decile and the subsequent standard alpha decile rank is statistically significant in each category for all three pricing models. This result

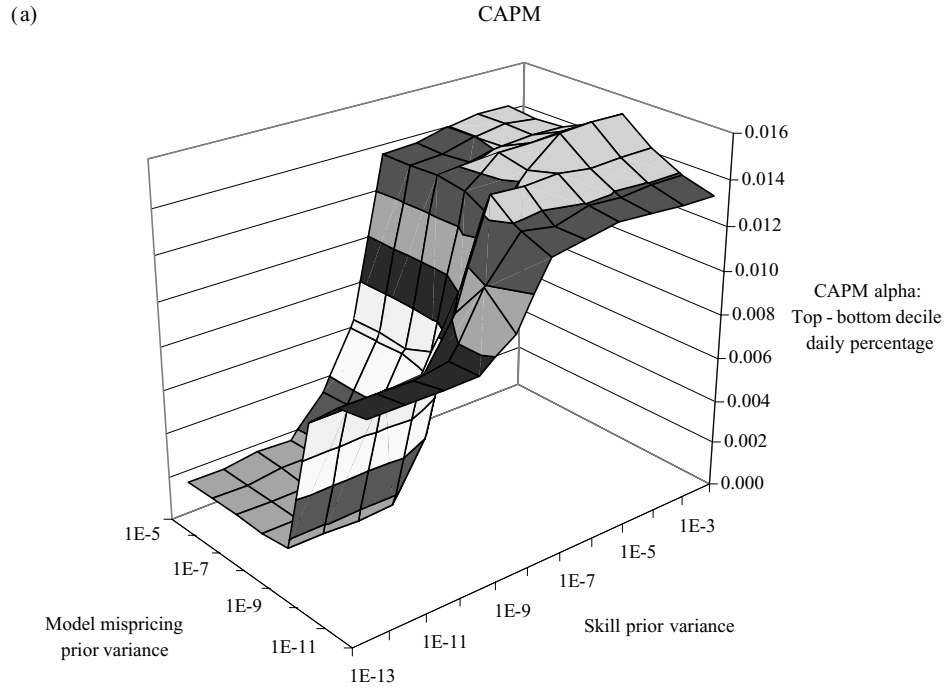
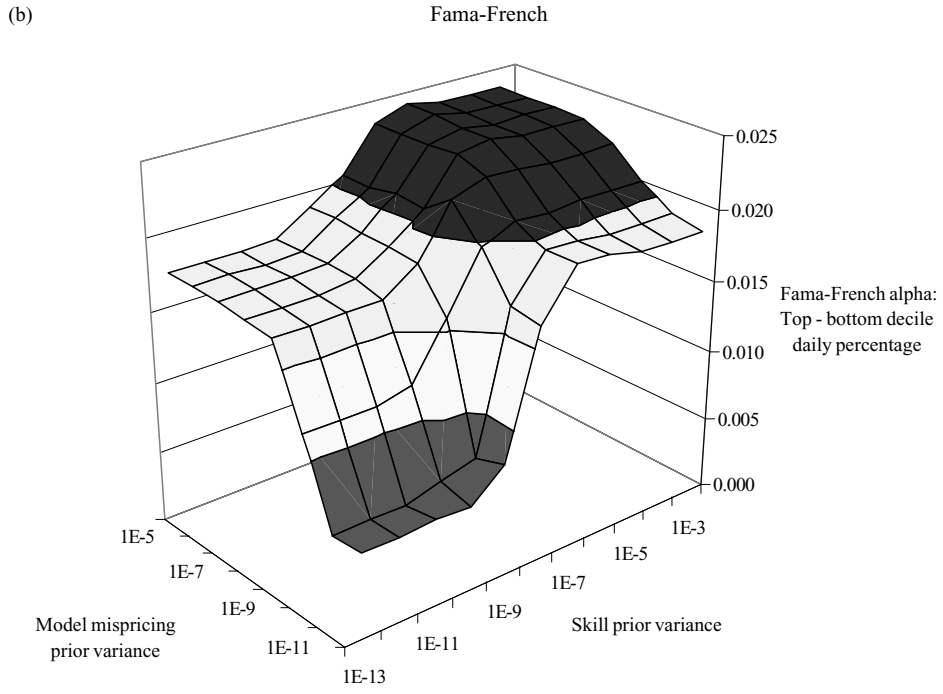


Figure 1. Performance predictability of Bayesian alphas with nonbenchmarks. The figure shows statistics that describe the extent to which Bayesian alphas predict subsequent standard alphas. For each combination of skill prior variance and model mispricing prior variance, we sort funds into deciles based on their Bayesian alpha posterior mean during an annual ranking period. For each decile, we compute the standard alpha over the following quarterly post-ranking period. The vertical axis in the figure shows the difference between the post-ranking-period standard alpha (daily percentage) for the top and bottom ranking-period deciles. We estimate the ranking-period Bayesian alphas using one benchmark (MKT) and seven nonbenchmarks (SMB, HML, UMD, CMS, and IND1–3) in Panel A, using three benchmarks (MKT, SMB, and HML) and five nonbenchmarks (UMD, CMS, and IND1–3) in Panel B, and using four benchmarks (MKT, SMB, HML, and UMD) and four nonbenchmarks (CMS and IND1–3) in Panel C. We estimate the post-ranking-period standard alphas using the same set of benchmarks that we use in the Bayesian alpha estimation, but no nonbenchmarks. We estimate the alphas using daily returns. We weight the standard alphas in the post-ranking-period deciles equally. The sample consists of 230 mutual funds. The mutual fund sample period is from January 2, 1985 to December 29, 1995. The factor sample period is from July 1, 1968 to December 29, 1995.



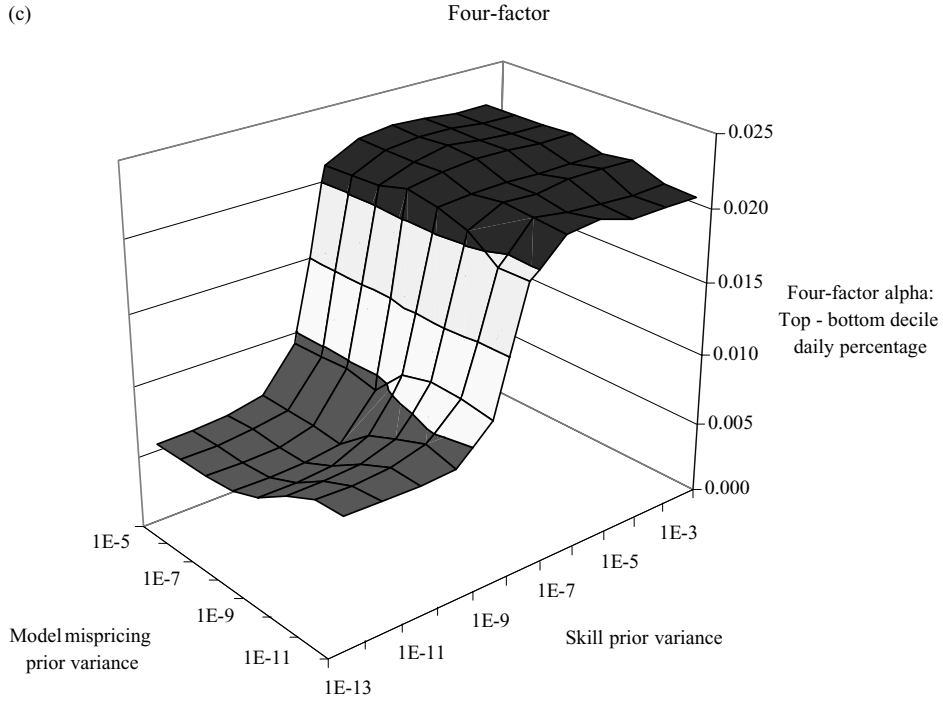


Figure 1—*Continued*

indicates a strong relationship between Bayesian alpha and subsequent risk-adjusted performance. The difference in average post-ranking-period standard alpha for funds in the highest Bayesian alpha mean decile and funds in the lowest Bayesian alpha mean decile (d1−d10) leads to a similar inference. Predictability is greater for the Fama–French and four-factor models than for the CAPM model. Multifactor models estimate managerial stock selection skill more accurately than the single-factor CAPM, since they remove more of the abnormal performance associated with passive assets. Thus, the greater predictability in the multifactor models might suggest that managerial stock selection skill persists more than the abnormal return associated with passive assets—a possibility we examine more closely later. For the CAPM and four-factor models, the precision-weighted results in Table I show greater predictability, especially for precise skill priors and diffuse model mispricing priors. For the Fama–French model, the precision-weighted results are mixed, with greater evidence of predictability with the Spearman rank correlations, but less with the d1−d10 differences.

Table I and Figure 1 indicate that, on average, Bayesian alphas more accurately predict future standard alphas for more diffuse skill priors. This result holds for all three models. Recall that as the skill prior becomes more diffuse, the Bayesian measure of managerial skill moves closer to a multifactor alpha (estimated using both benchmark and nonbenchmark assets). As the skill prior variance decreases, the Bayesian measure of managerial skill shrinks toward the prior mean of −1 multiplied by the fund's expense ratio. The greater predictability associated with more diffuse skill priors suggests that some amount of managerial skill exists among mutual fund managers and that it persists over time. The implication is that investing in funds with higher estimates of managerial skill produces higher future risk-adjusted performance than investing in funds with the lowest expense ratios (i.e., investing based on the precise prior that managers have no skill). Although predictability is greater for diffuse priors than for precise priors, the Spearman results indicate that predictability shows a slight improvement at moderate skill prior variances ranging from 10^{-6} to 10^{-8} in the Fama–French and four-factor models. This result suggests that investors should not completely disregard fund expenses.

The relationship between mispricing uncertainty and performance predictability depends on the pricing model. The CAPM results in Panel A of Table I and Figure 1 indicate that Bayesian alphas more accurately predict future standard alphas for precise model mispricing priors, that is, stronger priors that the nonbenchmark assets do not produce abnormal returns in a CAPM context. The Fama–French results in Panel B, however, indicate that Bayesian alphas better predict future standard alphas for more diffuse model mispricing priors. For the four-factor model in Panel C, the relationship between mispricing prior and performance predictability is mixed. The equal-weighted results in Table I and Figure 1 show slightly more predictability for precise model mispricing priors, and the precision-weighted results in Table I show slightly more predictability for more diffuse model mispricing priors.

The differences in the relationship between mispricing prior and predictability across the three models relate to differences in some of the factor risk premia

before and during the sample period. SMB, HML, and UMD have average daily returns of 0.0067%, 0.0293%, and 0.0409%, respectively, from July 1968 to December 1984; and −0.0121%, 0.0100%, and 0.0395%, respectively, from January 1985 to December 1995. Consequently, the average returns to the SMB and HML factors before the sample period differ considerably from their average returns during the sample period. UMD, however, shows large positive returns before and during the sample period. For the CAPM model, a diffuse prior allows for positive exposure to all the nonbenchmark assets—including SMB, HML, and UMD—to enhance a fund's Bayesian alpha, even though SMB and HML subsequently do not deliver returns consistent with their long-history means. For the Fama–French model, a diffuse prior allows for positive exposure to five nonbenchmark assets, including UMD, to enhance a fund's Bayesian alpha ranking, and consistently high returns to momentum support such a ranking. In the four-factor model, the nonbenchmark assets—CMS and IND1–3—generate relatively small average returns before and during the sample period.[8] Also, funds have relatively small loadings on CMS and IND1–3 during the sample period. CMS and IND1–3 thus play a small role in determining a fund's Bayesian alpha, and the mispricing prior has less impact.

The d1 results indicate that the precision-weighted post-ranking-period top decile alphas are greater than the equal-weighted top decile alphas. Averaged across all prior variance combinations, the annualized difference (precision-weighted vs. equal-weighted) is 0.35% per year. This result has one particularly important implication. The performance persistence literature generally focuses on two features in assessing persistence. First, post-period ranking tests determine whether relative performance is consistent across periods; the Spearman and d1−d10 statistics emphasize this specific type of test. Second, we look for significant positive performance in the top decile. The reason to check for positive top decile performance is to ensure that persistent rankings are not solely attributable to the persistence of high expense, low performance funds. The d1 statistic measures top decile performance.

In the precision-weighted results, the top decile is positive and significantly greater than zero for many combinations of priors. For the four-factor model, for example, the top decile post-ranking-period alpha across diffuse skill priors ranging from 10^{-6} to 10^{-3} averages 1.46% (1.83%) per year, with average t-statistics of 3.46 (4.23) when the post-ranking-period alphas are equally weighted (precision weighted). These results are not evident in Table I because the table reports averages across all prior combinations. The top decile for sorts based on standard frequentist alphas—the type of alpha typically examined in persistence studies—also increases. For the four-factor model, for example, the top decile abnormal performance is 1.47% (1.90%) per year, with a t-statistic of 3.37 (4.88) when the fund post-ranking-period alphas are equally weighted (precision weighted). In contrast, using two methodologies and monthly returns, Carhart (1997) finds insignificant post-ranking-period four-factor alphas

[8] Average daily returns from July 1968 through December 1984 (January 1985 through December 1995) for the CMS, IND1, IND2, and IND3 factors are −0.0104%, −0.0091%, 0.0014%, and −0.0061% (−0.0070%%, 0.0032%, −0.0121%, and −0.0055%), respectively.

of −0.12% and 0.02% per month for his top decile. Precision-weighted top decile CAPM and Fama–French model alphas experience similar increases for both Bayesian and frequentist measures. These results suggest that incorporating higher-order moments of fund performance estimates can improve the ability of investors to predict future abnormal performance.

B.2. Frequentist Alpha Predictability

The frequentist alphas in Panels A3, B3, and C3 of Table I also indicate a significant relationship between past and future performance. The degree of predictability depends on the pricing model and also on whether long-history passive assets are used. For the CAPM model in Panel A, the Bayesian alpha generally predicts future standard alpha more accurately than the standard CAPM alpha predicts future performance, except when the Bayesian alpha is estimated with precise priors against the existence of managerial skill.

A substantial advantage of the Bayesian alphas over standard alphas in predicting future standard alphas stems from the use of passive asset data that begin before the fund exists. By using the same passive asset data as the Bayesian measures, the long-history frequentist alphas level the playing field considerably. The long-history frequentist measures predict future performance better than the standard measures, but not as well as the Bayesian alphas for diffuse skill priors. For example, in results not completely evident in Table I, for skill priors ranging from 10^{-6} to 10^{-3}, the Spearman correlations between past Bayesian alpha decile ranking and subsequent decile performance ranking are greater than (equal to) the corresponding long-history frequentist Spearman correlation in 29 (three) of the 32 combinations of skill and model mispricing prior for the equally weighted deciles and in all 32 of the prior combinations for the precision-weighted deciles.

The frequentist alphas associated with the Fama–French model in Panel B and the four-factor model in Panel C predict future performance more accurately than the CAPM frequentist measures, which is the same pattern noted earlier with the Bayesian alphas. For the Fama–French and four-factor models, both frequentist measures predict about as well as the best category averages of Bayesian measures in the tables, and the standard alphas predict about as well as the long-history frequentist measures.[9]

[9] Since Table I reports averages across prior combinations, it is not evident in the table that a number of prior combinations exist in which the Fama–French or four-factor Bayesian alphas predict better than the standard or long-history frequentist measures. For example, for skill priors ranging from 10^{-7} to 10^{-3} combined with model mispricing priors ranging from 10^{-8} to 10^{-4}, all d1−d10 statistics associated with equal-weighted post-ranking-period deciles are greater for the Fama–French Bayesian sorts than for the standard or long-history frequentist sorts. As another example, for a moderate skill prior of 10^{-7} and a precise model mispricing prior of 10^{-10} or 10^{-11}, the Spearman correlation between the four-factor Bayesian alpha mean decile ranking and subsequent precision-weighted decile performance ranking is 1.000, compared to 0.988 for the frequentist measures.

The advantage of the Bayesian and long-history frequentist alphas over the standard alphas in predicting future performance drops off as the number of benchmark assets increases from one to three or four. This decrease occurs because, as the benchmark consists of more passive assets, fewer remain as nonbenchmark assets. With fewer nonbenchmark assets, the $\tilde{c}_{A,N}\tilde{\alpha}_N$ term in equation (8) decreases in importance, both because it consists of fewer nonbenchmark assets and also because it loses the most important passive assets, those that have been shown in the empirical asset pricing literature to best explain the cross-section of returns. In our sample, for example, CMS and IND1–3 contribute much less to the abnormal returns of mutual funds than do SMB, HML, and especially UMD.[10]

B.3. Fund Expenses

The results in the previous subsection suggest that the long-history passive assets are responsible for most of the predictive ability of the Bayesian alphas. However, the Spearman correlation results in Table I indicate that predictability peaks at a moderate skill prior variance. Since a moderate skill prior allows the Bayesian alpha to combine past fund performance with the prior mean of -1 multiplied by the expense ratio, this result suggests that the Bayesian alphas also benefit from the other piece of additional information that they use, that is, fund expenses.

To more closely examine whether expenses enhance the Bayesian alpha's ability to predict future performance, we first examine the simple relationship between expenses and subsequent performance. We sort funds into deciles based on the inverse of the expense ratio, using overlapping annual ranking periods similar to our Bayesian alpha tests. Next, we examine the average standard alpha for each decile during the following quarter. Table II, Panel A shows the results for all three models of standard alpha. The Spearman correlation coefficient between ranking-period expense decile and post-ranking-period performance decile is only marginally significant for the CAPM model, and insignificant for the Fama–French and four-factor models. However, d1$-$d10 is positive for all three models. Overall, the Bayesian alphas show greater predictive ability than expenses alone, especially for moderate to diffuse skill priors.

Using expenses together with past performance in order to predict performance provides another means for examining the effectiveness of expenses in predicting performance. We combine expenses with past alpha in two ways. Both involve estimating via OLS the linear combination of the past annual expense ratio and past annual standard alpha that best explains the subsequent quarterly standard alpha. The first method uses only the immediately preceding alpha and expenses. For example, the first method estimates the linear

[10] SMB, HML, UMD, CMS, IND1, IND2, and IND3 have daily CAPM alphas of -0.0088%, 0.0194%, 0.0317%, -0.0096%, -0.0051%, -0.0008%, and -0.0025%, respectively, from January 1985 through December 1995.

Table II

Performance Predictability of Expenses and Past Performance

The table reports statistics that describe the extent to which fund expenses and standard alpha predict subsequent standard alpha. We estimate the standard alphas using one (MKT), three (MKT, SMB, and HML), and four (MKT, SMB, HML, and UMD) benchmarks for the CAPM, Fama–French, and four-factor model, respectively. We sort funds into deciles during an annual ranking period. In Panel A, we sort based on 1/expense ratio, and in Panel B, we sort based on the linear combination of past expenses and past standard alpha that best explains the next standard alpha. Panel B1 (B2) uses past expenses and past standard alpha from one set (all sets) of prior periods to estimate the linear combination. For each decile, we compute the mean standard alpha over the following quarterly post-ranking period. We estimate all alphas using daily returns. A two-tailed Spearman test based on decile rankings is significant at the 10%, 5%, and 1% levels for correlations of 0.564, 0.648, and 0.794, respectively. The sample consists of 230 mutual funds. The mutual fund sample period is from January 2, 1985 to December 29, 1995.

Model	Spearman	d1−d10
Panel A: Sort by 1/Expenses		
CAPM	0.612	0.0085
Fama–French	0.188	0.0060
Four-factor	0.552	0.0090
Panel B: Sort by Linear Combination of Alpha and Expenses		
B1. Using Data from One Prior Period		
CAPM	−0.745	−0.0063
Fama–French	0.903	0.0158
Four-factor	0.939	0.0136
B2. Using Data from All Prior Periods		
CAPM	0.745	0.0077
Fama–French	0.927	0.0221
Four-factor	0.976	0.0189

combination of the annual expense ratio and standard alpha at time $t - 1$ that best explains the quarterly standard alpha at time t. The second method pools all preceding alpha and expense data to estimate the relationship between past annual alpha and expenses and subsequent quarterly standard alpha. For both methods, we then sort funds into deciles during the ranking period based on predicted performance using the estimated linear combination, and we examine the standard alphas of the deciles during the following quarterly post-ranking period.

Panel B of Table II shows the results. Panel B1 uses one set of standard alphas and expenses from a prior period to predict performance, while Panel B2 uses information from all prior periods. The results in Panel B1 are mixed. When we use the CAPM model, a significant negative relationship exists between predicted performance and actual future performance. With the Fama–French and four-factor models, however, the relationship is positive and statistically significant. In Panel B2, using all prior periods enhances the accuracy of the prediction, and in all three models a statistically significant positive relationship

exists between predicted and actual future performance. Only for the Fama–French model using all prior-period data, however, is the d1−d10 spread as large as it is when sorting on past alpha alone or when sorting on Bayesian alphas based on moderate to diffuse skill priors.[11] Thus, although expenses do not appear to be the primary predictor of future performance, these results suggest that they provide incremental information.

B.4. Without Nonbenchmark Assets

Table III repeats the analysis of Table I, except that it does not use nonbenchmark assets in the Bayesian alpha estimation. In the absence of nonbenchmark assets, no model mispricing prior exists. Comparing the CAPM model results to the earlier results based on nonbenchmark assets, we see two important differences. First, predictability is much smaller without nonbenchmarks than with nonbenchmarks. Second, without nonbenchmarks, predictability sharply peaks at a moderate skill prior variance. Recall that the Spearman correlations in Table I show a subtler peak at a moderate skill prior variance. We can best explain these contrasting results in the context of Stambaugh (1997). With no nonbenchmark assets, the Bayesian alpha in Table III can use no long-history passive assets to enhance the accuracy of the alpha estimation and no nonbenchmarks to improve the precision of the skill estimate. In contrast, the CAPM Bayesian estimate in Table I uses seven nonbenchmarks. Ultimately, predictability for the Bayesian alpha without nonbenchmarks peaks sharply at a moderate skill prior variance because a more diffuse prior weights the imprecise skill estimate too heavily. The moderate skill prior variance strikes the proper balance between the noisy skill estimate and the fund's expenses.

For the Fama–French model, the most important difference between using nonbenchmarks and not using nonbenchmarks is predictability when skill priors are precise. Predictability without nonbenchmarks is noticeably weaker than predictability with nonbenchmarks. With nonbenchmarks and a precise prior against managerial skill, the Bayesian alpha essentially combines −1 multiplied by fund expenses with the nonbenchmark alphas. For the Fama–French model, expenses combine with exposures to momentum and the less important nonbenchmark assets. When we exclude nonbenchmarks, however, funds sort largely according to expense ratio (low to high). Since momentum persists during our sample period, the estimates based on nonbenchmarks show greater predictability than the estimates without nonbenchmarks.

The differences between using and not using nonbenchmarks are less important for the four-factor model, where predictability depends on whether we equally weight or precision-weight the decile post-ranking-period standard alphas. The patterns associated with the four-factor model differ from the

[11] Bayesian alphas based on a number of prior combinations show greater predictability than the forecasts based on expenses and performance, including those noted earlier for the Fama–French model for skill priors ranging from 10^{-7} to 10^{-3} combined with model mispricing priors ranging from 10^{-8} to 10^{-4}.

Table III

Performance Predictability of Bayesian versus Frequentist Alphas, without Nonbenchmarks

The table reports statistics that describe the extent to which Bayesian and frequentist alphas predict subsequent standard alphas. In Panels A1, B1, and C1, for each skill prior variance, we sort funds into deciles based on their Bayesian alpha posterior mean during an annual ranking period. In Panels A2, B2, and C2, we sort funds into deciles based on their frequentist alpha during an annual ranking period. For each decile, we compute the standard alpha over the following quarterly post-ranking period. We estimate the Bayesian and frequentist alphas using one (MKT; Panel A), three (MKT, SMB, and HML; Panel B), or four (MKT, SMB, HML, and UMD; Panel C) benchmarks and no nonbenchmarks. We estimate all alphas using daily returns. The Spearman correlation measures the relationship between the ranking-period Bayesian alpha posterior mean or frequentist alpha decile ranking and the post-ranking-period standard alpha decile ranking. d1−d10 is the difference between the post-ranking-period standard alpha (daily percentage) for the top and bottom ranking-period deciles. In Panels A1, B1, and C1 (Panels A2, B2, and C2), we weight the standard alphas in the post-ranking-period deciles either equally or by the inverse of the posterior variance of the Bayesian alpha (variance of the frequentist alpha) from the ranking period ("Precision weighted"). A two-tailed Spearman test based on decile rankings is significant at the 10%, 5%, and 1% levels for correlations of 0.564, 0.648, and 0.794, respectively. The sample consists of 230 mutual funds. The mutual fund sample period is from January 2, 1985 to December 29, 1995. The factor sample period is from July 1, 1968 to December 29, 1995.

Prior or Data Series Length	Equal Weighted		Precision Weighted	
	Spearman	d1−d10	Spearman	d1−d10
Panel A: CAPM				
A1. Bayesian				
1E−13	0.685	0.0085	0.697	0.0073
1E−10	0.697	0.0091	0.612	0.0080
1E−8	0.903	0.0102	0.952	0.0094
1E−6	0.903	0.0066	0.867	0.0070
1E−3	0.903	0.0066	0.855	0.0070
A2. Frequentist				
Standard	0.867	0.0076	0.879	0.0084
Panel B: Fama–French				
B1. Bayesian				
1E−13	0.224	0.0057	0.345	0.0056
1E−10	0.224	0.0061	0.394	0.0061
1E−8	0.964	0.0213	0.952	0.0218
1E−6	0.927	0.0216	0.988	0.0223
1E−3	0.964	0.0216	0.988	0.0226
B2. Frequentist				
Standard	0.952	0.0213	0.988	0.0226
Panel C: Four-Factor				
C1. Bayesian				
1E−13	0.758	0.0089	0.842	0.0077
1E−10	0.745	0.0091	0.770	0.0081
1E−8	0.952	0.0214	0.952	0.0211
1E−6	0.939	0.0210	0.988	0.0209
1E−3	0.927	0.0210	0.988	0.0211
C2. Frequentist				
Standard	0.879	0.0226	0.988	0.0223

Fama–French model because, in the four-factor model, a large source of the predictable nonbenchmark asset return—the momentum factor—is no longer a nonbenchmark asset.

C. Methodology Modifications

C.1. Data Frequency and Ranking Period Interval

Our empirical methodology examines the relationship between Bayesian alphas estimated over a 1-year ranking period and subsequent standard quarterly alphas. We next examine the sensitivity of our results to the length of the ranking period and also to the frequency of the fund and factor returns. We compare the annual ranking period to a quarterly ranking period as well as to a 3-year ranking period. As with the annual ranking period analysis, the quarterly ranking period analysis uses daily frequency returns. The 3-year ranking period analysis uses daily returns as well as monthly returns, compounded from the daily returns. It would not be feasible to use monthly returns in conjunction with ranking periods shorter than approximately 3 years because the small number of monthly returns would preclude efficient estimation of alpha.

For the quarterly analysis, we use nonoverlapping quarterly ranking periods. For the 3-year analysis, we use overlapping 3-year ranking periods. For both alternative ranking period lengths, we maintain the nonoverlapping quarterly post-ranking period and use the same post-ranking-period standard alphas used earlier. In the 3-year monthly analysis, the frequency of the return data necessitates some slight modifications to the Bayesian estimation for proper comparison to the daily analysis. In particular, we multiply the skill prior variance and model mispricing prior variance by 21, the average number of trading days per month. Also, the skill prior mean equals -1 multiplied by the expense ratio divided by 12; in the daily analysis, we divide by 252.

Table IV shows that, regardless of the pricing model, the annual ranking periods produce Bayesian estimates that predict future performance better than either the quarterly or 3-year ranking periods. This result holds for nearly all prior variance categories using equally weighted deciles and for all prior variance categories using precision-weighted deciles. The advantage of the annual ranking period is often dramatic. Using a CAPM benchmark, precision-weighted deciles, and a diffuse skill prior, for example, the annual ranking period results more than double the quarterly ranking period results and are 30% to 40% greater than the 3-year results. Despite the absolute differences in performance predictability, all three alternatives produce similar patterns across different priors and benchmark models. Table IV also shows that the 3-year daily estimates generally show greater predictability than 3-year monthly estimates, except for diffuse skill priors, especially with the Fama–French and four-factor models.

In an effort to explain why the annual ranking periods dominate the quarterly and 3-year ranking periods, Table V compares the predictability of Bayesian skill and nonbenchmark factor loadings by return frequency and length of ranking period. Panel A shows that, with a few exceptions, annual ranking periods

predict skill better than quarterly or 3-year ranking periods. Panel B shows that annual ranking periods also produce better forecasts of post-ranking-period nonbenchmark factor loadings, with higher Spearman correlations between ranking and post-ranking-period loadings and larger d1−d10 differences.

A fund's factor loading is the weighted average of the individual factor loadings of its securities. Holdings are subject to turnover, and a longer ranking

Table IV

Performance Predictability of Bayesian Alpha, by Returns Frequency and Length of Ranking Period

The table reports statistics that describe the extent to which Bayesian alphas predict subsequent standard alphas for three different ranking period lengths. For each combination of skill prior variance and model mispricing prior variance, we sort funds into deciles based on their Bayesian alpha posterior mean during a one-quarter, 1-year, or 3-year ranking period. For each decile, we compute the standard alpha over the following quarterly post-ranking period. In Panel A, we estimate the Bayesian alphas using one benchmark (MKT) and seven nonbenchmarks (SMB, HML, UMD, CMS, and IND1–3). In Panel B, we estimate the Bayesian alphas using three benchmarks (MKT, SMB, and HML) and five nonbenchmarks (UMD, CMS, and IND1–3). In Panel C, we estimate the Bayesian alphas using four benchmarks (MKT, SMB, HML, and UMD) and four nonbenchmarks (CMS and IND1–3). In each panel, we estimate the post-ranking-period standard alphas using the same set of benchmarks that we use in the Bayesian alpha estimation, but no nonbenchmarks. The one-quarter and 1-year ranking periods use daily returns, and the 3-year ranking period uses either daily or monthly returns. The post-ranking periods use daily returns. The table reports d1−d10, the difference between the post-ranking-period standard alpha (daily percentage) for the top and bottom ranking-period deciles. Each statistic represents the mean associated with that parameter (skill prior variance in Panels A1, B1, and C1 or model mispricing prior variance in Panels A2, B2, and C2) over all combinations of the other parameter. We weight the standard alphas in the post-ranking-period deciles either equally or by the inverse of the posterior variance of the Bayesian alpha from the ranking period ("Precision weighted"). The sample consists of 230 mutual funds. The mutual fund sample period is from January 2, 1985 to December 29, 1995. The factor sample period is from July 1, 1968 to December 29, 1995.

	Equal Weighted				Precision Weighted			
	Daily			Monthly	Daily			Monthly
Prior	Quarter	Year	3 Years	3 Years	Quarter	Year	3 Years	3 Years
Panel A: CAPM								
A1. By skill prior variance								
1E−13	0.0059	0.0063	0.0041	−0.0027	0.0037	0.0076	0.0040	−0.0019
1E−10	0.0062	0.0065	0.0044	−0.0025	0.0042	0.0074	0.0046	−0.0017
1E−8	0.0085	0.0116	0.0068	−0.0012	0.0073	0.0130	0.0086	−0.0004
1E−6	0.0067	0.0138	0.0106	0.0105	0.0068	0.0142	0.0113	0.0095
1E−3	0.0072	0.0142	0.0106	0.0116	0.0067	0.0147	0.0113	0.0108
A2. By model mispricing prior variance								
1E−11	0.0075	0.0108	0.0084	0.0060	0.0065	0.0094	0.0065	0.0041
1E−10	0.0076	0.0111	0.0087	0.0059	0.0068	0.0101	0.0067	0.0041
1E–8	0.0065	0.0102	0.0071	0.0055	0.0048	0.0117	0.0086	0.0045
1E–6	0.0065	0.0099	0.0066	0.0007	0.0052	0.0115	0.0085	0.0027
1E–4	0.0066	0.0099	0.0065	0.0002	0.0052	0.0114	0.0084	0.0024

(continued)

Table IV—*Continued*

	Equal Weighted				Precision Weighted			
	Daily			Monthly	Daily			Monthly
Prior	Quarter	Year	3 Years	3 Years	Quarter	Year	3 Years	3 Years
Panel B: Fama–French								
B1. By skill prior variance								
1E–13	0.0133	0.0123	0.0082	0.0019	0.0104	0.0116	0.0050	0.0027
1E–10	0.0134	0.0125	0.0085	0.0020	0.0108	0.0115	0.0052	0.0028
1E–8	0.0190	0.0196	0.0174	0.0064	0.0174	0.0188	0.0145	0.0059
1E–6	0.0149	0.0221	0.0172	0.0208	0.0128	0.0212	0.0152	0.0186
1E–3	0.0144	0.0218	0.0170	0.0215	0.0122	0.0210	0.0151	0.0199
B2. By model mispricing prior variance								
1E–11	0.0094	0.0129	0.0105	0.0110	0.0078	0.0117	0.0087	0.0091
1E–10	0.0101	0.0129	0.0105	0.0109	0.0084	0.0124	0.0087	0.0091
1E–8	0.0171	0.0190	0.0145	0.0111	0.0146	0.0183	0.0117	0.0097
1E–6	0.0170	0.0199	0.0147	0.0100	0.0145	0.0192	0.0119	0.0113
1E–4	0.0170	0.0199	0.0148	0.0095	0.0145	0.0192	0.0119	0.0112
Panel C: Four-Factor								
C1. By skill prior variance								
1E–13	0.0048	0.0064	0.0064	0.0043	0.0035	0.0083	0.0048	0.0049
1E–10	0.0053	0.0072	0.0070	0.0044	0.0043	0.0085	0.0056	0.0050
1E–8	0.0116	0.0215	0.0120	0.0069	0.0114	0.0202	0.0117	0.0063
1E–6	0.0105	0.0222	0.0146	0.0174	0.0098	0.0217	0.0137	0.0149
1E–3	0.0106	0.0220	0.0147	0.0186	0.0093	0.0218	0.0139	0.0164
C2. By model mispricing prior variance								
1E–11	0.0105	0.0158	0.0120	0.0119	0.0092	0.0146	0.0105	0.0098
1E–10	0.0104	0.0159	0.0122	0.0118	0.0091	0.0147	0.0103	0.0098
1E–8	0.0082	0.0148	0.0109	0.0118	0.0074	0.0157	0.0096	0.0101
1E–6	0.0068	0.0150	0.0102	0.0088	0.0062	0.0164	0.0098	0.0101
1E–4	0.0067	0.0149	0.0102	0.0084	0.0061	0.0164	0.0098	0.0093

period would likely include a number of assets the fund no longer holds at the end of the period. Such turnover could bias factor loading forecasts. Consequently, among the three ranking period horizons that we consider, one might expect that a quarterly ranking period would best forecast a fund's post-ranking-period factor loadings and that a 3-year ranking period would forecast the worst. However, a trade-off exists between estimation accuracy and bias. One might expect smaller biases with a quarterly ranking period, but one might also expect larger measurement errors, at least compared to the annual ranking period. An annual ranking period consists of four times as many return observations as a quarterly ranking period and about seven times as many as a 3-year ranking period based on monthly returns. Thus, although it is not ex ante clear which ranking period length is optimal, the results in Table V suggest that the annual ranking period offers the best balance between the number of return observations and the time period length.

Table V

Predictability of Bayesian Skill and Nonbenchmark Factor Loadings, by Returns Frequency and Length of Ranking Period

The table reports statistics that describe the extent to which Bayesian skill and nonbenchmark factor loadings predict subsequent fund skill and nonbenchmark factor loadings. For each combination of skill prior variance and model mispricing prior variance, we sort funds into deciles based on their Bayesian posterior skill (Panel A) or nonbenchmark factor loading (Panel B) during a one-quarter, 1-year, or 3-year ranking period. For each decile, we compute skill or nonbenchmark factor loading over the following quarterly post-ranking period. In Panel A, we estimate skill using eight factors (MKT, SMB, HML, UMD, CMS, and IND1–3). In Panel B, SMB and HML are nonbenchmark assets in the CAPM model, UMD is a nonbenchmark asset in the CAPM and Fama-French models, and CMS and IND1–3 are nonbenchmark assets in the CAPM, Fama–French, and four-factor models. The one-quarter and 1-year ranking periods use daily returns, and the 3-year ranking period uses either daily or monthly returns. The post-ranking periods use daily returns. The Spearman correlation measures the relationship between the ranking-period Bayesian posterior skill or nonbenchmark factor loading decile ranking and the post-ranking-period skill or nonbenchmark factor loading decile ranking. d1−d10 is the difference between the post-ranking-period skill or nonbenchmark factor loading for the top and bottom ranking-period deciles. In Panel B, the statistics represent the mean over all skill prior variances. We equal weight the skill and factor loadings in the post-ranking-period deciles. A two-tailed Spearman test based on decile rankings is significant at the 10%, 5%, and 1% levels for correlations of 0.564, 0.648, and 0.794, respectively. The sample consists of 230 mutual funds. The mutual fund sample period is from January 2, 1985 to December 29, 1995. The factor sample period is from July 1, 1968 to December 29, 1995.

Prior or Passive Asset	Spearman				d1−d10			
	Daily			Monthly	Daily			Monthly
	Quarter	Year	3 Years	3 Years	Quarter	Year	3 Years	3 Years
Panel A: Skill								
1E–13	0.636	0.527	0.442	0.382	0.0056	0.0072	0.0070	0.0064
1E–10	0.576	0.430	0.564	0.382	0.0068	0.0077	0.0081	0.0063
1E–8	0.927	0.988	0.976	0.782	0.0152	0.0242	0.0159	0.0087
1E–6	0.952	0.976	0.964	0.539	0.0161	0.0272	0.0192	0.0230
1E–3	0.915	0.988	0.988	0.685	0.0159	0.0268	0.0193	0.0239
Panel B: Factor Loadings								
SMB	1.000	1.000	1.000	1.000	0.860	0.932	0.900	0.856
HML	1.000	1.000	1.000	1.000	0.841	0.925	0.892	0.666
UMD	0.990	1.000	0.999	0.984	0.316	0.380	0.343	0.265
CMS	0.968	0.981	0.994	0.924	0.170	0.237	0.218	0.221
IND1	0.990	0.989	0.994	0.974	0.098	0.124	0.118	0.065
IND2	0.972	0.990	0.991	0.941	0.076	0.096	0.107	0.053
IND3	0.981	1.000	0.990	0.989	0.059	0.084	0.073	0.050

C.2. Time-Varying Factor Loadings

Motivated by the results in Table V, we examine next whether Bayesian and frequentist models that directly incorporate time-varying factor loadings improve predictability. The Gibbs sampler estimation method required in the Bayesian time-varying parameter model is computer intensive, and estimation

using daily data is currently not feasible due to the large size of the matrices involved. However, we estimate the Bayesian time-varying model using monthly data. For comparison purposes, we also estimate the frequentist time-varying parameter model using monthly data. Comparing the time-varying results to constant-parameter estimates provides insight into the ability of a time-varying parameter model to improve performance predictability.

Table VI shows the time-varying parameter results. We estimate the time-varying parameter alpha, $\alpha_{A,T}$, in equation (11) using combinations of priors that span those used in the constant-parameter Bayesian model. We then compare the performance predictability of the monthly time-varying model with daily and monthly constant-parameter models, which we also show in Table VI.[12] With a few exceptions, the Bayesian time-varying results indicate greater predictability than the Bayesian constant-parameter results, regardless of the predictability statistic. In many instances, the improvement is dramatic. These results suggest that models which recognize the dynamic nature of mutual fund holdings improve predictability.

The frequentist results shown in Table VI are mixed. The time-varying parameter results show greater predictability using the d1−d10 measure for the CAPM and four-factor models, but less predictability using the Spearman measure. Thus, we are unable to draw any general conclusion regarding the superiority of a time-varying parameter model in a frequentist setting.

C.3. Length of Historical Factor Data

In addition to examining the sensitivity of the results to the length of the ranking period, we also examine the sensitivity of the results to the amount of historical factor data used. Thus far, we base the Bayesian estimates on the entire history of daily nonbenchmark factor returns, which begin in July 1968. We now examine the main predictability results for Bayesian alphas based on the past 2, 5, 10, and 20 years of historical factor data. Table VII, Panel A shows the Spearman and d1−d10 statistics for each historical factor length (including the entire history). Each statistic in the table corresponds to the average across all combinations of skill prior variance and model mispricing prior variance.

The results indicate that the range of historical factor data used to estimate the Bayesian alphas does affect how well the Bayesian measures predict future performance. Four of the six columns of results in Panel A show the greatest predictability statistic for estimates based on the entire history of nonbenchmark factor returns. For the other two columns, 10 and 20 years of historical data produce the highest statistics. In all cases, the statistics based on the entire history of historical data are greater than the statistics based on 2 years of historical data. Stambaugh (1997) indicates that more data is preferable because it permits a more precise estimate of nonbenchmark asset abnormal returns.

[12] Note that the Bayesian constant parameter results shown in Table VI are subsets of the results shown in Table IV. Since Table IV shows results that are averaged across multiple prior variances, the exact constant parameter results listed in Table VI cannot be directly observed in Table IV.

Table VI

Performance Predictability with Time-Varying Factor Loadings, Including Nonbenchmarks

The table reports statistics that describe the extent to which time-varying and constant-parameter Bayesian and frequentist alphas predict subsequent standard alphas. In Panels A1, B1, and C1, for various combinations of skill prior variance and model mispricing prior variance, we sort funds into deciles based on their time-varying or constant-parameter Bayesian alpha posterior mean during a 3-year ranking period. In Panels A2, B2, and C2, we sort funds into deciles based on their time-varying and constant-parameter frequentist alpha during a 3-year ranking period. For each decile, we compute the standard alpha over the following quarterly post-ranking period. We estimate the ranking-period alphas using one benchmark (MKT) and seven nonbenchmarks (SMB, HML, UMD, CMS, and IND1–3) in Panel A, using three benchmarks (MKT, SMB, and HML) and five nonbenchmarks (UMD, CMS, and IND1–3) in Panel B, and using four benchmarks (MKT, SMB, HML, and UMD) and four nonbenchmarks (CMS and IND1–3) in Panel C. We estimate all post-ranking-period standard alphas using the same set of benchmarks that we use in the Bayesian alpha estimation, but no nonbenchmarks. We estimate the time-varying (constant parameter) ranking-period alphas using monthly (daily or monthly) returns. We estimate all post-ranking-period standard alphas using daily returns. The Spearman correlation measures the relationship between the ranking-period Bayesian alpha posterior mean or frequentist alpha decile ranking and the post-ranking-period standard alpha decile ranking. d1−d10 is the difference between the post-ranking-period standard alpha (daily percentage) for the top and bottom ranking-period deciles. We weight the standard alphas in the post-ranking-period deciles equally. A two-tailed Spearman test based on decile rankings is significant at the 10%, 5%, and 1% levels for correlations of 0.564, 0.648, and 0.794, respectively. The sample consists of 230 mutual funds. The mutual fund sample period is from January 2, 1985 to December 29, 1995. The factor sample period is from July 1, 1968 to December 29, 1995.

Prior or Data Series Length		Time-Varying		Constant Parameter			
				Daily		Monthly	
Skill	Model	Spearman	d1−d10	Spearman	d1−d10	Spearman	d1−d10
			Panel A: CAPM				
A1. Bayesian							
1E–13	1E–11	0.927	0.0091	0.600	0.0067	0.079	−0.0000
1E–13	1E–4	0.661	0.0029	0.248	0.0022	−0.442	−0.0048
1E–8	1E–8	0.952	0.0126	0.903	0.0065	0.394	0.0007
1E–3	1E–11	0.988	0.0106	0.879	0.0098	0.794	0.0132
1E–3	1E–4	0.806	0.0072	0.927	0.0103	0.661	0.0087
A2. Frequentist							
Long-history		0.600	0.0145	0.903	0.0118	0.915	0.0117
			Panel B: Fama–French				
B1. Bayesian							
1E–13	1E–11	0.903	0.0162	0.248	0.0054	0.176	0.0009
1E–13	1E–4	0.952	0.0187	0.830	0.0095	0.636	0.0025
1E–8	1E–8	0.964	0.0213	0.879	0.0207	0.697	0.0063
1E–3	1E–11	0.903	0.0200	0.939	0.0148	0.782	0.0232
1E–3	1E–4	0.915	0.0193	0.915	0.0185	0.842	0.0184
B2. Frequentist							
Long-history		0.697	0.0093	0.976	0.0192	0.939	0.0188

(continued)

Table VI—*Continued*

Prior or Data Series Length		Time-Varying		Constant Parameter			
				Daily		Monthly	
Skill	Model	Spearman	d1−d10	Spearman	d1−d10	Spearman	d1−d10
Panel C: Four-Factor							
C1. Bayesian							
1E–13	1E–11	0.927	0.0168	0.515	0.0101	0.418	0.0060
1E–13	1E–4	0.867	0.0195	0.515	0.0037	0.552	0.0010
1E–8	1E–8	0.952	0.0197	0.952	0.0123	0.685	0.0074
1E–3	1E–11	0.891	0.0190	0.964	0.0141	0.636	0.0189
1E–3	1E–4	0.952	0.0187	0.952	0.0153	0.552	0.0188
C2. Frequentist							
Long-history		0.588	0.0200	0.952	0.0161	0.891	0.0192

The results in Panel A confirm that, overall, the Bayesian alphas predict future performance better when we estimate them using more historical factor data.

The range of historical factor data influences the Bayesian alpha via the nonbenchmark abnormal returns in equation (8). Consequently, the extent to which the abnormal returns of the nonbenchmark assets are predictable over various spans of time determines how well various estimates of the Bayesian alpha predict future performance. Thus, for each nonbenchmark asset, we estimate the relationship between the historical nonbenchmark alpha, $\hat{\alpha}_{N,t}$, and the future quarterly nonbenchmark alpha, $\hat{\alpha}_{N,t+1}$, for various lengths of historical nonbenchmark data, that is,

$$\hat{\alpha}_{N,t+1} = a + b\hat{\alpha}_{N,t} + u_t. \tag{13}$$

Since we use overlapping time series of nonbenchmark returns to estimate adjacent observations of the regressor, $\hat{\alpha}_{N,t}$, the time series of $\hat{\alpha}_{N,t}$ is autocorrelated and OLS estimation of equation (13) produces a biased estimate of b. To correct for the bias, we apply the bias correction of Kothari and Shanken (1997). Then, for each nonbenchmark asset, we compute the mean square error (MSE) R^2,

$$\text{MSE } R^2 = 1 - \left[\frac{\text{MSE}}{s^2(\hat{\alpha}_{N,t+1})}\right], \tag{14}$$

similar to Fama and French (1988). We compute the MSE in equation (14) between the time series of predicted and actual nonbenchmark alphas.

Table VII, Panel B shows the average MSE R^2 across all nonbenchmark assets associated with each pricing model. For all three models, the table shows that the MSE R^2 generally increases with the length of the time series of nonbenchmark factor returns. Similar to Panel A, the statistics based on the entire

Table VII

Performance Predictability of Bayesian Alpha, by Length of Historical Factor Returns

The table reports statistics that describe the extent to which Bayesian alphas estimated using various lengths of historical factor returns predict subsequent standard alphas. In Panel A, for each combination of skill prior variance and model mispricing prior variance, we sort funds into deciles based on their Bayesian alpha posterior mean during an annual ranking period. For each decile, we compute the standard alpha over the following quarterly post-ranking period. We estimate the ranking-period Bayesian alphas using three different pricing models. The CAPM model uses one benchmark (MKT) and seven nonbenchmarks (SMB, HML, UMD, CMS, and IND1–3). The Fama–French model uses three benchmarks (MKT, SMB, and HML) and five nonbenchmarks (UMD, CMS, and IND1–3). The four-factor model uses four benchmarks (MKT, SMB, HML, and UMD) and four nonbenchmarks (CMS and IND1–3). We estimate all post-ranking-period standard alphas using one benchmark (MKT), three benchmarks (MKT, SMB, and HML), and four benchmarks (MKT, SMB, HML, and UMD) for the CAPM, Fama–French, and four-factor models, respectively. We estimate all alphas using daily returns. The Spearman correlation measures the relationship between the ranking-period Bayesian alpha posterior mean decile ranking and the post-ranking-period standard alpha decile ranking. d1−d10 is the difference between the post-ranking-period standard alpha (daily percentage) for the top and bottom ranking-period deciles. Each statistic in Panel A represents the mean associated with all combinations of skill prior variance and model mispricing prior variance. We weight the standard alphas in the post-ranking-period deciles equally. A two-tailed Spearman test based on decile rankings is significant at the 10%, 5%, and 1% levels for correlations of 0.564, 0.648, and 0.794, respectively. Panel B shows the MSE R^2 averaged across all nonbenchmarks for each pricing model. We base the MSE on the difference between long-history forecasts of future nonbenchmark alphas and actual quarterly nonbenchmark alpha estimates, where we use predictive regressions to forecast future nonbenchmark alphas. The MSE R^2 is $1 - [\text{MSE}/s^2(\alpha_{N,t+1})]$. The sample consists of 230 mutual funds. The mutual fund sample period is from January 2, 1985 to December 29, 1995. The factor sample period is from July 1, 1968 to December 29, 1995.

Panel A: Predictability

Long-History Years	Spearman			d1−d10		
	CAPM	Fama–French	Four-Factor	CAPM	Fama–French	Four-Factor
2	0.549	0.833	0.594	0.0069	0.0144	0.0123
5	0.352	0.798	0.818	0.0038	0.0142	0.0130
10	0.632	0.851	0.878	0.0067	0.0165	0.0148
20	0.833	0.883	0.873	0.0093	0.0179	0.0153
All	0.851	0.876	0.869	0.0094	0.0182	0.0154

Panel B: Predictive Regression MSE R^2

Long-History Years	CAPM	Fama–French	Four-Factor
2	0.038	0.115	0.131
5	0.014	0.112	0.134
10	0.059	0.129	0.133
20	0.061	0.124	0.144
All	0.051	0.140	0.166

history of historical data are greater than the statistics based on 2 years of historical data. These results suggest that we can predict future nonbenchmark abnormal returns more accurately when we use long- rather than short-horizon abnormal returns.

D. Performance and Cash Flow

We next examine the relationship between Bayesian alphas and subsequent cash flow. Our primary goal is to determine which combinations of priors correspond most closely to actual fund cash flows. Are investors skeptical of manager skill, and do they simply invest in low-expense funds? Alternatively, do they focus on high performing funds regardless of expenses? Also, since our earlier results indicate different levels of performance predictability for different combinations of priors, it is interesting to verify whether priors that best predict future performance correspond to priors that best predict cash flows.

We rank funds into deciles according to raw return and the mean of their Bayesian posterior alpha distribution or frequentist alpha during a quarterly, annual, or 3-year ranking period. We base the quarterly and annual (3-year) measures on daily (monthly) fund returns. We estimate the alphas using the CAPM, Fama–French, and four-factor models. We then examine the normalized cash flows of the alpha-sorted decile portfolios during the following quarter. Table VIII reports statistics that assess the relationship between the alternative performance measures and subsequent cash flow. For the Bayesian measures, the table reports statistics for each value of skill prior variance and model mispricing prior variance averaged over all values of the other parameter.

The Spearman rank correlation tests the relationship between the ranking period performance decile and subsequent normalized cash flow decile. For raw returns as well as Bayesian and frequentist measures across all three models, the table indicates that investors behave as though they have diffuse managerial skill priors. That is, investors believe that managers can add value, and they invest in mutual funds with the highest estimates of past performance. The Spearman rank correlation coefficient between past performance and cash flow increases dramatically as the skill prior variance increases. The difference in quarterly post-ranking-period cash flows between the top and bottom deciles yields a similar inference. Furthermore, the table suggests that cash flows are less sensitive to mispricing uncertainty. Across most model types and ranking period lengths, Bayesian measures based on more diffuse model mispricing priors sort future cash flows slightly better than measures based on less diffuse priors. This result is consistent with slightly more cash flowing into funds with exposure to the nonbenchmark assets that produce positive risk-adjusted returns, such as UMD in the CAPM or Fama–French models.

For the Bayesian estimates, the panels indicate little difference in the cash flows across the alternative pricing models. Bayesian alphas from the Fama–French model appear to sort future cash flows more accurately than the CAPM or four-factor models, although differences among the models are small. Since investors appear indifferent between returns associated with managerial skill and returns attributable to style, managers could benefit from strategies that exploit known CAPM anomalies such as momentum.

The results for all three models indicate that long-term performance predicts cash flows better than short-term performance. For the four-factor model, for example, the difference in cash flows between the top and bottom deciles

Table VIII

Cash Flow Predictability of Bayesian versus Frequentist Alpha, by Returns Frequency and Length of Ranking Period, with Nonbenchmarks

The table reports statistics that describe the extent to which returns, Bayesian alphas, and frequentist alphas predict subsequent normalized cash flow. In Panel A, we sort funds into deciles based on their return during a ranking period. In Panels B1, B2, C1, C2, D1, and D2, for each combination of skill prior variance and model mispricing prior variance, we sort funds into deciles based on their Bayesian alpha posterior mean during a ranking period. In Panels B3, C3, and D3, we sort funds into deciles based on their frequentist alpha. The ranking period is one quarter, 1 year, or 3 years. For each decile, we compute the normalized cash flow over the following quarterly post-ranking period. We estimate the Bayesian and long-history frequentist alphas using one benchmark (MKT) and seven nonbenchmarks (SMB, HML, UMD, CMS, and IND1–3) in Panel B, using three benchmarks (MKT, SMB, and HML) and five nonbenchmarks (UMD, CMS, and IND1–3) in Panel C, and using four benchmarks (MKT, SMB, HML, and UMD) and four nonbenchmarks (CMS and IND1–3) in Panel D. In Panels B3, C3, and D3, we estimate the standard frequentist alphas using the same set of benchmarks that we use in the Bayesian and long-history frequentist alpha estimation, but no nonbenchmarks. The one-quarter and 1-year ranking periods use daily returns, and the 3-year ranking period uses monthly returns. The Spearman correlation measures the relationship between the ranking-period Bayesian alpha posterior mean or frequentist alpha decile ranking and the post-ranking-period normalized cash flow decile ranking. d1−d10 is the difference between the post-ranking-period normalized cash flow (percentage of assets) for the top and bottom ranking-period deciles. Each statistic in Panels B1, B2, C1, C2, D1, and D2 represents the mean associated with that parameter (skill prior variance or model mispricing prior variance) over all combinations of the other parameter. We equal weight the cash flows in the post-ranking-period deciles. A two-tailed Spearman test based on decile rankings is significant at the 10%, 5%, and 1% levels for correlations of 0.564, 0.648, and 0.794, respectively. The sample consists of 230 mutual funds over a sample period from January 2, 1985 to December 29, 1995. The factor sample period is from July 1, 1968 to December 29, 1995.

Prior or Data Series Length	Spearman			d1−d10		
	Quarter	Year	3 Years	Quarter	Year	3 Years
Panel A: Return						
–	0.988	0.988	0.988	4.40	8.73	8.95
Panel B: CAPM						
B1. Bayesian alpha by skill prior variance						
1E–13	0.645	0.703	0.012	1.30	1.81	0.47
1E–10	0.629	0.723	0.036	1.38	1.97	0.48
1E–8	0.936	0.988	0.448	3.30	6.54	2.00
1E–6	0.915	0.983	0.983	4.48	7.86	7.71
1E–3	0.917	0.983	0.991	4.55	7.93	8.01
B2. Bayesian alpha by model mispricing prior variance						
1E–11	0.635	0.777	0.615	2.73	4.78	3.86
1E–10	0.745	0.797	0.615	2.74	4.86	3.85
1E–8	0.845	0.903	0.654	3.20	5.37	4.11
1E–6	0.818	0.914	0.392	3.01	5.13	3.60
1E–4	0.818	0.906	0.361	3.00	5.12	3.44
B3. Frequentist Alpha						
Standard	0.939	0.988	0.952	4.45	8.33	9.45
Long-history	0.879	0.976	1.000	4.60	7.88	8.30

(continued)

Table VIII—*Continued*

Prior or Data Series Length	Spearman			d1−d10		
	Quarter	Year	3 Years	Quarter	Year	3 Years
Panel C: Fama–French						
C1. Bayesian alpha by skill prior variance						
1E–13	0.735	0.582	0.473	1.80	1.94	1.10
1E–10	0.755	0.605	0.483	1.81	2.12	1.12
1E–8	0.942	0.986	0.845	3.62	6.77	3.04
1E–6	0.941	0.952	0.965	4.63	7.78	8.24
1E–3	0.949	0.955	0.965	4.60	7.74	8.34
C2. Bayesian alpha by model mispricing prior variance						
1E–11	0.635	0.744	0.615	2.73	4.78	3.86
1E–10	0.675	0.779	0.615	2.75	4.80	3.86
1E–8	0.916	0.846	0.603	3.44	5.47	3.74
1E–6	0.934	0.810	0.882	3.50	5.22	5.20
1E–4	0.929	0.831	0.850	3.50	5.21	5.24
C3. Frequentist alpha						
Standard	0.952	0.988	0.903	4.89	7.66	9.06
Long-history	0.988	0.964	1.000	3.89	7.83	7.76
Panel D: Four-Factor						
D1. Bayesian alpha by skill prior variance						
1E–13	0.385	0.289	0.436	0.63	0.68	0.87
1E–10	0.403	0.447	0.433	0.71	0.79	0.86
1E–8	0.897	0.986	0.844	3.74	6.56	2.90
1E–6	0.961	0.944	0.970	3.88	7.70	8.09
1E–3	0.945	0.950	0.971	3.86	7.74	8.20
D2. Bayesian alpha by model mispricing prior variance						
1E–11	0.612	0.747	0.615	2.71	4.78	3.86
1E–10	0.612	0.679	0.615	2.67	4.77	3.86
1E–8	0.680	0.745	0.677	2.44	4.40	3.89
1E–6	0.761	0.743	0.871	2.35	4.49	4.99
1E–4	0.754	0.715	0.858	2.35	4.48	4.82
D3. Frequentist alpha						
Standard	0.806	0.988	0.927	4.54	7.43	8.32
Long-history	0.879	0.988	0.976	3.41	7.35	8.08

reaches a maximum of 8.20% of net assets when using a 3-year performance measure. This maximum is almost double the 4.23% maximum associated with a quarterly measure. These results suggest that investors focus on a performance horizon that is somewhat longer than one quarter when deciding where to invest. Furthermore, the Fama–French and four-factor models indicate that investors use as much as a 3-year performance record. The predictability results in Table IV suggest that a 3-year performance interval is too long; a 1-year horizon produces substantially greater predictability. The focus of investors on a longer performance horizon also suggests that managers should be cautious with respect to undertaking strategic risk behavior over an annual tournament

as in Brown, Harlow, and Starks (1996). A risky bet that loses could hamper multiyear performance and, consequently, adversely affect multiyear cash flows.

Since prior studies also document a strong relationship between various performance measures and subsequent cash flow, it is interesting to compare the Bayesian performance-flow relationship to that of raw return and both standard and long-history frequentist performance measures. Panels A, B3, C3, and D3 report the raw return and frequentist results. Consistent with prior studies, we find a strong positive relationship between standard measures and subsequent cash flow. Consistent with the Bayesian results, we find that investors focus on performance horizons that are greater than one quarter. For the CAPM and Fama–French models, the table suggests that cash flow responds more to the 3-year standard alpha than to any of the 3-year Bayesian or long-history frequentist measures. A main difference between the Bayesian or long-history frequentist measures and the standard frequentist measure is that the former use long-history nonbenchmark passive assets. The greater sensitivity of cash flow to standard alpha suggests that investors are more sensitive to the recent returns of the nonbenchmark assets, rather than to their long-run means. The raw return results provide a similar inference. During our sample period, this focus on standard alpha appears to be a reasonably sound strategy for the Fama–French or four-factor models, given the similarity in predictability between the standard and Bayesian or long-history frequentist estimates in Table I, Panels B and C. Under the CAPM model, however, investors would be better off using the long-history passive assets, as the predictability results in Panel A of Table I indicate.

Recall from Table I that over this sample period, Bayesian alpha means predict future performance best for moderate to diffuse skill priors. Since the cash flow evidence in Table VIII suggests that investors have more diffuse skill priors, this result is consistent with the notion that money is smart (Zheng (1999)). In other words, during this period, more skepticism would not have paid off.

Table I also shows that predictability, as a function of model mispricing priors, depends on the model: The CAPM (Fama–French) model predicts better with precise (diffuse) priors, and the predictability of the four-factor model is somewhat insensitive, on an average, to the model mispricing prior. In Table VIII, a more diffuse model mispricing prior sorts future cash flows best for the CAPM model. For this model, the evidence suggests that investors are attracted to certain nonbenchmark loadings that subsequently do not deliver. A more diffuse model mispricing prior also sorts future cash flows best for the Fama–French model. For this model, however, Table I, Panel B indicates that a diffuse prior is rewarded, probably due to consistently high payoffs to momentum strategies. Finally, no strong relationship exists between a model mispricing prior and subsequent cash flows for the four-factor model, which is consistent with the predictability pattern for Bayesian alphas in Table I, Panel C.

In results not reported, we examine the relationship between Bayesian alpha means and subsequent cash flows, where we exclude nonbenchmark assets in

the Bayesian estimation.[13] As with the earlier evidence, the results indicate that investors focus on performance over longer time periods.

IV. Conclusion

We use the techniques of PS (2002a) to compare Bayesian estimates of mutual fund abnormal performance with standard frequentist measures. When the returns on passive nonbenchmark assets are correlated with fund holdings, incorporating histories of these returns produces a powerful method for predicting future performance—a method superior to standard measures. Bayesian alphas based on the single-factor CAPM are particularly useful for predicting future standard CAPM alphas, since the CAPM does not account for a number of known anomalies.

We find that incorporating higher-order moments of performance estimates increases the predictability of abnormal performance. Investors who select funds based on performance can improve the future performance of their choices by weighting past performance by the precision of the estimate.

We also examine the ability of time-varying parameter models to improve predictability in both frequentist and Bayesian settings. Although the frequentist time-varying results are mixed, the Bayesian results show improved predictability as compared to the constant-parameter estimates. We conclude that recognizing the inherently dynamic nature of mutual fund holdings can improve performance predictability.

Given their sensitivity to investor prior beliefs, Bayesian measures are particularly useful for inferring these beliefs by examining which priors best predict future cash flows. We find that Bayesian estimates associated with diffuse prior beliefs in managerial skill predict subsequent investor cash flows more accurately than estimates based on more skeptical beliefs. That is, investors do not adhere to a strategy of investing in the lowest expense funds irrespective of performance. Instead, they focus on past performance net of expenses.

We find that daily fund returns dominate the more common monthly returns in the context of forecasting future performance. The advantages of daily data stem from their use in accurately estimating risk-adjusted performance over relatively short measurement intervals. Monthly returns necessitate estimating performance over longer intervals, intervals over which factor loadings and risk-adjusted performance estimates may not persist.

Appendix

PS (2002a) show that

$$(\tilde{\alpha}_N, \tilde{\beta}_N)' = (I \otimes (D + Z'Z)^{-1}Z'Z)\text{vec}(\hat{G}), \tag{A1}$$

[13] These results are available from the authors upon request.

where

$$D = \begin{bmatrix} \dfrac{s^2}{\sigma^2_{\alpha_N}} & 0 \\ 0 & 0 \end{bmatrix}, \tag{A2}$$

$$Z = (i \quad r_B), \tag{A3}$$

and

$$\hat{G} = (Z'Z)^{-1}Z'r_N. \tag{A4}$$

They also show that

$$\left(\delta_A, c'_A\right) = \left(\Lambda_0 + Z'_A Z_A\right)^{-1}\left(\Lambda_0\phi_0 + Z'_A r_A\right), \tag{A5}$$

where

$$\Lambda_0 = k\begin{bmatrix} \sigma^2_\delta & 0 \\ 0 & \Phi_c \end{bmatrix}, \tag{A6}$$

$$Z_A = (i \quad r_N \quad r_B), \tag{A7}$$

$$\phi_0 = (\delta_0, c'_0)', \tag{A8}$$

and k is an estimate of $E(\sigma^2_u)$, described in the text.

Further, PS (2002a) show that, conditional on the data,

$$\text{var}(\tilde{\alpha}_A) = \text{tr}(V_{\phi_A}V_d) + \tilde{d}'V_{\phi_A}\tilde{d} + \tilde{\phi}'_A V_d \tilde{\phi}_A, \tag{A9}$$

where

$$V_d = \begin{bmatrix} 0 & 0 & 0 \\ 0 & V_{\alpha_N} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{A10}$$

$$\tilde{d} = \begin{bmatrix} 1 \\ \tilde{\alpha}_N \\ 0 \end{bmatrix}, \tag{A11}$$

$$V_{\phi_A} = \tilde{\sigma}^2_u\left(\Lambda_0 + Z'_A Z_A\right)^{-1}, \tag{A12}$$

V_{α_N} denotes the covariance matrix of α_N, and $\tilde{\sigma}^2_u$ denotes the posterior variance of the residuals in equation (3). See PS (2002a) for further details.

REFERENCES

Baks, Klaas, 2006, On the performance of mutual fund managers, *Journal of Finance* (forthcoming).

Baks, Klaas, Andrew Metrick, and Jessica Wachter, 2001, Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, *Journal of Finance* 56, 45–85.

Banz, Rolf, 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Bollen, Nicolas, and Jeffrey Busse, 2005, Short-term persistence in mutual fund performance, *Review of Financial Studies* 18, 569–597.

Brown, Keith, W. Van Harlow, and Laura Starks, 1996, Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry, *Journal of Finance* 51, 85–110.

Brown, Stephen, and William Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679–698.

Busse, Jeffrey, 1999, Volatility timing in mutual funds: Evidence from daily returns, *Review of Financial Studies* 12, 1009–1041.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Cohen, Randolph, Joshua Coval, and Lubos Pástor, 2005, Judging fund managers by the company they keep, *Journal of Finance* 60, 1057–1096.

Connor, Gregory, and Robert Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: A new framework for analysis, *Journal of Financial Economics* 15, 373–394.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in stock returns, *Journal of Finance* 52, 1–33.

Davis, James, Eugene Fama, and Kenneth French, 2000, Characteristics, covariances, and average returns: 1929 to 1997, *Journal of Finance* 55, 389–406.

Elton, Edwin, Martin Gruber, and Christopher Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133–157.

Elton, Edwin, Martin Gruber, Sanjiv Das, and Matthew Hlavka, 1993, Efficiency with costly information: A reinterpretation of the evidence for managed portfolios, *Review of Financial Studies* 6, 1–22.

Fama, Eugene, and Kenneth French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

Fama, Eugene, and Kenneth French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene, and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Goetzmann, William, and Roger Ibbotson, 1994, Do winners repeat? Patterns in mutual fund performance, *Journal of Portfolio Management* 20, 9–18.

Grinblatt, Mark, and Sheridan Titman, 1992, The persistence of mutual fund performance, *Journal of Finance* 47, 1977–1984.

Gruber, Martin, 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance* 51, 783–810.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of performance, 1974–1988, *Journal of Finance* 48, 93–130.

Ippolito, Richard, 1989, Efficiency with costly information: A study of mutual fund performance, 1965–1984, *Quarterly Journal of Economics* 104, 1–23.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jensen, Michael, 1968, The performance of mutual funds in the period 1945–1964, *Journal of Finance* 23, 389–416.

Jensen, Michael, 1969, Risk, the pricing of capital assets, and the evaluation of investment portfolios, *Journal of Business* 42, 167–247.

Jones, Christopher, and Jay Shanken, 2005, Mutual fund performance with learning across funds, *Journal of Financial Economics* 78, 507–552.

Kothari, S.P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169–203.

Malkiel, Burton, 1995, Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance* 50, 549–572.

Mamaysky, Harry, Matthew Spiegel, and Hong Zhang, 2003, Estimating the dynamics of mutual fund alphas and betas, Working paper, Yale University.

Moskowitz, Tobias, and Mark Grinblatt, 1999, Do industries explain momentum? *Journal of Finance* 54, 1249–1290.

Ohlson, James, and Barr Rosenberg, 1982, Systematic risk of the CRSP equally-weighted common stock index: A history estimated by stochastic-parameter regression, *Journal of Business* 55, 121–145.

Pástor, Lubos, and Robert Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* 56, 335–381.

Pástor, Lubos, and Robert Stambaugh, 2002a, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63, 315–349.

Pástor, Lubos, and Robert Stambaugh, 2002b, Investing in equity mutual funds, *Journal of Financial Economics* 63, 351–380.

Sirri, Erik, and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Stambaugh, Robert, 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.

Sunder, Shyam, 1980, Stationarity of market risk: Random coefficients tests for individual stocks, *Journal of Finance* 35, 883–896.

Zheng, Lu, 1999, Is money smart? A study of mutual fund investors' fund selection ability, *Journal of Finance* 54, 901–933.