Cross-Sectional Learning and Short-Run Persistence in Mutual Fund Performance*

Joop Huij and Marno Verbeek[†]

This version: October 10, 2006

Abstract

Using monthly return data of more than 6,400 US equity mutual funds we investigate short-run performance persistence over the period 1984–2003. We sort funds into rank portfolios based on past performance, and evaluate the portfolios' out-of-sample performance. To cope with short ranking periods, we employ an empirical Bayes approach to measure past performance more efficiently. Our main finding is that when funds are sorted into decile portfolios based on 12-month ranking periods, the top decile of funds earns a statistically significant, abnormal return of 0.26 percent per month. This effect persists beyond load fees, and is mainly concentrated in relatively young, small cap/growth funds.

Keywords: mutual funds, performance persistence, Bayesian analysis

[†]Huij and Verbeek are at RSM Erasmus University. Corresponding author is Huij. Email addresses are jhuij@rsm.nl and mverbeek@rsm.nl.

Cross-Sectional Learning and Short-Run Persistence in Mutual Fund Performance

This version: October 10, 2006

Abstract

Using monthly return data of more than 6,400 US equity mutual funds we investigate short-run performance persistence over the period 1984–2003. We sort funds into rank portfolios based on past performance, and evaluate the portfolios' out-of-sample performance. To cope with short ranking periods, we employ an empirical Bayes approach to measure past performance more efficiently. Our main finding is that when funds are sorted into decile portfolios based on 12-month ranking periods, the top decile of funds earns a statistically significant, abnormal return of 0.26 percent per month. This effect persists beyond load fees, and is mainly concentrated in relatively young, small cap/growth funds.

Keywords: mutual funds, performance persistence, Bayesian analysis

I Introduction

Despite the large growth of the mutual fund industry during the last decades, empirical evidence does not appear to unambiguously indicate that some funds managers systematically outperform passive benchmarks. While studies by, among others, Hendricks et al. (1993), Goetzman and Ibbotson (1994), Elton et al. (1996), and Gruber (1996) report persistence in superior performance, Carhart (1997) demonstrates that most of this is explained by differences in exposures to common risk-factors, including size, book-to-market, and one-year momentum. These results do not support the existence of "skilled" managers who are able to outperform passive benchmarks after expenses. Nevertheless, using high frequency daily data Bollen and Busse (2005) find short-term persistence in superior performance beyond momentum. Their results indicate that performance persistence is observable using short measurement horizons. In the present paper, we investigate three issues regarding this phenomenon. First, do we also observe this predictability in a large sample of funds over a more extensive (and recent) time period? Second, is the persistence effect economically significant? And third, does persistence vary across the cross-section of mutual funds? To investigate these issues, we employ a large sample of monthly fund returns in combination with a Bayesian estimation method to measure performance.

The typical approach to investigate persistence in fund performance is to sort funds into rank portfolios based on a measure of past performance (e.g. three- or four-factor alphas), and evaluate the rank portfolios' subsequent performance. However, particularly when short horizons of monthly returns are used, pre-ranking alphas are hampered by potentially high levels of inaccuracy. With only a small number of observations available, it is notoriously difficult to separate managerial skill from simple luck. Funds that are less well-diversified and have higher levels of non-systematic risk experience a larger probability of ending up with an extreme ranking because the managers of these funds typically place larger bets. Consequently, the top and bottom deciles in performance rankings will to a large extent be attributable to simple luck rather than managerial skill. When subsequent performance of the ranked funds is analyzed, results may be biased towards finding no relation between past and future performance of these funds.

While it seems that the use of longer return histories leads to more accurate inferences when estimating a fund's performance, there are a number of disadvantages. First, fund

performance may vary over time, for example relating to a change of fund manager or the age of the fund. Second, the investment style of the fund may change, resulting in time varying exposures to common risk or benchmark factors, see e.g. Ferson and Schadt (1996). Third, longer histories are only available for funds that have been active over an extended period. Focusing upon, e.g., three- or five-year return histories substantially reduces the number of available funds and may lead to several survival-related biases as documented by Brown et al. (1992). Finally, a stream of literature suggests that even though managerial skill is heterogeneous across fund managers, the relation between past performance and subsequent fund flows as documented by Sirri and Tufano (1998) causes performance persistence to fade out quickly (see Berk and Green (2004) and Zhao (2004)).

On the other hand, the use of higher frequency data (e.g. daily returns) is hampered by its limited availability, leading to a substantial reduction in the cross-section of funds or the relevant time span that can be investigated. Consequently, it is useful to consider more accurate ways to estimate performance measures on the basis of short monthly return histories. Recently, several papers have advocated the use of Bayesian alphas to measure fund performance. Using this approach prior information related to funds' expenses, investors' beliefs about managerial skills, benchmark pricing abilities, or the returns on other mutual funds and benchmark factors are incorporated in the resulting estimates (see Baks et al. (2001) and Pastor and Stambaugh (2002b,a)). However, although these approaches allow for more efficient inference, Jones and Shanken (2005) and Busse and Irvine (2006) demonstrate that the predictive accuracy of Bayesian alphas is greatly affected by the investor's prior belief about managerial skill.

In this paper, we consider Bayesian alphas exploiting the entire cross-section of mutual fund returns. This approach does not require investors to explicitly formulate their beliefs about managerial skill, or to make assumptions about cross-sectional characteristics that drive performance. Rather, inference is based upon monthly returns only. As stressed by Jones and Shanken (2005) such an approach can be motivated by cross-sectional learning of investors. In this case, investors' prior beliefs are not independent across funds and an investor's expectation about the performance of a fund is partly a belief about the abilities of mutual fund managers as a group. On the other hand, the use of Bayesian estimators can be motivated purely on the basis of statistical arguments. When the cross-section of alpha

estimates is considered, the top and bottom decile are more than average subject to positive and negative estimation errors, respectively, and shrinking them towards the cross-sectional average may result in more accurate inferences.

The main advantage of our approach is that it is based on monthly return data only, which are much more readily available than daily frequency data. Based on the entire CRSP universe of more than 6,400 US equity funds over the past 20 year we find strong evidence that superior performance persists in the near future. When we rank funds on Bayesian alphas over the past 12 months, we find that the top decile of funds earns a statistically significant, superior return of 0.26 percent in the month after ranking. While these results confirm the findings of Bollen and Busse (2005) based on a sample of 230 mutual funds, it has not yet been investigated whether this effect is economically significant. It could be questioned whether superior performance persists beyond the transaction costs that are involved with a strategy of chasing winning funds, or that the effect is canceled out by the load fees that are charged by the majority of funds. Because we are able to study persistence across a large sample of funds, restricting our attention to a specific type leaves us with enough funds in the cross-section. When we only include strictly no-load funds in our sample, we find clear evidence that superior performance persists beyond load fees. Likewise, we are able to focus attention to subsamples of funds characterized by investment style or other characteristics such as age. Our findings indicate that superior performance varies across different types of funds, and is mainly concentrated in relatively young, small cap/growth funds.

The remainder of the paper is organized as follows. In the next section we briefly discuss the literature on Bayesian performance evaluation and learning priors, and introduce our approach. In Section 3, we describe our data. Section 4 presents the results of an analysis in which we analyze the efficiency of Bayesian alphas. In Section 5, we analyze performance persistence of US equity funds at different horizons, using monthly data from the CRSP Database over the period 1984–2003. We perform several robustness tests in Section 6. Finally, Section 7 concludes and discusses the main findings.

II Mutual fund performance measurement and Bayesian estimation

Mutual fund performance is often measured by a fund's alpha, defined as the intercept term in a regression of the fund's excess returns on the excess returns of one or more benchmark factors. Denoting excess returns of benchmark factor j in period t by x_{jt}, consider the following linear model

$$(1) \quad r_{it} = \alpha_i + \beta_{1i}x_{1t} + \beta_{2i}x_{2t} + \ldots + \beta_{ki}x_{kt} + \varepsilon_{it},$$

where r_{it} denotes the excess return of fund i in period t, β_{ji} denotes the sensitivity of fund i to factor j $(j = 1, ..., k)$, and ε_{it} denotes the residual return. The intercept term α_i measures the expected return of fund i in excess of a factor-mimicking benchmark portfolio. As a result, alpha can be thought of the portion of return that can be attributed to the fund manager. We can write the above model in matrix notation as

$$(2) \quad r_i = X_i \theta_i + \varepsilon_i,$$

where r_i denotes a vector of the excess returns of fund i, X_i denotes the matrix of excess returns of the benchmarks factors, including an intercept, θ_i denotes a $k{+}1$ vector of unknown parameters, containing alpha and the k factor sensitivities of fund i, and ε_i denotes the vector of the residual returns. The length of the evaluation period is denoted by T_i, which may be different across funds.

Traditionally, the above model is estimated by ordinary least squares (OLS). However, particularly with only a small number of observations available, the measurement error (or sampling error) of alpha estimates may be substantial. For example, consider the case where all true alphas are very close to zero. When a relatively high alpha estimate is reported for a given fund, this is likely to be an overestimate of the fund's true alpha. On the other hand, relatively low alpha estimates are likely to be underestimates of their true values. In the extreme case where there is no ability and all alphas are equal to zero, a large positive or negative alpha estimate is solely due to luck rather than managerial skill.

While OLS provides (under appropriate conditions) unbiased or asymptotically unbiased estimators for the underlying true alphas, it does not perform very well in reproducing the cross-section of true fund alphas. To obtain alpha estimates that more closely resemble

the true underlying distribution of managerial skills, it is advisable to shrink, in some way, the raw alpha estimates towards zero, or some other value for the (long-run) cross-sectional average. Shrinking individual estimates towards a common mean has been advocated in the 1970s for the estimation of security market betas of individual stocks, see Blume (1971) and Vasicek (1973). Currently, commercial information services like Bloomberg and Merrill Lynch provide simple routines to compute adjusted betas based upon these ideas. More recently, several papers have advocated the use of more advanced shrinkage procedures within a formal Bayesian framework to estimate mutual fund alphas, see, e.g., Baks et al. (2001), Pastor and Stambaugh (2002b,a), Jones and Shanken (2005), and Busse and Irvine (2006). A full Bayesian approach requires the specification of a prior distribution for the parameters of interest, which may be based upon different assumptions about investors' prior beliefs about managerial skill and fund characteristics that drive performance. For example, an investor might incorporate, among other things, the stylized fact that funds with relative high expenses generally underperform those with lower expenses, see e.g. Elton et al. (1993). Intuitively, estimates of alpha are shrunk towards the negative of the funds' expenses conditional on how strongly one beliefs in managerial skill in general.

However, such a set-up assumes that there is a negative and monotonic relationship between funds' alphas and expenses. While empirical analyses indicate that the worst performing funds indeed have substantially higher expenses than the average fund, it does not appear to be the case that the best performing funds have below average expenses (see e.g. Gruber (1996) and Carhart (1997)). In fact, when the worst performing funds are excluded, the relation between funds' expenses and performance seems rather more U-shaped than monotonically decreasing. More generally, the use of Bayesian alphas typically requires the formulation of prior, often subjective, beliefs, which may be inappropriate. While using informative priors may introduce a substantial subjective element in performance evaluation, using less informative priors makes it more difficult to identify the worst and best performing funds. For example, Jones and Shanken (2005) and Busse and Irvine (2006) demonstrate that the accuracy of Bayesian alphas is greatly affected by the prior beliefs that are attributed to the investor.

Another issue with the formulation of subjective priors is the following: Studies by among others Malkiel (1995) and Brown and Goetzmann (1995) suggest that the degree to which

managerial skill persists depends upon the time period observed. Consequently, a diffuse belief in managerial skill might dominate a sceptical belief over a certain time period, while it might yield inferior results over a different time period. These market dynamics make it necessary to allow beliefs in managerial skill to be time-varying.

In this study we incorporate the large cross-section of mutual fund alphas in measuring the skill of an individual fund manager. By allowing the prior to learn across other mutual funds, the resulting belief in managerial skill is no longer fully subjective, but entirely data-based. Furthermore, by incorporating cross-sectional information in this way, prior beliefs about alpha are 'adaptive' in the sense that they can vary over time and adjust to market dynamics and structural shifts.

To describe our Bayesian procedure, we start by specifying the cross-sectional distribution of funds' alphas and betas as normal

$$(3) \quad \theta_i \sim N\left(\mu, \Sigma\right),$$

where $\theta_i = (\alpha_i, \beta_{1i}, ..., \beta_{ki})'$, μ denotes a $(k+1)$-dimensional vector of cross-sectional means of alphas and factor sensitivities, and Σ denotes a $(k+1)$ by $(k+1)$ covariance matrix. Assuming that the error terms in (2) are $IIN(0, \sigma_i^2)$, the posterior distribution of θ_i is normal with expectation

$$(4) \quad E(\theta_i) = \left(\frac{1}{\sigma_i^2} X_i' X_i + \Sigma^{-1}\right)^{-1} \left(\frac{1}{\sigma_i^2} X_i' X_i \hat{\theta}_i + \Sigma^{-1} \mu\right),$$

where $\hat{\theta}_i$ denotes the OLS estimate, and σ_i^2 denotes the variance of ε_i (see for example Maddala et al. (1997)). The corresponding covariance matrix $V(\theta_i)$ is given by

$$(5) \quad V(\theta_i) = \left(\frac{1}{\sigma_i^2} X_i' X_i + \Sigma^{-1}\right)^{-1}.$$

Equation (4) shows that the posterior "estimates" of alpha and betas are a matrix-weighted average of the OLS estimates $\hat{\theta}_i$ and the prior μ. We can interpret this equation as a shrinkage formula, which shrinks the raw OLS estimates towards a common mean, where the degree of shrinkage depends on the precision of the OLS estimates and the cross-sectional dispersion reflected in Σ. For example, raw alpha estimates for funds with higher levels of non-systematic risk are shrunk more towards the overall mean than passively managed funds, such as index-trackers, with lower levels of non-systematic risk. Similarly, the degree of shrinkage is higher

for funds with short return histories. Note that shrinkage is applied to alphas as well as to the factor sensitivities, which may also be inaccurately estimated on the basis of short time series.

Bayesian alphas (and factor sensitivities) are determined using (4) and require a choice for the hyperparameters μ and Σ. Instead of fixing these parameters at some a priori values, we use the entire cross-section of funds to estimate μ, the cross-sectional mean, and Σ, the cross-sectional covariance matrix. A range of alternative approaches is available to implement this, including maximum likelihood estimation, (iterative) empirical Bayesian approaches, and Gibbs sampling (see Blattberg and George (1991), Casella and George (1992), Maddala et al. (1997), and Stern (2000)). Basically, the (iterative) empirical Bayesian approach initializes the (hyper)parameter with OLS estimates, and then re-estimates the (hyper)parameters until a pre-specified degree of convergence is reached. Using Gibbs sampling, conditional draws are simulated from the posterior distribution to generate a sample of this distribution rather than deriving it analytically. From Markov chain theory it follows that the simulated sample will eventually converge to the true density function of the posterior distribution. Parameter values are then estimated from this simulated sample.

III Data

Consistent with most of the literature, we focus on non-specialized US equity funds. Monthly return data are extracted from the December 2003 CRSP database, which goes back to 1962. Given the results of Elton et al. (2001), who report that the differences in returns reported by CRSP and Morningstar are marginal when data are used after the mid-1980s, we examine the period 1984–2003. To select funds in our data set we follow a procedure similar to the one used by Pastor and Stambaugh (2002b). We use the information that CRSP provides about classifications by Wiesenberger, Micropal/Investment Company Data, Inc., Strategic Insight, and the funds themselves. We exclude funds with unknown objectives, no expense, or load data in the annual summary at the end of each previous year. Additionally, we exclude flexible funds, bond funds, mortgage-backed funds, multi-manager funds, money market funds, balanced funds, funds that invest in precious metals, international funds, and specialized funds. From the remaining funds, we select funds that are classified as small cap/growth, growth, or growth & income. These are the same types of funds that are included

in the analyses of Carhart (1997) and Bollen and Busse (2005). In addition, we construct two subsamples of funds by whether or not they charge load fees. Exact details on our selection procedure are provided in the Appendix. Table 1 lists the number of available funds over the period 1984–2003. In total, our sample covers 6,429 equity funds that have more than 12 consecutive returns over the whole sample period. Since CRSP includes all funds that existed during this period, our data are free of survivorship bias as documented by Brown et al. (1992). Our subsamples cover 3,101 no-load, 2,134 small cap/growth, 3,349 growth, and 1,837 growth & income funds, respectively.

Throughout this study, fund performance is measured using the four-factor model employed by Carhart (1997):

$$(6) \quad r_{it} = \alpha_i + \beta_{1i}RMRF_t + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}UMD_t + \varepsilon_{it},$$

where r_{it} is the return of fund i in month t in excess of the one-month Treasury bill rate, $RMRF_t$ is the excess return on the value-weighted aggregate market proxy, SMB_t and HML_t are returns of value-weighted, zero-investment factor-mimicking portfolios for size and book-to-market (Fama and French, 1992, 1993), and UMD_t captures one-year momentum (Jegadeesh and Titman, 1993) in month t, respectively. The intercept α_i is a measure for the performance of fund i, while β_{ji} is the sensitivity of fund i for benchmark factor j. Monthly returns on the benchmark factors are obtained from Kenneth French's webpage, while one-month Treasury bill rates are obtained from Ibbotson and Associates.

IV Efficiency of Bayesian alphas

Below we report the results of a simulation study to evaluate the efficiency of Bayesian alphas and compare them to those of standard OLS estimates. Additionally, we compare the implications for fund rankings based on the two alternatives. We use actual data of mutual fund and factor returns to determine the parameter values for the return generating processes. This ensures that data are generated in a realistic way, while the experimental setting allows us to compare estimates with the underlying true values that are used to generate the returns.

First, we estimate the long-run performance for each fund in our sample over the period 1984–2003. To do so, we estimate alpha and the benchmark sensitivities in equation (6) using standard OLS and the entire return history of each fund. Sample averages of the resulting

parameter estimates and benchmark returns are reported in Table 2. Consistent with the current literature we find that the average fund underperforms the employed benchmarks with an average alpha of -0.22 percent per month. We use equation (6) to generate simulated samples of mutual fund returns. The simulated factor returns, fund alphas, and benchmark sensitivities are all drawings from an IID normal distribution with mean and standard deviation corresponding to the empirical estimates reported in Table 2. We also experiment with a GARCH(1,1) specification for the benchmark return variances. These results are qualitatively similar to those reported here, and are available upon request. The idiosyncratic returns (ε_{it}) are generated from a serially uncorrelated normal distribution with mean zero and variance σ_i^2. Following Brown et al. (1992), Hendricks et al. (1997), and ter Horst et al. (2001) we allow non-systematic risk to depend upon the fund's market beta β_{1i} as

$$(7) \quad \sigma_i^2 = \omega(1 - \beta_{1i})^2,$$

where ω captures the relation between non-systematic risk and market beta. We hereby impose that index funds, with a market beta close to one, typically have low values of non-systematic risk, while actively managed portfolios whose market betas deviate from one are generally less well diversified. In the simulations, we set the value for ω to 0.0050, which closely corresponds to the values reported by Brown et al. (1992) and ter Horst et al. (2001).

Simulated fund returns are generated for measurement horizons of 12, 24, 36 and 60 months. The number of funds in the cross-section is set to 1,000. We then estimate the factor models using standard OLS and our Bayesian estimator. Bayesian alphas are estimated using an empirical Bayes approach. Per simulation we evaluate the accuracy of OLS and Bayesian alphas, as well as their ranking ability. The number of replications for each measurement horizon is set to 1,000. The main question that we address is to what extent Bayesian alphas provide better estimates of (relative) fund performance than standard OLS estimates. An additional point of interest is whether this difference in accuracy remains substantial when more observations are available. Furthermore, to test whether our results are robust to non-normal errors, we run additional simulations where we generate the error terms from Student's t-distribution with 5 degrees of freedom, keeping all parameters at their previous values.

As a measure of accuracy we consider the cross-sectional average of the root mean squared error (RMSE) of both estimators. RMSE is the root of the average squared distance between the estimates and the true value of alpha, and can be interpreted as a standard error. $RMSE$

is computed as

$$(8) \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\alpha_i - \hat{\alpha}_i)^2},$$

where N is the number of funds in the cross-section, α_i is fund i's true alpha, and $\hat{\alpha}_i$ its estimate. To investigate the benefit of sorting on Bayesian alphas in the context of persistence analysis, we assign funds to decile portfolios based on their estimated rankings, and compute the average true alpha for the top and bottom deciles. The results of this exercise are reported in Table 3. Clearly, the average RMSEs of standard OLS are substantially higher than those of Bayesian alphas. For example, using a measurement horizon of 12 months, the RMSE of OLS is 0.67 percent, while this is 0.38 percent for Bayesian estimation. Hence, for this measurement horizon, Bayesian alphas are about 40 percent more accurate than standard OLS estimates. We find that the advantage of Bayesian estimation remains substantial when up to 60 observations are available. For measurement horizons of 24, 36, and 60 months, Bayesian alphas are about 30 percent, 25 percent, and 20 percent more accurate than standard OLS estimates.

Furthermore, the average true alpha of the top decile is higher for rankings on Bayesian alphas than for rankings on standard OLS alphas. From Table 3 it follows that the average true alpha of the top decile based on Bayesian alphas exceeds that of the top decile based on OLS alphas by 0.13 percent for measurement horizons of 12 months. For measurement horizons of 24, 36, and 60 months, these excess returns are 0.08 percent, 0.05 percent, and 0.03 percent, respectively. In a similar fashion, the results for the bottom decile indicate that Bayesian alphas are better in identifying poor performing funds. All differences are highly significant. In fact, the empirical p-values are zero for all simulations. The results of the simulations with t-distributed error terms are consistent with the results of the simulations with normal error terms in the sense that Bayesian alphas have lower RMSEs and produce better fund rankings. In fact, the difference between both estimators becomes larger. While the presence of "fat tails" increases the RMSE of standard OLS by about 30 percent, irrespective of the length of the estimation window, the corresponding increase for Bayesian alphas is only half. In an unreported experiment, we generate the error terms from Student's t-distribution with 3 degrees of freedom, keeping all parameters at their previous values. Compared to the results of the simulation with 5 degrees of freedom, we observe an increase in the difference between

both estimators. These results indicate that the advantage of using Bayesian alphas over standard OLS becomes larger when the data are more leptokurtic, e.g. exhibit "fatter tails".

V Empirical results

A Methodology

In this section, we employ Bayesian alphas to study short-run persistence in the performance of US equity mutual funds over the period 1984–2003. The typical setup is to rank funds based on a measure of past performance, and analyze the relation between rank and subsequent performance. Hendricks et al. (1993) study relative performance persistence for 165 no-load growth funds over the period 1974–1988 using quarterly returns. In each quarter the authors assign funds to octile portfolios based on funds' raw returns over the past one to eight quarters. Post-ranking alphas are estimated for the stacked time series of equally weighted portfolio returns using several benchmark sets. Their results indicate that fund performance tends to persist in the near future. Using a monthly sample of 188 funds over the period 1977–1993, Elton et al. (1996) rank funds on one- and three-year alphas with respect to several benchmark indices, and demonstrate that when fund rankings are based on risk-adjusted returns, performance persists over holding periods up to three years. However, Carhart (1997) demonstrates that almost all of this persistence is explained by expenses and the one-year momentum effect documented by Jegadeesh and Titman (1993). He replicates the methodology of Hendricks et al. (1993) using a survivorship-bias free data set containing monthly returns of 1,892 funds over the period 1962 to 1993. When post-ranking alphas are estimated using a four-factor model that includes momentum, the spread in alphas between the portfolios disappears. Additionally, portfolios are formed based on four-factor alphas over the previous three-years. While there appears to be some consistency in risk-adjusted rankings, these results are only significant for the lowest ranked funds. More recently, using daily return data on 230 funds over the period 1985–1995, Bollen and Busse (2005) find significant short-term persistence in fund performance beyond momentum when short measurement horizons are used. Every quarter, funds are sorted into decile portfolios based on three-month ranking periods, and the portfolios' performance is evaluated over the subsequent quarter. Using this framework, they find that the top decile earns a statistically significant positive alpha over the subsequent period. When they employ longer measurement horizons,

the superior performance disappears. These results suggest that superior performance is a short-lived phenomenon that is only observable using short measurement horizons.

We investigate short-run performance persistence using the entire CRSP universe of mutual funds. For the main tests of our empirical analysis, we form dynamic portfolios of mutual funds based on past performance, and evaluate the portfolios' post-ranking performance. Each month, funds are sorted into rank portfolios on their estimated alphas over the past period with respect to proxies for market, size, value, and momentum. Funds are required to have a complete return history available over the employed ranking period. Pre-ranking alphas are estimated using an empirical Bayes approach. We follow Hendricks et al. (1993) in assigning the funds to deciles based on their rank and calculate the equally weighted returns over the subsequent month for all rank portfolios. Portfolio weights are readjusted if a fund disappears after ranking. Finally, we estimate post-ranking alphas for the stacked time series of portfolio returns using OLS regression with Newey-West standard errors.

B Short-run persistence in superior performance

We investigate performance persistence using measurement horizons of 36 and 12 months. We not only build portfolios of funds based on Bayesian alphas, but also on raw returns and standard OLS alphas as in Hendricks et al. (1993) and Carhart (1997). More specifically, portfolios of mutual funds are formed from January 1984 to December 2003, yielding a time series of monthly portfolio returns over a period of 20 years. The analysis is conducted on our entire sample of funds. We start by sorting funds into decile portfolios based on 36-month ranking periods. In the average month, 980 funds are available ranging from 261 funds in 1984 to 2,982 funds in 2003. To measure post-ranking performance, we additionally estimate the portfolios' raw returns and Sharpe ratios. As we discuss in more detail in Section VI, using different performance measures over the pre- and post-ranking period is one way to prevent spurious persistence patterns arising due to a misspecification of the performance evaluation model. The results in Table 4 indicate a relation between past and future performance over measurement horizons of 36 months. Higher ranked funds seem to earn higher post-ranking returns. However, most of this predictability is explained by differences in the funds' exposures to common risk factors. When we employ the Sharpe ratio to measure post-ranking performance, the findings are only a weak indication of such relation. Further, when

we consider the deciles' post-ranking alphas, it appears that all predictability is concentrated in the bottom deciles. Only the worst performing funds earn post-ranking alphas that are significantly different from zero. The top decile of funds earns a post-ranking alpha that is economically and statistically indistinguishable from zero. Apparently, fund managers are unable to outperform the passive benchmarks over longer periods of time. When we consider the portfolios' sensitivities to the benchmark factors, it appears that higher ranked funds are more exposed towards growth stocks (negative loading on HML), while lower-ranked funds are more exposed to value stocks (positive loading on HML). Further, funds with more extreme rankings are more exposed to SMB. Also, the pattern in R^2-values points out that funds in the more extreme deciles take more non-systematic risk. All in all, our results indicate relative performance persistence at a three-year measurement horizon. However, this effect is mainly concentrated in persistently underperforming funds. The top decile of funds appear to earn back their expenses, while the bottom deciles strongly underperform the market. These results are very similar to those of Carhart (1997), and hold irrespective of which method of estimation is used to rank the funds.

Next, we rerun the analysis with a measurement horizon of 12 months. In the average month, 1,461 funds are available with at least 12 months of return history ranging from 306 funds in 1984 to 4,348 funds in 2003. The results of this analysis, reported in Table 5, indicate that short-run performance predicts future performance better than performance estimated over longer measurement horizons. The expected alpha of the top (bottom) decile for rankings based on 12-month measurement horizons is substantially higher (lower) than that of rankings based on 36-month horizons. Furthermore, the relation between past and future performance is close to monotonic. These results hold irrespective of which method of estimating alpha is used. Moreover, Bayesian alphas appear to capture persistent superior performance of the top decile of funds. We find a statistically significant, superior return of 0.26 percent per month for the top decile when funds are ranked on Bayesian alphas. In contrast, sorts on raw returns and standard OLS alphas yield post-ranking alphas that are not or only marginally significant. It should be stressed that in this analysis the different pre-ranking performance measures are only relevant to the extent that they result in a different composition of the top (and bottom) deciles. For example, the funds in the OLS and Bayesian top deciles have an overlap of slightly less than 75 percent. Further, we observe similar patterns for the deciles'

sensitivities to the benchmark factors: higher ranked funds are more growth oriented, and funds with extreme rankings are more exposed to the size factor.

C   Lifespan of persistence and fund age

Bollen and Busse (2005) find that the persistent outperformance of the top decile in their sample lasts over a period of slightly more than a month, while the bottom decile's underperformance declines at a constant rate over time. To investigate this issue in our sample, we rank funds on one-month lagged alpha estimates over the past 12 months, e.g. fund rankings in February 1984 are based on alphas estimated over the period January 1983 to December 1983. Consistent with Bollen and Busse, our results in Table 6 indicate that the superior performance of the top decile is only short-lived. While the top decile is still earning a positive alpha of 0.10 percent per month, this value is not significantly different from zero. We also find that the bottom decile's underperformance persists strongly. In fact, in tests not reported, we find that lagged alpha estimates over periods of 5, 11, and even 35 months are predictive for the bottom decile's future underperformance. This finding is consistent with persistent negative abnormal returns driven by funds' extremely high expenses as documented by Carhart (1997).

A potential explanation for the difference between the analyses with measurement horizons of 36 and 12 months is that the results are based on different samples of funds. In particular, the use of a measurement horizon of 36 months excludes relatively young funds from the analysis. This greatly influences the relevant cross-section of funds. By requiring funds to have 36 observations available over the past three years, on average more than one third of the existing funds are eliminated from our sample. When a full return history over the past year is required, less than 5 percent of the funds is left out our analysis. Furthermore, we find that almost 40 percent of the dead funds listed in our sample have been active for less than three years. Ignoring this group of funds might have substantial implications for inferences about the mutual fund industry as a whole, inducing some sort of survivorship bias. To investigate this issue, we conduct the analysis with the measurement horizon of 12 months on the same sample of funds that is used in the 36-month analysis. Likewise, we analyze performance persistence on a sample of only relative young funds (return history of 12 to at most 35 months). In the average month, 977 funds are classified as "old", and 480

funds as "young". The results of this analysis are presented in Table 7. When the analysis is conducted on older funds, the top decile of funds earns a statistically significant post-ranking alpha of 0.21 percent per month. When only relatively young funds are included in the analysis, the top decile's post-ranking alpha equals 0.35 percent per month, suggesting that the persistence effect is substantially stronger among for relatively young funds. It also seems that the bottom decile's underperformance is lower for younger funds. Most likely this can be attributed to the fact that a large part of the dead funds is left out of the analysis when a longer return history is required.

D   Economic significance and persistence across different types of funds

So far, we have demonstrated that the use of Bayesian estimation enables us to capture persistence in superior performance using monthly data. Our results, based on the entire CRSP universe of US equity funds over an extensive time period, complement the evidence provided by Bollen and Busse (2005) on the existence of short-run performance persistence. However, several additional questions are raised by these findings. For example, it is unclear whether the reported persistence is economically significant; a strategy of investing in past winners could involve significant transaction costs. Furthermore, it is not clear what the underlying economic reasons for these findings are. Empirical results on the relation between performance persistence and fund characteristics, such as investment style, may provide an insight into this phenomenon. To answer these questions, we repeat the analysis using particular subsets of funds, e.g. no-load funds. While doing so reduces the available cross-section of funds considerably, thereby exacerbating noise in portfolio returns, the benefit of our Bayesian approach in this context is evident. Because we only require monthly return data, which are much more readily available than daily frequency data, restricting attention to a specific type leaves us with a sufficiently large number of funds in the cross-section.

We first investigate whether superior performance persists beyond the transaction costs that are involved with a strategy of chasing winning funds. Only the 3,101 funds that do not charge load fees are included in the analysis. As before, funds are sorted into decile portfolios based on 12-month alpha estimates, and the rank portfolios post-ranking performance is evaluated. The results are presented in Panel A of Table 8, and provide clear evidence that the superior performance of the top decile (0.31 percent per month) persists beyond load fees.

Apparently, the top funds' superior performance is not canceled out by load fees that are involved with a strategy of chasing winners. In fact, it even seems that managerial skill plays a more important part in the performance for no-load funds than for otherwise similar funds. To investigate this issue more closely, we also perform the analysis on funds that do charge load fees. Our sample covers 4,025 load funds. The results of this analysis are in Panel B of Table 8. With a post-ranking alpha of 0.20 percent per month for load funds versus 0.31 percent per month for no-load funds, our findings add to the evidence that load funds tend to underperform no-load funds (see Morey (2003)).

Second, we analyze persistence across different investment styles. Funds are sorted into deciles on 12-month raw returns, standard OLS alphas, and Bayesian alphas. Table 9 lists the top deciles' performance for the different types of funds. The post-ranking alpha of the top decile of small cap/growth funds has a statistically significant value of 0.24 percent per month (using Bayesian estimation). For growth funds and growth & income funds, the top decile's post-ranking alphas have (insignificant) values of 0.20 and 0.08 percent per month, respectively. When funds are ranked on raw returns or OLS alphas, we find only marginal evidence of persistence in superior performance for small cap/growth funds; the type of fund among which this effect manifests itself most strongly.

Interestingly, none of the three style-specific top deciles outperforms the top decile based on the pooled sample of funds. This indicates that the style composition of the top decile in Table 5 varies over time and that part of its superior performance can be attributed to style timing, i.e. being in the right investment style at the right time. Since we observe more homogeneity in funds' factor exposures within samples of funds with the same classification, one could expect the prior for alpha to be more precise when this value is based on funds with the same classification than when based on the entire cross-section of funds. Therefore, we perform an analysis where funds' raw alphas are shrunk to the cross-sectional average of similar funds. Using this setup, not only is the value of the prior dependent on the funds' classifications, but also the degrees of shrinkage varies across different types of funds. The results of this analysis are presented in Table 10. Compared to the results in Table 5 the post-ranking alpha of the top decile increases to 0.28 percent per month, with a t-value of 2.41. Apparently, the benefits from using conditional priors are only marginal.

Up to this point, the empirical results indicate that the Bayesian approach works well in

capturing persistence in superior performance of the top decile of mutual funds. Even though the differences are sometimes small, Bayesian alphas are uniformly outperforming raw returns and standard OLS alphas; both in terms of economic and statistical significance. To compare the predictive power of standard OLS and Bayesian alphas formally, we perform a Diebold-Mariano test (Diebold and Mariano, 1995). This test basically involves a comparison of the average forecast error of both estimators, and is the standard approach in the econometric literature to compare forecasts. At the begin of each year during our sample period 1984–2003, we compute the root mean squared error (RMSE) of alpha forecasts over the preceding 12 months using both estimators, relative to the realized values over the subsequent year. Hereby, realized alphas are estimated using standard OLS. While OLS estimates obviously do not correspond to true alphas, they provide an unbiased measure of a fund's ex-post performance. We then consider the two sequences of forecasts errors, and test whether the average difference in RMSE between both estimators is equal to zero. The variance is computed using the Newey-West estimator. The analysis is performed on all subgroups of funds we defined earlier. Table 11 lists the results. It appears that the RMSE of Bayesian estimates is substantially (and significantly) lower than that of OLS alphas. When we consider the complete sample of funds, the average RMSE is 1.12 percent per month using OLS alphas, and 0.91 percent per month using Bayesian alphas. These outcomes are consistent and in the same order of magnitude across all subgroups of funds.

VI   Robustness

A   Model specification

In this subsection, we report the results of several robustness tests. We first investigate whether our findings of short-run persistence are robust to a potential misspecification of the factor model we employ to evaluate fund performance. Misspecification of the employed factor model can lead to spurious persistence results, since any misspecification is also likely to persist. In particular, we investigate the sensitivity of the found persistence patterns to industry effects. For this purpose, we employ the approach advocated by Pastor and Stambaugh (2002b), and Jones and Shanken (2005) to evaluate post-ranking performance. Using this approach, we employ principal components analysis (PCA) to extract factors that represent returns in specific industries that are not captured by the funds' sensitivities to the common

risk factors. More specifically, excess returns on the Fama and French 30 industry portfolios are regressed on a constant and the four Carhart (1997) factors. PCA is conducted on the time series of the residuals of each regression plus the intercept in that regression. Finally, the first three normalized components are taken as weights for the industry portfolios. The monthly return data on the 30 industry portfolios are obtained from Kenneth French's webpage. We augment our base four-factor model with the first three factor-mimicking principal components:

$$
(9) \quad r_{it} = \alpha_i + \beta_{1i}RMRF_t + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}UMD_t +
$$
$$
\beta_{5i}IP1_t + \beta_{6i}IP2_t + \beta_{7i}IP3_t + \varepsilon_{it},
$$

where $IP1$, $IP2$, and $IP3$ are the three industry factors.

Additionally, we investigate the sensitivity of our results to time-varying exposures to common risk factors. For this purpose we employ the conditional performance measure of Ferson and Schadt (1996) to evaluate the fund deciles' post-ranking performance. Using this approach, we extend our factor model with interaction terms of the benchmark returns with a set of pre-determined information variables:

$$
(10) \quad r_{it} = \alpha_i + \beta_{1i}RMRF_t + B'_{1i}[z_{t-1}RMRF_t] + \beta_{2i}SMB_t + B'_{2i}[z_{t-1}SMB_t] +
$$
$$
\beta_{3i}HML_t + B'_{3i}[z_{t-1}HML_t] + \beta_{4i}UMD_t + B'_{4i}[z_{t-1}UMD_t] + \varepsilon_{it},
$$

where z_{t-1} denotes a vector holding the lagged values of the information variables, and vector B'_{ji} captures the response of fund i's exposure to benchmark j to this information. Following Ferson and Schadt (1996), we employ the following information variables: (1) the lagged level of the one-month Treasury bill yield, (2) the lagged dividend yield of the S&P500, (3) a lagged measure of the slope of the term structure, and (4) a lagged default spread in the corporate bond market. The dividend yield is defined as the dividend per share as a percentage of the share price over the past 12 months for the index. The term spread is defined as a constant-maturity 10-year Treasury bond yield minus the 3-month Treasury bill yield. And the default spread is defined as Moody's BAA-rated corporate bond yield minus the AAA-rated corporate bond yield. One-month Treasury bill rates are obtained from Ibbotson and Associates, data on the dividend yield on the S&P500 are obtained from Thompson Financial, and the bond yield are obtained from the webpage of the Federal Reserve Bank. The resulting factor model includes 20 scaled factors, and an intercept. We evaluate the deciles post-ranking performance for fund rankings on 12-month raw returns, standard (four-factor) OLS alphas,

and (four-factor) Bayesian alphas.

The results of the analyses are in Table 12. The differences between the alternative factor models used for evaluation are typically small. Irrespective of the model that is employed, the top decile based on Bayesian alphas over the past 12 months delivers a statistically significant positive alpha of about 0.3 percent per month. These results indicate that our finding of abnormal returns for the top decile is robust to alternative factor models and cannot be explained by the conditional performance measure of Ferson and Schadt (1996). Further, it suggests that the reported performance persistence is unlikely to be caused by misspecification of the factor model.

B Persistence tests

An alternative methodology that is used to investigate mutual fund performance persistence are cross-sectional regressions, see e.g. Bollen and Busse (2005). In this section we explore the robustness of our findings with respect to this alternative approach. With the cross-sectional regressions, funds' post-ranking alphas are regressed on performance estimates over the pre-ranking period:

$$(11) \quad \hat{\alpha}_{i,post} = a_t + b_t Perf_{i,pre} + \epsilon_{i,t},$$

where $\hat{\alpha}_{i,post}$ is the post-ranking alpha estimate of fund i, and $Perf_{i,pre}$ is fund i's pre-ranking performance. The funds' post-ranking alphas are estimated using standard OLS over 12-month periods, and the funds' pre-ranking performance is estimated using raw returns, standard OLS alphas, and Bayesian alphas over 36-month and 12-month periods. The cross-sectional regression is estimated for all available funds at the beginning of every year, and the average of the coefficient estimates is taken over the entire sample period from 1984 to 2003. The standard deviations of the cross-sectional regression estimates are used to generate the sampling errors for these estimates (see Fama and MacBeth (1973)). A significant loading on the slope coefficient would be consistent with persistence in relative fund performance.

We perform these regressions for all subsamples of funds. The results of these analyses are reported in Table 13. Clearly, performance over the past year is much more informative about future relative performance than performance over the past three years: the slope coefficients are generally higher and more significant for regressions for one-year horizons than three-year horizons. However, note that these tests check the persistence in relative performance and

21

do not necessarily indicate that the winners provide abnormal positive returns. A positive loading on the slope coefficient could therefore be the result of persistent underperformance of a group of funds.

VII   Summary and concluding comments

When evaluating mutual fund performance, it is notoriously difficult to separate managerial skill from simple luck, particularly at short measurement horizons. In this paper, we employ a simple and straightforward empirical Bayes approach, based upon monthly returns, that is able to measure fund performance substantially more efficiently than standard OLS. Using Bayesian alphas, we investigate the persistence in mutual fund performance using the entire sample of US equity funds reported by CRSP over the period 1984–2003.

The results in this paper clearly support the idea that past performance of mutual funds has predictive power for future performance. When funds are ranked on Bayesian four-factor alphas, estimated over horizons of 36 or 12 months, the top deciles subsequently outperform the bottom deciles across all subsamples and using alternative ex post performance measures. More interestingly, we are able to capture short-run persistence in abnormal performance as documented by Bollen and Busse (2005) using monthly data for a much larger and more recent sample of US equity funds. When funds are sorted into decile portfolios based on Bayesian alphas over the previous 12 months, we find a monotonically decreasing spread in post-ranking alphas between the top and bottom deciles, the top decile earning a significant positive abnormal return of 0.26 percent in the first month after ranking. When raw returns or standard OLS alphas rather than Bayesian alphas are used, the findings are only a weak indication of such a relation. The predictive accuracy of OLS alphas is also significantly lower than that of Bayesian alphas.

Given the large cross-section of available funds and the limited demands imposed by our Bayesian procedure, we are able to investigate the persistence effect across different subsamples of funds. Accordingly, we investigate whether the superior performance of the top decile is canceled out by load fees. When we focus on the subsample of no-load funds, the persistence pattern is even stronger than for the entire sample, with a post-ranking alpha of the top decile of 0.31 percent per month. Apparently, the top deciles superior performance is not reduced by load fees that are involved with a strategy of chasing winners. In addition, we

form subgroups of different types of funds, and investigate whether the persistence effect is related to investment style. Indeed we find that persistence varies across styles, and manifests itself most strongly among relatively young, small cap/growth funds.

Appendix - Fund selection

To select funds from the 2003 CRSP database we follow a procedure similar to the one used by Pastor and Stambaugh (2002b). We use the information that CRSP provides about classifications by Wiesenberger (OBJ), Micropal/Investment Company Data, Inc. (ICDI OBJ), Strategic Insight (SI OBJ), and the funds themselves (POLICY). We exclude funds with unknown objectives, no expense, turnover, or load data in the annual summary at the end of each previous year. Additionally, we exclude flexible funds, bond funds, mortgage-backed funds, multi-manager funds, money market funds, balanced funds, funds that invest in precious metals, and international funds. Funds with the following classifications are excluded:

OBJ: BAL, BQ, BY, CBD, CGN, CHY, CIM, CSB, CSI, CSM, FLX, GB, GGN, GM, GOV, GPM, GS, GSM, IBD, ICA, IE, IFL, INT, IOH, IPA, MBD, MF, MGN, MHI, MHY, MIS, MMA, MMF, MMI, MNJ, MNY, MOH, MQ, MSS, MT, MTG, MTX, MVA, OI, OTH, SCU, SUT, TFG, TFM, TMM

ICDI OBJ: BL, BQ, BY, GB, GE, GM, GS, IE, MF, MG, MQ, MS, MT, MY, PM

SI OBJ: BAL, BGA, BGC, BGE, BGN, BGS, CGN, CHQ, CHY, CIM, CMQ, CPF, CPR, CSI, CSM, CVR, ECH, ECN, EGG, EGS, EGT, EGX, EID, EIG, EIS, EIT, EJP, ELT ,EPC, EPX, ERP, ESC, FLG, FLX, GBG, GBS, GGN, GIM, GLD, GMA, GMB, GSM, IAZ, ICA, ICO, ICT, IFL, IGA, IHI, IKS, IKY, IMA, IMD, IMI, IMN, IMT, IMX, INC, IND, INJ, INM, INY, IOH, IOR, IPA, ISC, ISD, ITN, ITX, IVA, IVT, IWA, IWV, JPN, LCA, LFL, LKY, LMA, LMD, LMI, LNC, LNY, LTN, LVA, MAL, MAR, MAZ, MCA, MCO, MCT, MDE, MFL, MGA, MGN, MHI, MHY, MIA, MID, MIL, MIM, MIN, MIS, MKS, MKY, MLA, MMA, MMD, MME, MMI, MMN, MMO, MMS, MMT, MNC, MND, MNE, MNH, MNJ, MNM, MNY, MOH, MOK, MOR, MPA, MPR, MRI, MSC, MSD, MSM, MTN, MTX, MUT, MVA, MVT, MWA, MWI, MWV, OPI, PAC, SBA, SBE, SBP, SBT, SBY, SCU, SIA,

SIE, SIP, SIT, SIY, SPE, SPR, SPY, SUA, SUT, TAL, TAZ, TBG, TCA, TCT, TFG, TFI, TFL, TGA, TMA, TMD, TMI, TMN, TNC, TNJ, TNY, TOH, TPA, TTN, TTX, TVA

POLICY: B & P, Bal, Bonds, C & I, Flex, GS, Hedge, Leases, MF, MM, TFM

We assign the remaining funds to the following investment styles, using the classifications provided by CRSP:

Small cap/growth: OBJ: MCG, SCG, AGG; ICDI OBJ: AG, AGG; SI OBJ: SCG, AGG

Growth: OBJ: G, G-S, S-G, GRO, LTG; IDCI OBJ: LG; SI OBJ: GRO

Growth & income: OBJ: GCI, G-I, G-I-S, G-S-I, I-G, I-G-S, I-S-G, S-G-I, S-I-G, GRI; ICDI OBJ: GI; SI OBJ: GRI

Funds are classified as no-load fund when they charge no front-end load fees, nor deferred sales charges and redemption fees (TOT LOAD = 0). Finally, we exclude funds with less than 12 consecutive returns over the whole sample period.

References

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

J. B. Berk and R. C. Green. Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112:1269–1295, 2004.

R. C. Blattberg and E. I. George. Shrinkage estimation of price and promotional elasticities: Seemingly unrelated equations. *Journal of the American Statistical Association*, 86:304–315, 1991.

M. E. Blume. On the assesment of risk. *Journal of Finance*, 26:1–10, 1971.

N. P. Bollen and J. A. Busse. Short-term persistence in mutual fund performance. *Review of Financial Studies*, 18:569–597, 2005.

S. J. Brown and W. Goetzmann. Performance persistence. *Journal of Finance*, 50:679–698, 1995.

S. J. Brown, W. Goetzmann, R. G. Ibbotson, and S. A. Ross. Survivorship bias in performance studies. *The Review of Financial Studies*, 5:553–580, 1992.

J. A. Busse and P. J. Irvine. Bayesian alphas and mutual fund persistence. *Journal of Finance, forthcoming*, 2006.

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

G. Casella and E. I. George. Explaining the Gibbs sampler. *The American Statistician*, 46: 167–174, 1992.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263, 1995.

E. J. Elton, M. J. Gruber, and C. R. Blake. The persistence of risk-adjusted mutual fund performance. *Journal of Business*, 69:133–157, 1996.

E. J. Elton, M. J. Gruber, and C. R. Blake. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and morningstar mutual fund databases. *Journal of Finance*, 56:2415–2429, 2001.

E. J. Elton, M. J. Gruber, S. Das, and M. Illavka. Efficiency with costly information: A reinterpretation of evidence from managed portfolios. *The Review of Financial Studies*, 6: 1–22, 1993.

E. F. Fama and K. R. French. The cross-section of expected stock returns. *Journal of Finance*, 47:427–465, 1992.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56, 1993.

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81:607–636, 1973.

W. E. Ferson and R. W. Schadt. Measuring fund strategy and performance in changing economic conditions. *Journal of Finance*, 51:425–461, 1996.

W. N. Goetzman and R. G. Ibbotson. Do winners repeat? *Journal of Portfolio Management*, 20:9–18, 1994.

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

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

D. Hendricks, J. Patel, and R. Zeckhauser. The J-shape of performance persistence given survivorship bias. *The Review of Economics and Statistics*, 79:161–166, 1997.

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

C. S. Jones and J. Shanken. Mutual fund performance with learning across funds. *Journal of Financial Economics*, 78:507–522, 2005.

G. S. Maddala, R. P. Trost, H. Li, and F. Joutz. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics*, 15:90–100, 1997.

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

M. R. Morey. Should you carry the load? A comprehensive analysis of load and no-load mutual fund out-of-sample performance. *Journal of Banking and Finance*, 27:1245–1271, 2003.

L. Pastor and R. F. Stambaugh. Investing in equity mutual funds. *Journal of Financial Economics*, 63:351–380, 2002a.

L. Pastor and R. F. Stambaugh. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, 63:315–349, 2002b.

E. R. Sirri and P. Tufano. Costly search and mutual fund flows. *Journal of Finance*, 53: 1589–1622, 1998.

S. Stern. *Simulation-Based Inference in Econometrics.* The University of Cambridge Press, 2000.

J. R. ter Horst, T. E. Nijman, and M. Verbeek. Eliminating look-ahead bias in evaluating persistence in mutual fund performance. *Journal of Empirical Finance*, 8:345–373, 2001.

O. A. Vasicek. A note on using cross-sectional information in bayesian estimation of security betas. *Journal of Finance*, 28:1233–1239, 1973.

X. Zhao. Why are some mutual funds closed to new investors? *Journal of Banking and Finance*, 28:1867–1887, 2004.

Table 1: Number of US equity funds over time. The table lists the number of available US equity mutual funds over the period 1984–2003. The data are extracted from the CRSP database. To be included in our sample, funds must be classified as a non-specialized US equity fund, and expense and load data of the funds should be available in the annual summary report at the end of the preceding year. In total, the samples cover 6,429 US equity funds, 3,101 no-load funds, 2,134 small cap/growth, 3,349 growth, and 1,837 growth & income funds that have more than 12 consecutive returns over the whole sample period.

	Equity	No-load	Small cap/growth	Growth	Growth & income
1984	362	171	91	182	89
1985	398	199	94	207	97
1986	456	233	90	272	94
1987	533	263	97	326	110
1988	601	302	102	372	127
1989	673	336	104	419	150
1990	681	336	96	457	128
1991	691	353	142	331	218
1992	730	374	153	358	219
1993	962	433	287	455	298
1994	1,226	574	323	567	337
1995	1,684	786	471	742	473
1996	1,903	896	571	807	529
1997	2,203	1,030	668	942	597
1998	2,858	1,375	883	1,245	738
1999	3,233	1,457	1,019	1,417	809
2000	3,634	1,593	1,133	1,616	895
2001	3,947	1,702	1,244	1,759	947
2002	4,621	1,925	1,497	2,041	1,084
2003	4,973	2,130	1,607	2,235	1,132

Table 2: Parameter values of the data generating process. Fund performance is measured using the four-factor Carhart (1997) model. Alphas and benchmark sensitivities are estimated using standard OLS and the entire return history of each fund. The sample is from 1984 to 2003 and covers 6,429 US equity funds. The table lists the benchmark factors' average returns (Panel A), and the cross-sectional means of funds' alphas and benchmark sensitivities (Panel B). The standard deviations are in parentheses.

Alpha	RMRF	SMB	HML	UMD
A. Return				
-	0.64% (4.57%)	0.01% (3.46%)	0.35% (3.31%)	0.86% (4.54%)
B. Sensitivity				
-0.22% (0.49%)	0.99 (0.25)	0.20 (0.38)	0.05 (0.39)	0.03 (0.18)

Table 3: Monte Carlo results. Simulated fund returns are generated using the Carhart (1997) model for measurement horizons of 12, 24, 36, and 60 months. The number of funds in the cross-section is set to 1,000. We then estimate the factor models using standard OLS and an empirical Bayes approach. The number of replications for each measurement horizon is set to 1,000. The table lists the average RMSE of alpha using standard OLS and Bayesian estimation (Panel A), and the average true alpha for the top (D1) and bottom (D10) deciles of funds resulting from rankings based on both estimators (Panel B and C).

| | 1. Using normal error terms | | | | 2. Using t-dist(5df) error terms | | | |
	OLS	Bayes	Difference	p-value	OLS	Bayes	Difference	p-value
A. RMSE								
12M	0.67%	0.38%	−0.28%	0.00	0.86%	0.45%	−0.42%	0.00
24M	0.41%	0.28%	−0.14%	0.00	0.53%	0.32%	−0.21%	0.00
36M	0.32%	0.24%	−0.09%	0.00	0.42%	0.27%	−0.14%	0.00
60M	0.24%	0.20%	−0.05%	0.00	0.32%	0.23%	−0.09%	0.00
B. Alpha D1 (top decile)								
12M	0.26%	0.39%	0.13%	0.00	0.18%	0.34%	0.16%	0.00
24M	0.42%	0.50%	0.08%	0.00	0.34%	0.46%	0.12%	0.00
36M	0.48%	0.54%	0.05%	0.00	0.42%	0.51%	0.09%	0.00
60M	0.54%	0.57%	0.03%	0.00	0.49%	0.54%	0.05%	0.00
C. Alpha D10 (bottom decile)								
12M	-0.70%	-0.83%	−0.13%	0.00	-0.61%	-0.77%	−0.16%	0.00
24M	-0.86%	-0.94%	−0.08%	0.00	-0.78%	-0.90%	−0.12%	0.00
36M	-0.93%	-0.98%	−0.05%	0.00	-0.86%	-0.95%	−0.09%	0.00
60M	-0.98%	-1.01%	−0.03%	0.00	-0.93%	-0.99%	−0.05%	0.00

Table 4: Rankings on three-year measurement horizons. Each month funds are sorted into equally weighted decile portfolios based on 36-month ranking periods. Pre-ranking performance is measured using raw returns (Panel A), standard OLS alphas (Panel B), and Bayesian alphas (Panel C). The sample is from 1984 to 2003 and covers 6,429 US equity funds. The table lists the deciles' post-ranking excess returns, Sharpe ratios, parameter estimates of the Carhart (1997) model, and R^2-values. The deciles' post-ranking alphas are estimated using standard OLS over the stacked time series of portfolio returns.

	Return	Sharpe	Alpha	Alpha-t	RMRF	SMB	HML	UMD	R^2
A. Rankings on raw returns									
D1	0.58%	0.10	-0.13%	-1.18	0.98	0.43	-0.26	0.20	0.93
D2	0.66%	0.13	-0.02%	-0.23	0.98	0.27	-0.14	0.11	0.96
D3	0.55%	0.12	-0.11%	-2.03	0.97	0.19	-0.05	0.05	0.98
D4	0.55%	0.12	-0.09%	-2.36	0.97	0.11	-0.00	0.02	0.98
D5	0.54%	0.12	-0.08%	-1.64	0.96	0.07	0.01	0.00	0.98
D6	0.53%	0.12	-0.09%	-1.74	0.97	0.08	0.06	-0.02	0.98
D7	0.53%	0.12	-0.10%	-1.54	0.97	0.08	0.08	-0.03	0.97
D8	0.50%	0.11	-0.11%	-1.33	0.97	0.06	0.11	-0.06	0.94
D9	0.50%	0.11	-0.11%	-1.16	0.99	0.13	0.14	-0.08	0.92
D10	0.35%	0.08	-0.21%	-2.01	0.96	0.20	0.11	-0.11	0.90
B. Rankings on OLS alphas									
D1	0.60%	0.11	-0.01%	-0.10	0.98	0.39	-0.16	0.04	0.95
D2	0.60%	0.13	-0.02%	-0.39	0.97	0.20	-0.05	0.02	0.97
D3	0.59%	0.13	-0.03%	-0.56	0.97	0.08	0.03	-0.01	0.97
D4	0.54%	0.12	-0.10%	-2.32	0.98	0.10	0.06	-0.01	0.98
D5	0.54%	0.12	-0.07%	-1.54	0.95	0.08	0.02	-0.00	0.98
D6	0.52%	0.12	-0.12%	-2.42	0.98	0.06	0.04	-0.00	0.98
D7	0.51%	0.11	-0.13%	-2.64	0.97	0.09	0.05	-0.00	0.98
D8	0.53%	0.12	-0.13%	-2.18	0.98	0.12	0.04	0.01	0.97
D9	0.51%	0.11	-0.16%	-2.28	0.99	0.18	0.04	0.03	0.97
D10	0.35%	0.07	-0.28%	-3.18	0.96	0.32	-0.01	0.02	0.94
C. Rankings on Bayesian alphas									
D1	0.62%	0.12	0.01%	0.11	0.99	0.30	-0.16	0.04	0.94
D2	0.58%	0.12	-0.03%	-0.44	0.98	0.17	-0.07	0.02	0.97
D3	0.56%	0.12	-0.08%	-1.35	0.98	0.15	-0.01	0.01	0.97
D4	0.51%	0.11	-0.10%	-1.88	0.95	0.15	0.02	-0.01	0.98
D5	0.58%	0.13	-0.03%	-0.64	0.95	0.11	0.02	-0.01	0.98
D6	0.52%	0.11	-0.13%	-2.26	0.99	0.11	0.04	0.01	0.98
D7	0.54%	0.12	-0.08%	-1.54	0.96	0.12	0.02	0.00	0.97
D8	0.52%	0.12	-0.12%	-1.73	0.97	0.11	0.06	-0.01	0.96
D9	0.50%	0.11	-0.17%	-2.13	0.98	0.17	0.05	0.02	0.96
D10	0.36%	0.08	-0.33%	-3.27	0.98	0.23	0.07	0.04	0.92

Table 5: Rankings on one-year measurement horizons. Each month funds are sorted into equally weighted decile portfolios based on 12-month ranking periods. Pre-ranking performance is measured using raw returns (Panel A), standard OLS alphas (Panel B), and Bayesian alphas (Panel C). The sample is from 1984 to 2003 and covers 6,429 US equity funds. The table lists the deciles' post-ranking excess returns, Sharpe ratios, parameter estimates of the Carhart (1997) model, and R^2-values. The deciles' post-ranking alphas are estimated using standard OLS over the stacked time series of portfolio returns.

	E(r)	Sharpe	Alpha	Alpha-t	RMRF	SMB	HML	UMD	R^2
A. Rankings on raw returns									
D1	0.97%	0.16	0.04%	0.36	0.96	0.57	-0.13	0.40	0.91
D2	0.77%	0.16	-0.04%	-0.51	0.97	0.36	-0.04	0.23	0.96
D3	0.68%	0.15	-0.06%	-1.13	0.97	0.21	-0.01	0.14	0.98
D4	0.59%	0.13	-0.12%	-3.22	0.97	0.13	0.03	0.08	0.98
D5	0.52%	0.12	-0.12%	-2.65	0.97	0.08	0.02	0.01	0.98
D6	0.48%	0.11	-0.12%	-2.22	0.97	0.04	0.04	-0.03	0.98
D7	0.47%	0.10	-0.09%	-1.53	0.96	0.04	0.03	-0.07	0.97
D8	0.38%	0.08	-0.13%	-1.60	0.96	0.07	0.04	-0.13	0.95
D9	0.30%	0.06	-0.17%	-1.63	0.97	0.08	0.06	-0.20	0.92
D10	0.12%	0.02	-0.21%	-1.54	0.94	0.15	0.03	-0.33	0.88
B. Rankings on OLS alphas									
D1	0.89%	0.16	0.24%	1.95	0.95	0.45	-0.15	0.11	0.90
D2	0.70%	0.14	0.03%	0.57	0.98	0.25	-0.04	0.06	0.97
D3	0.59%	0.13	-0.03%	-0.79	0.95	0.12	-0.00	0.02	0.98
D4	0.55%	0.13	-0.08%	-1.87	0.95	0.09	0.02	0.01	0.98
D5	0.54%	0.12	-0.08%	-1.72	0.95	0.05	0.03	-0.01	0.98
D6	0.51%	0.12	-0.11%	-2.29	0.96	0.07	0.03	-0.01	0.98
D7	0.46%	0.10	-0.16%	-2.86	0.96	0.09	0.04	-0.02	0.98
D8	0.43%	0.09	-0.19%	-3.32	0.97	0.13	0.04	-0.02	0.97
D9	0.42%	0.09	-0.21%	-2.79	0.99	0.18	0.04	-0.01	0.96
D10	0.18%	0.04	-0.44%	-4.22	0.98	0.30	0.07	-0.04	0.92
C. Rankings on Bayesian alphas									
D1	0.95%	0.18	0.26%	2.26	0.93	0.44	-0.15	0.16	0.89
D2	0.74%	0.15	0.08%	0.99	0.97	0.25	-0.09	0.09	0.96
D3	0.61%	0.13	-0.04%	-0.74	0.96	0.17	-0.04	0.05	0.98
D4	0.59%	0.13	-0.04%	-0.76	0.97	0.12	0.00	0.00	0.98
D5	0.50%	0.11	-0.11%	-2.09	0.95	0.09	0.01	-0.00	0.97
D6	0.49%	0.11	-0.12%	-2.00	0.97	0.09	0.05	-0.03	0.97
D7	0.44%	0.10	-0.16%	-2.39	0.97	0.10	0.04	-0.04	0.97
D8	0.40%	0.09	-0.21%	-2.92	0.97	0.12	0.05	-0.04	0.96
D9	0.36%	0.08	-0.28%	-3.21	0.99	0.15	0.08	-0.03	0.95
D10	0.19%	0.04	-0.43%	-3.70	0.98	0.20	0.13	-0.06	0.91

Table 6: The lifespan of performance persistence. Each month funds are sorted into equally weighted decile portfolios on 1-month lagged alpha estimates over the preceding 12 months. Pre-ranking performance is measured using Bayesian alphas. The sample is from 1984 to 2003 and covers 6,429 US equity funds. The table lists the deciles' post-ranking excess returns, Sharpe ratios, parameter estimates of the Carhart (1997) model, and R^2-values. The deciles' post-ranking alphas are estimated using standard OLS over the stacked time series of portfolio returns.

	Return	Sharpe	Alpha	Alpha-t	RMRF	SMB	HML	UMD	R^2
D1	0.78%	0.14	0.10%	0.92	0.94	0.43	-0.20	0.16	0.90
D2	0.71%	0.15	0.04%	0.55	0.97	0.26	-0.09	0.10	0.97
D3	0.58%	0.12	-0.06%	-1.29	0.98	0.17	-0.06	0.05	0.98
D4	0.58%	0.13	-0.04%	-0.85	0.95	0.12	-0.01	0.01	0.98
D5	0.53%	0.12	-0.10%	-1.85	0.96	0.11	0.04	-0.01	0.98
D6	0.51%	0.11	-0.11%	-1.80	0.98	0.09	0.05	-0.03	0.97
D7	0.47%	0.11	-0.15%	-2.14	0.97	0.11	0.06	-0.02	0.96
D8	0.46%	0.10	-0.14%	-2.18	0.96	0.12	0.06	-0.04	0.96
D9	0.40%	0.09	-0.22%	-2.60	0.98	0.14	0.08	-0.05	0.95
D10	0.27%	0.06	-0.34%	-3.08	0.96	0.18	0.14	-0.05	0.90

Table 7: Performance persistence of older versus younger funds. Each month funds are sorted into equally weighted decile portfolios based on 12-month ranking periods. Pre-ranking performance is measured using Bayesian alphas. The analysis is conducted on funds that have at least 36 observations available (Panel A), and on funds with 12 to at most 35 observations available (Panel B). The sample is from 1984 to 2003 and covers 6,429 US equity funds. In the average month, 977 funds are classified as "old", and 480 funds as "young". The table lists the deciles' post-ranking excess returns, Sharpe ratios, parameter estimates of the Carhart (1997) model, and R^2-values. The deciles' post-ranking alphas are estimated using standard OLS over the stacked time series of portfolio returns.

	Return	Sharpe	Alpha	Alpha-t	RMRF	SMB	HML	UMD	R^2
A. Old Funds									
D1	0.89%	0.17	0.21%	2.04	0.94	0.39	-0.14	0.15	0.92
D2	0.73%	0.15	0.05%	0.67	0.97	0.23	-0.06	0.09	0.96
D3	0.61%	0.13	-0.04%	-0.99	0.96	0.16	-0.03	0.06	0.98
D4	0.63%	0.14	-0.01%	-0.11	0.97	0.13	-0.01	0.02	0.98
D5	0.56%	0.12	-0.06%	-1.17	0.96	0.08	0.01	-0.00	0.97
D6	0.47%	0.10	-0.15%	-2.52	0.97	0.08	0.05	-0.03	0.97
D7	0.44%	0.10	-0.16%	-2.40	0.98	0.09	0.03	-0.04	0.97
D8	0.40%	0.09	-0.20%	-2.83	0.98	0.11	0.05	-0.05	0.96
D9	0.33%	0.07	-0.31%	-3.35	1.00	0.14	0.06	-0.03	0.95
D10	0.22%	0.05	-0.38%	-3.36	0.99	0.20	0.09	-0.07	0.91
B. Young Funds									
D1	1.03%	0.17	0.35%	2.02	0.91	0.53	-0.17	0.16	0.74
D2	0.81%	0.16	0.14%	1.38	0.98	0.29	-0.10	0.08	0.93
D3	0.62%	0.13	0.02%	0.31	0.93	0.20	-0.06	0.02	0.96
D4	0.49%	0.11	-0.12%	-1.91	0.95	0.12	0.02	-0.01	0.96
D5	0.47%	0.11	-0.14%	-2.12	0.94	0.12	0.02	-0.00	0.95
D6	0.46%	0.11	-0.12%	-1.43	0.92	0.09	0.05	-0.03	0.95
D7	0.41%	0.09	-0.15%	-2.10	0.93	0.09	0.04	-0.05	0.94
D8	0.39%	0.09	-0.19%	-2.18	0.94	0.15	0.04	-0.05	0.94
D9	0.33%	0.07	-0.30%	-3.43	0.96	0.17	0.10	-0.03	0.93
D10	0.16%	0.04	-0.53%	-3.96	0.97	0.22	0.24	-0.02	0.85

Table 8: **Performance persistence for no-load and load funds.** Each month funds are sorted into equally weighted decile portfolios based on 12-month ranking periods. Pre-ranking performance is measured using Bayesian alphas. The analysis is performed on strictly no-load funds (Panel A), and on fund that do charge sales loads (Panel B). The samples are from 1984 to 2003 and cover 3,101 no-load funds, and 4,025 load funds. The table lists the deciles' post-ranking excess returns, Sharpe ratios, parameter estimates of the Carhart (1997) model, and R^2-values. The deciles' post-ranking alphas are estimated using standard OLS over the stacked time series of portfolio returns.

	Return	Sharpe	Alpha	Alpha-t	RMRF	SMB	HML	UMD	R^2
A. No-load Funds									
D1	0.98%	0.18	0.31%	2.48	0.92	0.45	-0.18	0.17	0.88
D2	0.75%	0.16	0.10%	1.47	0.95	0.24	-0.09	0.08	0.96
D3	0.59%	0.13	-0.05%	-1.08	0.96	0.16	-0.01	0.03	0.97
D4	0.59%	0.13	-0.04%	-0.77	0.96	0.13	0.01	0.01	0.97
D5	0.53%	0.12	-0.08%	-1.37	0.95	0.09	0.04	-0.01	0.97
D6	0.52%	0.12	-0.10%	-1.56	0.96	0.09	0.08	-0.03	0.96
D7	0.46%	0.10	-0.16%	-2.40	0.96	0.10	0.08	-0.03	0.96
D8	0.41%	0.09	-0.17%	-2.27	0.94	0.12	0.07	-0.05	0.95
D9	0.37%	0.08	-0.26%	-2.96	0.98	0.15	0.10	-0.03	0.94
D10	0.24%	0.05	-0.41%	-3.30	0.97	0.21	0.20	-0.06	0.88
B. Load Funds									
D1	0.89%	0.17	0.20%	1.79	0.94	0.43	-0.14	0.16	0.89
D2	0.76%	0.15	0.09%	1.05	0.97	0.25	-0.09	0.09	0.95
D3	0.59%	0.13	-0.05%	-0.73	0.95	0.18	-0.06	0.06	0.97
D4	0.55%	0.12	-0.07%	-1.64	0.97	0.12	-0.02	0.01	0.97
D5	0.53%	0.12	-0.08%	-1.47	0.97	0.08	-0.00	-0.01	0.97
D6	0.45%	0.10	-0.17%	-2.67	0.98	0.08	0.02	-0.02	0.97
D7	0.44%	0.10	-0.16%	-2.36	0.97	0.10	0.02	-0.04	0.97
D8	0.39%	0.08	-0.23%	-3.01	1.00	0.12	0.04	-0.03	0.96
D9	0.34%	0.07	-0.31%	-3.17	1.00	0.15	0.06	-0.02	0.94
D10	0.15%	0.03	-0.44%	-3.82	0.98	0.21	0.07	-0.07	0.91

Table 9: Performance persistence across different investment styles. Each month funds are sorted into equally weighted decile portfolios based on 12-month ranking periods. Pre-ranking performance is measured using raw returns (Panel A), standard OLS alphas (Panel B), and Bayesian alphas (Panel C). The analysis is conducted on different types of funds. The samples are from 1984 to 2003 and cover 2,134 small cap/growth, 3,349 growth, and 1,837 growth & income funds. The table lists the top deciles' post-ranking excess returns, Sharpe ratios, parameter estimates of the Carhart (1997) model, and R^2-values. The deciles' post-ranking alphas are estimated using standard OLS over the stacked time series of portfolio returns.

	Return	Sharpe	Alpha	Alpha-t	RMRF	SMB	HML	UMD	R^2
A. Rankings on raw returns									
D1 Small cap/growth	1.06%	0.17	0.09%	0.73	0.98	0.67	-0.20	0.47	0.92
D1 Growth	0.92%	0.16	0.02%	0.20	0.99	0.42	-0.10	0.35	0.87
D1 Growth & income	0.74%	0.18	-0.03%	-0.44	0.90	0.09	0.09	0.18	0.95
B. Rankings on OLS alphas									
D1 Small cap/growth	0.93%	0.15	0.22%	1.85	1.01	0.65	-0.20	0.14	0.93
D1 Growth	0.81%	0.15	0.19%	1.42	0.95	0.31	-0.15	0.08	0.83
D1 Growth & income	0.65%	0.16	0.03%	0.46	0.92	-0.00	0.18	-0.03	0.93
C. Rankings on Bayesian alphas									
D1 Small cap/growth	0.99%	0.16	0.24%	2.01	1.00	0.65	-0.18	0.19	0.92
D1 Growth	0.87%	0.16	0.20%	1.57	0.95	0.29	-0.15	0.13	0.83
D1 Growth & income	0.71%	0.18	0.08%	1.20	0.91	-0.02	0.17	-0.02	0.93

Table 10: Rankings on one-year measurement horizons using conditional priors. Each month funds are sorted into equally weighted decile portfolios based on 12-month ranking periods. Pre-ranking performance is measured using Bayesian alphas. Funds' raw alphas are shrunk towards the cross-sectional average of funds with the same investment style. The sample is from 1984 to 2003 and covers 6,429 US equity funds. The table lists the deciles' post-ranking excess returns, Sharpe ratios, parameter estimates of the Carhart (1997) model, and R^2-values. The deciles post-ranking alphas are estimated using standard OLS over the stacked time series of portfolio returns.

	Return	Sharpe	Alpha	Alpha-t	RMRF	SMB	HML	UMD	R^2
D1	0.98%	0.18	0.28%	2.41	0.95	0.47	-0.17	0.17	0.89
D2	0.71%	0.15	0.05%	0.67	0.96	0.26	-0.09	0.09	0.96
D3	0.58%	0.13	-0.07%	-1.39	0.96	0.16	-0.03	0.04	0.98
D4	0.55%	0.12	-0.06%	-1.45	0.96	0.10	0.00	0.00	0.98
D5	0.51%	0.12	-0.12%	-2.40	0.95	0.07	0.04	-0.00	0.97
D6	0.52%	0.12	-0.08%	-1.43	0.95	0.06	0.04	-0.02	0.97
D7	0.43%	0.10	-0.18%	-2.89	0.97	0.09	0.06	-0.04	0.97
D8	0.41%	0.09	-0.20%	-2.92	0.98	0.12	0.05	-0.04	0.97
D9	0.36%	0.08	-0.26%	-2.93	0.98	0.15	0.06	-0.03	0.95
D10	0.19%	0.04	-0.43%	-3.75	0.98	0.25	0.11	-0.06	0.91

Table 11: Comparing forecasts of both estimators. Each year, the root mean squared error (RMSE) is computed for alpha forecasts based on the preceding 12 months, and realized values over the subsequent year. Forecasts are obtained using standard OLS and Bayesian estimation. Realized alphas are estimated using standard OLS. The analysis is conducted on different types of funds. The samples are from 1984 to 2003 and cover 6,429 US equity funds, 3,101 no-load funds, 2,134 small cap/growth, 3,349 growth, and 1,837 growth & income funds. The table lists both estimators' average RMSEs, and the Diebold-Mariano test on equality of the two.

	$RMSE_{OLS}$	$RMSE_{Bayes}$	Difference	Difference-t
All equity	1.12%	0.91%	-0.21%	-9.98
No-load	1.10%	0.89%	-0.21%	-9.85
Small cap/growth	1.24%	1.02%	-0.21%	-11.46
Growth	1.14%	0.92%	-0.22%	-8.91
Growth & income	0.74%	0.61%	-0.13%	-8.15

Table 12: Robustness-tests: Industry effects and time-varying benchmark sensitivities. Each month funds are sorted into decile portfolios based on 12-month ranking periods. Pre-ranking performance is measured using raw returns (Panel A), standard OLS alphas (Panel B), and Bayesian alphas (Panel C). The sample is from 1984 to 2003 and covers 6,429 US equity funds. The table lists the deciles' post-ranking excess returns, Sharpe ratios, and alphas. The deciles' post-ranking alphas are estimated using the Carhart (1997) model, the Carhart model augmented with three industry factors, and the Ferson and Schadt (1996) model with standard OLS over the stacked time series of portfolio returns.

	Return	Sharpe	1. Carhart		2. Carhart + industries		3. Ferson & Schadt	
			Alpha	Alpha-t	Alpha	Alpha-t	Alpha	Alpha-t
A. Ranking on raw returns
D1	0.97%	0.16	0.04%	0.36	0.10%	0.84	-0.00%	-0.02
D2	0.77%	0.16	-0.04%	-0.51	-0.02%	-0.21	-0.09%	-1.34
D3	0.68%	0.15	-0.06%	-1.13	-0.03%	-0.55	-0.07%	-1.40
D4	0.59%	0.13	-0.12%	-3.22	-0.08%	-2.08	-0.11%	-2.83
D5	0.52%	0.12	-0.12%	-2.65	-0.08%	-1.80	-0.09%	-2.48
D6	0.48%	0.11	-0.12%	-2.22	-0.07%	-1.64	-0.07%	-1.81
D7	0.47%	0.10	-0.09%	-1.53	-0.05%	-0.88	-0.06%	-1.16
D8	0.38%	0.08	-0.13%	-1.60	-0.09%	-1.07	-0.08%	-1.25
D9	0.30%	0.06	-0.17%	-1.63	-0.12%	-1.19	-0.09%	-1.05
D10	0.12%	0.02	-0.21%	-1.54	-0.15%	-1.08	-0.11%	-0.92
B. Ranking on OLS alphas
D1	0.89%	0.16	0.24%	1.95	0.32%	2.54	0.26%	2.02
D2	0.70%	0.14	0.03%	0.57	0.07%	1.12	0.05%	0.97
D3	0.59%	0.13	-0.03%	-0.79	0.00%	0.10	-0.02%	-0.39
D4	0.55%	0.13	-0.08%	-1.87	-0.05%	-1.11	-0.07%	-1.80
D5	0.54%	0.12	-0.08%	-1.72	-0.03%	-0.74	-0.05%	-1.32
D6	0.51%	0.12	-0.11%	-2.29	-0.06%	-1.39	-0.10%	-2.54
D7	0.46%	0.10	-0.16%	-2.86	-0.12%	-2.29	-0.16%	-3.54
D8	0.43%	0.09	-0.19%	-3.32	-0.16%	-2.80	-0.18%	-4.06
D9	0.42%	0.09	-0.21%	-2.79	-0.18%	-2.44	-0.18%	-2.70
D10	0.18%	0.04	-0.44%	-4.22	-0.39%	-3.89	-0.37%	-3.89
C. Ranking on Bayesian alphas
D1	0.95%	0.18	0.26%	2.26	0.33%	2.70	0.28%	2.31
D2	0.74%	0.15	0.08%	0.99	0.11%	1.45	0.09%	1.24
D3	0.61%	0.13	-0.04%	-0.74	-0.01%	-0.23	-0.03%	-0.55
D4	0.59%	0.13	-0.04%	-0.76	-0.00%	-0.05	-0.01%	-0.19
D5	0.50%	0.11	-0.11%	-2.09	-0.06%	-1.14	-0.09%	-2.05
D6	0.49%	0.11	-0.12%	-2.00	-0.08%	-1.35	-0.11%	-2.23
D7	0.44%	0.10	-0.16%	-2.39	-0.11%	-1.58	-0.15%	-2.67
D8	0.40%	0.09	-0.21%	-2.92	-0.16%	-2.42	-0.19%	-3.15
D9	0.36%	0.08	-0.28%	-3.21	-0.24%	-2.87	-0.24%	-3.36
D10	0.19%	0.04	-0.43%	-3.70	-0.39%	-3.51	-0.35%	-3.36

Table 13: Robustness-tests: Fama and MacBeth (1973) regressions. At the begin of each year, realized alphas over the subsequent 12 months are regressed on a measure of performance over the preceding 36 and 12 months. Fund performance is measured using raw returns (Panel A), standard OLS alphas (Panel B), and Bayesian alphas (Panel C). The analysis is conducted on different types of funds. The samples are from 1984 to 2003 and cover 6,429 US equity funds, 3,101 no-load funds, 2,134 small cap/growth, 3,349 growth, and 1,837 growth & income funds. The table lists the regressions' intercept, slope coefficients, and R^2-values. Realized alphas are estimated using standard OLS. The t-statistics are in parentheses.

	1. Three-year horizon			2. One-year horizon		
	a	b	R^2	a	b	R^2
A. Regressions on raw returns						
All equity	-0.14 (-2.14)	0.07 (1.93)	0.01	-0.11 (-1.73)	0.13 (3.28)	0.04
No-load	-0.15 (-2.55)	0.10 (2.44)	0.02	-0.11 (-1.90)	0.14 (3.65)	0.05
Small cap/growth	-0.20 (-2.18)	0.16 (1.93)	0.05	-0.14 (-1.66)	0.16 (4.59)	0.06
Growth	-0.10 (-1.52)	0.03 (0.71)	0.02	-0.09 (-1.56)	0.12 (2.39)	0.04
Growth & income	-0.08 (-1.30)	0.01 (0.12)	0.03	-0.16 (-2.37)	0.15 (2.27)	0.06
B. Regressions on OLS alphas						
All equity	-0.08 (-1.45)	0.15 (4.54)	0.01	-0.07 (-1.39)	0.18 (6.46)	0.04
No-load	-0.08 (-1.45)	0.19 (5.99)	0.02	-0.07 (-1.34)	0.19 (6.02)	0.06
Small cap/growth	-0.11 (-1.21)	0.29 (4.64)	0.05	-0.10 (-1.20)	0.19 (10.84)	0.05
Growth	-0.07 (-1.34)	0.07 (1.76)	0.01	-0.06 (-1.24)	0.16 (4.14)	0.04
Growth & income	-0.06 (-1.84)	0.09 (1.43)	0.02	-0.07 (-2.56)	0.15 (3.85)	0.07
C. Regressions on Bayesian alphas						
All equity	-0.08 (-1.49)	0.26 (3.60)	0.01	-0.07 (-1.42)	0.33 (5.34)	0.04
No-load	-0.07 (-1.35)	0.30 (3.31)	0.02	-0.06 (-1.36)	0.37 (4.97)	0.05
Small cap/growth	-0.12 (-1.30)	0.36 (2.89)	0.04	-0.10 (-1.26)	0.34 (8.77)	0.04
Growth	-0.07 (-1.36)	0.17 (2.34)	0.01	-0.06 (-1.29)	0.31 (3.47)	0.04
Growth & income	-0.06 (-2.12)	0.16 (1.81)	0.02	-0.07 (-2.33)	0.27 (4.36)	0.04