Mutual fund performance with learning across funds [☆]

Christopher S. Jones[a,*], Jay Shanken[b]

[a]Marshall School of Business, University of Southern California, Hoffman Hall 701, Los Angeles, CA 90089, USA

[b]Goizueta Business School, Emory University, 1300 Clifton Road, Atlanta, GA 30322, USA

Abstract

The average level and cross-sectional variability of fund alphas are estimated from a large sample of mutual funds. This information is incorporated, along with the usual regression estimate of alpha, in a (roughly) precision-weighted average measure of individual fund performance. Substantial "learning across funds" is documented, with significant effects on investment decisions. In a Bayesian framework, this form of learning is inconsistent with the assumption, made in the past literature, of prior independence across funds. Independence can be viewed as an extreme scenario in which the true cross-sectional distribution of alphas is presumed to be known a priori.
© 2005 Elsevier B.V. All rights reserved.

---

1. Introduction

With trillions of dollars invested in actively managed equity mutual funds, it is of great importance to investors to determine the optimal asset allocation to funds. Many studies, starting with Jensen (1968), have concluded that fund managers are unable to "beat the market," suggesting that investors might want to restrict their portfolios to passive index funds. Others have argued that, while the average manager may have no particular skill, ex ante variables such as past performance and manager characteristics can be used to identify investment skill.[1] More recently, papers by Baks et al. (2001) and Pastor and Stambaugh (2002a,b) have explored the role of prior information in analyzing the performance of mutual funds and making investment decisions.[2] Henceforth, we refer to these studies as BMW and PS, respectively.

Since the standard measure of fund performance, "alpha," is typically not estimated with much precision, prior beliefs can have a substantial impact. We would argue, however, that another important source of information about fund performance has been overlooked up to now, both in traditional studies and in the more recent Bayesian analyses. The neglected information is the history of returns on *other* funds. Ignoring these returns might seem reasonable at first glance—what could the returns on Vanguard's Windsor fund possibly tell us about the skill of the managers of Fidelity's Magellan fund? Our answer is that aggregating data on Windsor and all other funds yields important information about the abilities of fund managers as a group. Since Fidelity is a member of this group, this general knowledge should have some bearing on our beliefs about Fidelity.

There are currently more than 5,000 equity funds in the U.S. It is natural to think of the true alphas of these funds as a large sample from an underlying population. For simplicity, assume they are independent draws from a normal distribution with mean μ_α and standard deviation σ_α. Thus, μ_α reflects the average level of skill in the universe of fund managers and σ_α captures the variability around that average level for different funds. We could refine the analysis further by conditioning on individual fund or manager characteristics. However, a careful analysis of the unconditional case, in which each fund is a random draw from the overall population, seems to us an appropriate starting point. It will also facilitate comparison with the earlier literature.

Consider *XYZ* Investments, a hypothetical fund of interest. Suppose we had returns data for other funds, but not *XYZ*. If the number of funds were sufficiently large and residual fund returns independent, we would effectively be able to infer the true cross-sectional distribution of the alphas from the ordinary least squares estimates, i.e., we would know the actual values of the "hyperparameters" μ_α and σ_α.

[1]See Chevalier and Ellison (1999) and Carhart (1997), for instance.

[2]Perhaps the first important application of the Bayesian perspective in investment research was Merton's (1980) paper on estimating the market expected return. Bayesian methods were first used in testing asset pricing relations in Shanken (1987). Kandel and Stambaugh (1996) examine aggregate return predictability in a Bayesian framework. Their paper has stimulated much recent research.

In the absence of any other information about XYZ, it would be natural to take μ_α as our best guess or "estimate" of α_{XYZ}, and view σ_α as indicative of the precision of this estimate. After all, we know the cross-sectional distribution of the true alphas in this setting, but we don't know where in that distribution XYZ's alpha lies.

Suppose, now, that we *did* have returns for XYZ and were able to compute the conventional regression estimate of alpha and its associated standard error. Would we discard the information about μ_α and σ_α in this case? Intuition suggests that the OLS estimate should be combined with the μ_α estimate. If data on XYZ were limited, the OLS standard error might be quite large, perhaps much larger than σ_α. Then, presumably, μ_α would still be given considerable weight in the estimation of α_{XYZ}. But if the OLS estimate were very precise, it would be weighted more heavily. This is essentially what happens using the estimation approach developed in this paper.

We use the Fidelity Magellan Fund to illustrate this idea. Magellan had an impressive OLS alpha estimate of 10.4% per annum (standard error 1.9%) over the 1963 to 2001 period. Aggregating the evidence for all funds in our sample produces fairly precise estimates of μ_α and σ_α, both (coincidentally) about 1.5% before costs. To a close approximation, our learning model for Magellan amounts to taking a precision-weighted average of the OLS estimate 10.4% and $\mu_\alpha = 1.5\%$. Here, the precisions are based on the OLS standard error of 1.9% and $\sigma_\alpha = 1.5\%$, respectively. Since the standard error is larger in this case, greater weight is placed on μ_α. The resulting alpha estimate is 4.8%, substantially below the OLS estimate. For funds with shorter return histories or high residual variance, the weight placed on the aggregate estimate can be even greater, and alphas can increase as well as decrease.

In general, estimation of alpha in our model will depend on the returns on all other funds through the estimates of μ_α and σ_α. We refer to this as "learning across funds." The learning arises because each fund alpha is recognized as an observation from an underlying population. In this sense, all alphas are linked, and so whatever we learn about the population feeds back into the estimation for any given fund. This sort of econometric specification is often referred to as a random coefficients model (e.g., Swamy, 1971). Our use of Gibbs sampling techniques in Bayesian estimation of the model is appealing in that the measures of precision obtained for the individual fund alphas incorporate estimation error in the hyperparameters, μ_α and σ_α. This contrasts with commonly used two-stage estimation procedures. It also allows us to accommodate prior information, as we now discuss.[3]

Consider a prior on alpha for our fund, XYZ Investments. The prior is a subjective belief that we bring to the problem before observing the data. What

---

[3]Shrinkage is an important feature of the Bayes-Stein estimates of Jorion (1986) and others, though the economic motivation and estimation approaches are quite different. In Jorion's paper, expected returns are shrunk towards the mean of the minimum variance portfolio. The degree of shrinkage is computed based on an empirical Bayes approach that examines the cross-sectional dispersion in sample means. The use of empirical Bayes methods in finance goes back at least to Vasicek (1973) who was concerned with estimating betas. See Pastor and Stambaugh (1999) for a more recent application. Empirical Bayes methods, while they could be applied to the other regression parameters of our model, are not appropriate in the context of our goal of estimating the cross-sectional distribution of (true) alphas, particularly given the relatively low precision with which individual alphas are generally estimated.

values, we ask, might the true alpha of *XYZ* take, and how plausible are these values? For now, let's think about alpha before trading costs and various fund expenses. As mentioned above and consistent with earlier Bayesian studies, we do not condition our belief about alpha on any observed characteristics of *XYZ*, though this would be a natural extension of our framework. Therefore, we are really thinking about a ''generic'' fund here—i.e., a random draw from the universe of funds. In this setting, the prior on α_{XYZ} just reflects our belief about the abilities of fund managers as a group and amounts to an initial assessment of the cross-sectional distribution of fund alphas. Thus, it is determined by our prior beliefs about the values of μ_α and σ_α. In particular, the mean of our prior for α_{XYZ} is just the mean of the prior distribution for μ_α and reflects our view of the average level of skill in the fund universe.

Consistent with this perspective, past research has adopted identical (marginal) priors on management skill for each fund in the given sample. These studies, however, have gone one step further and specified a *joint* prior distribution in which the beliefs about alphas are independent across funds.[4] It appears that this assumption has been adopted for reasons of mathematical tractability, rather than some underlying intuition or principle. Although this setting provides a natural starting point for papers breaking new ground, we now use a thought experiment to argue that the independence assumption is intuitively implausible in the present context.

Imagine that the *true* values of alpha for thousands of other funds were somehow revealed before you even examine *XYZ*'s returns. Would this information affect your belief about α_{XYZ}? Consider two specific scenarios. In the first, the true fund alphas are all exactly equal to zero, i.e., there is no evidence that fund managers have any skill. In the second scenario, half of the funds have alphas that exceed 3% per annum. So, in one scenario it looks like the market is extremely efficient and beating the benchmark may well be an impossible task for professional money managers. In the other, superior performance is quite common, suggesting an inefficient market in which mispriced securities are not so hard to identify.

Our intuition is as follows. In the first scenario, the strong ''evidence'' that the market is efficient would largely eclipse whatever we initially believed about alphas, as represented by our marginal prior. After all, that belief was merely a preliminary assessment of the cross-sectional distribution of alphas, and now we've seen the actual values of alpha for a large sample of funds. Similarly, in the second scenario, even if we started out extremely skeptical about the abilities of fund managers, we'd have to acknowledge that skill is fairly common and, therefore, we would revise our belief about *XYZ* accordingly.

In short, the prior for a given fund's alpha will be affected by conditioning on the values of other fund alphas. Mathematically, this is a statement that the conditional prior differs from the marginal prior, i.e., the priors are not independent. Our model accommodates this sort of prior dependence, as beliefs about different fund alphas

---

[4]PS decompose alphas into two components, one related to skill and the other to model misspecification. The skill components are taken to be independent across funds.

are linked through the common prior belief about the hyperparameters μ_α and σ_α. Even if we have "diffuse priors" or "vague" prior information about μ_α and σ_α, estimation of individual fund alphas is affected in a manner similar to that in the Magellan example above. Readers who prefer not to think in terms of prior distributions may want to focus on the diffuse prior results, which include some of our most striking observations.

The independent prior assumption can actually be nested in our model, providing additional perspective on its restrictive nature. In the model, prior independence amounts to a "dogmatic" belief about the cross-sectional distribution of fund skill, i.e. a belief in which there is no perceived uncertainty about the values of μ_α and σ_α. Right or wrong, this individual is certain that the true values are, say, $\mu_\alpha = m_\alpha$ and $\sigma_\alpha = s_\alpha$. The prior for each alpha is then $N(m_\alpha, s_\alpha^2)$, and there is independence across funds, as the "common factor"—uncertainty about the characteristics of the underlying alpha population—has been eliminated. Since there is no updating of beliefs about this population, the returns on other funds will have no impact on the estimation of a given fund's alpha (provided the return residuals are independent). In other words, there is no longer any learning across funds. In particular, if the prior for alpha is diffuse ($s_\alpha \to \infty$), Bayesian estimation reduces to standard OLS regression estimation.

Before going on to present the details of the model, we highlight two additional ways in which Bayesian analysis with learning across funds differs fundamentally from that in earlier studies. First is the issue of survivorship bias. BMW and PS rely on prior independence, in conjunction with other assumptions, to justify ignoring data on the nonsurvivors in their asset allocation analyses. However, it is clear that this is not possible in our framework since posterior beliefs about a given fund's alpha depend on other fund returns through the estimates of μ_α and σ_α. Naturally, excluding the "losers" results in an estimate of μ_α that is biased upwards. Later, we estimate this bias to be 50 to 60 basis points per annum.[5]

The second issue concerns the behavior of the maximum posterior alpha estimate across all funds. BMW focus on the important question of whether *any* active investment in mutual funds is warranted. A sufficient condition for some active management is that the maximum posterior mean alpha is positive after subtracting expenses. BMW conclude that the maximum is indeed positive, even when investors have priors that are very skeptical about the existence of skilled managers. While we also find that the maximum mean alpha is positive, agreement between the procedures need not occur in general. This is clear from the following disturbing implication of prior independence.

With independent residuals, the maximum OLS alpha estimate becomes unbounded as the number of funds approaches infinity, even when the true alphas

---

[5]Recent independent work by Stambaugh (2003) also explores survivorship issues in the context of prior dependence. Our framework differs from his in that we allow for uncertainty in both the mean (μ_α) and variance (σ_α^2) of the cross-sectional distribution of alphas, rather than just the mean, which is Stambaugh's focus. We find that learning about σ_α is crucial for such issues as the plausibility of extreme alphas. Stambaugh's analysis of survivorship issues extends to the case in which nonsurvivors' returns are not observed, which is a more realistic assumption for the universe of hedge funds.

are all zero. This follows from standard properties of order statistics under independent sampling. Therefore, given prior independence, no matter how high the initial degree of prior skepticism about superior performance, with enough funds the data will eventually dominate and favor active investment in the funds with extreme regression estimates. Of course, this need not occur with a fixed number of funds, but the limiting case suggests that, more generally, extreme performance due to chance will be given too much weight.

In contrast, in our model with learning across funds, all alpha estimates are shrunk toward the pooled estimate of μ_α, which will tend toward zero if fund managers have no skill. There is greater shrinkage as the number of funds grows large since the estimate of σ_α also approaches zero in this context. Given this countervailing force, the maximum Bayesian estimate of alpha need not be high and, in particular, need not even be positive after deducting expenses. We present simulation evidence for $N = 10,000$ funds that supports this conclusion. This sort of evaluation of the behavior of different priors under hypothetical circumstances can be helpful in the process of eliciting a prior with which one can identify. Here, it underlines the importance of incorporating dependence in the joint prior on the vector of alphas.

The remainder of this paper is organized as follows. Section 2 introduces our model with learning across funds and provides an overview of the estimation procedure. Simulations are used in Sections 3 and 4 to examine the properties of the Bayesian estimators in repeated sampling. Specifically, Section 3 focuses on the estimation of μ_α and σ_α, while Section 4 compares estimates of alpha based on our learning model with those obtained under prior independence. Empirical results using returns from the CRSP Mutual Funds data file are presented in Section 5 and robustness along several dimensions is explored in Section 6. Section 7 summarizes our results and discusses implications of our basic framework for future work on asset pricing tests.

2. The model with continuous learning priors for alpha

In our initial exploration of prior dependence, we adopt the simplest features of both BMW and PS. Like PS, we posit a model in which beliefs about fund alphas are represented by continuous densities. In contrast, BMW truncate the distribution and place positive mass at a negative value of alpha that reflects the average loss of an unskilled manager to superior managers. In our empirical application, skill is defined relative to the CAPM, the Fama and French (1993) three-factor model, an expanded model that includes the Carhart (1997) momentum factor motivated by the work of Jegadeesh and Titman (1993), and a seven-factor model that includes, in addition to the previous four factors, three factors constructed to explain industry return covariation orthogonal to the other four factors. PS go further by identifying a subset of the passive assets as pricing model benchmarks and incorporating prior beliefs about model mispricing as well as skill. Like BMW, we only consider beliefs about skill.

2.1. Model and prior specification

We assume that excess returns have a linear factor structure,

$$r_{j,t} = \alpha_j + \beta_j' F_t + \varepsilon_{j,t}, \tag{1}$$

where $\varepsilon_{j,t} \sim N(0, \sigma_j^2)$. Following BMW, we assume that factor model residuals are cross-sectionally uncorrelated. Pastor and Stambaugh (2002a,b) impose this condition after "non-benchmark" passive assets have been included in the factor model. Their assumption is roughly equivalent to that made in our seven-factor model. We explore the effects of weakening the residual independence assumption toward the end of the paper. The vector of factors F_t is assumed to be observable. In our applications, it is taken as some vector of excess returns on benchmark portfolios.

The investor views true alphas as random draws from a normal distribution with unknown mean μ_α and unknown standard deviation σ_α. Formally, this is the *conditional* prior for each α_j given μ_α and σ_α. Prior beliefs about μ_α and σ_α then determine the marginal priors for the alphas. Because all alphas depend on these same two parameters, the alphas are not independent of one another in the prior. In addition, the marginal prior of each alpha is non-Gaussian since it is a mixture of normals. Priors for μ_α and σ_α are assumed independent and are represented by a normal distribution for μ_α and an inverted gamma distribution for σ_α. The numerical values used in these priors are given in the next section.

In contrast, the priors for betas and residual variances are diffuse (proportional to $1/\sigma_j$), independent of the alphas, and independent across managers. While informative dependent priors could be introduced for these parameters as well, the greater precision with which these parameters are estimated makes such an extension less interesting. Later, we allow the prior for alpha to be conditioned on a fund's residual variance.

2.2. Overview of the estimation procedure

In this section, we briefly discuss the main features of our estimation procedure. Further details are given in the appendix. To simplify the computation, we use a hierarchical approach in which parameters are divided into sets, some global and some fund-specific. The global parameters, which affect all funds, consist of μ_α and σ_α. Fund-specific parameters include all the α_j, β_j, and σ_j. Using the Gibbs sampler, we can characterize the joint posterior of all these parameters by analyzing only one set at a time. By cycling repeatedly through draws of each parameter conditional on the remaining parameters, the Gibbs sampler produces a Markov chain of parameter draws whose joint distribution converges to the posterior.[6]

The Gibbs sampling approach that we use divides the parameters into four blocks, each of which consists of a draw from a known conditional posterior distribution.

1. σ_α conditional on α_j ($j = 1, \ldots, M$) and μ_α.
2. μ_α conditional on α_j ($j = 1, \ldots, M$) and σ_α.

---

[6]See Casella and George (1992) for an introduction to the Gibbs sampler.

3. σ_j and β_j conditional on F, r_j, and α_j for all $j = 1, \ldots, M$.
4. α_j conditional on μ_α, σ_α, F, r_j, β_j, and σ_j for all $j = 1, \ldots, M$.

While the appendix describes each draw in detail, we outline each step briefly here. As shown in the appendix, any parameters not conditioned on are irrelevant for that draw.

In step 1, given μ_α and all the α_j, the conditional distribution of σ_α combines the normal likelihood of the α_j with the inverted gamma prior for σ_α. It is well known in this case that the conditional distribution of σ_α is also an inverted gamma. Step 2 then combines the normal likelihood of the α_j with the normal prior for μ_α. The draw of μ_α is therefore normal as well.[7]

Step 3 replicates traditional linear regression analysis using conjugate priors. Since priors on β_j and σ_j are flat and independent of α_j, we may simply subtract α_j from fund j's excess returns and proceed with the draws of σ_j from its inverted gamma distribution and β_j from its Student-t distribution.

Standard conjugate analysis is also used in step 4, where a normal likelihood for each α_j (conditional on β_j and σ_j) is combined with a normal prior with mean μ_α and standard deviation σ_α. In this case the conditional distribution of α_j is normal as well.

2.3. Frequentist properties of Bayesian procedures

A distinctive feature of Bayesian inference is that the probabilistic analysis is conditioned entirely on the given data. This differs from the classical or frequentist approach, which considers the average behavior of statistics under hypothetical repetitions of the experiment on new data sets—data that is not actually observed. Frequentist properties can still be of interest to a Bayesian from a *pre-experimental* perspective, however. As Berger (1985, p. 26) explains, before looking at the data, one can only measure how well a statistical procedure "is likely to perform through a frequentist measure, but after seeing the data one can give a more precise final measure of performance."[8]

In Section 3, we conduct a frequentist analysis by repeatedly applying our Bayesian methodology to panels of randomly simulated mutual fund data and tracking the average behavior of various characteristics of the posterior distribution of the alphas. We examine sensitivity to the number of funds in the panel as well as different levels of prior skepticism about the magnitude of managerial skill. In Section 4, we make comparisons that highlight the role of prior dependence in forming posterior beliefs about alphas.[9] Besides enhancing our insight into the potential performance of various procedures on actual data, an analysis of this sort can play an important role in the process of eliciting a satisfactory prior. If repeated

---

[7] Note that in many similar Bayesian settings, the draw of σ_α would not condition on μ_α. Our setting differs in that the prior on μ_α has a fixed standard deviation rather than one that is proportional to σ_α. Since this prior is not fully conjugate, our setting requires the additional conditioning argument.

[8] Savage (1962) makes this distinction between initial precision and final precision.

[9] Stambaugh (1997) and Jones (2003) also explore the frequentist properties of Bayesian procedures in financial contexts.

application of a given prior to *hypothetical* data reveals properties that are inconsistent with one's intuition about how a properly specified procedure *should* behave, then it may be time to go back and modify the prior specification so as to better reflect one's actual belief. Of course, all of this exploration and refining of priors should occur, in principle, before making any inference or decision with the actual data.

The priors on μ_α and σ_α that we use reflect different views on the level of skill in the population of fund managers. Three versions of our learning prior are considered, namely, high skepticism, some skepticism, and no skepticism. The no-skepticism prior is taken to be diffuse for both μ_α and σ_α (proportional to $1/\sigma_\alpha$). In this case, the data will dominate our beliefs. The other priors for μ_α and σ_α are informative. The μ_α priors are normally distributed with mean zero and standard deviation 0.25% (high skepticism) or 1% (some skepticism). All numbers given are annualized monthly figures. With high skepticism, σ_α has an inverted gamma prior centered around 0.75%, with 100 degrees of freedom. With some skepticism, the values are 3% and 10, respectively. Thus, in these priors, greater skepticism is associated with a stronger belief that μ_α is close to zero, as well as greater confidence that the true alphas will be close to μ_α. However, one can also imagine plausible scenarios in which cross-sectional variation in skill would be perceived as quite high, even if there were a strong belief that the *average* level of skill is close to zero.

3. Simulation results with learning priors

Now we study the distribution of beliefs that investors with the priors above would arrive at given different data sets. First, we consider a world in which managers have no skill at all, and then we consider one in which the average fund manager is skilled. For each experiment, we run 1,000 Monte Carlo simulations. Let M equal the number of funds in our hypothetical panel of returns. We consider values of M ranging from 10 to 10,000 in order to get a sense of the rate at which investors learn about the true parameter values. All funds are assumed to exist over the same 77-month sample period. The actual number of funds in the empirical sample analyzed later in the paper is 5136, with an average life of 77.3 months.

Fund returns are generated under the factor model in Eq. (1) assuming a single factor with a monthly mean excess return of 0.005 and a standard deviation of 0.045. The β and σ parameters for each fund are drawn randomly and independently of each other and of other funds. β is drawn from a normal distribution with mean 1 and standard deviation 0.29, while $\ln(\sigma)$ is normal with mean -3.7 and standard deviation 0.5, a distribution that implies a mean σ of 0.028 with a standard deviation of 0.015 (also expressed on a monthly basis). Both distributions conform closely with the OLS estimates of these parameters obtained from the empirical sample used later in the paper. When linear factor pricing does not hold and managers may be skilled ($\alpha \neq 0$), the alphas are also drawn independently from a normal distribution with annualized values specified for the mean μ_α and standard deviation σ_α.

3.1. Simulations when managers have no skill ($\alpha_j = 0$)

Results are presented in plots that display the sampling moments of various posterior means or functions of posterior means. Fig. 1 shows that the initial prior can have a significant effect on beliefs about μ_α, the mean of the population from which the true alphas are drawn. Panel A of that figure indicates that inferences about μ_α are correct on average (across 1,000 simulations) regardless of the sample size, which is not surprising given that the priors are centered around this value. The qualitative patterns observed in the rest of the figure follow from a few basic principles. Since the (true) expected value of each alpha estimator is zero in our no-skill population, with residual independence, the cross-sectional average of the alpha
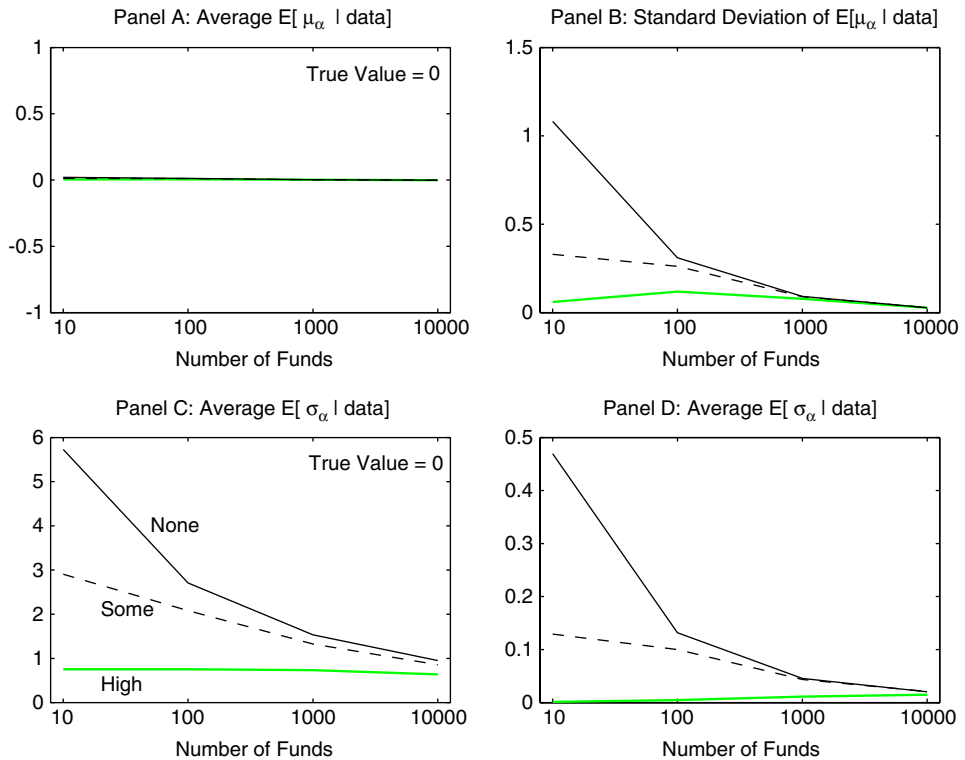

Fig. 1. Monte Carlo averages and standard deviations of μ_α and σ_α when managers are unskilled. Learning priors assume that the M fund alphas are random draws from a normal distribution with mean μ_α and standard deviation σ_α. Each panel displays an average or standard deviation of posterior moments across 1,000 Monte Carlo simulations of hypothetical data. For each fund, a sample of 77 monthly fund returns is generated from a one-factor model with true alpha equal to zero. Analyses are performed under three degrees of prior skepticism—none, some, and high—denoted, respectively, by solid, dashed, and grey lines. The highly and somewhat skeptical normal priors for μ_α have mean zero and standard deviations 0.25 and 1.0, respectively; the corresponding inverted gamma priors for σ_α are centered around 0.75 and 3.0, with degrees of freedom 100 and 10. The unskeptical priors are diffuse.

estimates must converge to zero as $M \to \infty$. For large M, the influence of the prior becomes negligible as well. Consequently, for each of our priors, the standard deviation (across 1,000 simulations) of posterior means of μ_α approaches zero when M is sufficiently large. This is observed in Panel B, which shows that the dispersion of beliefs about μ_α is quite small for M of 1,000 or more. Thus, investors become increasingly convinced that their prior mean was correct, i.e., that managers have no skill on average.

In general, we can think of the posterior mean as roughly a weighted combination of a cross-sectional average of the alpha estimates and zero, the prior mean of μ_α. In other words, the average estimate is shrunk toward zero in forming the posterior mean of μ_α. Shrinkage is greatest when M is low (little data) and when the prior is very precise (high skepticism). In the extreme, when $M = 0$, the mean of μ_α is just the prior mean of zero and there is no variability at all. Thus, there are two offsetting effects of increasing M: Higher M increases data precision, which reduces dispersion across simulations, but increasing M also reduces shrinkage, which tends to increase dispersion. Initially, the shrinkage effect is dominant, but eventually the data precision effect takes over. Since shrinkage is greatest for the high-skepticism prior, it takes longer for the data precision effect to dominate and, as a result, dispersion in the posterior means increases in going from $M = 10$ to 100, as is evident in Panel B.

By similar reasoning, since the informativeness of the data is held constant when M is fixed, we would expect dispersion to increase as shrinkage is reduced in going from highly skeptical to unskeptical priors. This effect should be greatest when M is small and shrinkage is substantial. The patterns in Panel B confirm these ideas.

Panels C and D of Fig. 1 present results for the posterior means of σ_α under the same scenarios. When $M = 10$, the locations of the first two distributions largely reflect the assumptions about σ_α in the informative priors. Increasing M does not have much impact in the high-skepticism case, as the data are apparently never given much weight. With some skepticism, the means for σ_α decline from around 3% with $M = 10$ to about 1% with $M = 10,000$. Investors learn very gradually that not only is there no skill on average ($\mu_\alpha = 0$), but there is no skill *at all* ($\mu_\alpha = 0$ and $\sigma_\alpha = 0$) in this population. The learning is more pronounced with the no-skepticism diffuse prior, which is not anchored at any particular value. The large posterior mean σ_α of about 6% in this case, with $M = 10$, may in part reflect the considerable uncertainty about the location of the mean.

3.2. Simulations when managers have some skill ($\alpha_j \neq 0$)

We now summarize a similar simulation experiment in which the true alpha of each fund is drawn randomly from a normal distribution with $\mu_\alpha = 0.6\%$ and $\sigma_\alpha = 1.5\%$, a draw that is independent of the draws of β_j and σ_j and of the draws for other funds. In panels A and B of Fig. 2, we see again that the average simulated posterior mean for μ_α converges toward the true value, with considerable learning occurring by the time M equals 1,000, especially for the less skeptical priors. Similarly, the lower panels show that by $M = 10,000$, investors are likely to conclude that σ_α is close to the true value 1.5%. In the case of high skepticism, however, the
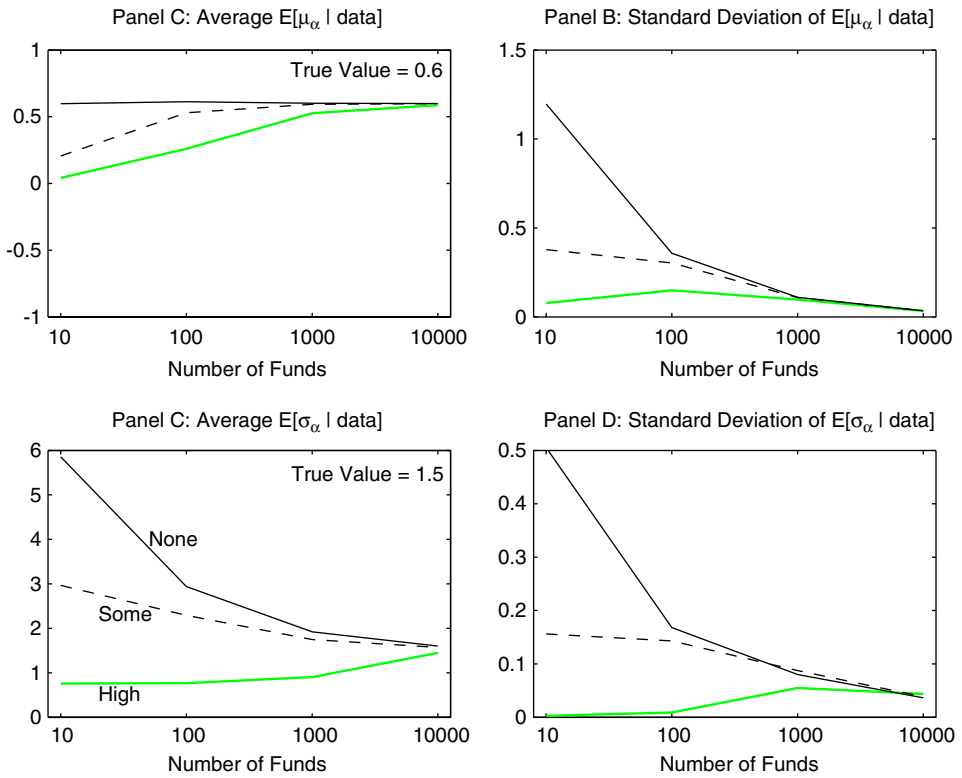
Fig. 2. Monte Carlo means and standard deviations of μ_α and σ_α when managers have skill. Learning priors assume that the M fund alphas are random draws from a normal distribution with mean μ_α and standard deviation σ_α. Each panel displays an average or standard deviation of posterior moments across 1,000 Monte Carlo simulations of hypothetical data. For each fund, a sample of 77 monthly fund returns is generated from a one-factor model with true alpha drawn from a normal distribution with annualized mean 0.6% and standard deviation 1.5%. Analyses are performed under three degrees of prior skepticism—none, some, and high—denoted, respectively, by solid, dashed, and grey lines. The highly and somewhat skeptical normal priors for μ_α have mean zero with standard deviations 0.25 and 1.0, respectively; the corresponding inverted gamma priors for σ_α are centered around 0.75 and 3.0, with degrees of freedom 100 and 10. The unskeptical priors are diffuse.

prior largely dominates the belief about σ_α for M as large as 1,000. The more diffuse investor beliefs naturally adjust more quickly.

4. The impact of learning across funds: Simulation results

Having explored the basic properties of our model with learning priors, we now compare simulation results based on our model with those based on a model with prior independence across funds. To highlight the impact of learning across funds, the *marginal* priors are taken to be the same whether we incorporate dependence or

not. These marginal priors for alpha are the unconditional "mixtures" implied by the three joint priors for μ_α and σ_α considered above, given the assumption that alphas are drawn from the normal distribution $N(\mu_\alpha, \sigma_\alpha^2)$. Our objectives are to determine whether incorporating dependence has much of an effect on posterior beliefs and to evaluate the extent to which the different beliefs approximate the true underlying population. As with size and power calculations in classical statistics, this is done separately for each hypothesis—here, our no-skill and some-skill worlds.

The marginal priors are obtained by simulation using the fact that the density equals the expected conditional density given μ_α and σ_α. Many values of μ_α and σ_α are drawn from their prior distributions, and the densities implied by each pair are averaged at each point in a grid of alpha values to obtain the implied prior for the alphas. We find that the "somewhat skeptical" prior distribution is leptokurtic, implying a higher probability of very large and small alphas than would a normal distribution. Deviations from normality are more difficult to detect for the highly skeptical prior, whose tighter distributions for μ_α and σ_α imply a more homogeneous mixture of normals.

For each simulated data sample, we form posterior means of the alphas using both the "learning" prior considered previously and the "no-learning" prior that imposes independence across fund alphas. Inference under the latter prior is simplified by the fact that each fund can be treated separately. The non-Gaussian nature of this prior requires, however, that these posterior means be computed numerically. We make use of the fact that the no-learning posterior density for each alpha can be written (up to a constant of proportionality) as the product of the marginal prior on alpha and the posterior density of alpha that would have been obtained under flat priors, or[10]

$$p_{\text{no-learning}}(\alpha_j \,|\, r_j, F) \propto p_{\text{flat}}(\alpha_j \,|\, r_j, F) \times p_{\text{no-learning}}(\alpha_j). \tag{2}$$

We focus on three aspects of the cross-sectional (across M funds) distribution of posterior means of the alphas, namely, their average, standard deviation, and maximum. Again, it is the sampling distributions of these quantities, based on 1,000 simulations, that we examine, first in a world without skill and then in one with. The cross-sectional average and standard deviation will give us a general feel for the differences between posterior beliefs with and without learning across funds. The maximum is of interest in addressing the question of whether *any* active investment in mutual funds is warranted, as in BMW. A maximum in excess of transaction costs is sufficient to warrant some active investment in an optimal portfolio when the investment universe consists of a market index (and other benchmark assets, if any), the mutual funds, and a riskless asset.

---

[10]More specifically, we use the fact that the marginal posterior for α_j can be obtained from the joint posterior by integrating out β_j and σ_j. Given the prior independence between α_j and the other parameters, the prior for α_j can be factored out of the integral. Since it is well known that the flat prior implies a Student-t distribution for the posterior of α_j, both terms on the right-hand side are known. We numerically integrate once to obtain the normalizing constant, then integrate again to calculate the posterior mean of the α_j under the no-learning prior.

To gain some intuition for the effect of prior dependence, consider the posterior distribution of the Mth fund's alpha, given the entire data set of returns. By a standard Bayesian result, the posterior for that fund's alpha can be decomposed as

$$p(\alpha_M \mid F, r) \propto p(r_M \mid F, \alpha_M)p(\alpha_M \mid F, r_1, r_2, \ldots, r_{M-1}), \tag{3}$$

where the second term may be regarded as an "effective prior" on α_M.[11] This term represents the investor's belief about α_M before observing the returns on fund M, but after combining the initial prior with all other fund data. Under learning priors, this effective prior evolves as $M \to \infty$, eventually converging to the true cross-sectional distribution of the alphas (as long as the assumed distributional forms are correct). Under the no-learning prior, however, the other $M - 1$ funds are irrelevant, and the effective prior on the Mth fund's alpha is simply that fund's marginal prior. For a given fund, the learning prior therefore leads to a more "data-based" conclusion, since the data affect the second term in the posterior as well as the first.

More formally, since each alpha is a random draw from a $N(\mu_\alpha, \sigma_\alpha^2)$ distribution under the learning prior, the mean of the effective prior in (3) equals the posterior mean of μ_α and its variance is the posterior variance of μ_α plus the posterior mean of σ_α^2, both based on the $M - 1$ fund returns.[12] Without learning across funds, it is the *marginal prior* moments that matter. Thus, with no-learning priors, a fund's alpha estimate is shrunk toward zero while, under learning priors, there is shrinkage toward the ($M - 1$ fund) *posterior* mean of μ_α. The latter incorporates some shrinkage toward zero as well.

Because the effective prior will eventually converge to the true distribution of the alphas, the learning prior must eventually (as $M \to \infty$) lead to more accurate inferences, on average, than any no-learning prior that is not exactly equal to the true cross-sectional distribution. In finite samples, however, the relative performance of the two priors depends on how "right" the marginal prior happens to be—a prior with a mean equal to the true value and with a small enough standard deviation will naturally imply posteriors that are closer to the truth. Put differently, from a frequentist perspective there are two sources of error in the effective prior, conventional estimation errors and the error of choosing a prior that does not conform to the truth. The no-learning prior mitigates the first error by giving less weight to the data, but it is utterly vulnerable to the second. In the simulations summarized in the next section, the marginal priors are all centered around the true value of zero. In the most skeptical case, the prior is extremely tight around that value, and hence is expected to perform relatively well.

4.1. Simulations when managers have no skill

Fig. 3 presents sampling distribution means (across 1,000 simulations), under the assumption that all managers are unskilled, for three functions of the posterior

---

[11]Earlier we spoke of priors conditioned on the true values of some alphas. Here, we are conditioning on some of the data.

[12]The former follows from the law of iterated expectations while the latter is based on the variance decomposition formula.

means of the alphas. For each of our three priors, results are given first without and then with learning across funds (prior dependence). For brevity, we denote the posterior mean of the alpha of fund j as $\bar{\alpha}_j$. Panels A and B depict the sampling distribution means of the cross-sectional average $\bar{\alpha}_j$, illuminating one dimension of the performance of learning and no-learning priors. As in Fig. 1, since the lack of skill in any of the Monte Carlo samples is consistent with prior beliefs, both learning and no-learning priors result in inferences that are correct, at least on average.

Arguing as earlier, with independent informative priors, the prior variance of the average alpha ($1/M$ times the marginal variance of alpha) approaches zero as M increases. Although the prior variance still declines with M under learning priors, it does not approach zero since *prior* uncertainty about the common μ_α component is unaffected.[13] Consequently, there will be more shrinkage toward the prior mean in the no-learning case, resulting in less dispersion for the average $\bar{\alpha}_j$. If the prior correctly "guesses" the true population mean, as in Fig. 3, this is a benefit. Of course, the situation will be quite different when we simulate a world with skill but our priors continue to reflect a belief that there is none.

Next, we consider results for the cross-sectional standard deviation of the $\bar{\alpha}_j$, shown in the middle panels of Fig. 3. In general, when there is no learning across funds, the standard deviations are unaffected on average as M increases. This makes sense in that the posterior mean for each fund is an i.i.d. draw with no learning, so increasing M simply results in more precise estimates of the same underlying standard deviation of the $\bar{\alpha}_j$, a typical sampling result. As in Fig. 1, there's not much effect of learning with the high-skepticism prior. With the less skeptical learning priors, the standard deviations decline sharply and are much lower than the no-learning standard deviations. This is consistent with the earlier observations about σ_α. In short, the investor with a learning prior becomes increasingly convinced of the reality that the alphas are all zero, while her no-learning counterpart seems capable only of confirming that the *average* alpha is zero. Thus, the overall belief about the set of fund alphas is quite sensitive to the learning/dependence assumption.

The key is that with learning, the data is pooled, which permits a conclusion to be drawn about the nature of the latent population from which alphas are drawn. Upon seeing that *all* of the alpha estimates are statistically "close" to zero, for a large set of funds, the investor with a learning prior perceives the world as one in which skill is unlikely to exist and markets are efficient, which thereby informs his belief about the next fund's alpha. The investor with a no-learning prior does not recognize such a link and views the evidence for each fund in isolation. As a result, the maximum $\bar{\alpha}_j$, examined in the bottom two panels of Fig. 3, increases with M under no learning. This is to be expected in light of the well-known properties of order statistics under independent sampling. Given enough funds, there will virtually always be some fund with an extremely large alpha *estimate* and associated posterior mean, even when the true alphas are all zero.

---

[13]Conditional (on μ_α and σ_α) independence under learning priors implies that the prior variance of the average alpha is the prior variance of μ_α plus $1/M$ times the prior mean of σ_α^2. When $M = 1$, this is just the marginal variance of alpha.
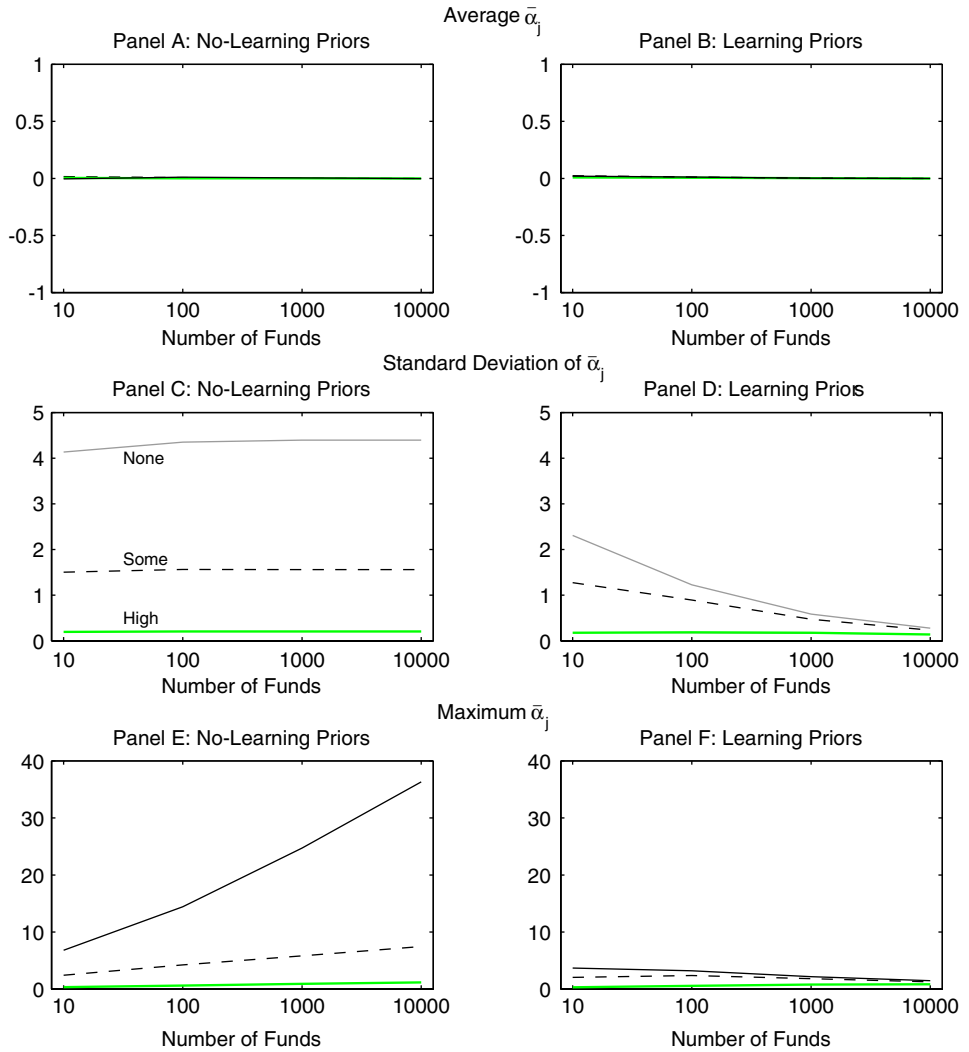
Average $\bar{\alpha}_j$

Fig. 3. Monte Carlo means when managers are unskilled. Learning priors assume that the M fund alphas are random draws from a normal distribution with mean μ_α and standard deviation σ_α. No-learning priors are independent across funds, but have the same marginal distributions as the learning priors. Each panel displays an average statistic over 1,000 Monte Carlo simulations of hypothetical data. The statistics are based on posterior mean alphas for a cross-section of M funds. For each fund, a sample of 77 monthly fund returns is generated from a one-factor model with true alpha equal to zero. Analyses are performed under three degrees of prior skepticism—none, some, and high—denoted, respectively, by solid, dashed, and grey lines. The highly and somewhat skeptical normal priors for μ_α have mean zero with standard deviations 0.25 and 1.0, respectively; the corresponding inverted gamma priors for σ_α are centered around 0.75 and 3.0, with degrees of freedom 100 and 10. The unskeptical priors are diffuse.

The situation is quite different with our less skeptical learning priors. Rather than increase, as in the no-learning case, the sampling distribution average of the maximum $\bar{\alpha}_j$ actually declines slightly as M increases. Under the no-skepticism (diffuse) prior with $M = 10,000$, a maximum as large as 40% is often observed with no learning, whereas the values with learning cluster around 1.5%. This is another manifestation of the fundamentally different perspective attained by incorporating prior dependence. With learning across funds, each fund's alpha is shrunk toward the posterior mean of μ_α, which converges to zero with M when managers have no skill (see Fig. 1). Simultaneously, shrinkage increases with M as σ_α, and hence the variance of the effective prior in Eq. (3), approaches zero (again, see Fig. 1). These effects combine to keep the posterior alphas from getting too large. More intuitively, if the returns of all other funds have convinced us that mutual fund alphas are generally close to zero, then the given fund's alpha estimate will have relatively less impact on its posterior mean.

4.2. Simulations when managers have some skill

Fig. 4 presents simulation results paralleling those in Fig. 3 for a world in which $\mu_\alpha = 0.6\%$ and $\sigma_\alpha = 1.5\%$, the same values used in Fig. 2. Since the true alphas are no longer zero, they are subtracted from the posterior means before computing the average, standard deviation, or maximum "alpha error." This facilitates the evaluation of how closely the posterior means approximate reality.

The beliefs about alphas based on the *informative* no-learning priors are anchored at the prior mean of zero. This would be true even with an infinite sample of funds, since shrinkage is not affected by adding funds under prior independence. As a result, the average error in Fig. 4 Panel A is consistently negative for these priors, whereas it is roughly zero on average under the diffuse no-learning prior.

In contrast, with the learning priors, the average error is much closer to zero, at least for $M > 10$. As seen in Panels C and D, the standard deviations under learning are less than half those without learning when priors are uninformative (no skepticism). The advantage is reduced with some skepticism and the differences are very small with high skepticism. From a mean-square error (squared mean error plus variance) perspective, therefore, the posterior mean alphas based on the learning priors are clearly superior in this world with skill. Results for the maximum in Fig. 4 under the no-skepticism prior are again striking, especially for $M = 10,000$, with distributions centered around 36% and 5% for no learning and learning, respectively. Again, these differences reflect shrinkage toward an aggregate alpha estimate with learning across funds.

5. Empirical application

Given our understanding of the behavior of learning and no-learning priors under simulated data, we now turn to an application on actual U.S. equity mutual fund data.
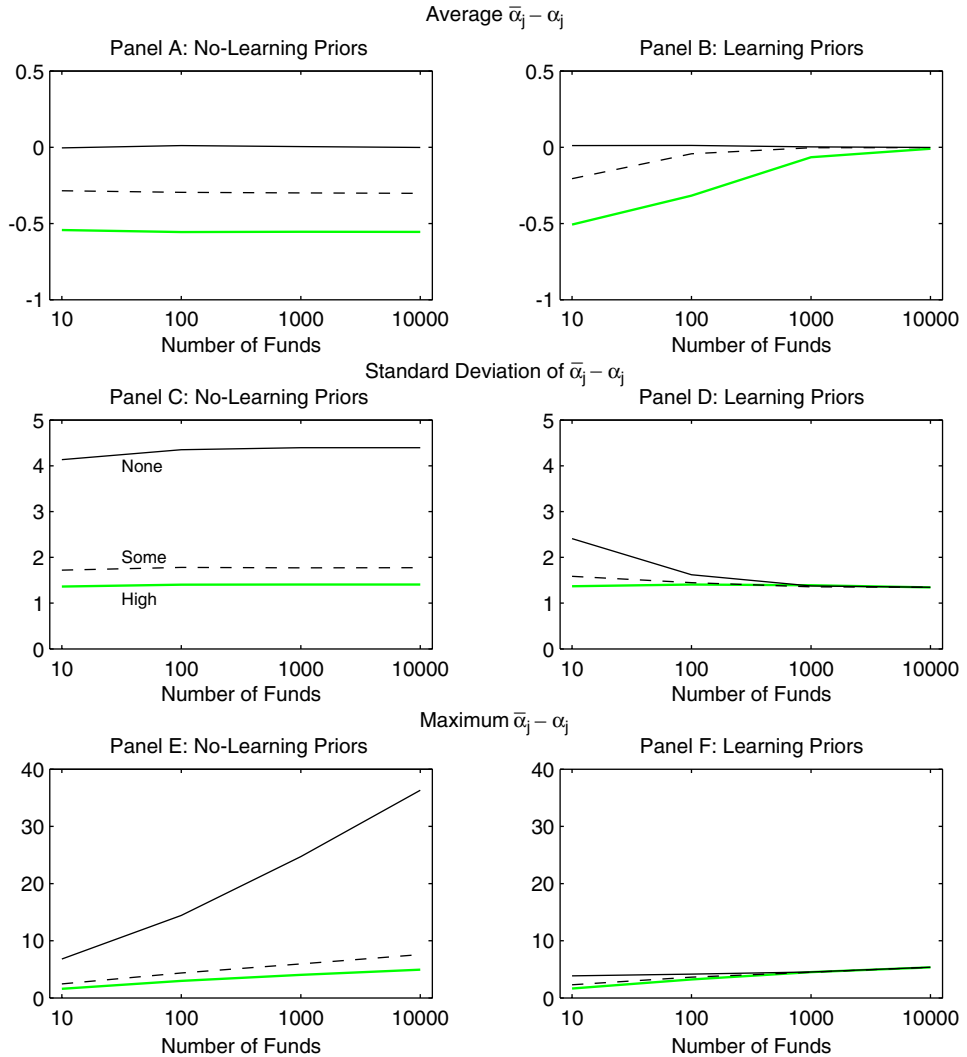
Fig. 4. Monte Carlo means when managers have skill. Learning priors assume that the M fund alphas are random draws from a normal distribution with mean μ_α and standard deviation σ_α. No-learning priors are independent across funds, but have the same marginal distributions as the learning priors. Each panel displays an average statistic over 1,000 Monte Carlo simulations of hypothetical data. The statistics are based on posterior mean alphas for a cross-section of M funds. For each fund, a sample of 77 monthly fund returns is generated from a one-factor model with true alpha drawn from a normal distribution with annualized mean 0.6% and standard deviation 1.5%. Analyses are performed under three degrees of prior skepticism—none, some, and high—denoted, respectively, by solid, dashed, and grey lines. The highly and somewhat skeptical normal priors for μ_α have mean zero with standard deviations 0.25 and 1, respectively; the corresponding inverted gamma priors for σ_α are centered around 0.75 and 3, with degrees of freedom 100 and 10. The unskeptical priors are diffuse.

5.1. Data

Our source for all mutual fund data is the 2001 CRSP Mutual Funds data file, which contains mutual fund returns from January 1961 to June 2001. To focus solely on the sample of domestic equity funds, we follow the selection procedure of BMW to eliminate funds that are likely to have made substantial allocations to other asset classes.[14] In addition, we require that the fund have at least 12 months of returns data available. This results in a sample of 5,136 funds with an average of 77.3 months of monthly return observations.

As in BMW, we focus on returns *before* fees and expenses, with the justification that it is the returns on the underlying stocks themselves that are most likely to conform with the linear pricing model. Since the mutual fund returns reported by CRSP are net of both these costs, we add them back to the reported returns. As BMW note, however, only the management fees are reported by CRSP—the transactions costs incurred by each fund are unknown. Following BMW, we assume these costs amount to six basis points per month. Unlike BMW, we include all equity mutual funds in our sample rather than just those that still existed at the end of our sample. In some cases, however, we compare these results with inferences based solely on the 3,844 funds that survived to the end of the sample.

We employ four sets of benchmark returns in our empirical work, specifically, the excess market return factor (RMRF) motivated by the CAPM, the three-factor model of Fama and French (1993) (adding SMB and HML), a four-factor model that augments the Fama–French factors with the momentum spread portfolio (MOM) of Carhart (1997), and a seven-factor model that also includes three industry factors. Our primary motivation for including the industry factors is to better approximate the assumption of residual independence. The industry factors are constructed in a manner similar to that in Pastor and Stambaugh (2002b).

First, excess returns on 30 industry-sorted portfolios are regressed on a constant, the three Fama–French factors, and the momentum factor. The unexplained part of the industry return is then defined as the residual of each regression plus that regression's intercept. These unexplained components can thus be viewed as returns on zero-investment positions. A principal components analysis is performed on these 30 time series, and the first three principal components are taken as portfolio weights for the three industry portfolios. Given concerns that there might be some sort of spurious correlation between the weights and the realized industry returns over the period, we modify the PS procedure and compute principal components separately for odd and even months. The portfolio weights applied to returns in odd months are then obtained from the even month data and vice versa.

5.2. Results with learning priors

Table 1 contains posterior means and standard deviations for μ_α and σ_α computed under various learning priors for samples of all funds and surviving funds only. It is

---

[14]We are grateful to Klaas Baks for providing the code used to construct this data set.

Table 1
Posterior means and standard deviations of μ_α and σ_α under learning priors

	Highly skeptical priors	Somewhat skeptical priors	Unskeptical priors
		K = 1 (*RMRF*)	
μ_α—all funds	1.40 (0.04)	1.47 (0.05)	1.48 (0.05)
μ_α—surviving funds only	1.92 (0.05)	2.08 (0.06)	2.11 (0.06)
σ_α—all funds	1.00 (0.07)	1.40 (0.06)	1.50 (0.06)
σ_α—surviving funds only	0.82 (0.05)	1.24 (0.06)	1.36 (0.06)
		K = 3 (*RMRF, SMB, and HML*)	
μ_α—all funds	1.30 (0.05)	1.38 (0.05)	1.38 (0.05)
μ_α—surviving funds only	1.77 (0.06)	1.92 (0.06)	1.95 (0.07)
σ_α—all funds	1.99 (0.07)	2.21 (0.07)	2.26 (0.07)
σ_α—surviving funds only	1.92 (0.08)	2.24 (0.08)	2.30 (0.08)
		K = 4 (*RMRF, SMB, HML, and MOM*)	
μ_α—all funds	1.33 (0.04)	1.37 (0.05)	1.39 (0.05)
μ_α—surviving funds only	1.73 (0.05)	1.85 (0.05)	1.87 (0.06)
σ_α—all funds	1.52 (0.06)	1.77 (0.06)	1.84 (0.06)
σ_α—surviving funds only	1.38 (0.07)	1.74 (0.07)	1.81 (0.07)
		K = 7 (*RMRF, SMB, HML, MOM, and industry factors*)	
μ_α—all funds	1.71 (0.05)	1.80 (0.05)	1.81 (0.05)
μ_α—surviving funds only	2.18 (0.05)	2.34 (0.06)	2.37 (0.06)
σ_α—all funds	2.07 (0.07)	2.27 (0.07)	2.32 (0.07)
σ_α—surviving funds only	2.07 (0.07)	2.27 (0.07)	2.32 (0.07)

Estimation is based on monthly fund returns over the period January 1961 to June 2001. All numbers are in annualized percentage terms. The highly and somewhat skeptical normal priors for μ_α have mean zero with standard deviations 0.25 and 1.0, respectively; the corresponding inverted gamma priors for σ_α are centered around 0.75 and 3.0, with degrees of freedom 100 and 10. The unskeptical priors are diffuse. The factors are the excess market return RMRF, small-big market cap SMB, high-low book-to-market equity HML, and past one-year momentum MOM. The industry factors are the normalized first three principal components of the four-factor model residuals.

immediately apparent that there is strong evidence of skill in the population of equity mutual funds. Posterior means of μ_α from the sample of all funds are generally between 1.3% and 1.8% per year and are somewhat sensitive to the choice of benchmark portfolios, with posterior standard deviations that are extremely small. Thus, the typical fund outperforms all benchmarks considered by a fairly substantial amount, at least before fees and costs.

When considering asset allocation issues, as we do later, one is necessarily restricted to existing or surviving funds. BMW address the question of whether this imparts some sort of survivorship bias on the Bayesian analysis. They make the following interesting observation. Suppose that the probability of survival is a function solely of past fund returns, with no separate dependence on the fund parameters—a seemingly reasonable assumption. In this case, posterior beliefs for the surviving funds will not be altered by conditioning on the ex post information about survival. Together with the assumptions of prior and residual independence, this implies that the posterior distribution for a surviving fund's parameters depends only on its own returns.

The situation is more subtle when prior dependence is introduced. It is easy to show that the information about survival is still redundant if we condition on the returns for *all* funds (and factors), not just survivors. However, our belief about one fund's parameters will generally depend on the returns of other funds, including the disappearing funds. The dependence arises because these other returns convey information about the average level of skill in the population, as measured by μ_α. Ignoring these returns can be likened to throwing out one tail of the sample distribution when estimating a population mean.[15]

It is clear in Table 1 that including only those funds that survived to the end of the sample results in a posterior mean for μ_α that is higher by about 40–60 basis points per year. Therefore, ignoring survivorship, while irrelevant under no-learning priors, can substantially inflate alphas computed under learning priors. The survival sample also tends to generate posteriors for σ_α that are a little closer to zero. Both of these effects are to be expected, as the survival sample is likely to exclude those funds whose alphas are in the left tail of the cross-sectional distribution. Eliminating these funds increases the mean and slightly reduces the dispersion in the sample.

Less intuitive are the patterns related to the use of different asset pricing models. Posteriors of μ_α are fairly similar across the one-, three-, and four-factor models. However, we are surprised to find that adding industry factors substantially increases the posterior mean of μ_α by about 40 basis points. The estimates of μ_α are even higher (40 to 50 basis points) than those reported in Table 1 when, as in PS, we compute the industry components using the full data sample. In principle, any passive zero-investment position should be a legitimate factor in this context, so the results are puzzling.

---

[15]In principle, one could compute the conditional posterior density $p(\theta_j \,|\, r_j, F)$ based on a censored sample if that were the only information available. However, the computation would be complicated considerably by the fact that the density (likelihood) function describing the data generating process must now reflect the censoring procedure. Stambaugh (2003) considers this issue in detail.

The various models produce much more diversity in their estimates of σ_α, with posterior means ranging from 1% to 2.3% for surviving funds. The multifactor models sometimes yield posterior means twice those of the CAPM. Thus, under the Fama–French model, for instance, there are a significant number of funds with very high or very low alphas, even if the average alpha is not much different from that of a CAPM world.

Finally, the effects of differing degrees of prior skepticism are relatively small, consistent with the greater weight given to the aggregate data under learning priors. Posterior means of μ_α under highly skeptical and unskeptical priors never differ by more than 13 basis points for all funds and 19 basis points for surviving funds only. Although inferences about σ_α are somewhat more sensitive, the degree of prior skepticism is still not as relevant as the choice of asset pricing model.

5.3. Results for individual fund alphas

The same calculations also produce posteriors for each fund's alpha. In Table 2, we compare summary statistics for the alpha posterior means under learning priors with those computed under comparable no-learning priors. The sample contains all funds, including those funds that did not survive to the end of the sample.

In general, learning and no-learning priors result in very different inferences. All versions of the learning prior yield average alpha posterior means of 1.3% to 1.8% per year, consistent with the posterior means of μ_α, while average alphas for the no-learning prior may be much higher or much lower depending on the degree of skepticism imposed. Highly skeptical no-learning priors produce average alphas no greater than 17 basis points per year, while unskeptical priors imply average alphas of over 3% for the seven-factor model. The lower values for the highly skeptical no-learning priors are consistent with the greater shrinkage toward the prior mean of zero.

Dispersion in alpha posterior means also varies greatly across learning and no-learning priors, particularly for the extreme cases of high and no skepticism. Dispersion is closely related to the degree of shrinkage, which depends on the prior standard deviations under no-learning, but on the estimates of σ_α under learning. In Table 1, those estimates all lie between the skeptical prior standard deviations of 0.75% and 3%. With diffuse priors, there's no shrinkage at all under no-learning, though shrinkage toward a "grand mean" under learning still reduces dispersion substantially.

With the Fama–French model, for example, the standard deviation of the alpha posterior means is just 1.3% for the learning prior, but over 8% for the no-learning prior. Under highly skeptical priors, the ordering is reversed, with the learning prior yielding a standard deviation nearly four times that of the no-learning prior (1.13% versus 0.33%). Here, shrinkage toward zero dominates under no-learning due to the low prior standard deviation of just 0.75%. As in Table 1, dispersion is heavily dependent on the asset pricing model as well.

One curious result from Table 2 is that for unskeptical priors, there is a large difference between average alpha means computed under no learning and learning.

Table 2
Summary statistics on posterior means of individual fund alphas (all funds)

		Highly skeptical priors	Somewhat skeptical priors	Unskeptical priors
		$K = 1$ (RMRF)		
Average α posterior mean	No learning	0.10	0.65	2.84
	Learning	1.41	1.47	1.48
Standard deviation of mean α	No learning	0.25	1.72	9.17
	Learning	0.32	0.53	0.59
		$K = 3$ (RMRF, SMB, and HML)		
Average α posterior mean	No learning	0.12	0.78	2.48
	Learning	1.32	1.38	1.38
Standard deviation of mean α	No learning	0.33	2.22	8.31
	Learning	1.13	1.28	1.31
		$K = 4$ (RMRF, SMB, HML, and MOM)		
Average α posterior mean	No learning	0.13	0.80	2.55
	Learning	1.34	1.37	1.39
Standard deviation of mean α	No learning	0.34	2.11	7.87
	Learning	0.76	0.93	0.97
		$K = 7$ (RMRF, SMB, HML, MOM, and industry factors)		
Average α posterior mean	No learning	0.17	1.03	3.03
	Learning	1.73	1.80	1.81
Standard deviation of mean α	No learning	0.37	2.27	8.41
	Learning	1.23	1.38	1.41

Estimation is based on monthly fund returns over the period January 1961 to June 2001. All numbers are in annualized percentage terms. The highly and somewhat skeptical normal priors for μ_α have mean zero with standard deviations 0.25 and 1.0, respectively; the corresponding inverted gamma priors for σ_α are centered around 0.75 and 3.0, with degrees of freedom 100 and 10. The unskeptical priors are diffuse. The factors are the excess market return RMRF, small-big market cap return SMB, high-low book-to-market equity return HML, and past one-year momentum MOM. The no-learning priors are independent across funds with the same marginal distribution as in the learning case. The industry factors are based on principal components of the four-factor model residuals.

With the CAPM benchmark, for example, the average posterior mean is 2.84% under no-learning priors but only 1.48% under learning priors. It appears that the higher average of the no-learning alphas is due to the presence of a number of recently introduced funds that happened to perform well in the late 1990s.[16]

5.4. Alphas after fees and costs

The central question addressed by BMW is whether *any* investment in actively managed mutual funds can be justified. They demonstrate that a necessary and sufficient condition such investment is that the posterior mean of the alpha for some fund be greater than the fees and transactions costs required to invest in that fund. Accordingly, in this section we examine mutual fund alphas computed after fees and costs.

The results reported in Table 3 differ from those of Table 2 in several ways. First, alphas are calculated net of the assumed trading costs and actual management fees that prevailed at the end of the sample (2001). These two components are simply subtracted from the posterior means computed previously. Like BMW, in asset allocation decisions we assume that future fees are equal to the last fee observed for each fund.

Table 3 also differs from Table 2 in that it reports statistics only on the alphas of those funds that survived to the end of the sample. This is done for comparability with BMW and is motivated by the fact that these are the only funds in which an investor could potentially allocate assets.[17] Note that although only survivors' alphas are summarized, all funds are used to compute posteriors under the learning prior. As discussed earlier, this is necessary in order to avoid survivorship bias when priors are dependent across funds.

The numbers in the table again indicate large differences between learning and no-learning priors. Under learning priors, alphas net of fees and costs are on average around −60 to −70 basis points per annum for the one- to four-factor models and between −13 and −23 basis points per annum for the seven-factor model. In all cases, this is roughly 2% below the corresponding levels before fees and costs. Thus, any advantage that fund managers have in terms of superior skill is apparently more than offset, from the investor perspective, by the surplus the managers reap and other fund/trading costs.[18] Without learning, average alphas are sometimes below −2% or above +2%, depending mostly on the degree of prior skepticism but also on

---

[16]Since these funds have fairly short track records and tend to have large residual standard deviations, they contribute relatively little to the posterior mean of μ_α which, as noted earlier, incorporates a sort of weighted least squares estimate. Under unskeptical learning priors, there is substantial shrinkage of these estimates toward μ_α. With unskeptical no-learning priors, however, there is no shrinkage, and as a result these estimates substantially raise the average alpha posterior mean.

[17]In fact, some of these mutual funds may not have survived past the end of the sample, making them uninvestable, too.

[18]Taking into account costs that might be incurred in implementing the benchmark strategies would improve relative fund performance somewhat. The expense ratio for Vanguard's Total Stock Market index fund, for example, is currently 20 basis points.

Table 3
Summary statistics on posterior means of surviving fund alphas *after* fees and costs

		Highly skeptical priors	Somewhat skeptical priors	Unskeptical priors
		K = 1 (*RMRF*)		
Average α posterior mean	No learning	−2.03	−1.17	2.32
	Learning	−0.69	−0.59	−0.57
Standard deviation of mean α	No learning	0.76	1.71	8.87
	Learning	0.75	0.84	0.87
Maximum α posterior mean	No learning	0.74	6.87	86.19
	Learning	1.72	2.96	3.23
		K = 3 (*RMRF, SMB*, and *HML*)		
Average α posterior mean	No learning	−2.00	−1.06	1.34
	Learning	−0.68	−0.60	−0.58
Standard deviation of mean α	No learning	0.81	2.33	8.62
	Learning	1.31	1.44	1.46
Maximum α posterior mean	No learning	0.63	44.54	88.13
	Learning	4.12	4.51	4.58
		K = 4 (*RMRF, SMB, HML*, and *MOM*)		
Average α posterior mean	No learning	−1.99	−1.05	1.40
	Learning	−0.71	−0.65	−0.62
Standard deviation of mean α	No learning	0.82	2.22	8.06
	Learning	1.03	1.15	1.19
Maximum α posterior mean	No learning	1.21	39.36	81.26
	Learning	2.96	3.26	3.28
		K = 7 (*RMRF, SMB, HML, MOM*, and *industry factors*)		
Average α posterior mean	No learning	−1.94	−0.79	1.86
	Learning	−0.23	−0.14	−0.13
Standard deviation of meanα	No learning	0.85	2.38	8.51
	Learning	1.39	1.51	1.54

Table 3 (*continued*)

		Highly skeptical priors	Somewhat skeptical priors	Unskeptical priors
Maximum α posterior mean	No learning	1.59	39.01	96.72
	Learning	4.21	4.79	4.97

Estimation is based on monthly fund returns over the period January 1961 to June 2001. All numbers are in annualized percentage terms. The highly and somewhat skeptical normal priors for μ_α have mean zero with standard deviations 0.25 and 1.0, respectively; the corresponding inverted gamma priors for σ_α are centered around 0.75 and 3.0, with degrees of freedom 100 and 10. The unskeptical priors are diffuse. The factors are the excess market return RMRF, small-big market cap return SMB, high-low book-to-market equity return HML, and past one-year momentum MOM. The no-learning priors are independent across funds with the same marginal distribution as in the learning case. The industry factors are based on principal components of the four-factor model residuals.

the benchmark portfolios used. Standard deviations are also sensitive to the choice of priors, as in Table 2.

In addition, Table 3 shows the maximum posterior mean for every combination of prior and set of benchmark portfolios. In each case, this maximum posterior mean is positive, indicating that there is always at least one fund whose alpha, net of fees and costs, is greater than zero. Our results therefore support BMWs conclusion that *some* allocation to actively managed funds is likely warranted.[19]

While the maximum alpha mean, net of fees and costs, is always positive, its magnitude is frequently far different under the two priors. Using the Fama–French factors, for example, the highest no-learning mean alpha is just 63 basis points for highly skeptical priors, but an enormous 88.13% with no skepticism. Mean alphas under learning priors in the same two cases are 4.12% and 4.58%, respectively. While investment in these funds is positive in all of these cases, the extent of this investment would vary widely.

5.5. Examples

Looking at some specific examples should be helpful in synthesizing what we have learned in this research. We focus on before-cost alphas from the one-factor model. One of the top-performing funds under the no-learning unskeptical prior was Schroder Capital's Ultra Fund, a "micro cap" fund with (annualized) posterior mean alpha of 65%.[20] There are only 44 monthly returns for this fund and residual risk is 5.6%. Our second example is the well-known Fidelity Magellan Fund, with 457 monthly returns and a lower residual risk of 3.4%.

---

[19]Recall that BMW also model a probability q that a manager is skilled and explore the impact of different values for this additional parameter.

[20]Among all funds with at least three years of returns data, Schroder was the top performer (under unskeptical no-learning priors) both before and after fees and costs. The posterior mean of Schroder's alpha *after* fees and costs was 62.1%.

Table 4
Alpha posterior means and standard deviations for two mutual funds under the one-factor model

	Degree of prior skepticism		
	High	Some	None
Schroder Ultra			
No-learning	0.2	8.1	65.0
	(0.8)	(11.1)	(10.4)
Learning	1.7	2.1	2.2
	(1.0)	(1.4)	(1.6)
Fidelity Magellan			
No-learning	1.5	8.5	10.4
	(0.8)	(1.9)	(1.9)
Learning	3.3	4.6	4.8
	(1.0)	(1.1)	(1.3)

"High" and "Some" skepticism denote priors on μ_α that are normal with zero mean and standard deviation 0.25% and 1%, respectively. Corresponding priors for σ_α are centered around 0.75% and 3%, respectively. Results for "None" are based on diffuse priors.

The before-cost annualized alpha means and standard deviations (in %) under various priors are as reported in Table 4. As in our other tables, the degree of skepticism about the magnitude of skill declines from left to right. Also, as earlier, the entire panel of fund data is used in the estimation under learning priors.

First, consider the results for no-learning priors. Under the unskeptical (diffuse) prior, the posterior mean is just the OLS regression estimate and the posterior standard deviation is equal to the OLS standard error apart from a slight degrees of freedom adjustment. The enormous estimate of 65% for the Ultra Fund is shrunk very close to the prior mean of zero under the high-skepticism prior. On the other hand, the Magellan Fund, with a much lower OLS estimate of 10.4%, has a *higher* posterior mean under high skepticism. The reason is that with a much longer time series and lower residual risk, the Magellan estimate is far more precise, resulting in less shrinkage toward the prior mean. The greater precision also accounts for the lower posterior standard deviations of the Magellan Fund alphas.

To further illustrate these ideas, we can approximate the calculation for Magellan under high skepticism as follows. Since the marginal priors under no-learning are taken to be the same as those under learning, the prior variance for alpha is roughly equal to the squared mean of σ_α plus the variance of μ_α, or $0.0075^2 + 0.0025^2$, and the precision is the reciprocal, or 16,000. Given the standard error of about 1.9%, the precision of

the OLS estimate for Magellan is $1/0.019^2 \approx 2,770$ and, therefore, the posterior weight on this estimate is $2770/18,770 \approx 0.15$. The resulting precision-weighted average of the prior mean and the OLS estimate is $0.85(0) + 0.15(10.4\%) = 1.5\%$, which is equal to the posterior mean under no-learning in Table 4.

Now, consider the learning prior results. All alpha posterior means reflect shrinkage toward the posterior means for μ_α, which range from 1.4% to 1.5% for the one-factor model in Table 2. For the reason just discussed, shrinkage is again much greater for the Ultra Fund, resulting in alphas that are uniformly lower than those for Magellan. As another illustration, consider the Magellan calculation under no skepticism (none). Since the posterior standard deviations in Table 1 are relatively small, the effective prior under learning is roughly $N(\mu_\alpha, \sigma_\alpha^2)$ with $\mu_\alpha = 1.48\%$ and $\sigma_\alpha = 1.5\%$. Basing the "prior" precision on this value of σ_α, and recalling the 1.9% OLS standard error for Magellan, the precision-weighted average alpha is now $0.62(1.48\%) + 0.38(10.4\%) = 4.9\%$, close to the posterior mean of 4.8% in Table 4.

For each fund, learning across funds results in larger alphas under high skepticism but smaller alphas otherwise. With tight (skeptical) priors, this is the result of shrinkage toward zero (prior mean) under no-learning, but toward 1.4% (mean μ_α) under learning. The ordering of alphas reverses as the prior becomes more diffuse (less skeptical) since shrinkage under no-learning declines, while the data-based shrinkage under learning remains substantial.[21]

Differences in posterior standard deviations of alpha under learning and no-learning priors can be understood in a similar manner. Under high skepticism, greater shrinkage toward zero reduces the standard deviation under no-learning. With less skepticism, the data play a greater role and shrinkage toward the pooled estimate of μ_α lowers posterior variability under learning priors. The nonmonotonic behavior of the Schroder fund's standard deviations as we vary the degree of skepticism under no-learning is surprising and reflects the bimodal nature of the posterior distribution for alpha in this case.

Finally, note that alphas need not be lower under learning priors. The three-factor alpha of the Dreyfus Premier Aggressive Growth Fund, for example, has a posterior mean of -8.5% before fees and costs under somewhat skeptical no-learning priors. With the same degree of prior skepticism, learning priors imply a posterior mean of -1.8%.

5.6. Optimal asset allocation

While a complete analysis of optimal investment in equity mutual funds is beyond the scope of this paper, we continue the previous examples and ask how variations in prior beliefs affect the allocations to a particular mutual fund. Specifically, we consider an investor who is able to allocate assets to the value-weighted market index, a risk-free asset yielding 6% interest, and *either* the Schroder Capital Ultra Fund or the Fidelity Magellan Fund.

---

[21]To keep things relatively simple, we do not incorporate the link between residual variance and the prior standard deviation of skill used by BMW and PS. Such a link, considered in Section 6.2, would make the Ultra prior less precise and reduce shrinkage to the prior mean somewhat, as compared to Magellan.

Following Kandel and Stambaugh (1996), the investor is assumed to maximize the expectation of a power utility function of the form

$$U(W_{T+1}) = \begin{cases} \dfrac{W_{T+1}^{1-A}}{1-A} & \text{for } A > 0 \text{ and } A \neq 1 \\ \ln(W_{T+1}) & \text{for } A = 1, \end{cases} \tag{4}$$

where $A = 1$, 2, or 5. Given a \$1 investment at time T, the end of the sample, the investor's end-of-period wealth is given by $W_{T+1} = 1 + r_{f,T+1} + w_j r_{j,T+1} + w_m r_{m,T+1}$. Here, r_f is the riskless return, while r_j (w_j) and r_m (w_m) are excess returns on (allocations to) the fund and the market index, respectively. These returns are net of annualized fees and costs, which are estimated to be 2.7% for Schroder and 1.6% for Magellan.

The expectation of $U(W_{T+1})$ is taken with respect to the investor's predictive distribution for $r_{j,T+1}$ and $r_{m,T+1}$, which incorporates posterior parameter uncertainty. This is given by

$$p(r_{j,T+1}, r_{m,T+1} \mid r, r_m) = \int p(r_{j,T+1}, r_{m,T+1} \mid \theta_j, \theta_m) p(\theta_j, \theta_m \mid r, r_m) \, d\theta_j \, d\theta_m, \tag{5}$$

where θ_m and θ_j denote, respectively, the parameters of the distributions of r_m and r_j given r_m. The vector of all fund returns is denoted as r. As is often the case in regression models, with independent priors for θ_m and θ_j, the posterior distribution can be factored as

$$p(\theta_j, \theta_m \mid r, r_m) \propto p(\theta_j \mid r, r_m) p(\theta_m \mid r_m), \tag{6}$$

where $p(\theta_j \mid r, r_m)$ has been the object of our study thus far. Yet to be examined is $p(\theta_m \mid r_m)$, which, despite its irrelevance for inferences about θ_j, is important for determining allocations to the market portfolio.

Our approach to computing $p(\theta_m \mid r_m)$ is standard. Letting $\theta_m \equiv \{\mu_m, \sigma_m\}$, we assume that $r_m \sim$ i.i.d. $N(\mu_m, \sigma_m^2)$. Given the diffuse prior $p(\mu_m, \sigma_m) \propto 1/\sigma_m$, the posterior distribution of σ_m is an inverted gamma and the posterior of μ_m is a Student-t. Using monthly excess value-weighted market returns from January 1961 to June 2001, we find the posterior distribution of μ_m to have a mean of 0.47% and a standard deviation of 0.21%. The posterior of σ_m has a mean of 4.48% and a standard deviation of 0.15%.

Ten thousand draws from the predictive distribution (5) are simulated by first drawing θ_j and θ_m at random from their respective posteriors and then drawing returns from

$$p(r_{j,T+1}, r_{m,T+1} \mid \theta_j, \theta_m) = p(r_{j,T+1} \mid r_{m,T+1}, \theta_j) p(r_{m,T+1} \mid \theta_m). \tag{7}$$

Optimal portfolio weights are solved numerically by maximizing the sample average of $U(W_{T+1})$, taken across the 10,000 draws. As Kandel and Stambaugh (1996) note, expected power utility may equal $-\infty$ when the total allocation to risky assets is 100% or when short sales are allowed. We therefore impose the constraints that $w_j + w_m \leqslant 0.99$, $w_j \geqslant 0$, and $w_m \geqslant 0$. For brevity, we report only the optimal allocation to the mutual fund.

Table 5
Optimal portfolio allocations to two mutual funds under the one-factor model

| | | Degree of prior skepticism | | |
		High	Some	None
		Schroder		
No-learning	$A = 1$	0.000	0.501	0.990
	$A = 2$	0.000	0.316	0.990
	$A = 5$	0.000	0.136	0.990
Learning	$A = 1$	0.000	0.000	0.000
	$A = 2$	0.000	0.000	0.000
	$A = 5$	0.000	0.000	0.000
		Fidelity		
No-learning	$A = 1$	0.443	0.990	0.990
	$A = 2$	0.042	0.990	0.990
	$A = 5$	0.011	0.529	0.672
Learning	$A = 1$	0.990	0.990	0.990
	$A = 2$	0.669	0.925	0.945
	$A = 5$	0.268	0.371	0.379

"High" and "Some" skepticism denote priors on μ_α that are normal with zero mean and standard deviation 0.25% and 1%, respectively. Corresponding priors for σ_α are centered around 0.75% and 3%, respectively. Results for "None" are based on diffuse priors.
A denotes the investor's coefficient of relative risk aversion.

Broadly speaking, fund allocations tend to track the alphas reported above once fees and transactions costs are taken into account. The alphas of these funds are higher with less skeptical priors, particularly in the no-learning case, and the allocations reflect this finding. With no-learning priors, as for the alphas themselves, allocations to the same fund can differ dramatically between investors with different degrees of prior skepticism. With learning, the information extracted from other fund returns reduces the influence of the prior. As a result, investors with learning priors tend to arrive at the same view of managerial skill regardless of their initial belief, and so their allocations are not very sensitive to the degree of prior skepticism.

The differences between allocations under no-learning and learning, as displayed in Table 5, are extreme in some cases. With no skepticism, an investor with no-learning priors would allocate 99%, the maximum, to the Schroder Fund for all levels of risk aversion considered, while investors with learning priors always allocate zero to Schroder. The latter reflects the fact that fees and costs of 2.7% for Schroder exceed the mean alphas with learning in Table 4. Allocations for the Fidelity Fund

sometimes differ substantially as well. A somewhat risk-averse ($A = 2$) investor with highly skeptical learning priors would allocate 67% to Magellan, while a comparable investor with no-learning priors would allocate only 4.2%.

6. Alternative specifications

6.1. Residual independence

Thus far, our analysis has assumed that factor model residuals are cross-sectionally uncorrelated. This simplification is driven, in part, by the complexity in estimating the millions of residual correlations, which is even more problematic than usual given the differences in sample sizes and sample periods across funds.

As a first step toward addressing this issue, we replace the assumption of residual independence with the somewhat stylized yet substantially more general assumption that the residual covariance matrix has a linear factor structure, i.e.,

$$\text{Cov}(\varepsilon_t, \varepsilon_t') = \delta\delta' + \Omega, \tag{8}$$

where Ω is diagonal and ε_t is the vector, from (1), of factor model errors that would exist if all M funds were simultaneously observed. The "residual factors" are assumed to be orthogonal and, without loss of generality, to have unit variance.

Estimation of this model follows the algorithm of Geweke and Zhou (1996) almost exactly by augmenting the data with latent residual factors, denoted G_t. Rewriting $\varepsilon_{j,t}$ as $\delta_j' G_t + \xi_{j,t}$, the factor model (1) becomes

$$r_{j,t} = \alpha_j + \beta_j' F_t + \delta_j' G_t + \xi_{j,t}, \tag{9}$$

where $\xi_{j,t} \sim N(0, \Omega_{j,j})$ is cross-sectionally independent. The latent factors have mean zero and are orthogonal to the benchmark factors F_t. We maintain the assumption that the benchmark is correctly specified, so our previous interpretation of α_j as a measure of skill is unaffected. Given our normality assumptions, Geweke and Zhou show that the distribution of G_t conditional on the parameters and all other time-t quantities is Gaussian. Therefore, for each t, we draw G_t in a separate block of the Gibbs sampler. Conditional on the full time series of G_t, we may draw values of δ_j together with β_j, which are jointly multivariate normal. Details are provided in the appendix.

We consider three different factor specifications for (8). The first assumes that all residuals load on one aggregate residual factor, so that δ is an $M \times 1$ vector. The second specification holds that there are seven orthogonal residual factors, one for each of the mutual fund objectives identified by Pastor and Stambaugh (2002b). The errors of each fund classified as having a particular objective load on that objective's residual factor only. The third specification combines the first two. Fund residuals load both on the residual factor related to their objective and to the aggregate residual factor. In this last case, δ is an $M \times 8$ matrix, but only two elements of each row are nonzero.

Posterior means and standard deviations for μ_α and σ_α are presented in Table 6. For brevity, we report only the results for the market model specification ($K = 1$)

Table 6
Effects of alternative assumptions about residual covariances

	Uncorrelated residuals	One aggregate residual factor	Objective-based residual factors	Aggregate and objective-based factors
μ_α	1.48	1.70	1.11	0.94
	(0.05)	(0.05)	(0.06)	(0.10)
σ_α	1.50	1.78	1.69	1.76
	(0.06)	(0.07)	(0.07)	(0.07)

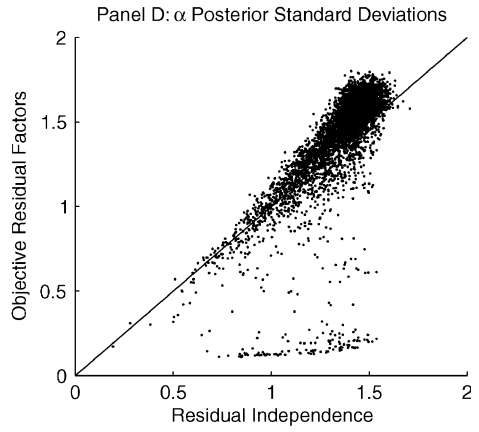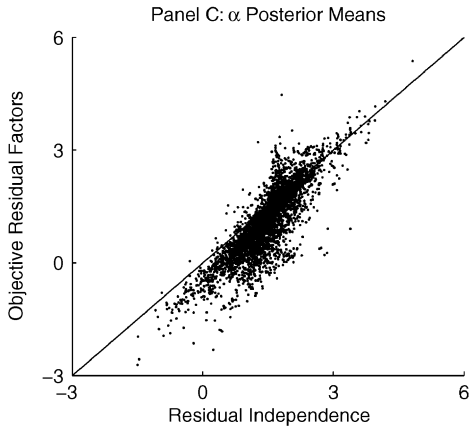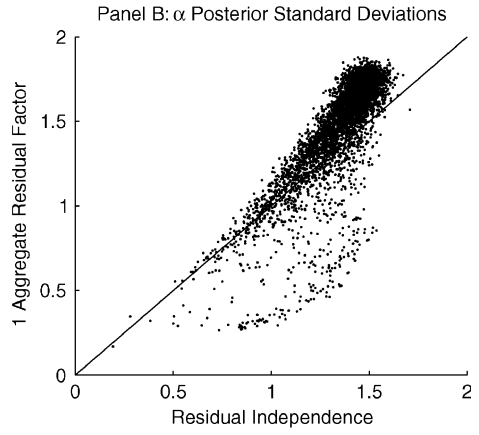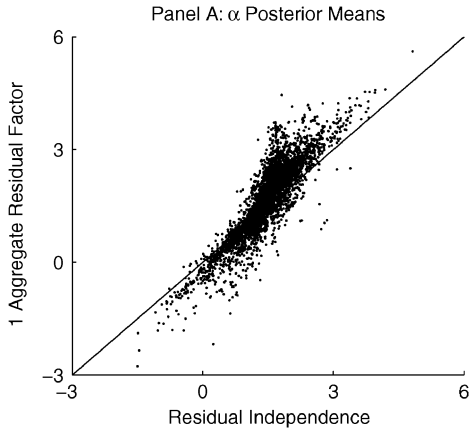μ_α and σ_α represent the cross-sectional mean and standard deviation, respectively, of the population of mutual fund alphas.

under unskeptical learning priors. Overall, richer covariance structures are consistent with lower posterior means and higher posterior standard deviations for μ_α. The latter would follow if the covariation captured is predominantly positive. Higher values are also found for σ_α. Both effects should lead to less shrinkage in the individual alphas themselves. To see if this is the case, Fig. 5 plots the posterior means and standard deviations for alphas estimated under the different residual models. A 45° line is also shown.

Panel A of the figure shows the relation between alpha posterior means computed under residual independence and those estimated with a single aggregate residual factor. The relation, while strong, is far from perfect. The two sets of alpha means have a correlation coefficient of 0.89, and there is somewhat more dispersion in the posterior means that result from the residual factor model. Greater dispersion in alphas is also apparent in Panel B, where posterior standard deviations computed using the residual factor are larger for most funds.

The remainder of Fig. 5 is qualitatively similar, with alphas generally more dispersed for the specifications involving a residual factor structure. One surprising result in panels D and F is the existence of a small number of funds whose alpha posterior standard deviations are fairly close to zero when computed with residual factors. It turns out that all 95 funds with alpha posterior standard deviations below 0.25% in Panel F are classified "Growth and Income" funds, even though those funds comprise only 23% of the entire sample. At least some of the 95 appear to be index funds, and their posterior mean alphas, which average 1.68% with

Fig. 5. Posterior moments of α under different residual covariance assumptions. This figure compares posterior mean alphas calculated under residual independence (horizontal axes) with those computed under a factor structure for the residual covariance matrix (vertical axes). In panels A and B, we assume there is only a single residual "factor." In panels C and D, each fund loads on a residual factor that is common to all funds with the same fund objective. Panels E and F combine these objective-based factors with an aggregate residual factor. Estimation is based on monthly fund returns over the period January 1961 to June 2001 using the market portfolio as the sole benchmark asset. Numbers are in annualized percentage terms and result from unskeptical learning priors.

Panel A: α Posterior Means

Panel B: α Posterior Standard Deviations

Panel C: α Posterior Means

Panel D: α Posterior Standard Deviations

Panel E: α Posterior Means

Panel F: α Posterior Standard Deviations
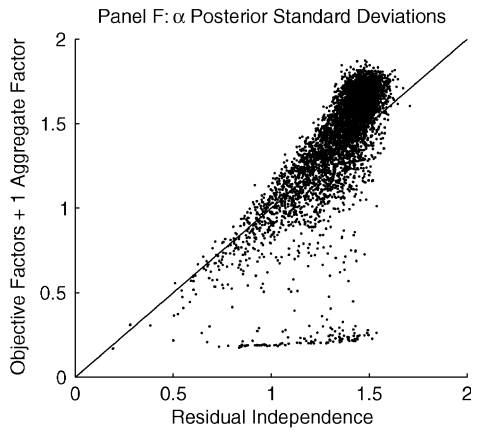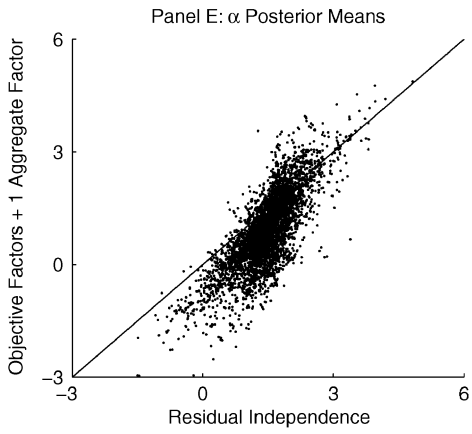
independent residuals, drop to just 0.23% on average when aggregate and objective residual factors are introduced.

6.2. Prior dependence between α_j and σ_j

In considering a large cross-section of stocks, MacKinlay (1995) argues that, other things equal, if α_j and σ_j are unrelated, then there will exist portfolios with more extreme Sharpe ratios. Both Pastor and Stambaugh (2002a,b), building on their earlier work, and Baks et al. (2001) incorporate this observation into their prior specifications for fund parameters, letting the prior variance of α_j, conditional on σ_j, be proportional to σ_j^2.

This idea can be incorporated in our framework as well. With the priors on μ_α and σ_α specified as before, we let the conditional prior for α_j be normal with mean μ_α, but with a variance of $(\sigma_j^2/\sigma_0^2)\,\sigma_\alpha^2$. In this specification, σ_α can be interpreted as the conditional prior standard deviation of the alpha for a fund with residual variance equal to σ_0^2. To maintain comparability with earlier results, we set σ_0 equal to the median OLS estimate of σ_j. Thus, σ_α^2 can now be interpreted as the population variance of mutual fund alphas for funds that have the "typical" amount of residual noise. Note that σ_0 decreases with K, as the addition of more factors reduces the magnitude of σ_j.

Alternatively, σ_α^2 can be thought of as the population variance of $\sigma_0(\alpha_j/\sigma_j)$. Thus, we can now think in terms of the "information" or "appraisal" ratio of Treynor and Black (1973), which considers the risk-adjusted reward to investing in asset j in relation to the residual risk borne. As they show, the increase in the Sharpe ratio attained by optimally tilting the benchmark portfolio(s) toward asset j is determined by both this ratio and the Sharpe ratio of the benchmark.

Posterior computations using this prior are slightly simpler than those for our previous priors. Conditional on μ_α and σ_α, distributions for the α_j, β_j, and σ_j follow Pastor and Stambaugh's (2002b) results exactly.[22] Posterior results for μ_α and σ_α incorporating the prior link between α_j and σ_j are presented in Table 7. In general, we see small changes in both the mean and standard deviations of μ_α, but imposing the link tends to increase posterior means of σ_α, sometimes by as much as 50%. While the interpretations of σ_α with and without the prior link are not identical, these results suggest an important difference between the two classes of priors. This is more than confirmed in Fig. 6, which demonstrates huge differences between individual alphas computed under the two priors when $K = 3$ and there is no prior skepticism. In this case, there is much greater variation in the alphas when the link is imposed, with alpha posterior means ranging anywhere from −20% to 30%.

Inspection of OLS results for individual funds suggests a culprit for this disparity. We find a small number of funds whose returns are almost perfectly explained by

---

[22]Gibbs draws of μ_α and σ_α, conditional on α_j and σ_j, are easily derived following the representation of α_j as $\mu_\alpha + (\sigma_j/\sigma_0)\sigma_\alpha\eta_j$, where $\eta_j \sim$ i.i.d. $N(0,1)$. Thus, conditional on the other parameters, μ_α and σ_α may be estimated using standard regression techniques since this representation implies $(\sigma_0/\sigma_j)\alpha_j = (\sigma_0/\sigma_j)\mu_\alpha + \sigma_\alpha\eta_j$, which is a homoskedastic linear regression.

Table 7
Posterior means and standard deviations of μ_α and σ_α with and without a link between α_j and σ_j

	Highly skeptical priors	Somewhat skeptical priors	Unskeptical priors
	K = 1 (*RMRF*)		
μ_α—without link	1.40 (0.04)	1.47 (0.05)	1.48 (0.05)
μ_α—with link ($\sigma_0 = 0.027$)	1.33 (0.04)	1.35 (0.04)	1.36 (0.04)
σ_α—without link	1.00 (0.07)	1.40 (0.06)	1.50 (0.06)
σ_α—with link ($\sigma_0 = 0.027$)	1.43 (0.10)	2.01 (0.09)	2.09 (0.09)
	K = 3 (*RMRF, SMB, and HML*)		
μ_α—without link	1.30 (0.05)	1.38 (0.05)	1.38 (0.05)
μ_α—with link ($\sigma_0 = 0.020$)	1.20 (0.04)	1.24 (0.04)	1.24 (0.04)
σ_α—without link	1.99 (0.07)	2.21 (0.07)	2.26 (0.07)
σ_α—with link ($\sigma_0 = 0.020$)	2.80 (0.07)	2.97 (0.07)	3.00 (0.07)
	K = 4 (*RMRF, SMB, HML, and MOM*)		
μ_α—without link	1.33 (0.04)	1.37 (0.05)	1.39 (0.05)
μ_α—with link ($\sigma_0 = 0.019$)	1.36 (0.04)	1.39 (0.04)	1.40 (0.04)
σ_α—without link	1.52 (0.06)	1.77 (0.06)	1.84 (0.06)
σ_α—with link ($\sigma_0 = 0.019$)	2.30 (0.07)	2.52 (0.07)	2.56 (0.07)
	K = 7 (*RMRF, SMB, HML, MOM, and industry factors*)		
μ_α—without link	1.71 (0.05)	1.80 (0.05)	1.81 (0.05)
μ_α—with link ($\sigma_0 = 0.017$)	1.65 (0.04)	1.70 (0.04)	1.70 (0.04)
σ_α—without link	2.07 (0.07)	2.27 (0.07)	2.32 (0.07)
σ_α—with link ($\sigma_0 = 0.017$)	2.67 (0.07)	2.84 (0.07)	2.87 (0.07)

Results in this table labeled "without link" are identical to those reported in Table 2 for learning priors using all funds. Results labeled "with link" also use learning priors and all available funds but assume that the conditional prior for α_j is normal with mean μ_α and variance $(\sigma_j^2/\sigma_0^2)\sigma_\alpha^2$, where σ_j^2 is the residual variance for fund j and σ_0 is the cross-sectional median OLS estimate of σ_j. Estimation is based on monthly fund returns over the period January 1961 to June 2001. All numbers are in annualized percentage terms. The highly and somewhat skeptical normal priors for μ_α have mean zero with standard deviations 0.25 and 1.0, respectively; the corresponding inverted gamma priors for σ_α are centered around 0.75 and 3.0, with degrees of freedom 100 and 10. The unskeptical priors are diffuse. The factors are the excess market return RMRF, small-big market cap return SMB, high-low book-to-market equity return HML, and past one-year momentum MOM. The no-learning priors are independent across funds with the same marginal distribution as in the learning case. The industry factors are based on principal components of the four-factor model residuals. Factors and priors for μ_α and σ_α are identical for specifications with and without the "link".
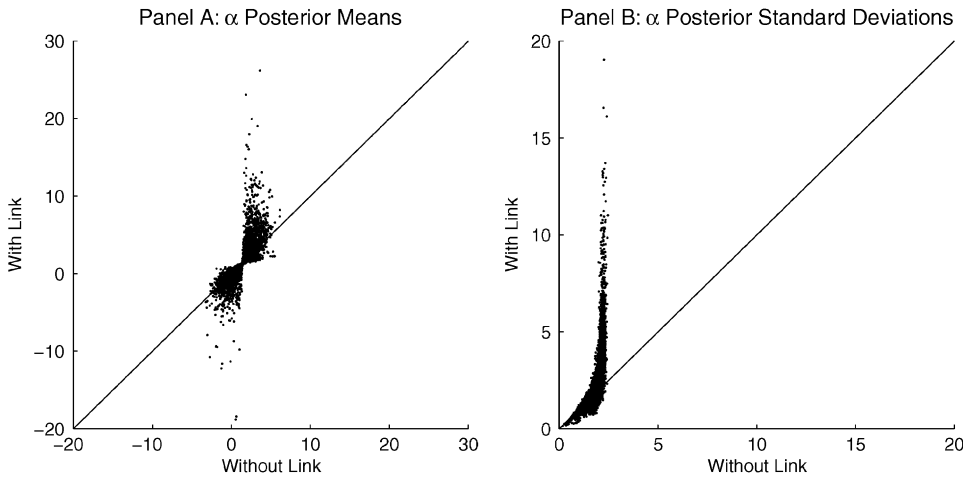
Fig. 6. Posterior moments of α with and without the link between α and σ. This figure compares posterior means and standard deviations of alphas with (vertical axes) and without (horizontal axes) a prior link between each fund's alpha and residual standard deviation. Without the link, alphas are assumed distributed as $N(\mu_\alpha, \sigma_\alpha^2)$, as in most of the paper. With the link, α is normal with mean μ_α and standard deviation $(\sigma_j/\sigma_0)\sigma_\alpha$, where σ_0 is equal to the median OLS estimate of σ_j. Estimation is based on monthly fund returns over the period January 1961 to June 2001. Numbers are in annualized percentage terms and result from applying unskeptical learning priors to the three-factor model ($K = 3$) of returns.

three or more factors, but which have OLS alphas that are small (approximately two basis points per month) yet statistically significant. Given our earlier discussion, the large information ratios of these funds can be justified only by a large value for σ_α. It follows that the (conditional) prior for a fund with high σ_j is highly dispersed, resulting in little shrinkage towards the sample mean for many funds.

Is it more desirable to impose the prior link between α_j and σ_j or not? In the mutual fund context, substantive issues arise that go beyond the asymptotic arbitrage arguments made previously when considering the pricing of individual stocks. BMW note that, other things equal, increasing the amount of cash held by a fund lowers both its alpha and residual risk. However, differences in residual risk across funds and the empirical association between alpha and residual risk will be driven by many other factors related to variation in funds' investment strategies. One important consideration is the nature of the insight that skilled fund managers possess. Are astute managers able to detect mispricing across a wide range of stocks at a given point in time and, therefore, able to deliver large alphas with relatively low residual variance? Or, do managers with the highest alphas focus on particular industries or market sectors, thereby forgoing some diversification? A prior belief in the latter hypothesis is consistent with a positive link between α_j and σ_j, while the former suggests a negative relation. In light of the uncertainties, assuming independence in the joint prior for α_j and σ_j seems to us to be as reasonable as the alternatives, but the ultimate choice lies with the reader.

6.3. Correlation among the alphas

The identification of mispriced stocks can be viewed as a zero-sum game in which the buyer achieves an abnormal return at the expense of the seller, or vice versa. Thus, if all equities were held through mutual funds, one might want to treat μ_α as known and equal to zero. However, this scenario implies not only that the expected alpha is zero, but also the further restriction that the *realized* or actual (size-weighted) sum of all alphas equals zero. This is inconsistent with our modeling of the alphas as independent draws from an underlying population, but could be accommodated by allowing for negative correlation among those draws.

According to the Investment Company Institute, stock mutual funds held just 21% of all publicly traded U.S. stocks at the end of 2003, even after the huge growth of funds in the 1990s. Thus, incorporating negative correlation is not likely to be important in this context. An initial exploration of this issue via simulations reinforces that impression. We abstract from some of the complexities of the actual data and simplify the computations by simulating funds of equal size, history length, and (known) residual variance. In this symmetric context, it is easy to show that a pairwise correlation of $-1/(M-1)$ between fund alphas would ensure that the sum of alphas has zero variance and hence equals zero when μ_α is zero.

A given fund may achieve abnormal returns at the expense of non-fund investors as well as the other $M-1$ funds, however. The question then is how to model the prior for these non-fund alphas. For simplicity, we treat the non-fund investment as if it were generated by M other investment vehicles that are not observed but are treated symmetrically with our M funds.[23] Now, since μ_α is the population mean of the fund alphas, the mean non-fund alpha must be $-\mu_\alpha$. If we further assume that the correlation among all investment vehicles, funds and non-funds, equals $-1/(2M-1)$, then the sum of *all* the alphas must be zero.

We find that imposing the negative correlation in this manner has a trivial effect on the posterior means of the hyperparameters and the individual fund alphas in simulations with $N = 10$, 100, 1,000 or 10,000 funds. We do see a noticeable reduction in the posterior variance for μ_α, however. This makes sense in that the variance of the average alpha is reduced by imposing the negative correlation. This has little effect on the posterior means of the alphas, however, since shrinkage of the regression estimates toward μ_α depends mainly on the level of σ_α. It would be interesting to study the impact of imposing the exact "adding-up constraint" when we consider the role of prior dependence in asset pricing tests, where the weighted-average alpha across all assets is zero when a market factor is included in the model.

7. Summary and conclusions

This paper is based on a simple intuitive premise: If the true measures of performance (alphas) for a large set of mutual funds were somehow revealed to an

---

[23]Using M, as opposed to something closer to $4M$ non-funds, as the 21% figure mentioned above might suggest, is conservative in that the role of non-fund investment is given greater weight.

investor, it would affect her belief about the likely degree of abnormal performance for some other fund not in that set. Mathematically, this is a statement that *prior* beliefs for different funds are dependent. They are dependent insofar as an investor's expectation about the performance of a fund depends on his belief about mutual fund managers as a group and, more generally, the degree to which financial markets are efficient.

We accommodate this perspective by assuming that the true alphas are random draws from a distribution with hyperparameters μ_α and σ_α, the average level of skill and the standard deviation of skill, respectively. Investors learn about these parameters by pooling data for all funds and this feeds back into the estimation of individual fund alphas, a phenomenon we refer to as "learning across funds." Highly efficient numerical techniques are developed for evaluating posterior moments in this context. Simulations are then used to explore the beliefs an investor might arrive at under different assumptions about actual management skill, an investor's initial level of skepticism about abnormal performance, and the number of funds observed.

Of central interest are the differences in estimates that arise as a result of incorporating "learning across funds." Two sorts of shrinkage factors emerge as relevant for understanding the differences observed. First, whereas estimates of a given fund's alpha are based solely on that fund's returns in the traditional approach, learning gives rise to a *data-based* shrinkage factor, with each fund's estimate tilted toward the pooled estimate of μ_α, to a degree also determined by the data. Second, when prior information is incorporated, there is also shrinkage toward the prior mean, which we take to be zero. This attenuation is stronger without learning since the data are perceived as less informative about a given fund's alpha in this case.

With learning across funds, estimates of the hyperparameters gradually converge to the true values as M, the number of funds, increases. The convergence tends to be slower for σ_α than for μ_α in the simulations we perform. Ideally, deviations between the estimates and the true fund alphas would be tightly centered about zero. With learning, data-based shrinkage does result in an average error that approaches zero, with very good results when M is 1,000 or higher. In contrast, the "bias" induced by shrinkage of each fund's alpha toward the prior mean is fixed under a no-learning prior, as there is no data-based effect to offset it. Hence, the average error does not decline with M and is zero only if the prior mean happens to coincide with the actual value of μ_α. This is a fundamental difference between the two approaches.

Our empirical application with actual monthly fund returns is based on a set of over five thousand funds with an average history of about 77 months of data. The estimates of μ_α and the average fund alpha are usually around 1.3% to 1.4% per annum (before expenses) but, surprisingly, are about 40 basis points higher when industry factors are included. This suggests that managers do indeed possess some skill in selecting stocks, though not enough to offset the typical expenses of about 2%. Estimates of σ_α mostly range between 1.5% and 2.3%, depending on the prior and the benchmark model.

The empirical implications of incorporating "learning across funds" are conveyed most dramatically by focusing on the fund with the largest posterior mean alpha.

Using returns net of expenses, the maximum is typically between 2% and 7.3% with learning. Under no-learning priors, the maximum can be as high as 44% with some skepticism about the magnitude of alpha or 92% if prior information is ignored (diffuse priors). These results reflect the data-based shrinkage that occurs with learning across funds and the absence of this effect when one imposes prior independence.

While we document a significant impact on the cross-sectional distribution of estimated alphas, the implications of learning across funds for asset allocation remain largely unknown. Although two examples demonstrate the substantial effects of learning on the allocation to a particular fund, the implications for asset allocation *across* funds remain unexplored. The additional layer of cross-sectional dependence in the perceived or "predictive" distribution of returns that arises with learning makes this an interesting and challenging issue for future work. It would also be desirable to integrate uncertainty about pricing model misspecification in our framework, building on Pastor and Stambaugh (2002a,b).

In addition, our simple model with μ_α and σ_α might be extended to reflect conditional dependence through additional hyperparameters related to fund characteristics. For example, Baks (2003) considers the common effect of a given manager or fund organization on fund alphas. Alternatively, dependence could be related to fund holdings data in a Bayesian version of the recent Cohen et al. (2005) approach. For a given fund, this would provide a basis for optimally combining its own alpha estimate with that of funds overall and, additionally, with the alphas of funds following similar investment strategies.

Prior dependence will likely play a significant role in other cross-sectional contexts as well, such as the testing and evaluation of asset pricing models. For example, one might doubt the adequacy of the CAPM, a priori, because the theory fails to incorporate hedging demands, taxes, or behavioral biases, to name just a few of the many possibilities. Analogous to our argument for mutual fund alphas, knowing the true deviations from the CAPM for a large set of stocks would affect our belief about the adequacy of the model in general. This would inform our *prior* belief about the deviations for other stocks, though perhaps through a more complicated specification that ultimately incorporates the covariances between securities and other stock characteristics. While these natural extensions of our basic framework are beyond the scope of this paper, we look forward to exploring them in future work.

Appendix A. the MCMC sampling procedure

The data generating process (for excess fund returns) is assumed to be the linear factor model

$$r_{j,t} = \alpha_j + \beta_j' F_t + \varepsilon_{j,t}, \tag{A.1}$$

where $\varepsilon_{j,t} \sim N(0, \sigma_j^2)$ are uncorrelated across funds and through time.

Priors are represented as

$$\mu_\alpha \sim N(m_\alpha, V_\alpha),\tag{A.2}$$

$$\sigma_\alpha \sim IG(s_\alpha, N_\alpha),\tag{A.3}$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2),\tag{A.4}$$

and

$$p(\beta_j, \sigma_j) \propto 1/\sigma_j.\tag{A.5}$$

There are several standard parameterizations of the inverted gamma distribution. The parameterization we use is easily interpreted as the posterior distribution that would result (under the "diffuse" prior $p(\sigma_\alpha) \propto 1/\sigma_\alpha$) following the observation of the alphas of N_α funds, given a known value of μ_α. The parameter s_α is simply the standard deviation of this sample computed using the known mean μ_α and using N_α rather than $N_\alpha - 1$ to divide the sum of squared errors (so that the prior remains well defined even when $N_\alpha = 1$). The density of the inverted gamma distribution in (A.3) therefore satisfies

$$p(\sigma_\alpha \,|\, s_\alpha, N_\alpha) \propto \frac{1}{\sigma_\alpha^{N_\alpha}} \exp\left(-\frac{N_\alpha s_\alpha^2}{2\sigma_\alpha^2}\right),\tag{A.6}$$

with a mean approximately equal to s_α when N_α is large.

As outlined in Section 2.2, the Gibbs sampling approach we use consists of four blocks. We consider each in turn.

A.1. Drawing σ_α conditional on α and μ_α

For shorthand, let α denote the vector of all fund alphas. Given μ_α and α, the conditional posterior of σ_α can be written using Bayes' Rule as

$$p(\sigma_\alpha \,|\, \mu_\alpha, \alpha, s_\alpha, N_\alpha) \propto p(\alpha \,|\, \mu_\alpha, \sigma_\alpha)p(\sigma_\alpha \,|\, s_\alpha, N_\alpha),\tag{A.7}$$

which is proportional to

$$\frac{1}{\sigma_\alpha^M} \exp\left(-\frac{M\hat{\sigma}_\alpha^2}{2\sigma_\alpha^2}\right) \frac{1}{\sigma_\alpha^{N_\alpha}} \exp\left(-\frac{N_\alpha s_\alpha^2}{2\sigma_\alpha^2}\right) = \frac{1}{\sigma_\alpha^{M+N_\alpha}} \exp\left(-\frac{M\hat{\sigma}_\alpha^2 + N_\alpha s_\alpha^2}{2\sigma_\alpha^2}\right),\tag{A.8}$$

where $\hat{\sigma}_\alpha^2 = (1/M)(\alpha - \mu_\alpha \iota)'(\alpha - \mu_\alpha \iota)$ and ι is an $M \times 1$ vector of ones. The conditional distribution of σ_α is therefore

$$IG\left(\sqrt{(M\hat{\sigma}_\alpha^2 + N_\alpha s_\alpha^2)/(M + N_\alpha)}, M + N_\alpha\right).\tag{A.9}$$

A.2. Drawing μ_α conditional on α and σ_α

In this case we have

$$p(\mu_\alpha \,|\, \sigma_\alpha, \alpha, m_\alpha, V_\alpha) \propto p(\alpha \,|\, \mu_\alpha, \sigma_\alpha) p(\mu_\alpha \,|\, m_\alpha, V_\alpha), \tag{A.10}$$

which is proportional to

$$\exp\left(-\frac{(\mu_\alpha - \hat{\mu}_\alpha)^2}{2\sigma_\alpha^2 / M}\right) \exp\left(-\frac{(\mu_\alpha - m_\alpha)^2}{2V_\alpha}\right) = \exp\left(\frac{-(\mu_\alpha - \bar{\mu}_\alpha)^2}{2(M/\sigma_\alpha^2 + 1/V_\alpha)^{-1}}\right), \tag{A.11}$$

where $\hat{\mu}_\alpha = (1/M)\Sigma_j \alpha_j$ and

$$\bar{\mu}_\alpha = (\hat{\mu}_\alpha M/\sigma_\alpha^2 + m_\alpha/V_\alpha)(M/\sigma_\alpha^2 + 1/V_\alpha)^{-1}. \tag{A.12}$$

Thus, the conditional distribution of μ_α is

$$N(\bar{\mu}_\alpha, (M/\sigma_\alpha^2 + 1/V_\alpha)^{-1}). \tag{A.13}$$

A.3. Drawing σ_j and β_j conditional on F, r_j, and α_j for all $j = 1, \ldots, M$

Given α_j, we may rearrange the return equation as

$$r_{j,t} - \alpha_j = \beta_j' F_t + \varepsilon_{j,t}. \tag{A.14}$$

Assume there are T_j return observations for fund j. Let $\mathbf{r}_j$ denote the vector of those returns, $\boldsymbol{\varepsilon}_j$ the vector of fund-j residuals, and $\mathbf{F}_j$ the matrix of contemporaneous factor realizations.[24]

Since both residuals and priors for σ_j and β_j are independent across funds, each fund may be treated separately. Given diffuse priors on β_j and σ_j, conditional posteriors of these parameters for each fund follow easily from standard results for the linear regression model under normality. Specifically,

$$\sigma_j \sim IG(\hat{\sigma}_j, T_j) \tag{A.15}$$

and

$$\beta_j \,|\, \sigma_j \sim N(\hat{\beta}_j^{\text{OLS}}, \sigma_j^2 (\mathbf{F}_j' \mathbf{F}_j)^{-1}), \tag{A.16}$$

where $\hat{\sigma}_j^2 = (1/T)\hat{\boldsymbol{\varepsilon}}_j' \hat{\boldsymbol{\varepsilon}}_j$, $\hat{\boldsymbol{\varepsilon}}_j = \mathbf{r}_j - \alpha_j - \hat{\beta}_j^{\text{OLS}\prime} \mathbf{F}_j$, and $\hat{\beta}_j^{\text{OLS}}$ is obtained from a regression with no intercept.

A.4. Drawing α_j conditional on μ_α, σ_α, F, r_j, β_j and σ_j for all $j = 1, \ldots, M$

Given β_j, we may rearrange the return equation as

$$r_{j,t} - \beta_j' F_t = \alpha_j + \varepsilon_{j,t}. \tag{A.17}$$

---

[24]$\mathbf{F}_j$ requires a fund-specific subscript because different funds are observed over different time periods.

Conditional on μ_α and σ_α, we have complete independence across funds, so we may again consider each fund separately. Applying Bayes' Rule and eliminating irrelevant conditioning arguments, we have

$$p(\alpha_j \mid \mu_\alpha, \sigma_\alpha, \mathbf{F}_j, \mathbf{r}_j, \beta_j, \sigma_j) \propto p(\mathbf{r}_j \mid \mathbf{F}_j, \alpha_j, \beta_j, \sigma_j) p(\alpha_j \mid \mu_\alpha, \sigma_\alpha), \tag{A.18}$$

which is itself proportional to

$$\exp\left(-\frac{(\alpha_j - \hat{\alpha}_j)^2}{2\sigma_j^2/T_j}\right) \exp\left(-\frac{(\alpha_j - \mu_\alpha)^2}{2\sigma_\alpha^2}\right) = \exp\left(-\frac{(\alpha_j - \tilde{\alpha}_j)^2}{2(T_j/\sigma_j^2 + 1/\sigma_\alpha^2)^{-1}}\right), \tag{A.19}$$

where $\hat{\alpha}_j = (1/T)\Sigma_t(r_{j,t} - \beta_j F_t)$ and

$$\tilde{\alpha}_j = (\hat{\alpha}_j T_j/\sigma_j^2 + \mu_\alpha/\sigma_\alpha^2)(T_j/\sigma_j^2 + 1/\sigma_\alpha^2)^{-1}. \tag{A.20}$$

This implies that the conditional distribution of α_j is

$$N(\tilde{\alpha}_j, (T_j/\sigma_j^2 + 1/\sigma_\alpha^2)^{-1}). \tag{A.21}$$

A.5. Convergence of the Gibbs chain

Since the support of each draw is unbounded over the region on which the parameters are defined, the Gibbs chain is irreducible and therefore convergent (see Tierney, 1994) given a long enough sequence of draws.

The Gibbs chains run in this paper all consist of 11,000 "cycles" through the above four draws. The first 1,000 are discarded to negate the effects of initial conditions. Visual analysis of the autocorrelations of the remaining draws suggests that convergence is fairly rapid.

A.6. Modifications to the Gibbs sampler when residuals have a factor structure

Little of the above changes when a factor structure in the residuals is introduced. Writing the data-augmented model as

$$r_{j,t} = \alpha_j + \beta_j' F_t + \delta_j' G_t + \xi_{j,t}, \tag{A.22}$$

then conditional on G_t (the "residual factors") we may use the previous results simply by replacing β_j with $\beta_j^* = [\beta_j', \delta_j']'$, replacing F_t with $F_t^* = [F_t', G_t']'$, and reinterpreting σ_j as the variance of $\xi_{j,t}$ instead of $\varepsilon_{j,t}$.

One additional block in the sampler is required to draw each $L \times 1$ latent variable G_t. If M_t fund returns are observed at time t, then let r_t and α_t denote the $M_t \times 1$ vectors formed by stacking the excess returns and alphas for these funds. Let β_t denote the $M_t \times K$ matrix formed by stacking the β_j' and let δ_t denote the $M_t \times L$ matrix formed by stacking the δ_j'. Then following Geweke and Zhou (1996), we find that G_t is conditionally normally distributed with mean

$$\delta_t'(\delta_t\delta_t' + \Omega_t)^{-1}(r_t - \alpha_t - \beta_t F_t) \tag{A.23}$$

and covariance matrix

$$I - \delta_t'(\delta_t\delta_t' + \Omega_t)^{-1}\delta_t, \tag{A.24}$$

where Ω_t is the diagonal covariance matrix of the $\xi_{j,t}$ for the funds observed at time t.

A.7. A simplified approximate sampling procedure

As described previously, each iteration of the Gibbs sampler involves four steps (assuming residual independence):

1. σ_α conditional on α_j $(j = 1, \dots, M)$ and μ_α.
2. μ_α conditional on α_j $(j = 1, \dots, M)$ and σ_α.
3. σ_j and β_j conditional on F, r_j, and α_j for all $j = 1, \dots, M$.
4. α_j conditional on μ_α, σ_α, F, r_j, β_j, and σ_j for all $j = 1, \dots, M$.

Running the sampler for 11,000 iterations is sufficient for obtaining very accurate estimates of posterior moments, and this process may take as long as two hours on a fast personal computer. We find, however, that the algorithm may be easily modified to produce results that are approximately correct in about 30 s. These modifications involve steps 1 to 3; step 4 remains unaffected. They are sensible only under unskeptical (diffuse) priors.

Previously, step 1 involved drawing σ_α conditional on the α_j, where the distribution of σ_α was an inverted gamma centered around the sample standard deviation of the alphas. With a large number of funds, however, the variation in this draw is very low, so we instead simply set σ_α equal to the sample standard deviation of the current values of the alphas. The modification to step 2 is similar: Rather than drawing μ_α from its appropriate normal distribution, we simply set it to the sample average of the current alpha values.

Previously, step 3 involved drawing σ_j from an inverted gamma distribution and β_j from a normal. Instead, we simply set σ_j equal to its OLS estimate, obtained by regressing $r_{j,t} - \alpha_j$ on the factors F_t (with no intercept). Given this value for σ_j, the ensuing draw of β_j is the same as it is for the full algorithm.

This simplified algorithm is faster primarily because it avoids generating random variables with inverted gamma distributions, which is a relatively slow computation. For the same reason, it is conceptually somewhat simpler, especially for users unfamiliar with the inverted gamma distribution. The approximate posterior means it generates are also very accurate. For a three-factor model, the approximate means (standard deviations) of μ_α and σ_α are 1.37 (0.03) and 2.17 (0.05), respectively, which are close to the true posterior means (standard deviations) under unskeptical learning priors, which are 1.38 (0.05) and 2.26 (0.07). True and approximate posterior means for the individual alphas are compared in Panel A of Fig. 7. In general, differences between the two are small and unrelated to the magnitudes of the alphas. Panel B shows that the approximate and true posterior standard deviations are similar as well.
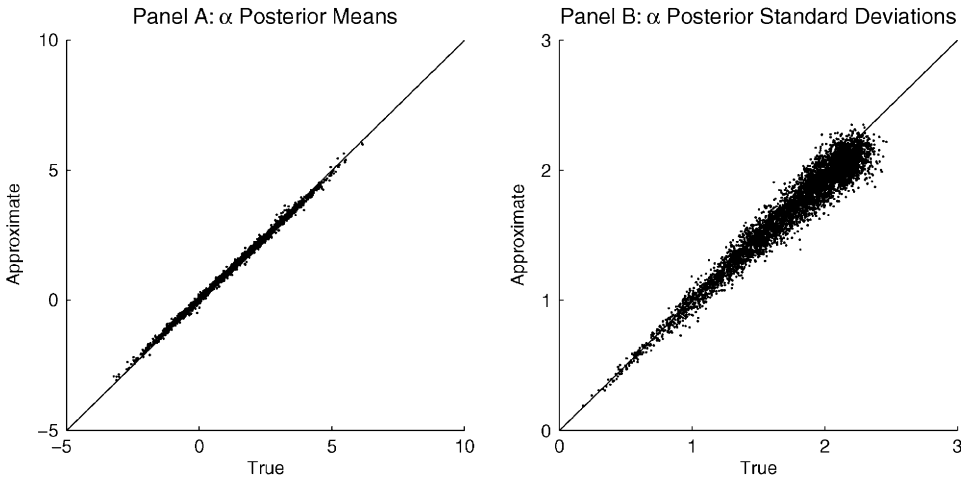
Panel A: α Posterior Means   Panel B: α Posterior Standard Deviations

Fig. 7. True and approximate posterior moments of α for the three-factor model under unskeptical learning priors. Approximate posterior means and standard deviations of the alphas are compared with true moments computed under the three-factor model ($K = 3$) with unskeptical learning priors. The approximations, described in Appendix A.7, ignore sampling error in all quantities except the fund alphas and betas. Estimation is based on monthly fund returns over the period January 1961 to June 2001. Numbers are in annualized percentage terms.

A.8. Incorporating negative correlation among the alphas

Imposing a prior correlation ρ among the alphas requires the modification of three of the blocks of the Gibbs sampler, namely $p(\mu_\alpha \mid \alpha, \beta, \sigma, \sigma_\alpha, r, F)$, $p(\sigma_\alpha \mid \alpha, \beta, \sigma, \mu_\alpha, r, F)$, and $p(\alpha \mid \beta, \sigma, \mu_\alpha, \sigma_\alpha, r, F)$.

Drawing from the first two blocks, $p(\mu_\alpha \mid \alpha, \beta, \sigma, \sigma_\alpha, r, F)$ and $p(\sigma_\alpha \mid \alpha, \beta, \sigma, \mu_\alpha, r, F)$, requires a relatively straightforward adjustment for correlation. To do so, we write the conditional prior for α as $\alpha \sim N(\mu_\alpha \iota, \sigma_\alpha^2 C)$, where C is the prior correlation matrix, a matrix with ones on the diagonal and all off-diagonal elements equal to ρ. If $C^{-1/2}$ is the inverse of the Cholesky decomposition of C, then

$$C^{-1/2}\alpha \sim N(\mu_\alpha C^{-1/2}\iota, \sigma_\alpha^2 I). \tag{A.25}$$

We use the fact that elements of the vector $C^{-1/2}\alpha - \mu_\alpha C^{-1/2}\iota$ are i.i.d. $N(0, \sigma_\alpha^2)$ to draw σ_α conditional on α and μ_α. This step is identical to that of Section A.1 except that $\alpha - \mu_\alpha \iota$ is replaced by $C^{-1/2}\alpha - \mu_\alpha C^{-1/2}\iota$. To draw μ_α, we simply regress the elements of $C^{-1/2}\alpha$ on the corresponding elements of $C^{-1/2}\iota$, where the residual standard deviation is equal to σ_α and there is no intercept. Priors on μ_α are imposed as priors on the regression slope coefficient.

For the last block, Bayes' Rule implies that

$$p(\alpha \mid \beta, \sigma, \mu_\alpha, \sigma_\alpha, r, F) \propto p(\alpha \mid \beta, \sigma, \mu_\alpha, \sigma_\alpha)p(r \mid \beta, \sigma, \mu_\alpha, \sigma_\alpha, \alpha, F). \tag{A.26}$$

Given residual independence, the joint likelihood function $p(r \mid \beta, \sigma, \mu_\alpha, \sigma_\alpha, \alpha, F)$ can be decomposed as the product of individual fund likelihoods, each of which is

proportional to a normal density for α_j with mean $\hat{\alpha}_j$, as defined in section A.4, and variance σ_j^2/T_j.

Under the simplifying assumption that σ_j and T_j are identical for all funds, the joint likelihood is therefore proportional to a multivariate normal density for the vector α with mean $\hat{\alpha}$ and covariance matrix $(\sigma^2/T)I$. The conditional prior $p(\alpha\,|\,\beta, \sigma, \mu_\alpha, \sigma_\alpha)$ is by assumption $N(\mu_\alpha \iota, \sigma_\alpha^2 C)$, and the product of these two densities is therefore also multivariate normal with covariance matrix

$$[(T/\sigma^2)I + \sigma_\alpha^{-2}C^{-1}]^{-1} \tag{A.27}$$

and mean vector

$$[(T/\sigma^2)I + \sigma_\alpha^{-2}C^{-1}]^{-1}[(T/\sigma^2)\hat{\alpha} + \sigma_\alpha^{-2}C^{-1}\mu_\alpha \iota]. \tag{A.28}$$

This is therefore the conditional distribution of the vector of alphas.

References

Baks, K.P., 2003. On the performance of mutual fund managers. Unpublished working paper, Emory University.

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

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

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

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

Chevalier, J., Ellison, G., 1999. Are some mutual fund managers better than others? Cross sectional patterns in behavior and performance. Journal of Finance 54, 875–899.

Cohen, R., Coval, J., Pastor, L., 2005. Judging fund managers by the company they keep. Journal of Finance, forthcoming.

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

Geweke, J., Zhou, G., 1996. Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies 9, 557–587.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 93–130.

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

Jones, C.S., 2003. Nonlinear mean reversion in the short-term interest rate. Review of Financial Studies 16, 793–843.

Jorion, P., 1986. Bayes-Stein estimation for portfolio analysis. Journal of Financial and Quantitative Analysis 21, 279–292.

Kandel, S., Stambaugh, R.F., 1996. On the predictability of stock returns: an asset allocation perspective. Journal of Finance 51, 385–424.

MacKinlay, A.C., 1995. Multifactor models do not explain deviations from the CAPM. Journal of Financial Economics 38, 3–28.

Merton, R.C., 1980. On estimating the expected return on the market. Journal of Financial Economics 8, 323–361.

Pastor, L., Stambaugh, R.F., 1999. Costs of equity capital and model mispricing. Journal of Finance 54, 67–121.

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

Pastor, L., Stambaugh, R.F., 2002b. Mutual fund performance and seemingly unrelated assets. Journal of Financial Economics 63, 313–349.

Savage, L.J., 1962. The Foundations of Statistical Inference: A Discussion. Methuen, London.

Shanken, J., 1987. A Bayesian approach to testing portfolio efficiency. Journal of Financial Economics 19, 195–215.

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

Stambaugh, R.F., 2003. Inference about survivors. Unpublished working paper, University of Pennsylvania.

Swamy, P.A.V.B., 1971. Statistical Inference in Random Coefficient Regression Models. Springer, New York.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762.

Treynor, J.L., Black, F., 1973. How to use security analysis to improve portfolio selection. Journal of Business 46, 66–88.

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