MUTUAL FUND PERFORMANCE WITH
LEARNING ACROSS FUNDS

Christopher S. Jones
Jay Shanken

**ABSTRACT**

This paper is based on the premise that knowledge about the alphas of one set of funds will influence an investor's beliefs about other funds. This will be true insofar as an investor's expectation about the performance of a fund is partly a belief about the abilities of mutual fund managers as a group and, more generally, a belief about the degree to which financial markets are efficient. We develop a simple framework for incorporating this "prior dependence" and find that it can have a substantial impact on the cross-section of posterior beliefs about fund performance as well as asset allocation. Under independence, the maximum posterior mean alpha increases without bound as the number of funds increases and "extremely large" estimates are randomly observed. This is true even when fund managers have no skill. In contrast, with prior dependence, investors aggregate information across funds to form a general belief about the potential for abnormal performance. Each fund's alpha estimate is shrunk toward the aggregate estimate, mitigating extreme views. An additional implication is that restricting the estimation to surviving funds, a common practice in this literature, imparts an upward bias to the average fund alpha.

Christopher S. Jones

Marshall School of Business

University of Southern California

Hoffman Hall 701

Los Angeles, CA 9089

christopher.jones@marshall.usc.edu

Jay Shanken

Goizueta Business School

Emory University

1300 Clifton Road

Atlanta, GA 30322

and NBER

jay_shanken@bus.emory.edu

**Mutual Fund Performance with Learning Across Funds**

## 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 like past performance and manager characteristics can be used to identify investment skill.[1] It is well known among academics that the standard measure of fund performance, "alpha," is typically not estimated with much precision. In this context, an investor's *prior* beliefs about market efficiency and the possibility of superior investment performance can play an important role in the asset allocation decision.

Recent papers by Baks, Metrick, and Wachter (2001) and Pastor and Stambaugh (2002) analyze these issues in a Bayesian statistical framework.[2] The BMW paper focuses on the question of whether *any* investment in actively managed funds can be justified. They conclude that, even when investors are initially quite skeptical about the possibility of management skill, some active investment is appropriate. The PS paper evaluates the optimal portfolio of funds for an investor with given beliefs about investment skill as well as the ability of asset pricing models to correctly "price" passive investment assets.

In this project, we build on these important studies by developing a richer representation of investor prior beliefs about management skill. While this might initially sound like a minor mechanical extension, we believe it is more fundamental than that. Although prior beliefs are by their nature subjective, it is important to ask whether the properties of a given prior or family of priors are indeed consistent in essential respects

---

[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) estimation of 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. See Shanken and Tamayo (2001) for additional references.

with the actual learning process of investors or researchers. At issue here is the manner in which fund returns data are processed in deriving conclusions about performance. Should the alpha estimates of other funds influence our belief about the likely performance of a given fund or not? Technical issues aside, that is the basic question addressed here.[3]

As with any joint distribution, one can talk about the prior for one set of (true) fund alphas conditional on the alphas for other funds. Though BMW and PS model priors in different ways, a common feature is that the prior belief about a given fund's alpha is taken to be independent of the prior for all other fund alphas. In conjunction with the assumption that fund residual returns are independent, this greatly simplifies the analyses. While such simplification is natural in papers breaking new ground, we would argue that it is important to consider dependence across funds in these priors.

To develop some intuition about this issue, consider the following hypothetical scenario. Assume that a *marginal* prior distribution has been formulated for a typical fund. In the BMW model, this would include a prior probability that a random manager is skilled, as well as a standard deviation that relates to the possible range of alphas for a skilled manager. Now, suppose an investor is considering a particular fund and is given the (prior) information that the *true* alphas of all other funds in the market are identically zero, i.e., none of those fund managers is able to beat the market. Alternatively, the information might be that half of all funds have positive alphas in excess of 5% per annum. In this context, the prior independence assumption across funds implies that none of this information would affect the investor's belief about the given fund's alpha. More formally, the conditional prior for the fund, given other fund alphas, equals the marginal prior.

It seems more plausible, to us, that a typical investor's beliefs would be affected by such information about other funds. The underlying premise is that an investor, or a researcher, comes to the situation with a basic uncertainty about the possibility and extent of management skill *in general*. Such a belief would be tied to a perception about the degree to which financial markets are efficient. Information about the universe of funds

---

[3] Readers who are interested in fund performance, but are uncomfortable with thinking in terms of prior distributions, may want to simply focus on our results for "diffuse" priors, which let the data completely dominate the analysis.

will affect this general belief which, in turn, will inform one's a priori perception of any given fund.

To formalize this idea, it is convenient to think of the true fund alphas as random draws from a common distribution with unknown mean $\mu_\alpha$ and variance $\sigma_\alpha^2$ – the hyperparameters of our model. All of our key results follow from this simple assumption. A large positive value of $\mu_\alpha$, for example, indicates the presence of considerable skill in the general population of managers, while the variance reflects heterogeneity across managers. In particular, a strictly positive variance means that there is additional uncertainty about the alpha for a given manager, no matter how many other alphas have been "observed" in the sense of the thought experiment above. Knowledge of the other alphas is relevant, however, insofar as it affects our belief about $\mu_\alpha$, the true mean of the population from which the given fund's alpha is drawn. Thus, there is an important role for prior dependence in this framework.

To get some idea of the manner in which prior dependence will affect the interpretation of empirical evidence, it is helpful to consider an extreme case. Suppose that $\sigma_\alpha$ is zero, so that all managers have the same alpha, $\mu_\alpha$. This parameter would be estimated, roughly speaking, by *pooling* the regression evidence for all funds[4]. The posterior belief about any fund's alpha would then be based on this pooled estimate and the initial prior belief. At the other end of the spectrum, with prior independence, as in BMW and PS, the posterior belief about alpha depends only on the given fund's data and the returns on the benchmark assets in the factor model.[5] Our general model with positive $\sigma_\alpha$ incorporates pooling to estimate $\mu_\alpha$, but reflects a degree of independence as well in that the draws from the underlying alpha population are independent across funds.

The dependent priors in our model are sometimes referred to as "learning priors," since an investor considering a given fund learns something of relevance by examining the properties of other funds. We explore the impact of this alternative modeling of priors on posterior beliefs about mutual fund performance. In one striking example, the pooling of fund returns implicit in the learning approach reduces the posterior mean of alpha from 42.1% to 2.8%. Beyond such important normative issues, studying the

---

[4] Cohen, Coval, and Pastor (2002) independently develop a pooled estimator of performance, though not in a Bayesian setting.
[5] More precisely, PS decompose alphas into separate components related to managerial skill and model misspecification, and only the skill component is independent across managers.

hyperparameters, $\mu_\alpha$ and $\sigma_\alpha$, is of interest in itself, in that these parameters can be viewed as convenient summary measures of the overall evidence on fund performance and indications of the degree of semi-strong form efficiency of financial markets.

Our approach also has important implications concerning survivor bias. When considering asset allocation issues, one is necessarily restricted to investing in existing or surviving funds. Previous Bayesian analyses have computed posterior beliefs for restricted samples of this sort, appealing to assumptions under which there is no bias. While the mathematical justification is undoubtedly clever, it has always seemed to us to be counterintuitive. We show that with prior dependence, there is indeed a bias, estimated to be 40-50 basis points per annum. A more detailed discussion is given in the next section. Recent independent work by Stambaugh (2002) also explores survival issues in the context of prior dependence.

## 2. Survivorship issues

Before going on to describe the details of our model and associated methodology, we discuss another important general implication of allowing for prior dependence. When considering asset allocation issues, one necessarily is 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 simple, but significant 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 subtler when prior dependence is introduced. Let $\theta_j$ be the vector of regression parameters (alpha, beta, and residual variance) and $r_j$ the vector of returns for fund j. Let $\theta \equiv (\theta_1, \theta_2, \ldots, \theta_N)$ the parameters and $r \equiv (r_1, r_2, \ldots r_N)$ returns for all N funds, surviving as well as disappearing ones. Finally, let F denote the vector of factor or benchmark asset returns. BMW assume that

$$p(survival_j \mid r, F, \theta_j) = p(survival_j \mid r, F), \tag{1}$$

where survival$_j$ is a 0-1 indicator variable. It then follows directly from Bayes law that the posterior beliefs satisfy

$$p(\theta_j \mid r, F, survival_j) = p(\theta_j \mid r, F). \qquad (2)$$

With prior and residual independence, it also follows that

$$p(\theta_j \mid r, F) = p(\theta_j \mid r_j, F), \qquad (3)$$

so that other funds can be ignored in forming a belief about the skill of fund j.

With prior *dependence*, (2) continues to hold, but (3) is no longer valid. Thus, (2) says that knowing whether a fund survived or not provides no additional information, given that we observe the return histories. However, our belief about fund j's parameters will, in general, 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 $\mu_\alpha$. Ignoring these returns can be likened to throwing out one tail of the sample distribution when estimating a population mean.[6] To summarize, the implication of our discussion of survivorship issues is that the returns of all funds, surviving and disappearing, will impact our beliefs about any given set of funds and, ultimately, will influence asset allocation decisions too when prior dependence is entertained.

## 3. 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-French (1993) three-factor model, an expanded model that includes the Carhart (1997) momentum factor, and a seven-factor model that

---

[6] One could, in principle, still compute the conditional posterior density $p(\theta_j \mid r_j, F)$ 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.

includes, in addition to the previous four factors, three factors constructed to explain industry return covariation orthogonal to the other four factors.[7] 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.

*3.1 Model and prior specification*

We assume that excess returns have a linear factor structure,

$$r_{j,t} = \alpha_j + \beta_j' F_t + \varepsilon_{j,t}, \tag{4}$$

where $\varepsilon_{j,t} \sim N(0, \sigma_j^2)$. As noted previously, these residuals are assumed to be uncorrelated across funds. 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 $\mu_\alpha$ and unknown standard deviation $\sigma_\alpha$. Therefore, prior beliefs about $\mu_\alpha$ and $\sigma_\alpha$, imply priors for the alphas and, 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 $\mu_\alpha$ and $\sigma_\alpha$ are assumed independent and are represented by a normal distribution for $\mu_\alpha$ and an inverted gamma distribution for $\sigma_\alpha$. 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.[8] While informative priors could be introduced for these parameters as well, the greater precision with which these parameters are estimated makes such an extension less interesting.

---

[7] The three industry factors are constructed in a manner similar to those in Pastor and Stambaugh (2002). 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. A principal components analysis is then performed on these 30 time series, and the first three principal components, once normalized, are taken as portfolio weights for the three industry portfolios.

[8] PS and BMW condition the prior for alpha on a fund's residual variance. Our independence assumption simplifies the analysis, but could be generalized.

*3.2 Overview of the estimation procedure*

In this section, we briefly discuss the main features of our estimation procedure. Further details will be given in an 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 $\mu_\alpha$ and $\sigma_\alpha$. Fund-specific parameters include all the $\alpha_j$, $\beta_j$, and $\sigma_j$. 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.[9]

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

1. $\sigma_\alpha$ conditional on $\alpha_j$ (j=1,…,M) and $\mu_\alpha$

2. $\mu_\alpha$ conditional on $\alpha_j$ (j=1,…,M) and $\sigma_\alpha$

3. $\sigma_j$ and $\beta_j$ conditional on F, $r_j$, and $\alpha_j$ for all j=1,…,M

4. $\alpha_j$ conditional on $\mu_\alpha$, $\sigma_\alpha$, F, $r_j$, $\beta_j$ and $\sigma_j$ for all j=1,…,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 $\mu_\alpha$ and all the $\alpha_j$, the conditional distribution of $\sigma_\alpha$ combines the normal likelihood of the $\alpha_j$ with the inverted gamma prior for $\sigma_\alpha$. It is well known in this case that the conditional distribution of $\sigma_\alpha$ is also an inverted gamma. Step 2 then combines the normal likelihood of the $\alpha_j$ with the normal prior for $\mu_\alpha$. The draw of $\mu_\alpha$ is therefore normal as well.[10]

Step 3 replicates traditional linear regression analysis using conjugate priors. Since priors on $\beta_j$ and $\sigma_j$ are flat and independent of $\alpha_j$, we may simply subtract off $\alpha_j$ from fund j's excess returns and proceed with the draws of $\sigma_j$ from its inverted gamma distribution and $\beta_j$ from its student-t distribution.

---

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

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

Standard conjugate analysis is also used in step 4, where a normal likelihood for each $\alpha_j$ (conditional on $\beta_j$ and $\sigma_j$) is combined with a normal prior with mean $\mu_\alpha$ and standard deviation $\sigma_\alpha$. In this case the conditional distribution of $\alpha_j$ is normal as well.

### 3.3 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) 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."[11]

In Section 4, 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 5, we make comparisons that highlight the role of prior dependence in forming posterior beliefs about alphas.[12] . 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 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, in principle, occur before making any inference or decision with the actual data.

---

[11] Savage (1962) makes this distinction between initial precision and final precision.
[12] Stambaugh (1997) and Jones (2002) also explore the frequentist properties of Bayesian procedures in financial contexts.

The priors on $\mu_\alpha$ and $\sigma_\alpha$ 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: high skepticism, some skepticism, and no skepticism. The no-skepticism prior is taken to be diffuse for both $\mu_\alpha$ and $\sigma_\alpha$ (proportional to $1/\sigma_\alpha$). In this case, the data will dominate our beliefs. The other priors for $\mu_\alpha$ and $\sigma_\alpha$ are informative. The $\mu_\alpha$ 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, $\sigma_\alpha$ 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, greater skepticism, as modeled here, implies a stronger belief that $\mu_\alpha$ is close to zero, as well as greater confidence that the true alphas will be close to $\mu_\alpha$.

## 4. Simulation results with learning priors

Now, we study the distribution of beliefs that investors with the priors above would arrive at on 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. 1,000 Monte Carlo simulations are performed for each experiment. 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 equation (4) assuming a single factor with a monthly mean of 0.005 and a standard deviation of 0.045. The $\beta$ and $\sigma$ parameters for each fund are drawn randomly and independently of each other and of other funds. $\beta$ 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 $\sigma$ of 2.8% with a standard deviation of 1.5% (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

11

independently from a normal distribution with annualized values specified for the mean $\mu_\alpha$ and standard deviation $\sigma_\alpha$.

### 4.1  Simulations when managers have no skill ($\alpha_j = 0$)

Results are presented in histograms that display the sampling distributions of various posterior means or functions of posterior means. In order to distinguish between plots that appear similar and may have different scales, we also include the mean and standard deviation of each distribution. The mean is in the top left corner of each plot, while the standard deviation is in the top right. Figure 1 shows that the initial prior can have a significant effect on beliefs about $\mu_\alpha$, the mean of the population from which the true alphas are drawn. The qualitative patterns observed in 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 average of the alpha estimates must converge to zero as $M \rightarrow \infty$. For large M, the influence of the prior becomes negligible as well. Consequently, for each of our priors, the frequency distribution (across 1,000 simulations) of posterior means of $\mu_\alpha$ becomes more concentrated around zero when M is sufficiently large. Thus, investors become increasingly convinced that their prior belief was correct, 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 $\mu_\alpha$. In other words, the average estimate is shrunk toward zero in forming the posterior mean of $\mu_\alpha$. Shrinkage is greatest when M is low (little data) and when the prior is very precise. In the extreme, when $M = 0$, the mean of $\mu_\alpha$ 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 $M = 100$.

By similar reasoning, since the informativeness of the data is held constant when M is fixed, we would expect dispersion to increase as we go from left to right in Figure 1, with prior precision and shrinkage toward zero declining. This effect should be greatest when M is small and shrinkage is substantial. The patterns in Figure 1 confirm these ideas.

Figure 2 presents results for the posterior means of $\sigma_\alpha$ under the same scenarios as in Figure 1. When M = 10, the locations of the first two distributions largely reflect the assumptions about $\sigma_\alpha$ in the informative priors. Increasing M does not have much impact in the high-skepticism case, as the data are apparently not given much weight. With some skepticism, the means for $\sigma_\alpha$ 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 toward any particular value. The large posterior mean $\sigma_\alpha$ of about 6% in this case, with M = 10, may in part reflect the considerable uncertainty about the location of the mean.

*4.2 Simulations when managers have some skill ($\alpha_j \neq 0$)*

We now summarize a similar simulation experiment in which $\mu_\alpha = 0.6\%$ and $\sigma_\alpha = 1.5\%$. In this case the true alpha of each fund is drawn randomly from a normal distribution with these moments, a draw that is independent of the draws of $\beta_j$ and $\sigma_j$ and of the draws for other funds. The dotted lines in Figures 3 and 4 identify the location of the true population values. In Figure 3, we see again that the average simulated posterior mean for $\mu_\alpha$ converges toward the true value, with considerable learning occurring by the time M equals 1,000, especially for the less skeptical priors. Similarly, Figure 4 shows that by M = 10,000, investors are likely to conclude that $\sigma_\alpha$ is close to the true value 1.5%. In the case of high-skepticism, though, the prior largely dominates the belief about $\sigma_\alpha$ for M as large as 1,000. The more diffuse investor beliefs naturally adjust more quickly.

## 5. 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, 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 $\mu_\alpha$ and $\sigma_\alpha$ 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 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. Many values of $\mu_\alpha$ and $\sigma_\alpha$ are drawn from their prior distributions, and the densities implied by each pair are averaged to obtain the implied prior for the alphas. These priors are plotted in Figure 5 along with a normal density with the same mean and variance. It is apparent that the "somewhat skeptical" prior is not Gaussian. The fatter tails of its leptokurtic distribution imply a higher probability of very large and small alphas than would a normal. Deviations from normality are more difficult to detect for the highly skeptical prior.

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 must 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

$$p_{\text{no-learning}}(\alpha_j \mid r_j, F) \propto p_{\text{flat}}(\alpha_j \mid r_j, F) \times p_{\text{no-learning}}(\alpha_j). \qquad (5)$$

Since it is well known that the flat prior implies a student-t distribution for the posterior of $\alpha_j$, 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 $\alpha_j$ under the no-learning prior.

We focus on three aspects of the cross-sectional (across M funds) distribution of posterior means of the alphas – 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. 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.

To gain some intuition for the effect of prior dependence, consider the posterior distribution of the $M^{th}$ 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

$$p(\alpha_M \mid F, r) \propto p(r_M \mid F, \alpha_M)\, p(\alpha_M \mid F, r_1, r_2, \ldots, r_{M-1}), \qquad (6)$$

where the second term may be regarded as a "conditional prior" on $\alpha_M$.[13] This term represents the investor's belief about $\alpha_M$ before observing the returns on fund M, but after combining the initial prior with all other fund data. Under learning priors, this conditional 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 conditional prior on the $M^{th}$ 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 conditional prior in (6) equals the posterior mean of $\mu_\alpha$ and its variance is the posterior variance of $\mu_\alpha$ plus the posterior mean of $\sigma_\alpha^2$,

---

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

both based on the M-1 fund returns.[14]  Without learning, 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 $\mu_\alpha$.  The latter incorporates some shrinkage toward zero as well.  Because the conditional 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 conditional 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.

*5.1  Simulations when managers have no skill*

Figures 6-8 present sampling distributions for the three functions of the posterior means of the alphas when the true alphas are all zero.  For each of our three priors, results are given first without, and then with learning (prior dependence).  Figure 6 presents the sampling distribution of the average posterior mean alpha, one dimension of the performance of learning and no-learning priors.  For the learning prior, the behavior of these averages mirrors the behavior of $\mu_\alpha$ plotted in Figure 1, and the intuition provided for that figure is helpful for understanding how dispersion changes as M increases.  There is less dispersion in the distribution of posterior means for $\mu_\alpha$, however, because $\mu_\alpha$ is a population mean rather than a "sample" mean, which is what Figure 6 displays estimates

---

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

of.[15]  For M ≥ 100, the distinction becomes minor.  It is also irrelevant with unskeptical (diffuse) priors, for which the posterior mean of $\mu_\alpha$ is almost exactly equal to the average of the posterior means of the alphas.

Now, comparing the learning with the no-learning results, we see that the latter often performs better in Figure 6, sometimes even with cross sections as large as 10,000 funds.  As we will see later, the fact that the no-learning distributions are centered around the true value is mostly a matter of the marginal prior fortuitously being "right".  To gain further insight into the effect of learning across funds, let's think about the prior on the average alpha for each approach.

Arguing as earlier, with independent 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 $\mu_\alpha$ component is unaffected.[16]  Consequently, there will be more shrinkage toward the prior mean in the no-learning case, resulting in less dispersion for the posterior means of the average alpha.  If the prior correctly "guesses" the true population mean, as in Figure 6, this is a benefit.  Of course, the situation will be quite different when we simulate a world with skill.

The shrinkage argument just made is not relevant when the priors are diffuse.  In fact, we see in Figure 6 that dispersion is now lower *with* learning in the unskeptical prior case.  We suspect that this is related to the fact that, with residual heteroskedasticity across funds, the aggregate alpha estimate implicit in the posterior means of the alphas (and $\mu_\alpha$) is akin to a weighted-least-squares average.  The improved efficiency of this average, as compared to the simple average taken across funds with the no-learning approach, may be the source of the lower dispersion.

Next, we consider results for the standard deviation of mean alphas, shown in Figure 7.  In general, when there is no learning, the standard deviations become much more concentrated around a fairly stable central value as M increases.  This makes sense

---

[15]We show below that the prior variance of the average alpha exceeds the prior variance of $\mu_\alpha$ suggesting more shrinkage toward zero for the posterior mean $\mu_\alpha$

[16] Conditional (on $\mu_\alpha$ and $\sigma_\alpha$) independence under learning priors implies that the prior variance of the average alpha is the prior variance of $\mu_\alpha$ plus 1/M times the prior mean of $\sigma_\alpha^2$. When M=1, this is just the marginal variance of alpha.

in that the posterior mean for each fund is an i.i.d. draw with no learning, so we're getting more precise estimates of the same underlying standard deviation of the posterior mean, a typical sampling result. As in Figure 2, 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 $\sigma_\alpha$. In short, the investor with a learning prior becomes increasingly convinced 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, permitting 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, informing 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 posterior mean, examined in Figure 8, 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.

The situation is quite different with our less skeptical learning priors. Rather than shift to the right, as in the no-learning case, the distribution of the maximum mean alpha actually shifts a bit to the left in Figure 8 and becomes much more concentrated as M increases. Under the no-skepticism (diffuse) prior with M=10,000, a maximum as large as 50% 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. Under learning, each fund's alpha is shrunk toward the posterior mean of $\mu_\alpha$, which converges to zero with M when managers have no skill (see Figure 1). This keeps the posterior alphas from getting too large. Shrinkage increases with M as $\sigma_\alpha$, and hence the variance of the conditional prior in equation (6), approaches zero (see Figure 2). Intuitively, if the returns of all other funds have

18

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.

*5.2  Simulations when managers have some skill*

Figures 9-11 present simulation results paralleling those in Figures 6-8, for a world in which $\mu_\alpha = 0.6\%$ and $\sigma_\alpha = 1.5\%$.  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 toward the prior mean 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 Figure 9 is consistently negative for these priors, whereas it approaches zero under the diffuse no-learning prior.

In contrast, with the learning priors, the average error is always centered much closer to zero, at least for M > 10, with dispersion narrowing substantially for higher values of M and less skepticism (less shrinkage).  As discussed earlier, this is because the learning prior aggregates the information about all funds in arriving at a belief about any given fund.  The standard deviations in Figure 10 are also smaller under the less skeptical learning priors, about 1-2% versus 4-5% with no learning in the no-skepticism case.  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 Figure 11 under the no-skepticism prior are even more 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 under learning.

## 6.  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 US equity mutual fund data.

*6.1 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.[17] 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: 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 whose construction was described previously. Our primary motivation for including the industry factors is to better approximate the assumption of residual independence.

*6.2 Results with learning priors*

Table 1 contains posterior means and standard deviations for $\mu_\alpha$ and $\sigma_\alpha$ computed under various learning priors for samples of all funds and surviving funds only. It is immediately apparent that there is strong evidence of skill in the population of equity mutual funds. Posterior means of $\mu_\alpha$ from the sample of all funds are generally between

---

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

1.3% and 2.2% 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.

It is also clear that including only those funds that survived to the end of the sample results in a posterior mean for $\mu_\alpha$ that is higher by about 40 to 60 basis points per year. Therefore, survival bias, while irrelevant under no-learning priors, can substantially inflate alphas computed under learning priors. The survival sample also tends to generate posteriors for $\sigma_\alpha$ 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 $\mu_\alpha$ are fairly similar across the one, three, and four-factor models, though adding industry factors substantially increases the posterior mean of $\mu_\alpha$. The various models produce much more diversity in their estimates of $\sigma_\alpha$, 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 a CAPM world.

Finally, the effects of differing degrees of prior skepticism are relatively small. Posterior means of $\mu_\alpha$ 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 $\sigma_\alpha$ are somewhat more sensitive, the degree of prior skepticism is still not as relevant as the choice of asset pricing model.

*6.3 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 to 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 2.25% per year, consistent with the posterior means of $\mu_\alpha$, 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 20 basis points per year, while unskeptical priors imply average alphas of over three percent 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. With the Fama-French model, for example, under the unskeptical prior 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 implying a standard deviation nearly four times that of the no-learning prior (1.13% versus 0.33%). The unskeptical comparison is driven by the shrinkage of alpha estimates toward a "grand mean" with the learning prior, while the highly skeptical comparison reflects the greater shrinkage toward zero with no learning.[18] As in Table 1, dispersion is heavily dependent on the asset pricing model as well.

Given the results in Table 1, these findings are not particularly surprising. Recall from the earlier discussion of (6) that the posterior for a given fund under the learning prior can be viewed as the result of combining that fund's data with a "conditional prior" that approximates the true alpha distribution when M is large. Thus, posteriors of the alphas are centered near $\mu_\alpha$, with dispersion determined by $\sigma_\alpha$. From Table 1, we can see that the values of $\sigma_\alpha$ supported by the data are somewhat smaller than 3%, which is approximately the mean of $\sigma_\alpha$ under the somewhat skeptical learning prior, but often much larger than .75%, the mean of $\sigma_\alpha$ under the highly skeptical prior. Therefore, loosely speaking, learning results in an effective marginal prior with a positive mean and dispersion that lies between that of the two skeptical no-learning priors.

---

[18] Although not reported in section 5.2, these results are similar to those obtained in the simulations in which managers were assumed to be skilled on average. In these simulations, as in the empirical results, highly skeptical priors shrink alpha posterior means very close to zero, with the no-learning prior effectively applying greater shrinkage. This results in very little dispersion in no-learning posterior means.

One notable result from Table 2 is that for unskeptical priors, there remains a large difference between average alpha means computed under no learning and learning. 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. 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 $\mu_\alpha$ which, as noted earlier, incorporates a sort of weighted-least-squares estimate. The result is a downward shift in the distribution of $\mu_\alpha$, which in turn leads to substantially lower alpha estimates for these high-performing funds under the learning prior.

Figures 12 and 13 plot some of the relationships between the alpha posterior means, for all funds, computed under different priors. For brevity, we plot results only for the seven-factor model and we omit results for the unskeptical priors. Figure 12 shows the relation between learning and no-learning alpha posterior means computed under highly skeptical priors. While the higher mean and higher dispersion reported in Table 2 for the learning prior is clearly evident, the figure further reveals that posterior means under the two priors are highly correlated. High correlation makes sense, even with large differences in the two sets of mean alphas, insofar as the differences are driven by differing degrees of shrinkage toward the fixed prior mean. Interestingly, the correlation seems stronger for funds with low alphas. Figure 13, computed under somewhat skeptical priors, displays the higher dispersion with the no-learning prior. The figure shows that without learning or very skeptical beliefs, some funds are likely to have extremely large alphas, with one no-learning alpha in excess of 25% per year.

The effects of different priors on posterior precision are investigated in Figures 14 and 15, again for all funds. The posterior standard deviation of each fund's alpha is compared across learning and no-learning priors. In Figure 14, we see that a highly skeptical no-learning prior effectively puts a cap on the standard deviation of each fund's alpha. In effect, the precision of this no-learning prior is so strong that, for many funds, the data offer little in terms of added precision. The posterior standard deviations are therefore not much less than that of the prior, which in this case was roughly 1%. Figure 15 shows that a lower degree of prior skepticism leads to a much different result. When

priors are not too skeptical, the effect of learning is to make the "conditional prior" in (6) more precise for each fund. This results in more data-based shrinkage and brings posterior standard deviations into a much narrower range for the learning prior.

Figures 16 and 17 focus on the alpha posterior means of the funds that survived to the end of the sample. These posterior means are computed under learning priors using either the surviving funds only or the entire sample, but only displaying results for the survivors. In the latter case, though we focus on a select group of funds, our inferences about those funds should not be biased. Again, only results for high and some skepticism are displayed. ("No-skepticism" results are similar to "some skepticism.") Figure 16 shows that under high skepticism, survival bias impacts the worst performers most significantly, with better-performing funds approximately unaffected. In Figure 17, however, which is computed under less skeptical priors, survival bias affects funds more homogeneously. The results appear consistent with the 40-60 basis point effects on the corresponding estimates of $\mu_\alpha$ in Table 1.

*6.4 Alphas after fees and costs*

The central question addressed by BMW is whether *any* investment in actively managed mutual fund can be justified. They demonstrate that a necessary and sufficient condition for this to be the case 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. In this section, we therefore 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 end-of-period trading costs and management fees, i.e., these two components are just subtracted off from the posterior means computed previously. As noted before, trading costs are unknown, so we follow BMW and specify costs of six basis points per month. Management fees are known, and like BMW, 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 that an investor could

potentially allocate assets to.[19]    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 survivor 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 20 and 35 basis points per annum for the seven-factor model.    In all cases, this is roughly 2% below the corresponding levels before fees and costs.    Without learning, average alphas are sometimes below −2% or above +2%, depending mostly on the degree of prior skepticism but also on 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.[20]

While the maximum alpha mean, net of fees and costs, is always positive, its size 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.

6.5 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 1-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

---

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

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

65%.[21]  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 lower residual risk, at 3.4%.  The annualized alpha means and standard deviations (in %) under various priors are:

Table 3

Alpha posterior means and standard deviations
for two mutual funds under the 1-factor model

| **Schroder** | High | Some | None |
|---|---|---|---|
| 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** | High | Some | None |
| 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) |

As in our other figures, the degree of skepticism about the magnitude of skill declines from left to right.

First, consider the results for no-learning priors.  Under the unskeptical (diffuse) prior, the posterior mean is just the OLS regression estimate.  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 much 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.

---

[21] Schroder was the top performer out of all funds with at least three years of returns data, and was seventh overall.  The posterior mean of Schroder's alpha *after* fees and costs, 62.1%, was also seventh best.

Now, consider the learning prior results. All alpha posterior means reflect shrinkage toward the posterior means for $\mu_\alpha$, which range from 1.4% to 1.5% for the 1-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.

For each fund, learning 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 $\mu_\alpha$) 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.[22] 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 $\mu_\alpha$ lowers posterior variability under learning priors. The non-monotonic behavior of the Schroder fund's standard deviations as we vary the degree of skepticism under no-learning is surprising and reflects the bi-modal nature of the posterior distribution for alpha in this case.

6.6  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.

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

---

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

$$U(W_{T+1}) = \begin{cases} \dfrac{W_{T+1}^{1-A}}{1-A} & \text{for } A > 0 \text{ and } A \neq 1 \\[2mm] \ln(W_{T+1}) & \text{for } A = 1, \end{cases} \tag{7}$$

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.

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

$$p(r_{j,T+1},\, r_{m,T+1} \mid r, r_m) = \int p(r_{j,T+1},\, r_{m,T+1} \mid \theta_j, \theta_m)\, p(\theta_j, \theta_m \mid r, r_m)\, d\theta_j\, d\theta_m, \tag{8}$$

where $\theta_m$ and $\theta_j$ denote, respectively, the parameters of the distributions of $r_m$ and of $r_j$ given $r_m$. As is often the case in regression models, with independent priors for $\theta_m$ and $\theta_j$, the posterior distribution can be factored as

$$p(\theta_j, \theta_m \mid r, r_m) \propto p(\theta_j \mid r, r_m)\, p(\theta_m \mid r_m), \tag{9}$$

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 $\theta_j$, is important for determining allocations to the market portfolio.

Our approach to computing $p(\theta_m \mid r_m)$ is standard. We assume that $\theta_m = \{\mu_m, \sigma_m\}$ and 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 $\sigma_m$ is inverted gamma and the posterior of $\mu_m$ is Student-t. Using monthly excess value-weighted market returns from January 1961 to June 2001, we find the posterior distribution of $\mu_m$ to have a mean of .47% and a standard deviation of .21%. The posterior of $\sigma_m$ has a mean of 4.48% and a standard deviation of 0.15%.

Ten thousand draws from the predictive distribution (8) are simulated by first drawing $\theta_j$ and $\theta_m$ at random from their respective posteriors and then simulating from

$$p(r_{j,T+1},\, r_{m,T+1} \mid \theta_j, \theta_m) = p(r_{j,T+1} \mid r_{m,T+1}, \theta_j)\, p(r_{m,T+1} \mid \theta_m). \tag{10}$$

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 $w_j + w_m \le .99$, $w_j \ge 0$, and $w_m \ge 0$. For brevity, we report here only the optimal allocation to the mutual fund.

Table 4

Optimal portfolio allocations to two mutual funds under the 1-factor model

| Schroder | | High | Some | None |
|---|---|---|---|---|
| No-learning | A=1 | 0.000 | 0.518 | 0.990 |
| | A=2 | 0.000 | 0.390 | 0.990 |
| | A=5 | 0.000 | 0.220 | 0.990 |
| | | | | |
| Learning | A=1 | 0.002 | 0.098 | 0.070 |
| | A=2 | 0.064 | 0.099 | 0.091 |
| | A=5 | 0.030 | 0.043 | 0.040 |
| | | | | |
| Fidelity | | High | Some | None |
| | | | | |
| No-learning | A=1 | 0.000 | 0.990 | 0.990 |
| | A=2 | 0.186 | 0.990 | 0.990 |
| | A=5 | 0.238 | 0.818 | 0.919 |
| | | | | |
| Learning | A=1 | 0.990 | 0.990 | 0.990 |
| | A=2 | 0.950 | 0.990 | 0.990 |
| | A=5 | 0.381 | 0.435 | 0.442 |

The differences between allocations under no-learning and learning are extreme in some cases. With log utility (A=1), for example, a highly skeptical investor with no-learning priors would invest nothing in the Fidelity fund, while a similar investor with learning priors would allocate 99% of his portfolio, the maximum. In the case of the Schroder fund, an investor with unskeptical no-learning priors would allocate the maximum of 99%, while a comparable investor with learning priors would allocate less than 10%.

Broadly speaking, fund allocations tend to track the alphas reported above. The alphas of these funds tend to be higher with less skeptical priors, particularly in the no-learning case, and the allocations generally reflect this finding. The exception is found in

the allocations to the Schroder fund under learning priors. While Schroder's mean alphas are highest for the least skeptical priors, allocations to the Schroder fund are largest under the "somewhat" skeptical prior. It appears, in this case, that the appeal of a higher mean alpha is more than offset by the "estimation risk" associated with the unskeptical investor's greater uncertainty about alpha (see Table 3). In general, the important message to take away from these examples is that learning can have a huge impact on asset allocation.

## 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 magically revealed to an investor, it would affect her belief about the likely degree of abnormal performance for a given 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 is partly a belief about mutual fund managers as a group and, more generally, a belief about the degree to which financial markets are efficient.

We introduce prior dependence by assuming that the true alphas are random draws from a distribution with hyperparameters $\mu_\alpha$ and $\sigma_\alpha$, the average level of skill and standard deviation of skill, respectively. 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 beliefs that arise as a result of incorporating dependence or "learning" in the priors for alpha. To evaluate this effect, beliefs about the cross-section of alphas are calculated for independent "no-learning" priors and compared to those based on our learning model. Two sorts of shrinkage factors emerge as relevant for understanding the differences observed. First, whereas beliefs about a given fund's alpha are based solely on that fund's returns under prior independence, with learning, an aggregate estimate across funds determines the belief about $\mu_\alpha$. This, in turn, affects beliefs about the individual fund alphas. In other words, learning gives rise to a *data-based* shrinkage factor, with each fund's estimate tilted

toward the average, to a degree also determined by the data. Second, when the priors are informative, there is also shrinkage toward the prior mean of 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 a learning prior, an investor's posterior belief about the hyperparameters gradually converges to the true distribution as M, the number of funds, increases. The convergence tends to be slower for $\sigma_\alpha$ than for $\mu_\alpha$ in the examples we explore. Ideally, deviations between the posterior means and the true fund alphas would be tightly centered about zero. With the learning prior, 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 $\mu_\alpha$ (a zero probability event for continuous priors). This is a fundamental difference between the two approaches. In addition to this bias, the cross-sectional standard deviation of the alpha errors and the maximum error are both much higher under the no-learning prior. In fact, the expected maximum under no learning is unbounded in M, while it is fairly stable with the learning prior.

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. Under learning priors, all funds must be considered in forming posterior beliefs, not just the surviving funds. Using all funds, the posterior means for $\mu_\alpha$ and the average fund alpha are usually around 1.3 to 1.4% per annum (before expenses), but are 70 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%. Results based on the subsample of surviving funds suggest a survival bias of 40-60 basis points. Posterior means for $\sigma_\alpha$ mostly range between 1.5% and 2.3%, depending on the prior and benchmark model. Both $\mu_\alpha$ and $\sigma_\alpha$ are estimated fairly precisely, at least under the residual independence assumption.

Under no-learning priors, average alphas can be much lower (around 10 basis points) with highly skeptical priors or much higher (over 2.5%) with diffuse priors. The former results from strong shrinkage toward the prior mean, while the latter reflects the

*absence* of data-based shrinkage with no-learning. Standard deviations of alpha posterior means exhibit similar behavior. The effect of incorporating dependence is most evident in the maximum posterior mean alphas across funds. 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 or 92% with no skepticism (diffuse priors).

While we have documented substantial effects on the cross-sectional distribution of posterior beliefs about alphas, the implications of prior dependence for asset allocation remain largely unknown. Although two examples demonstrated the substantial effects of learning on the allocation to a particular fund, the implications of prior dependence for asset allocation *across* funds remain unexplored. We believe that the additional layer of cross-sectional dependence introduced by learning priors makes this an interesting and challenging issue for future work. In addition, our simple model of prior dependence might be extended to reflect conditional dependence related to fund characteristics.[23] Alternatively, dependence could be related to fund holdings data in a Bayesian version of the recent Cohen, Coval, and Pastor (2002) approach.

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.

---

[23] For example, Baks (2002) considers the common effect of a given manager or fund organization on fund alphas.

**References**

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

Berger, James O., 1985, *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.

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

Casella, George and Edward I. George, 1992, "Explaining the Gibbs sampler," *American Statistician* 46, 167-174.

Chevalier, Judith and Glenn Ellison, 1999, "Are some mutual fund managers better than others? Cross sectional patterns in behavior and performance," *Journal of Finance* 54, 875-899.

Cohen, Randolph, Joshua Coval, and Lubos Pastor, 2002, "Judging Fund Managers by the Company They Keep," Working paper, University of Chicago.

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

Jensen, Michael C., 1968, "The performance of mutual funds in the period 1945-64," *Journal of Finance* 23, 389-416.

Jones, Christopher S., 2002, "Nonlinear mean reversion in the short-term interest rate," forthcoming in *Review of Financial Studies*.

Pastor, Lubos and Robert F. Stambaugh, 2002, "Investing in equity mutual funds," *Journal of Financial Economics* 63, 351-380.

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

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

Stambaugh, Robert F., 2002, "Inference about survivors," Working paper, Wharton.

Table 1

Posterior means and standard deviations of $\mu_\alpha$ and $\sigma_\alpha$ under learning priors

| | Highly skeptical priors | | Somewhat skeptical priors | | Unskeptical priors | |
|---|---|---|---|---|---|---|
| | | | *K = 1 (RMRF)* | | | |
| $\mu_\alpha$ – all funds | 1.40 | (0.04) | 1.47 | (0.05) | 1.48 | (0.05) |
| $\mu_\alpha$ – surviving funds only | 1.92 | (0.05) | 2.08 | (0.06) | 2.11 | (0.06) |
| $\sigma_\alpha$ – all funds | 1.00 | (0.07) | 1.40 | (0.06) | 1.50 | (0.06) |
| $\sigma_\alpha$ – surviving funds only | 0.82 | (0.05) | 1.24 | (0.06) | 1.36 | (0.06) |
| | | | *K = 3 (RMRF, SMB, and HML)* | | | |
| $\mu_\alpha$ – all funds | 1.30 | (0.05) | 1.38 | (0.05) | 1.38 | (0.05) |
| $\mu_\alpha$ – surviving funds only | 1.77 | (0.06) | 1.92 | (0.06) | 1.95 | (0.07) |
| $\sigma_\alpha$ – all funds | 1.99 | (0.07) | 2.21 | (0.07) | 2.26 | (0.07) |
| $\sigma_\alpha$ – surviving funds only | 1.92 | (0.08) | 2.24 | (0.08) | 2.30 | (0.08) |
| | | | *K = 4 (RMRF, SMB, HML, and MOM)* | | | |
| $\mu_\alpha$ – all funds | 1.33 | (0.04) | 1.37 | (0.05) | 1.39 | (0.05) |
| $\mu_\alpha$ – surviving funds only | 1.73 | (0.05) | 1.85 | (0.05) | 1.87 | (0.06) |
| $\sigma_\alpha$ – all funds | 1.52 | (0.06) | 1.77 | (0.06) | 1.84 | (0.06) |
| $\sigma_\alpha$ – surviving funds only | 1.38 | (0.07) | 1.74 | (0.07) | 1.81 | (0.07) |
| | | | *K = 7 (RMRF, SMB, HML, MOM, and Industry Factors)* | | | |
| $\mu_\alpha$ – all funds | 2.11 | (0.05) | 2.23 | (0.05) | 2.24 | (0.05) |
| $\mu_\alpha$ – surviving funds only | 2.63 | (0.06) | 2.83 | (0.06) | 2.86 | (0.06) |
| $\sigma_\alpha$ – all funds | 2.15 | (0.07) | 2.37 | (0.06) | 2.42 | (0.06) |
| $\sigma_\alpha$ – surviving funds only | 2.00 | (0.08) | 2.29 | (0.07) | 2.34 | (0.08) |

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 | 0.20 | 1.25 | 3.39 |
| | learning | 2.14 | 2.23 | 2.24 |
| standard deviation of mean α | no learning | 0.37 | 2.25 | 8.51 |
| | learning | 1.30 | 1.47 | 1.50 |

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 | -1.91 | -0.55 | 2.18 |
| | learning | 0.20 | 0.31 | 0.34 |
| standard deviation of mean α | no learning | 0.85 | 2.34 | 8.54 |
| | learning | 1.43 | 1.57 | 1.59 |
| maximum α posterior mean | no learning | 1.69 | 24.67 | 92.27 |
| | learning | 6.47 | 7.59 | 7.31 |

Figure 1
Posterior means of $\mu_\alpha$ –– no skill
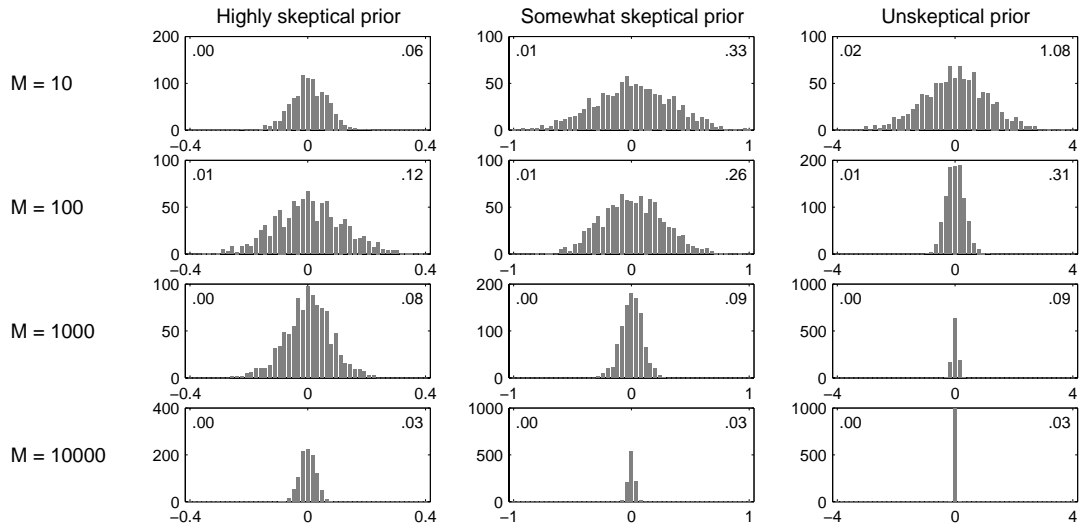


Figure 2
Posterior means of $\sigma_\alpha$ –– no skill

37
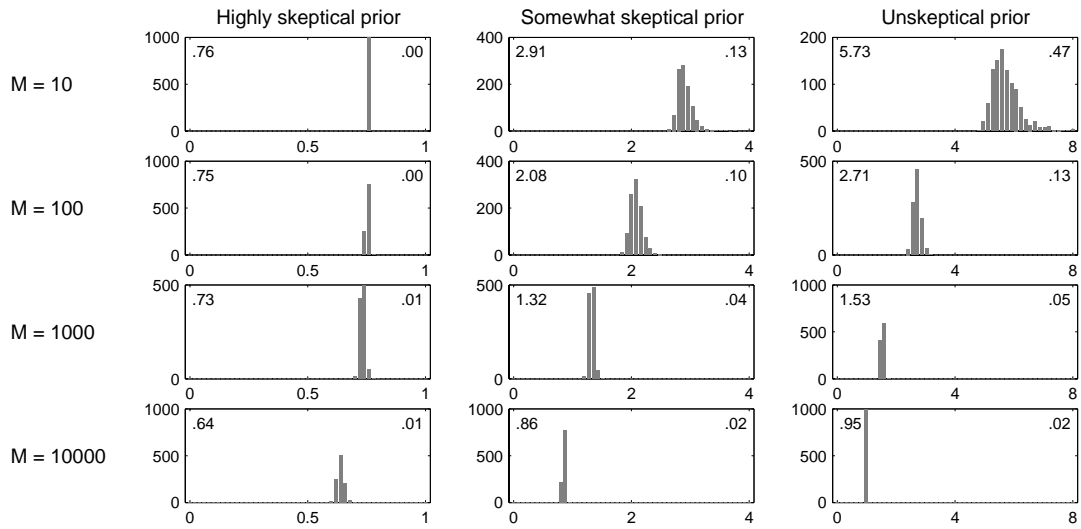
**Figure 3**
Posterior means of $\mu_\alpha$ –– true value = .6%

|  | Highly skeptical prior | Somewhat skeptical prior | Unskeptical prior |
|---|---|---|---|

M = 10

M = 100

M = 1000

M = 10000
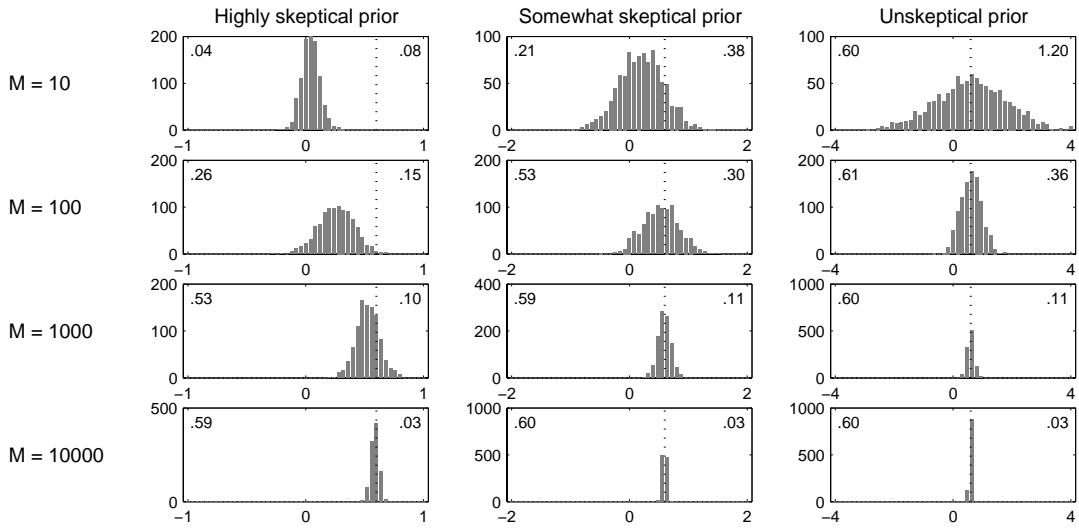
**Figure 4**
Posterior means of $\sigma_\alpha$ –– true value = 1.5%

|  | Highly skeptical prior | Somewhat skeptical prior | Unskeptical prior |
|---|---|---|---|

M = 10

M = 100

M = 1000

M = 10000

38
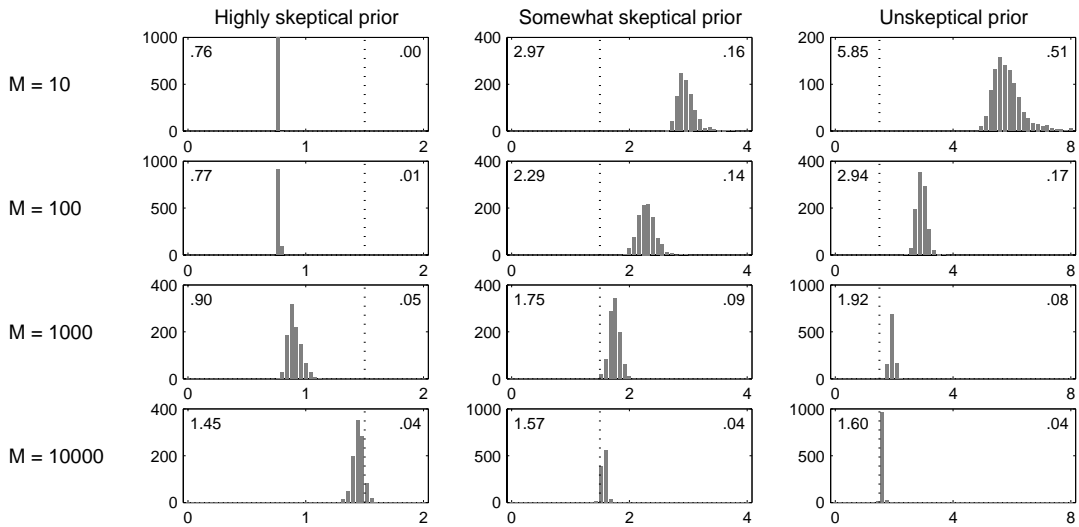
Figure 5
Marginal priors for $\alpha$ (solid) and a normal density with the same mean and variance (dotted)

Figure 6
Average mean α –– no skill

| | Highly skeptical priors | | Somewhat skeptical priors | | Unskeptical priors | |
|---|---|---|---|---|---|---|
| | No learning | With learning | No learning | With learning | No learning | With learning |
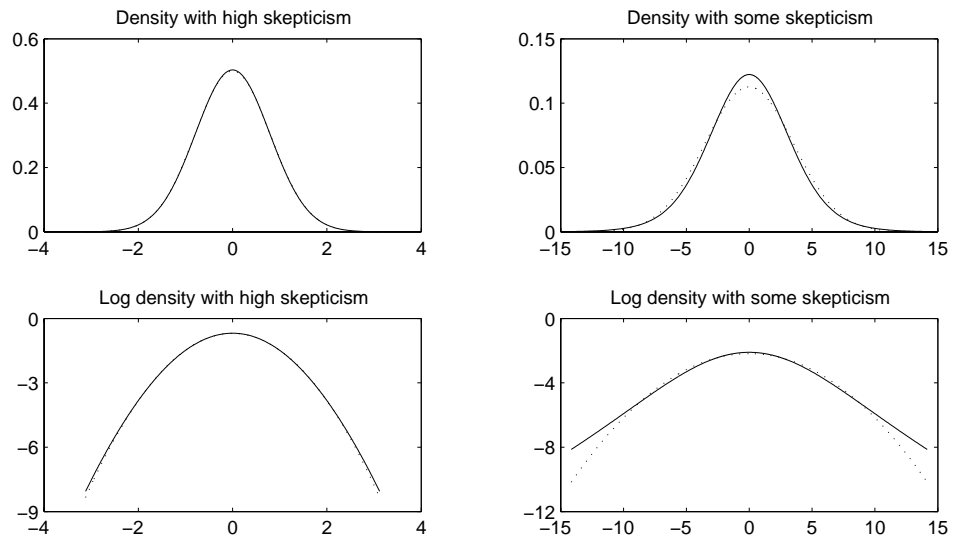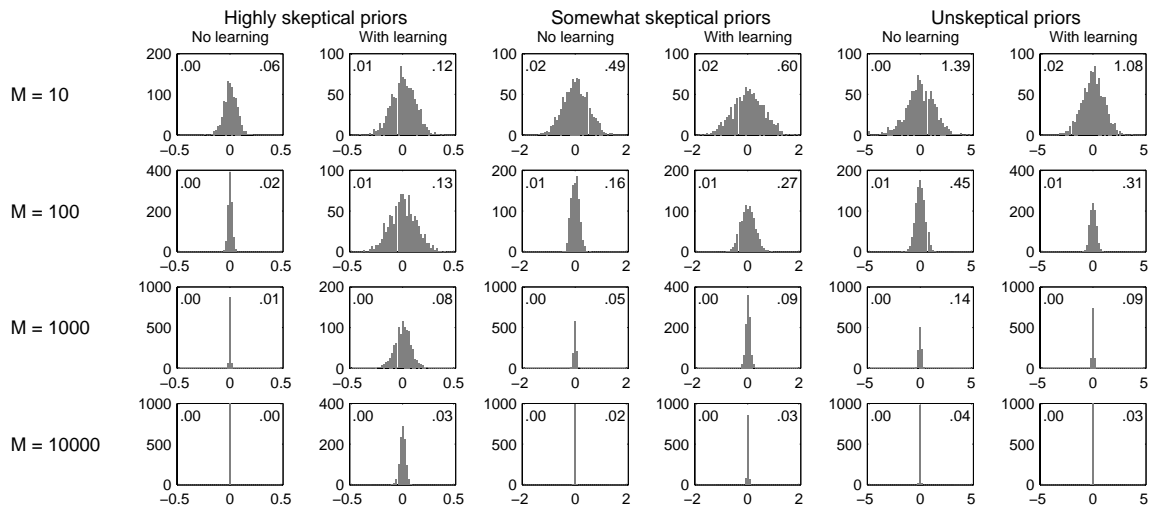


Figure 7
Standard deviation of mean α –– no skill

| | Highly skeptical priors | | Somewhat skeptical priors | | Unskeptical priors | |
|---|---|---|---|---|---|---|
| | No learning | With learning | No learning | With learning | No learning | With learning |



40
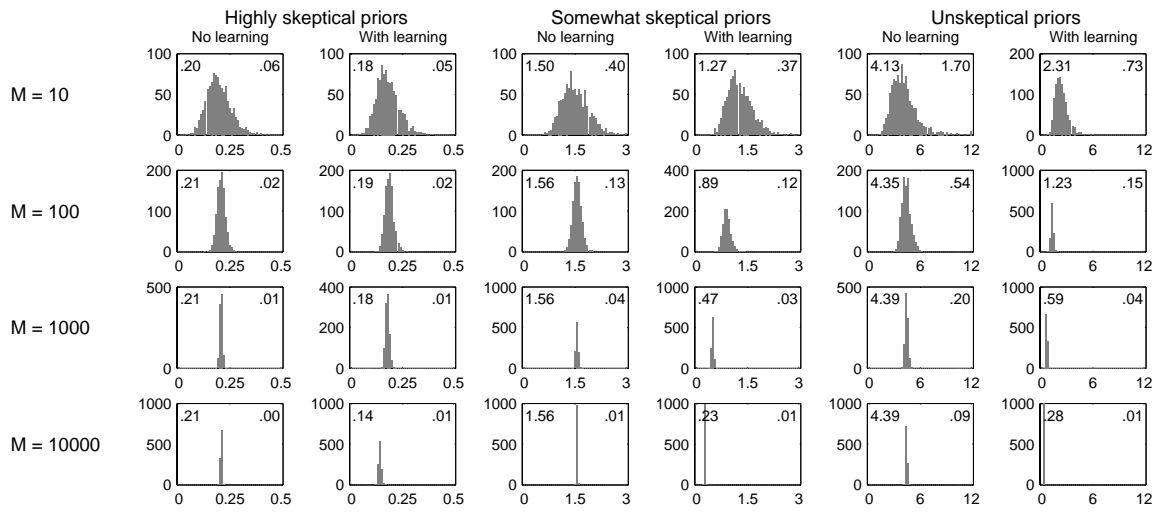
Figure 8
Maximum mean α –– no skill
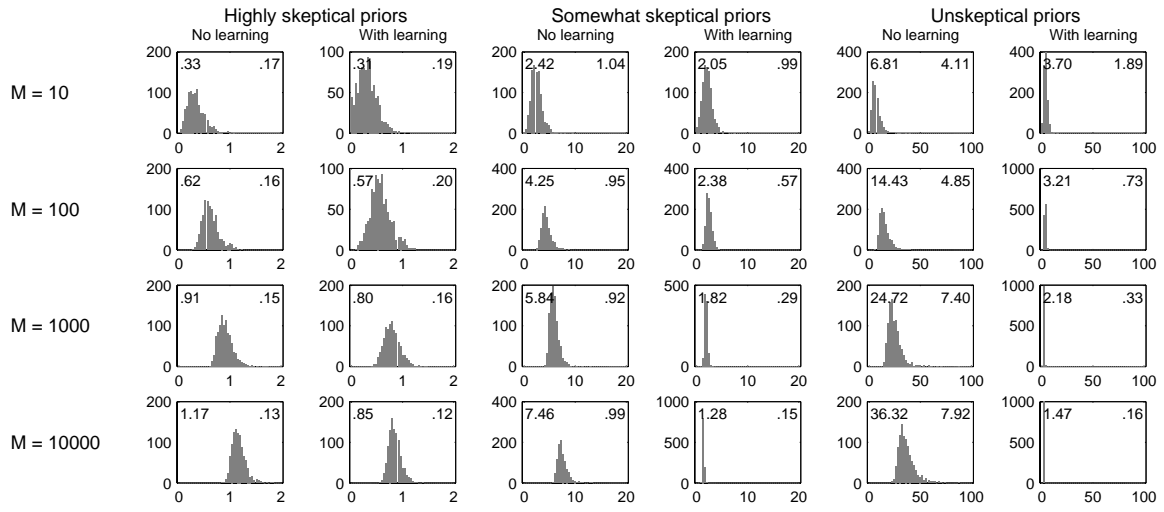
Figure 9
Average mean $\alpha$ error (mean $\hat{\alpha}$ − true $\alpha$) −− $\mu_\alpha$ = .6%, $\sigma_\alpha$ = 1.5%

| Highly skeptical priors | | Somewhat skeptical priors | | Unskeptical priors | |
|---|---|---|---|---|---|
| No learning | With learning | No learning | With learning | No learning | With learning |

M = 10

M = 100

M = 1000

M = 10000
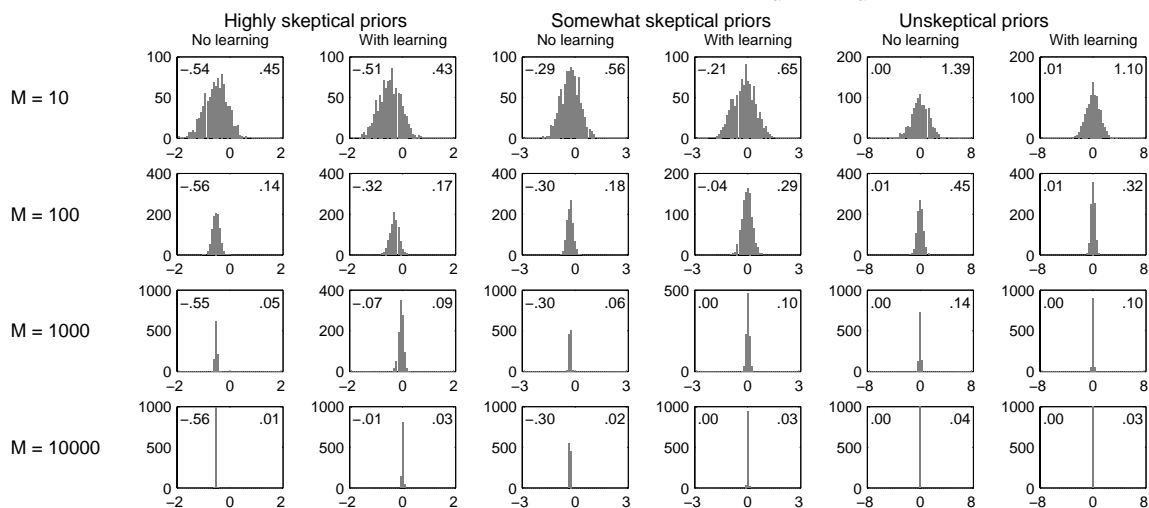
Figure 10
Standard deviation of mean $\alpha$ error (mean $\hat{\alpha}$ − true $\alpha$) −− $\mu_\alpha$ = .6%, $\sigma_\alpha$ = 1.5%

| Highly skeptical priors | | Somewhat skeptical priors | | Unskeptical priors | |
|---|---|---|---|---|---|
| No learning | With learning | No learning | With learning | No learning | With learning |

M = 10

M = 100

M = 1000

M = 10000

42
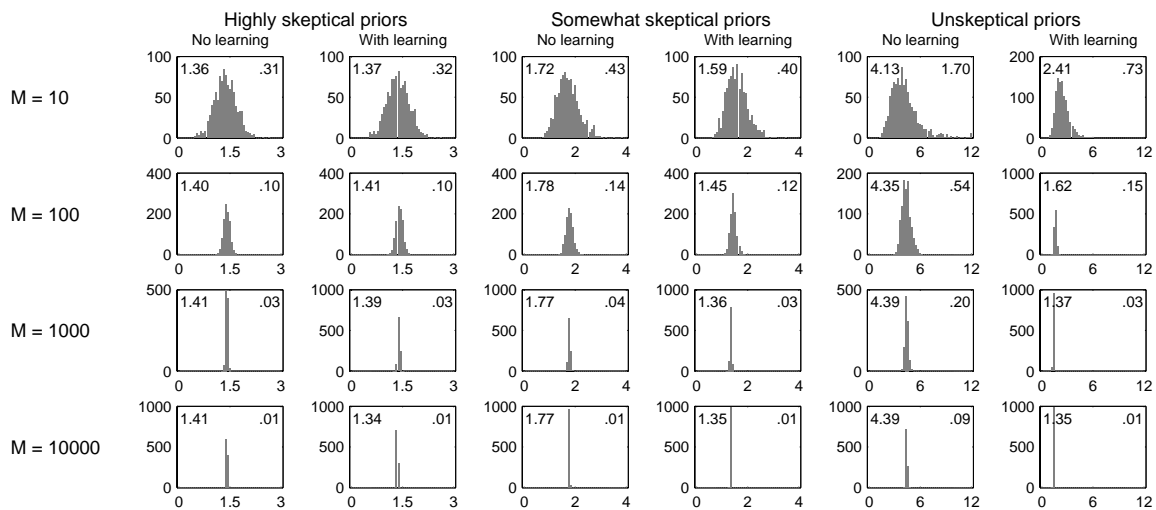
Figure 11
Maximum mean $\alpha$ error (mean $\alpha$ – true $\alpha$) –– $\mu_\alpha$ = .6%, $\sigma_\alpha$ = 1.5%
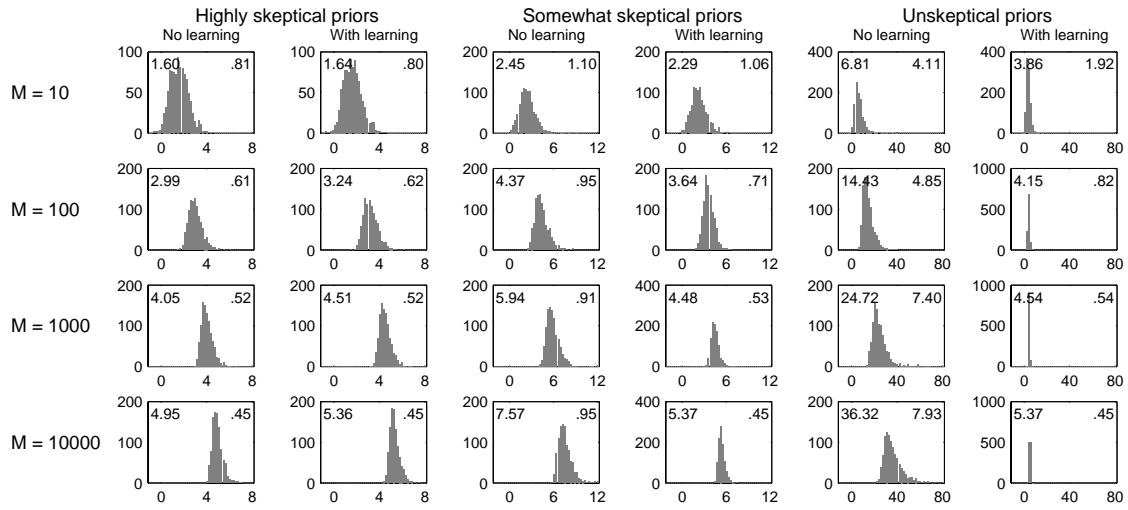
Figure 12
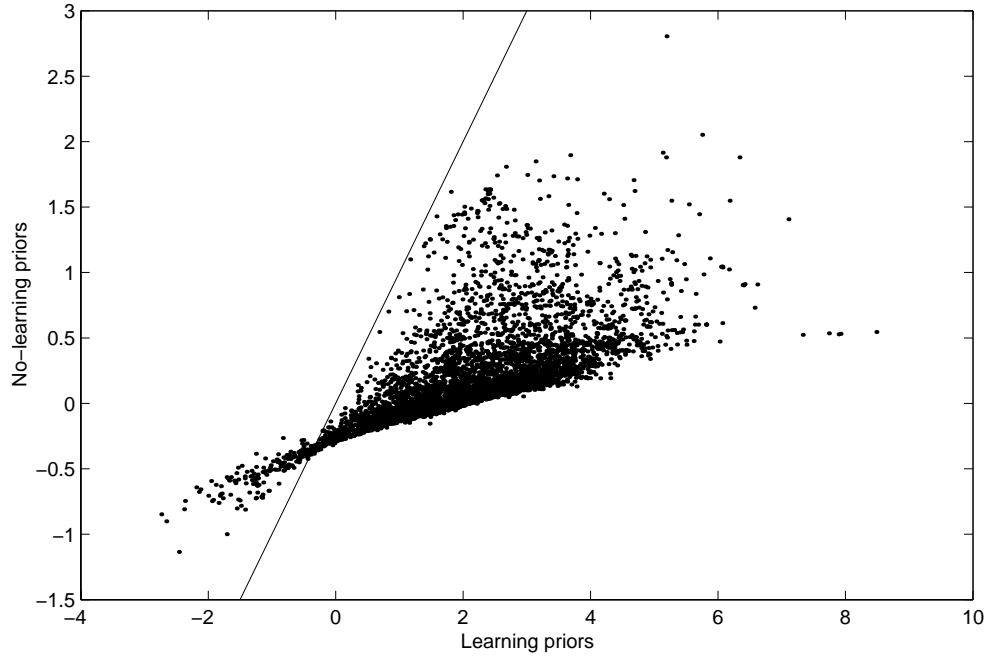Posterior means of individual fund alphas with high skepticism (K = 7)



Figure 13
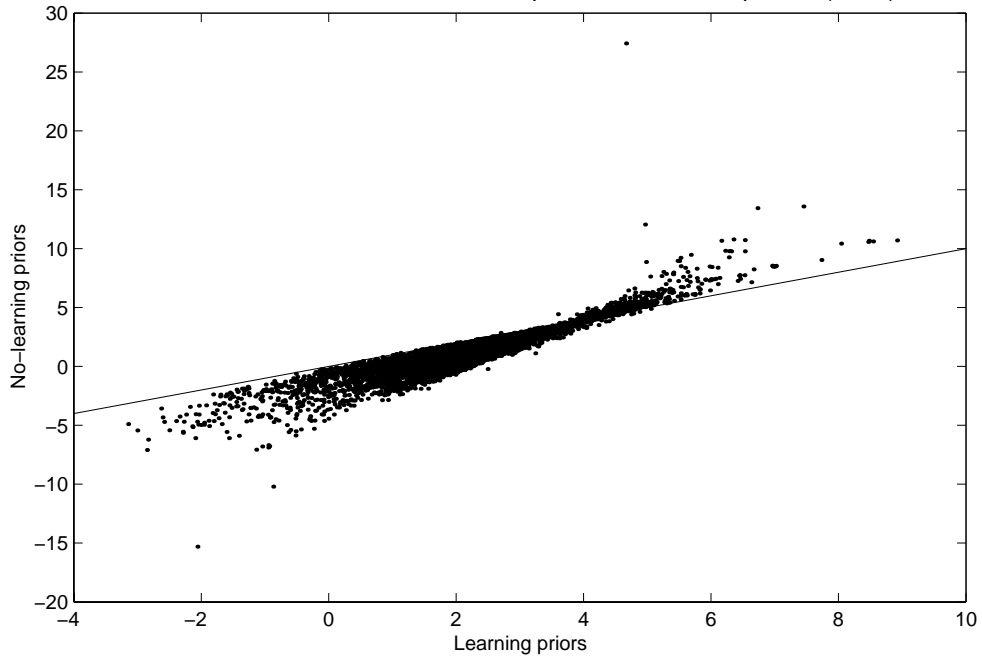Posterior means of individual fund alphas with some skepticism (K = 7)

44

Figure 14
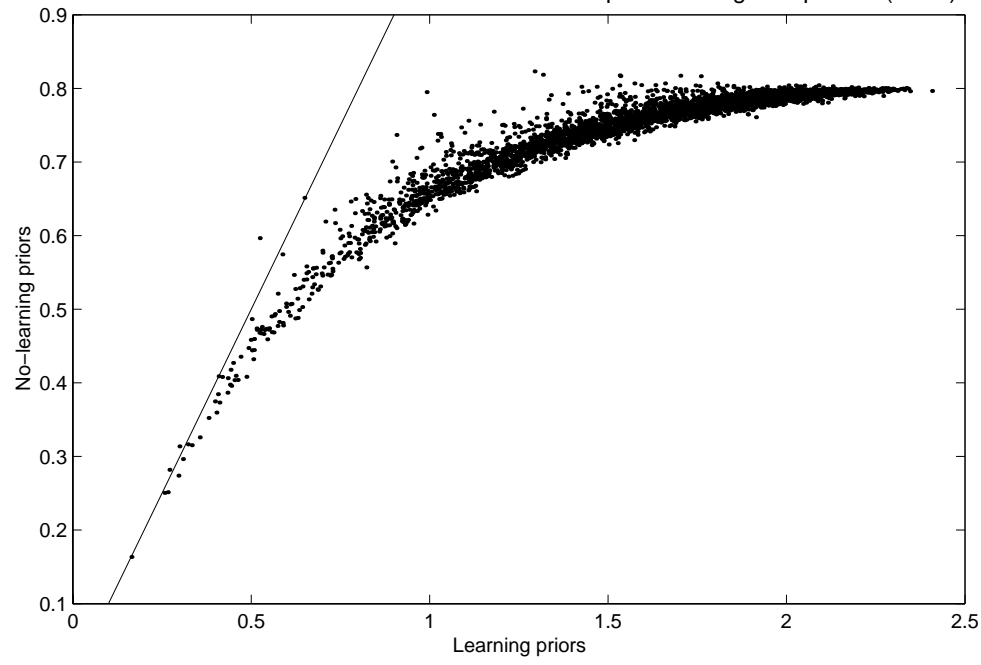Posterior standard deviations of individual fund alphas with high skepticism (K = 7)



Figure 15
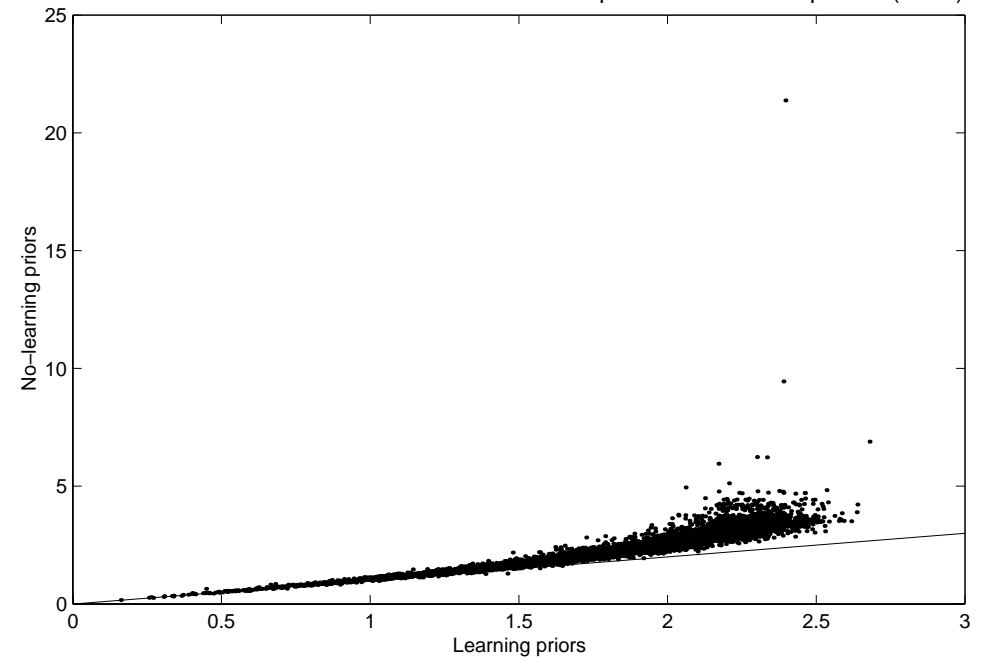Posterior standard deviations of individual fund alphas with some skepticism (K = 7)

Figure 16
Posterior means of individual fund alphas with high skepticism under learning priors (K = 7)
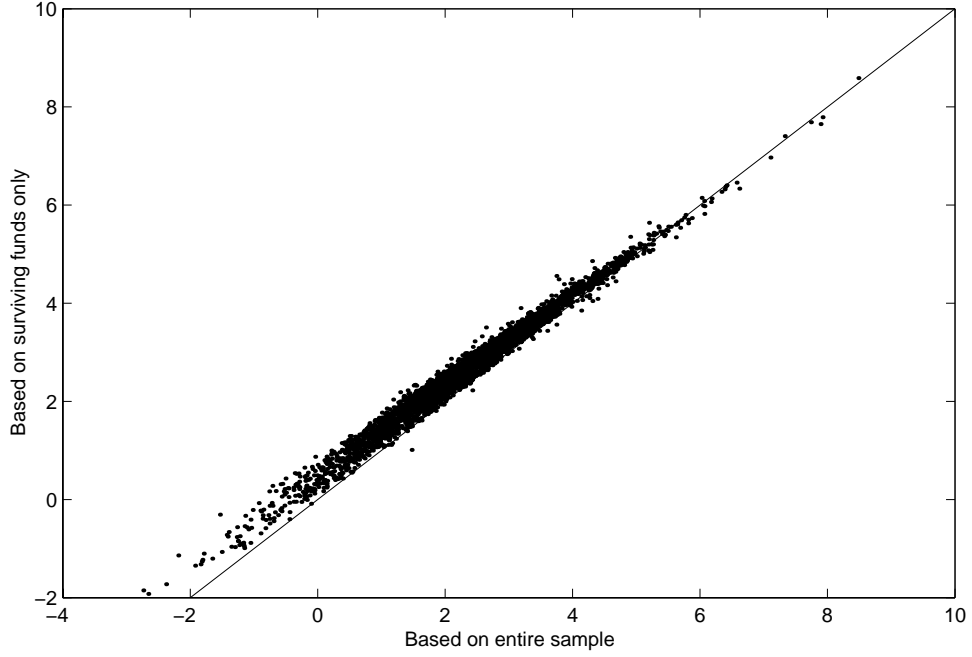


Figure 17
Posterior means of individual fund alphas with some skepticism under learning priors (K = 7)