The Cross-Section of Investing Skill

Ravi Sastry

Columbia Business School

rsastry13@gsb.columbia.edu
www.columbia.edu/∼rvs2103/

this version: November 21, 2011

Abstract

Building on insights from the economics of superstars, I develop an efficient method for estimating the skill of mutual fund managers. Outliers are especially helpful for disentangling skill from luck when I explicitly model the cross-sectional distribution of managerial skill using a flexible and realistic function. Forecasted performance is dramatically improved relative to standard regression estimates: an investor selecting (avoiding) the best (worst) decile of funds would improve risk-adjusted performance by 2% (3%) annually. Forecasted performance also helps to predict fund flows, consistent with the smart money effect. The distribution of skill is found to be fat-tailed and positively skewed; its shape helps to explain the convexity of fund flows.

1 Introduction

With over $3 trillion in aggregate assets under management, the scale of the active mutual fund industry is a testament to the conviction among investors that skilled managers exist and can be identified in advance. Yet the unpredictability of good performance remains the principal stylized fact in the mutual fund literature, and is often interpreted to imply the absence of skill among fund managers and the inability of investors to benefit from active mutual funds.[1] In

[1] Berk and Green (2004) propose a model that reconciles unpredictability with the existence of skill, but all benefits from skill accrue to fund managers rather than their investors.

this paper, I will show that this "fact" is the consequence of strong, unrealistic, and often implicit assumptions about the distribution of "alpha"—namely that alphas are independent across funds, or that alpha is normally distributed. I propose a hierarchical model of investment performance that relaxes both assumptions, and I provide empirical support for alternative "facts": alphas are predictable, and extreme performance is not always luck.

The assumption that alphas are independent across funds is what justifies any fund-by-fund approach to performance evaluation.[2] Although seemingly reasonable, Jones and Shanken (2005) show that this is equivalent to assuming, counterfactually, that the distribution of alphas across managers is known. In the realistic case where the distribution of alphas is acknowledged to be unknown, the returns of all funds intuitively help to illuminate the abilities of fund managers *as a group*, which in turn enables improved inference for any particular fund. Jones and Shanken present a hierarchical model that formalizes this cross-sectional "learning across funds".

They also impose normality on the cross-sectional distribution of alpha, but there is little evidence that alpha follows this convenient form.[3] Rather, it is a modeling choice intended to improve inference by reducing the influence of outliers. When extreme alpha estimates are thought to be spurious, reasonable remedies include shrinkage towards a normal distribution, or still more cautious parameter filters that simply drop funds with estimated alphas outside some allowable range.[4] The surviving funds' alphas can be interpreted with greater confidence, but little is learned about the dropped funds. This is a disadvantage of employing such sharp, and inherently subjective, parameter filters. But the larger cost of treating outliers as problematic statistical artifacts is that the extreme alphas might instead be the most interesting and informative cases—economic rather than statistical outliers.

The "superstars" literature documents convex returns to talent across a range of disciplines. Rosen (1981) describes the phenomenon: "In certain kinds of economic activity there is concentration of output among a few individuals, marked skewness in the associated distributions of income, and very large

[2] Carhart (1997) uses OLS, but more recent methods allow substantially improved inference. See, e.g., Pástor and Stambaugh (2002), Bollen and Busse (2005), and Mamaysky et al. (2007).

[3] Kosowski et al. (2006) and Fama and French (2010) employ distinct bootstrap techniques and document non-normalities in the alpha distribution. Kosowski et al. find more pronounced non-normalities than do Fama and French. They also find stronger evidence of positive alphas net of fees.

[4] Mamaysky et al. (2007) use a Kalman filter to estimate alphas and market betas, and then drop funds where $\left(\widehat{\alpha}, \widehat{\beta}_{mkt}\right)$ falls outside a predefined region.

rewards at the top." Subsequent work explores the pronounced disparities in, for example: CEO compensation, rock concert revenues, television ratings for NBA games, and citation counts for finance academics.[5] Across diverse fields of endeavor, observed outcomes are distinctly non-normal and positively skewed, regardless of the distribution of underlying, and unobservable, talent. Viewed through this lens, extreme alphas are the natural result of differences in innate investment skill.

This perspective requires some semantic clarification. *Alpha* is often considered synonymous with investing *skill*, but these are distinct concepts. Investment skill itself—like the related attribute of intelligence—cannot be directly observed, although it is associated with observable manager characteristics.[6] Alpha is a measure of investment performance, an *outcome* that we expect to be increasing in managerial skill. The economics of superstars shows that the mapping between talent and outcomes is often convex, and that extreme alphas may be the rule rather than the exception. The approach to performance evaluation in this paper incorporates these insights.[7] (For the sake of consistency with the mutual fund literature, *alpha* and *skill* are used interchangeably in the rest of this paper.)

I present a hierarchical model where the cross-sectional distribution of alpha is described by a mixture-of-normals, and each fund manager represents a draw from this distribution. The mixture-of-normals is a flexible and realistic modeling choice, allowing—but not imposing—both skewness and fat tails. The data will dictate whether these features are present.

Out-of-sample tests reveal significant short-term alpha predictability. The "mixture" alphas are striking improvements over the estimates from OLS, which in turn outperform alphas from a hierarchical model with a normal cross-sectional distribution: imposing the wrong structure on the cross-section of alpha can be worse than imposing no structure at all. The bottom decile portfolio of mutual funds formed using the mixture-of-normals distribution is more than twice as bad as the OLS bottom decile portfolio, while the top "mixture" portfolio is nearly twice as good as the OLS top portfolio. (The alphas of the top and bottom "mixture" decile portfolios are +5.4% and -6.1%, respectively.)

[5] These four examples are from Gabaix and Landier (2008), Krueger (2005), Hausman and Leonard (1997), and Chung and Cox (1990), respectively.

[6] See, e.g., Chevalier and Ellison (1999), Cohen et al. (2008), and Coval and Moskowitz (2001) who relate managerial performance to undergraduate institution quality, social ties to corporate board members, and geographic proximity, respectively.

[7] The distinction between alpha and skill is also central to the model of Berk and Green.

I find that the true distribution of alpha is fat-tailed and positively skewed, consistent with the economics of superstars. I also find that fund flows are sensitive to these "mixture" alphas—evidence that investors respond to, and benefit from, the skills of mutual fund managers. Finally, I show that the non-normality of the alpha distribution can help to explain the convexity of fund flows.[8]

The rest of the paper proceeds as follows. Section 2 describes the hierarchical model in detail. Section 3 explains the data used for the analysis. Section 4 presents the main empirical results. Section 5 discusses these results and their relation to the mutual fund literature. Section 6 concludes.

2 Model

I adopt a hierarchical Bayesian model, similar in spirit to Jones and Shanken (2005), that allows for skewness and fat-tails in the cross-sectional distribution of managerial skill. The parameters of this distribution are determined by the data, including "outliers". Skill is measured relative to a four-factor model including the Fama and French (1993) factors and the momentum factor of Carhart (1997). Prior distributions are weakly informative—proper but diffuse.

2.1 Background

The (semi-strong) efficient markets hypothesis implies that abnormal mutual fund returns should not persist. Carhart (1997) provides strong supporting evidence: adjusting for a momentum factor in addition to the three Fama-French (1993) factors produces ranked portfolios of mutual funds without significant persistence among top performing funds. However, Berk and Green (2004) showed that if fund returns are decreasing functions of assets under management, the absence of persistence in factor-model skill estimates is not inconsistent with the existence of skill.[9] Furthermore, studies using shorter holding periods than the one-year holding periods employed by Carhart have found evidence of persistence.[10] Skill may indeed exist, but it appears difficult to identify.

[8] Fund flows are much more responsive to high performance than to poor performance.

[9] Berk and Green (2004) also assume rational investor learning in the context of a common-knowledge prior on the distribution of skill. This is an unrealistic feature of their model.

[10] See, e.g., Bollen and Busse (2004).

Imperfect empirical methodologies may explain the difficulty. Jones and Shanken (2005) observe that "...the maximum OLS alpha estimate [among a population of funds] becomes unbounded as the number of funds approaches infinity, even when the true alphas are all zero."[11] Mamaysky et al. (2007) demonstrate that "...sorting on the estimated alphas populates the top and bottom deciles not with the best and worst funds, but with those having the greatest estimation error." Carhart's result appears less damning upon reconsideration: alpha may seem non-persistent simply because it is poorly measured.

One "solution" is to drop funds with suspiciously extreme alphas. Another approach is to use shrinkage estimators.[12] Bayesian inference automatically incorporates shrinkage towards a prior distribution. Inferences are made on all funds, but the accuracy of the posterior estimates depends very strongly on the shape of the prior.

Baks et al. (2001) employ a prior on skill consisting of the right tail of a truncated normal distribution and a probability mass at the point of truncation. They provide a procedure for eliciting the parameters of this prior from the investor. Jones and Shanken use a normal prior on skill but do not specify its mean or variance, instead estimating these along with all fund-level parameters in a hierarchical procedure. Both priors are simple, but they do not reflect the richness of the true distribution of skill.

Kosowski et al. (2006) document a complex non-normal distribution of mutual fund alphas. They note that "the separation of luck from skill becomes extremely sensitive to the assumed joint distribution from which the fund alphas are drawn" and motivate their non-parametric bootstrap approach by invoking the "intractability of parametrically modeling the joint distribution of mutual fund alphas across several hundred funds." This paper tackles the same problem head-on. The prevailing cross-sectional distribution of skill among mutual fund managers is indeed complex, but it is not intractable.

2.2 Model specification

Excess fund returns are assumed to follow a standard linear factor structure,

$$R_t^j - R_t^f = \alpha^j + F_t\beta^j + \varepsilon_t^j, \tag{1}$$

[11] They assume a standard factor model with independent residuals, as in (1).
[12] See Efron and Morris (1977) for an introduction to shrinkage estimation.

where $\varepsilon_t^j \sim N\left(0, V_\varepsilon^j\right)$, factor loadings β^j and residuals ε^j are cross-sectionally uncorrelated, and the $T \times K$ matrix of factor returns F is observable.

The investor assumes that true manager alphas are independent draws from a finite mixture of two normal distributions, with density

$$f\left(\alpha^j\right) = \sum_{i=1}^{N} w_i \cdot \frac{1}{\sqrt{V_i}} \cdot \phi\left(\frac{\alpha^j - \mu_i}{\sqrt{V_i}}\right), \tag{2}$$

where $N = 2$, $\phi(x)$ is the standard normal density, μ_i and V_i are the *unknown* mean and variance of mixture component i, and the w_i are *unknown* non-negative weights that sum to one.[13] This conditional distribution of alpha is simple and flexible, allowing both skewness and excess kurtosis in the cross-section of managerial skill but imposing neither.[14] It also suggests an intuitive interpretation of the two mixture components as "good" and "bad" manager types.

Greater flexibility could be achieved at the expense of parsimony. Increasing the number of normal mixture components is straightforward, although interpretation is problematic for $N \geq 4$. One could alternately employ finite mixtures of Student's t or Laplace distributions—imposing excess kurtosis—or a more general unimodal skewed distribution.[15]

In the absence of strong justification for any specific alternative distribution, however, it is reasonable to begin with the simplest choice that allows some variability in skewness and kurtosis. The distribution in (2) fits the bill, while minimizing computational complexity and the impact of specification searching.

2.3 Bayesian approach

The hierarchical model defined by (1) and (2) is conceptually simple, yet estimation with classical methods is not feasible. Consider the sheer number of parameters: for each of the J funds there are $K + 2$ parameters, in addition

[13] Without additional restrictions, the mixture of normals in (2) is not identified. An additional identifying restriction, $\mu_2 \geq \mu_1$, is imposed without loss of generality to allow estimation. Stephens (2000) describes the problem in more detail and proposes an alternative resolution.

[14] The distribution is conditional because it depends on the unknown population hyperparameters (μ_i, V_i, w_i) that will be estimated along with the fund-level parameters $\left(\alpha^j, \beta^j, V_\varepsilon^j\right)$.

[15] See Fernández and Steel (1998) for a general approach to modeling skewed unimodal distributions in a Bayesian context. Their method allows skewness and kurtosis to vary independently; an advantage over the simpler model in (2).

to 5 population-level hyper-parameters.[16] For any reasonable values of J and K, the prospects for successfully implementing a high-dimensional numerical optimization, as required by MLE or GMM, are dim.

The Markov chain Monte Carlo (MCMC) approach to the estimation of Bayesian models breaks this curse of dimensionality. Rather than dealing directly with a single intractable $[J \cdot (K+2)+5]$-dimensional joint posterior distribution, we can equivalently work with the $[J \cdot (K+2)+5]$ individual 1-dimensional conditional posterior distributions, which together fully characterize the joint posterior distribution. This equivalence is guaranteed by the Hammersley-Clifford theorem.[17]

2.4 Prior specification

A complete Bayesian model requires both likelihoods and priors. The former are defined by (1) and (2). The latter are described below.

An ideal prior distribution (i) accurately represents the investor's prior information and beliefs while (ii) enabling easy sampling from the posterior distribution. The second condition is often automatically satisfied by a conjugate prior, which gives rise to a known analytical posterior distribution from the same family as the prior. For example, a normal prior and a normal likelihood result in a normal posterior. Conjugacy is a valuable property, but is not always consistent with (i). In such cases, I relax (ii) in favor of (i).

A common criticism of Bayesian inference is that the inherently subjective prior distribution can exert substantial influence on posterior parameter estimates. I minimize subjectivity by choosing intentionally *uninformative* priors. By this, I mean diffuse proper distributions. These are sometimes also called *weakly informative* priors to distinguish them from improper priors.[18]

Putting aside complex approaches to uninformative priors that are themselves subject to criticism,[19] I follow the pragmatic lead of Leamer (1978):

[16] The $K+2$ parameters are the managerial skill and the idiosyncratic variance, in addition to K factor loadings.

[17] See, e.g., Robert and Casella (2004).

[18] Any valid probability density must have a finite integral over the relevant parameter space. An improper prior is a function whose integral is *infinite*, hence the impropriety. Typical improper priors would be $f(\theta) \propto 1$ for a location parameter, or $f(\sigma) \propto \frac{1}{\sigma}$ for a scale parameter. While superficially appealing, such priors do not always lead to valid posteriors. (See, e.g., Lindley (1972).) They are not even uninformative. The improper prior $f(\theta)$ appears to be uninformative with respect to θ, but it also suggests that $|\theta|$ is extraordinarily large and that $\frac{1}{\theta}$ is very close to zero.

[19] For example, the Jeffreys prior or Zellner's g-prior.

> "... we can find a prior distribution that will have little impact on the posterior. It is important to understand that this is not a prior intended to represent ignorance... In practice, the sample may dominate the prior information and the posterior distribution may be inconsequentially different from a posterior distribution corresponding to an improper [uninformative] prior. *A prior density that is relatively uniform where the likelihood function attains its maximum is likely to imply such a posterior.*"[20]

The priors on μ_i and *each* β_k^j are independent normal distributions with means of zero and large variances, with the exception of β_{mkt}^j which has a prior mean of one. The priors on V_i and V_j^ε are independent gamma distributions with shape parameters of one and large scale parameters.[21] The prior on $w = [w_1, ..., w_N]$ is a Dirichlet distribution with an $N \times 1$ vector of ones as the parameter.[22] The prior on α^j is already defined by (2). See Appendix A for a detailed presentation of the prior and posterior distributions for all model parameters.

2.5 Hierarchical structure

Having specified a complete Bayesian model in some detail, it may seem self-defeating to select priors that convey no information. Such a view overlooks the importance of the hierarchical structure of the model. The power of this structure to influence posterior parameter estimates is not undone by arbitrarily diffuse priors, as long as they are proper.[23] Indeed, the chief justification for choosing diffuse priors is to ensure that the estimation results are driven by the model itself rather than any specific choice of priors.

If we altered the model to eliminate (2) and estimated (1) with diffuse priors, this would be equivalent to OLS. However, the estimates generated by the full model are radically different from those given by OLS—these differences must be attributed to the hierarchical structure itself.

[20] pp. 61-63, emphasis added.

[21] A more common choice would be inverse-gamma distributions, which have the virtue of conjugacy. However, these are not consistent with condition (i) or Leamer's rule of thumb, especially when the posterior parameter estimates are likely to be close to zero. See Gelman (2006).

[22] This results in a marginal prior for each w_i that is beta distributed with parameters 1 and $N - 1$. For the degenerate case where $N = 2$ that is employed here, this is simply the uniform distribution on the unit interval.

[23] Improper priors are another story, since they can effectively preclude the cross-sectional learning that is integral to hierarchical model estimation.

3 Data

Mutual fund data are obtained from the CRSP Survivor-Bias-Free US Mutual Fund Database, which begins in 1961. Monthly returns are net holding period returns—all estimated alphas are *after* costs and management fees.[24] This takes the perspective of an investor considering allocating capital to active fund managers, rather than that of a philosophical investigation of the existence of skill. Alpha only matters to the extent that it exceeds costs.

Monthly factor returns are obtained from Kenneth French's website via WRDS. I use a four-factor model, consisting of the three Fama-French factors and a momentum factor, throughout the analysis.

The CRSP data are incomplete in several key ways. Multiple share classes of a single fund appear as separate funds in CRSP, with no foolproof way of identifying and grouping them. Addressing this shortcoming is especially important in this analysis, since double-counting funds would impair estimation of the cross-sectional distribution of skill. Furthermore, CRSP data regarding fund objectives and fund managers are not comprehensive. Each of these is addressed in turn.

3.1 Multiple share classes

The cross-sectional distribution of skill in (2) can only be estimated if each fund in the sample represents a distinct draw from the distribution. Many mutual funds have multiple share classes, and each class appears in the CRSP database as a separate fund. To proceed with the analysis, these "funds" need to be identified in order to avoid double-counting. Unfortunately, there is no explicit share class variable in the database.

CRSP suggests matching "funds" that share the same portfolio of holdings, but portfolio data are not available until 1998. The only alternative is to examine management company and fund names, matching as deemed appropriate. This latter approach is employed here.

Funds without fund name data are consequently excluded from the analysis. This affects approximately 10% of the funds at the beginning of the sample in 1961 but generally disallows less than 5% of all funds.

[24] To reconstruct gross returns, one should add back management expenses and trading costs. Expense ratios are available for most funds, so these could be added back to the reported net returns. However, trading costs would need to be estimated in order to compute approximate gross returns. See, e.g., Wermers (2000) and French (2008).

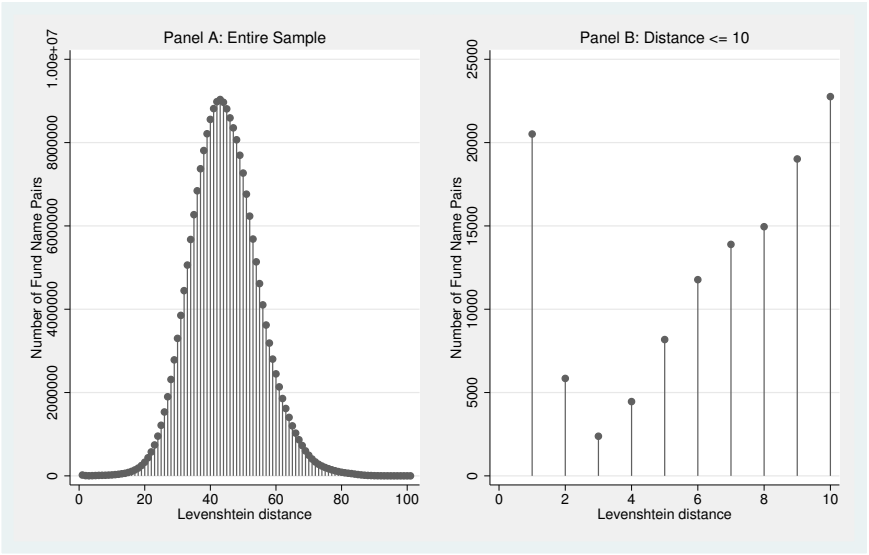

Figure 1: Levenshtein distances of all pairs of fund names. Panel A shows the entire CRSP sample. The mode is quite large – between 40 and 50. Panel B zooms in to show only distances less than or equal to 10, clearly indicating the presence of a secondary mode at 1.

Fund names of the remaining funds are filtered to remove all non-alphabetic characters and extra spaces. Based on the general pattern that share classes of the same fund will have the same name followed by a suffix denoting the share class, funds are then grouped according to their names' pairwise Levenshtein distances.[25] Panel A of Figure 1 shows a histogram of all pairwise Levenshtein distances for the entire sample of 20,000+ funds. The mode is between 40 and 50, and it appears that very few pairs of funds have similar names. Panel B of Figure 1 shows the same histogram only for distances less than or equal to 10. Here we can see that there is a secondary mode for distances equal to 1. Although most random pairs of funds have names that are nothing alike, some pairs have names that are nearly identical.

Funds with Levenshtein distances less than 3 are assumed to be share classes of a single actual fund. From each grouping of fund share classes, the class with the most assets under management is retained to represent the fund. Figure 2 shows the total number of "funds" in the CRSP database, the number of funds

[25]The Levenshtein distance between two strings is the minimum number of edits needed to transform one string into the other, with the allowable edit operations being insertion, deletion, or substitution of a single character. See Levenshtein (1966).

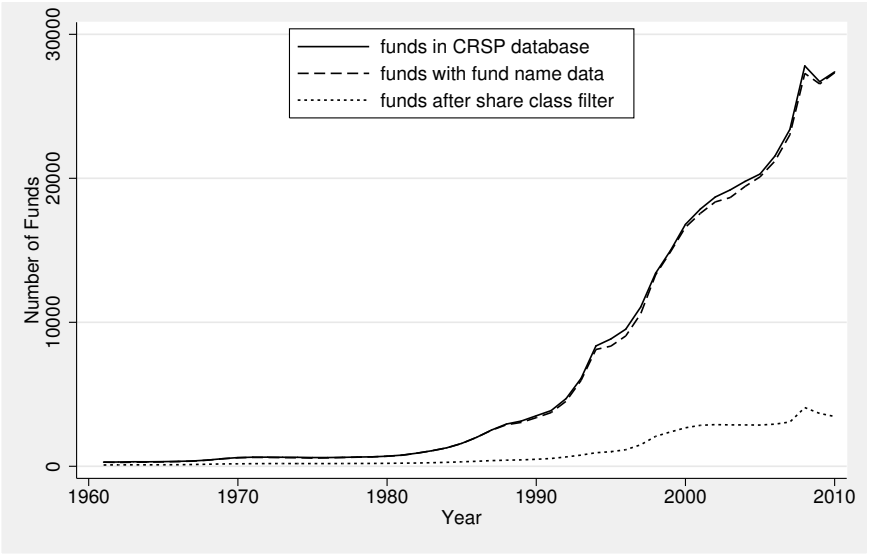

Figure 2: Number of funds in the sample before and after the share class filter. Essentially all funds have fund names recorded in the database (dashed line), but the majority do not survive the share class filter (dotted line).

that have fund name data, and the number of true funds that survive the share class filter.

3.2 Fund objectives

This analysis is confined to actively managed domestic equity mutual funds. Fund names and Lipper class (where available) are screened to drop money market, bond, balanced, and international funds. Index funds are identified by screening fund names for explicit references to indexing. Closet indexers cannot be identified in this way, and any such funds remain in the sample. This corresponds to the challenge faced by investors in prospectively identifying closet indexers. Furthermore, such funds will in general have estimated alphas close to zero, less their trading costs and management fees. They are unlikely to populate the highest or lowest deciles of ranked mutual funds, which are the basis of the key results of this study.

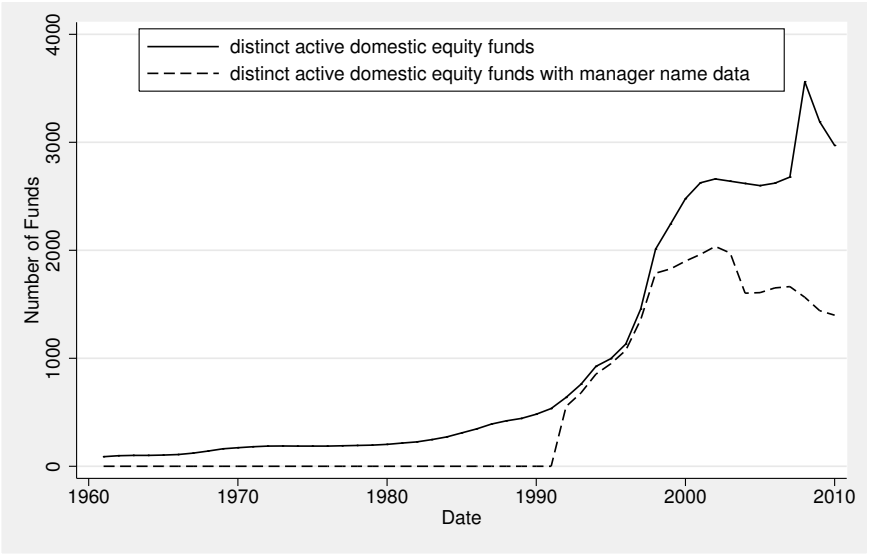

Figure 3: Effect of requiring fund manager information on sample size. The solid line represents the fund sample used in the rest of the analysis. The sudden jump in the number of funds in 2007 corresponds to a change in the data sources used by CRSP.

3.3 Fund managers

Data on fund managers in the CRSP database are not comprehensive. No manager name data is available before 1998. Even after 1998, in many cases, the fund manager field is missing or recorded as *team*.[26] Utilizing individual fund manager data is therefore not feasible without impoverishing the sample. The full impact of requiring individual manager data is depicted in Figure 3.

One consequence of the lack of reliable manager data is that the funds themselves become the objects of analysis. Ideally, when a fund changes its portfolio manager, the new prior on its alpha would depend on the new manager's previous fund's returns. Since the data do not allow managers to be tracked in this way, I use funds' returns as if no management changes have occurred.[27] This does not introduce any bias but does weaken the model's predictive ability for funds that have changed managers during the evaluation period.

[26] Massa et al. (2010) discuss why fund management companies may elect to have anonymously managed funds.

[27] This approach assumes fund returns are generated according to (1) and depend only on the fund manager. If the fund family makes a difference, then future returns are related to past returns even after a managerial change. Brown and Wu (2011) estimate such a model in a Bayesian setting.

4 Results

For each month, alphas are estimated for all funds with at least 12 monthly observations during the 2-year window $[t-23, t]$. Funds are evaluated using each of three methods: OLS, hierarchical Bayes with a mixture-of-normals prior, and hierarchical Bayes with a "standard" normal prior.[28] All funds are ranked according to their estimated alphas and assigned to deciles. Funds that are missing during month $t+2$ are dropped; equally-weighted portfolios are formed from the remaining funds.[29] These portfolios are held for one month and then rebalanced according to the updated alpha estimates.

Following Busse and Irvine (2006), the alpha of each of these decile portfolios is estimated using standard OLS regressions. When estimating the alpha of a *portfolio* of funds, there is likely to be little difference between the OLS measure and any reasonable Bayesian measure based on the same model. Using the OLS post-ranking-period measure avoids any bias that could result from using the Bayesian measure for both the ranking-period and the post-ranking-period.

4.1 Parameters of the mixture-of-normals distribution

Posterior means and 95% confidence intervals of the parameters in (2) are shown in Figure 4 and Figure 5. Each rolling estimation window is 24 months long, so the estimates depicted for month t correspond to fund performance during the window $[t-23, t]$.

Figure 4 shows posterior estimates of both mixture component means, μ_1 and μ_2, and the weight on the second mixture component, w_2. By construction, and without loss of generality, the mean of the second component is greater than or equal to the mean of the first component. Panel A shows that μ_1 is generally indistinguishable from zero, although it does diverge from zero towards the end of the sample – positively during the very early 2000s and negatively since then. Panel B shows that μ_2 follows a pattern similar to μ_1, essentially always statistically equal to zero. Panel C shows that the proportion of managers drawn from the second, high skill, component has always been low and has steadily decreased over time – falling from 20% at the start of the sample to 10% at the end.

[28] This last method follows Jones and Shanken (2005) and allows a direct assessment of the value of using a sufficiently flexible prior distribution.

[29] A one-month gap is introduced between the ranking-period and the post-ranking-period to eliminate spurious microstructure-induced correlations.

Figure 5 shows posterior estimates of the standard deviations of both mixture components. Far more managers are drawn from the first component, so the confidence intervals on $\sqrt{V_1}$ are much tighter than those on $\sqrt{V_2}$. Nevertheless, it is clear that $\sqrt{V_2}$ is an order of magnitude larger than $\sqrt{V_1}$.

The overall distribution has substantial excess kurtosis because the individual variances are unequal; in this case $V_2 \gg V_1$. This, plus the fact that $\mu_2 > \mu_1$ gives rise to positive skewness as well. Both findings are consistent with the economics of superstars.

Although one virtue of the mixture-of-normals distribution was the potentially straightforward interpretation of the components as good and bad manager types, the high dispersion of the second mixture component suggests a similar but more nuanced interpretation. Managers with extremely high *or* low relative alphas are likely to be drawn from the second mixture component, consistent with positively correlated alpha and risk-taking: managers drawn from this high-alpha, high-variance component are taking aggressive risks that pay off on average but that expose their funds to significant losses.

4.2 Comparison of skill distribution moments, by prior

Posterior moments of the distribution of skill for all three "priors" are shown in Figure 6.[30] Moments for both hierarchical Bayes priors—normal and mixture-of-normals—are computed directly from their posterior parameter estimates. Moments for the OLS distribution are computed from the estimated fund alphas.

Panel A compares the estimated means of the prevailing distribution of skill from OLS and the mixture model. The estimated mean from the normal hierarchical model is nearly identical to that of the mixture model and is not depicted. The posterior mean from the mixture model behaves as we might expect: it generally follows the OLS mean but shrinkage mitigates the larger fluctuations.

Panel B compares the estimated standard deviations of the prevailing distribution of skill. The susceptibility of OLS to the influence of outliers is clear. The normal hierarchical model generates the lowest estimates—shrinkage in its strongest form. In between lies the estimated standard deviation from the

[30] OLS may not be an explicit Bayesian prior, but OLS estimates will exactly match Bayesian estimates if the priors are dogmatically diffuse; e.g. normal distributions with infinite variances. If such strictly uninformative, improper priors are specified, there can be no shrinkage and the posterior alpha estimates will depend only on the likelihood. Thus, Bayesian estimation would yield the maximum-likelihood estimates for a regression model such as (1), which are equivalent to the OLS estimates.

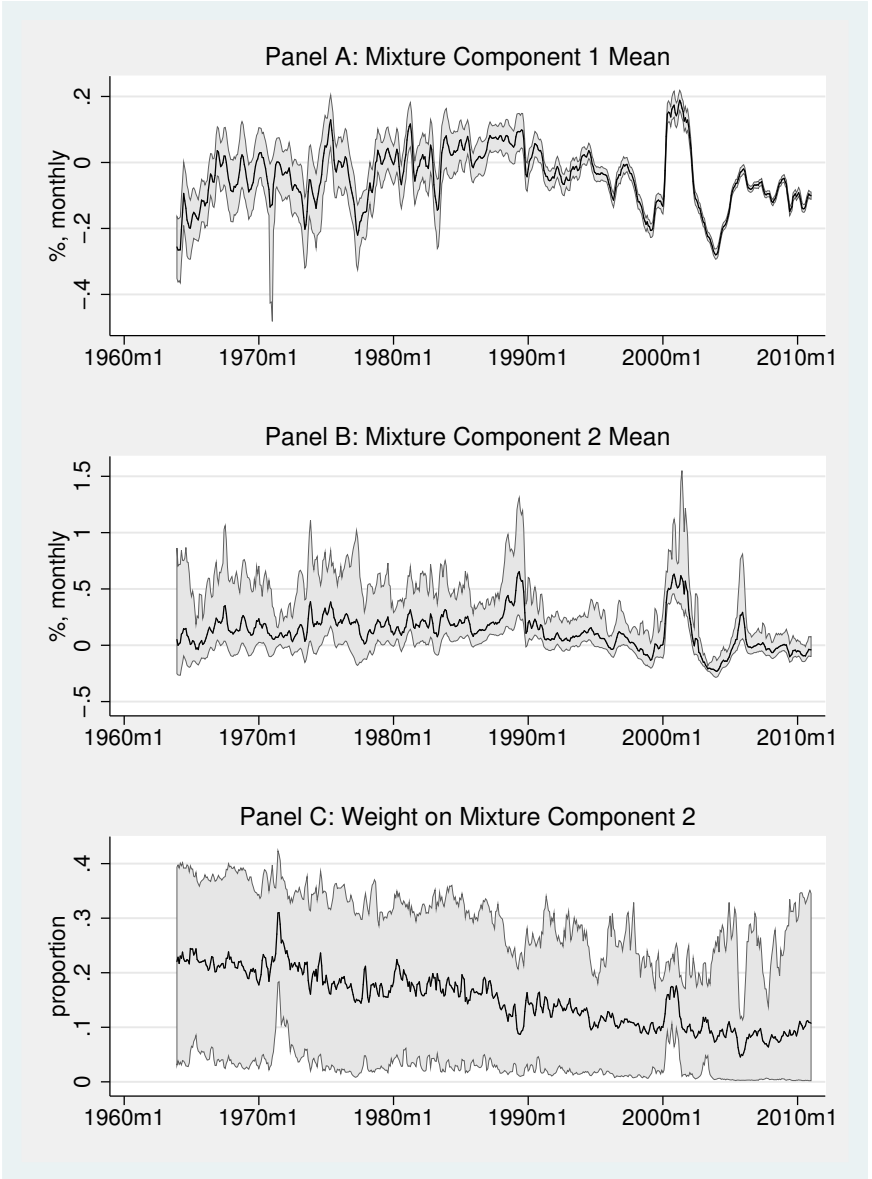

Figure 4: Rolling posterior estimates and 95% confidence intervals for the means of both mixture components and the weight on the second mixture component in (2). The mean of the second component is constrained to be higher than the mean of the second component.

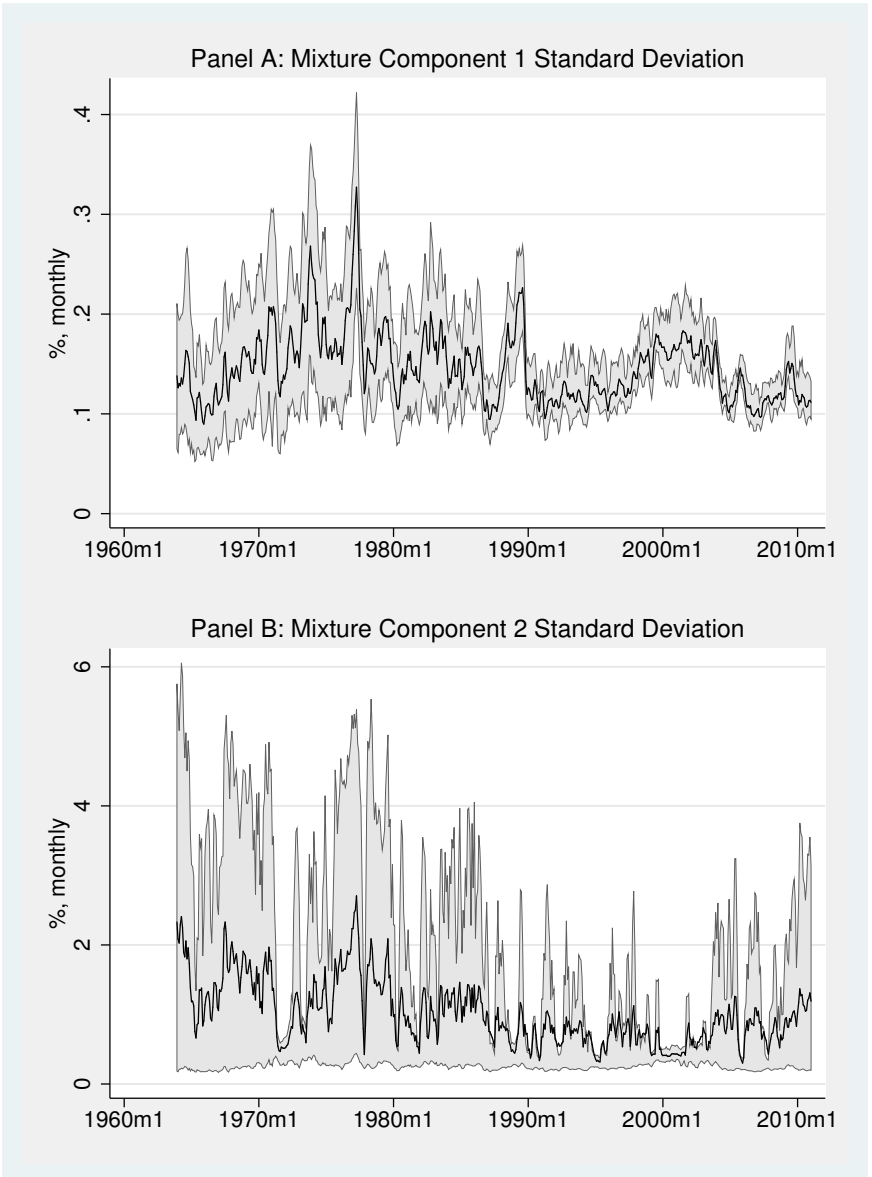

Figure 5: Rolling posterior estimates and 95% confidence intervals for the standard deviations of both mixture components in (2).

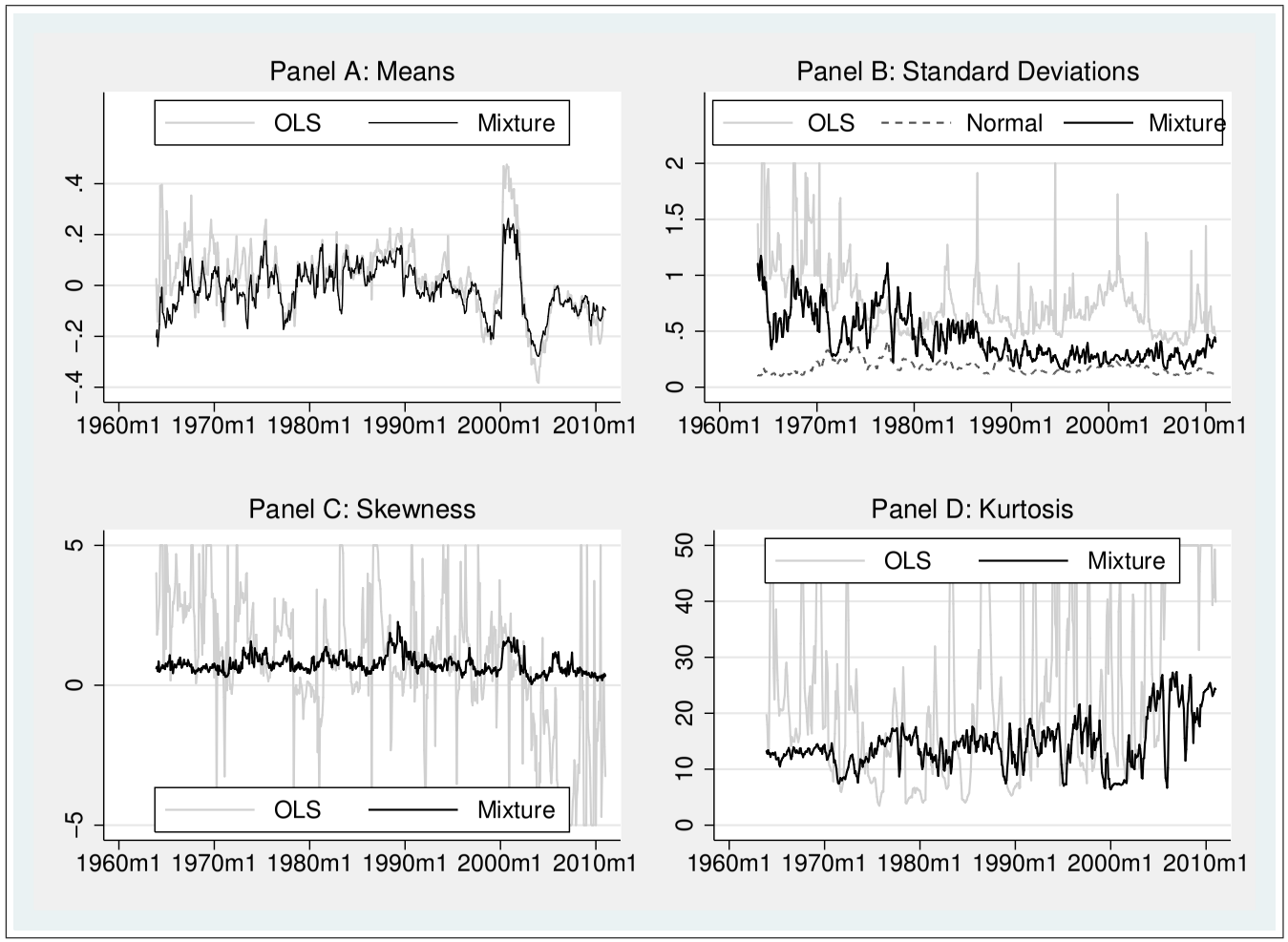

Figure 6: Comparison of the moments of the distribution of skill over time, by model. In Panel A, the mean for a normal prior is almost identical to that for a mixture of normals, and is omitted for the sake of clarity. In Panel C, all skewness values with magnitude greater than 5 are set to +5 or -5, as appropriate, to prevent outliers from dominating the figure. Similarly, in Panel D, all kurtosis values greater than 50 are set to 50.

mixture hierarchical model. This model is flexible enough to take account of extreme alphas without being dominated by them.

Panel C compares the estimated skewness of the prevailing distribution of skill from OLS and the mixture model. The normal model, of course, is constrained to have zero skewness. OLS cannot decide whether skill is positively or negatively skewed. Indeed, the series is so volatile and the excursions from zero are so large (in both directions) that it cannot possibly correspond to reality. The mixture model produces relatively stable and comprehensible estimates of mild positive skewness.

Panel D tells a similar story via the kurtosis estimates. The normal model is constrained to have a kurtosis of three, while both OLS and the mixture model "agree" that kurtosis is likely to be much higher—although they diverge regarding how much higher. Here is clear evidence that the default Bayesian approach, with a normal prior on alpha, over-shrinks the estimated alphas in this case. Fitting to a normal distribution assumes away large outliers, while a more flexible model can learn from them.

4.3 Estimated cross-sectional distribution of alpha

The moments of the distribution of alpha are revelatory, but the distribution itself completes the story. The overall distribution of skill is shown in Figure 7. This view makes the distribution appear normal, but the pivotal deviations from normality are in the tails.

Understanding the tail behavior requires an examination of the individual mixture components. Both components, scaled to reflect their relative weights, are shown in Figure 8. The two components differ dramatically in terms of relative weighting and dispersion. Component 1 represents approximately 85% of all managers and requires that alpha is fairly close to zero. Component 2, representing the remaining 15% of fund managers, allows for extreme alphas in both directions.

Figure 9 illustrates this extreme tail behavior. Although component 1 assigns approximately zero probability to alphas larger than 0.6% monthly, component 2 assigns small positive probabilities to alphas as high as 3% monthly. Relative to a normal model of alpha, which completely discounts the possibility of "superstar" fund managers capable of dramatic outperformance, the mixture-of-normals allows the upper echelon of managers to exhibit superior alphas.

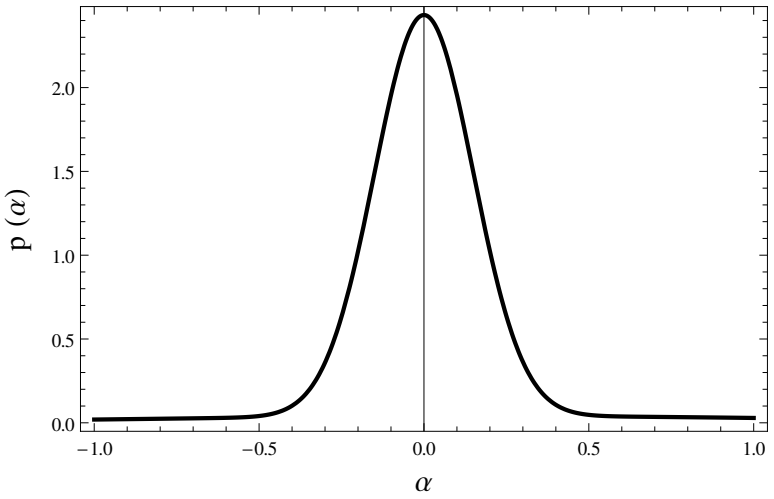

Figure 7: Overall estimated cross-sectional distribution of alpha from 2004, summing the individual components shown in Figure 8. Alpha is monthly and measured in percent. The distribution appears generally normal in this zoomed-out view, since the weight on mixture component 1 is much larger than the weight on mixture component 2. A closer examination of the tails, depicted in Figure 9, reveals pronounced non-normality.

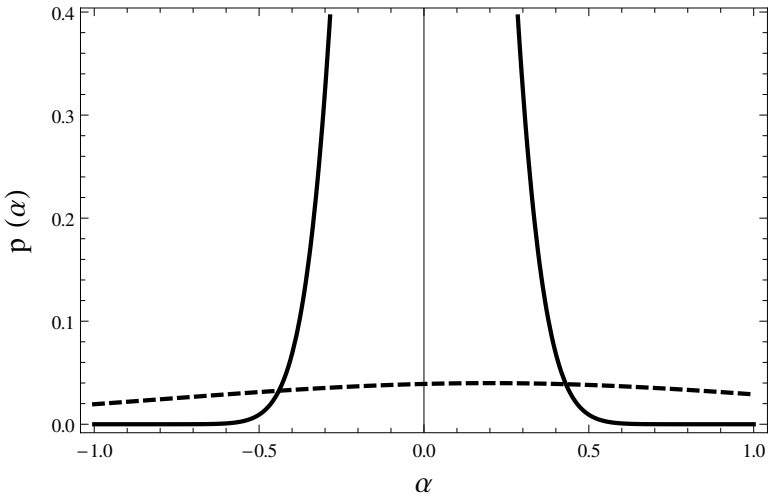

Figure 8: Estimated mixture components from 2004, chosen to be representative. Component 1 is depicted by the solid line, while the higher-mean component 2 is depicted by the dashed line. Alpha is monthly and measured in percent. Component 2 has a slightly higher mean and a much higher variance than component 1.

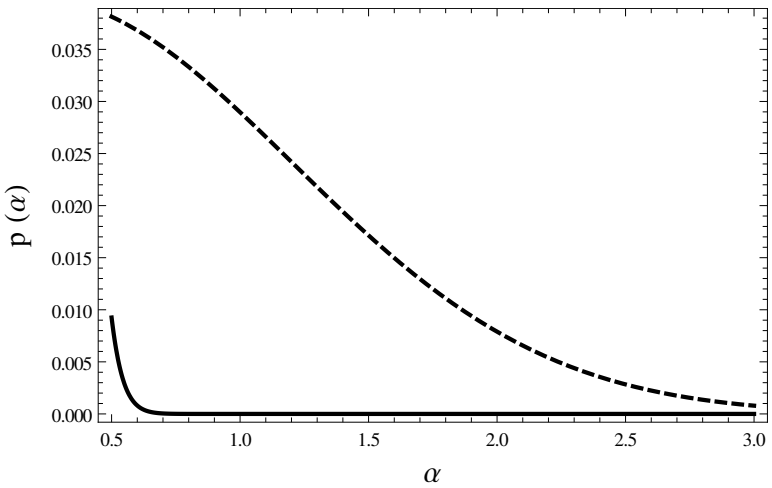

Figure 9: Tail behavior in the cross-sectional distribution of skill from 2004—a zoomed-in view of the distributions in Figure 8. Component 1 is depicted by the solid line, while the higher-mean component 2 is depicted by the dashed line. Alpha is monthly and measured in percent. The probability of observing extremely high alphas is several orders of magnitude higher than it would be in a model that imposed normality on alpha, a result of allowing for the second component.

4.4 Decile portfolio performance

Decile portfolios are formed using all three ranking methods and their abnormal returns are estimated using OLS. The power of the mixture-of-normals prior is made clear in this test. Numerical results are given in Table 1 and shown graphically in Figure 10.

Even OLS reveals substantial short-term predictability. The worst funds (deciles 1 and 2) continue to under-perform and the best funds (deciles 9 and 10) continue to outperform. Carhart documented persistent poor performance over a longer time frame, attributing it to high fees. Bollen and Busse, among others, have more recently documented persistent outperformance in the short term using OLS. Thus, the results from the OLS decile portfolios are just as expected.

Results using rankings from the normal hierarchical model are telling—they are substantially weaker than those from OLS. The performance of the bottom decile portfolio is *better* than the OLS bottom decile portfolio, and the performance of the top decile is *worse*. The normal prior does a worse job of sorting

Out-of-Sample Monthly Alphas of Decile Portfolios			
Decile	OLS	Normal Prior	Mixture Prior
1	-0.21*	-0.14*	-0.51*
	(0.05)	(0.04)	(0.03)
2	-0.12*	-0.16*	-0.37*
	(0.04)	(0.04)	(0.04)
3	-0.07	-0.09*	-0.34*
	(0.04)	(0.04)	(0.04)
4	-0.05	-0.07	-0.18*
	(0.03)	(0.04)	(0.04)
5	-0.03	-0.02	-0.10*
	(0.03)	(0.04)	(0.04)
6	0.00	0.04	0.25*
	(0.03)	(0.08)	(0.05)
7	0.03	0.08	0.25*
	(0.03)	(0.05)	(0.05)
8	0.06	0.06	0.32*
	(0.04)	(0.04)	(0.04)
9	0.10*	0.15*	0.36*
	(0.04)	(0.05)	(0.04)
10	0.27*	0.15*	0.45*
	(0.07)	(0.04)	(0.04)

Table 1: Out-of-sample monthly alphas of decile portfolio by ranking method, 1961-2010. At the end of each month t, alphas are estimated for each mutual fund using each method and the past 24 monthly returns. Funds are ranked, and equally-weighted decile portfolios are formed and held during month $t + 2$. Alphas for these portfolios are estimated using OLS, with bootstrapped standard errors.

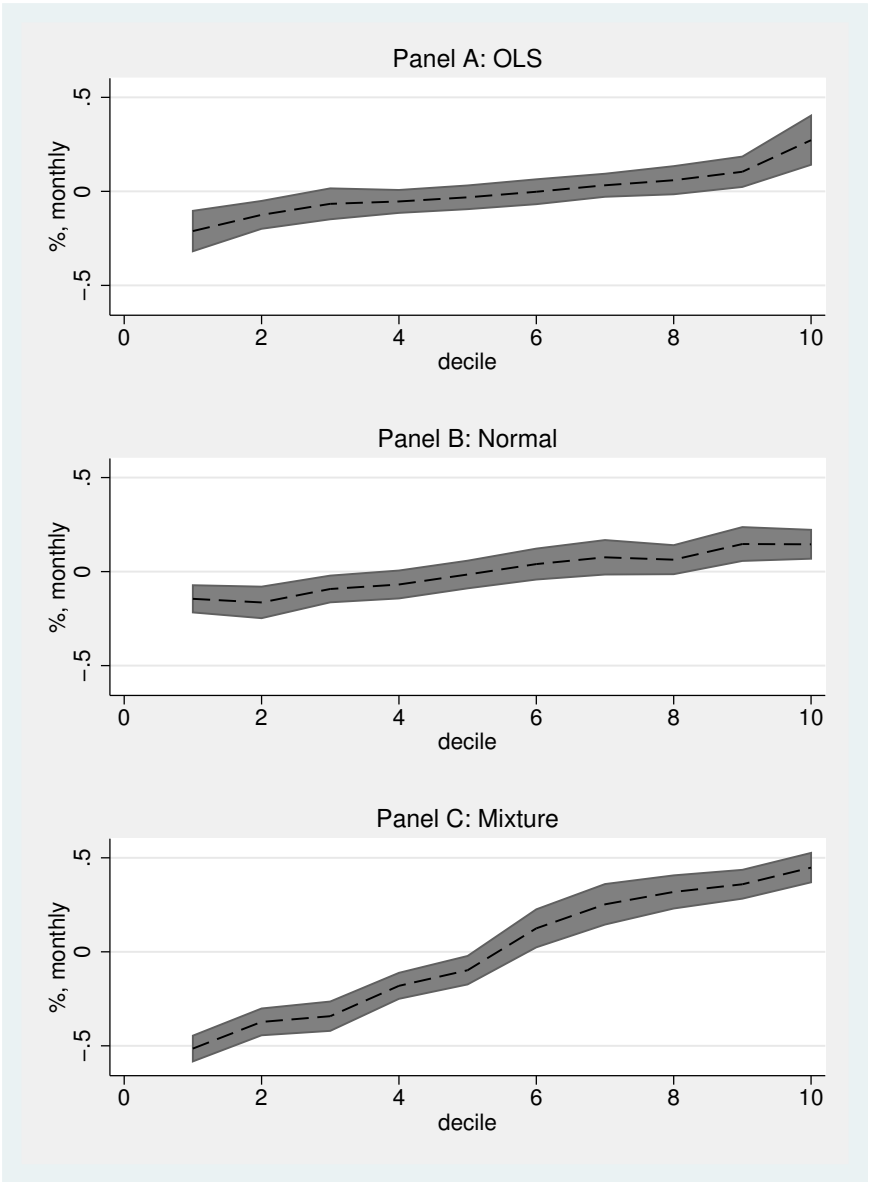

Figure 10: Out-of-sample OLS alphas and 95% confidence intervals for decile portfolios, by ranking method, 1961-2010. Data are taken from Table 1. Panel A shows decile portfolios formed from funds ranked on their OLS alphas. Panel B shows decile portfolios formed from funds ranked on their hierarchical Bayes alphas with a normal prior. Panel C shows decile portfolios formed from funds ranked on their hierarchical Bayes alphas with a mixture-of-normals prior.

funds. There is valuable information in the extreme alpha estimates that the normal model is forced to disregard.

The results from the mixture hierarchical model are striking: all portfolios have statistically significant alphas, decile 1 under-performs OLS decile 1 by 30 basis points and decile 10 outperforms OLS decile 10 by 18 basis points. When annualized, these translate to 360 basis points and 216 basis points, respectively. The differences relative to the normal hierarchical estimates are even starker. In absolute terms, the mixture decile 1 portfolio posts an annual alpha of -6.1% and the mixture decile 10 portfolio posts an annual alpha of +5.4% during the fifty-year period 1961-2010. The mixture bottom decile portfolio is more than twice as bad as the OLS bottom decile portfolio, while the mixture top decile portfolio is nearly twice as good as the OLS top decile portfolio.

4.5 Fund-by-fund alpha correlations

The mixture model generates superior portfolios of funds; its fund-level estimates are better as well. Using the same two-year estimation windows, we can compare the estimated fund alphas for the interval $[t-23, t]$ to the estimated alphas for the non-overlapping interval $[t+2, t+25]$.[31] Let the vector of estimated fund alphas generated by a particular model k on the interval $[t_a, t_b]$ be denoted by $\widehat{\alpha}^{k}_{[t_a,t_b]}$, and define the correlation of the alphas on the first interval with the alphas on the second interval to be

$$\rho_t^{k_1,k_2} = correlation\left(\widehat{\alpha}^{k_1}_{[t-23,t]}, \widehat{\alpha}^{k_2}_{[t+2,t+25]}\right). \tag{3}$$

Funds that do not appear in both intervals are dropped before the correlation is computed. I expect that the typical fund used in this part of the analysis is consequently better than average. Even if the level of alpha is biased upwards, this does not imply that the serial correlation is biased.

When comparing the performance of two models, k_1 and k_2, in predicting the estimates generated by model k_3, it is also useful to define:

$${}^{k_1,k_3}\Delta_t^{k_2,k_3} = \rho_t^{k_1,k_3} - \rho_t^{k_2,k_3}. \tag{4}$$

[31] By evaluating the estimation methods in this way, I assume that the short-term predictability in fund alphas is not completely offset by any un-modeled longer-term negative correlation.

Panel A: Serial Correlation of Estimated Fund Alphas, by Method				
$\rho_t^{OLS,OLS}$	$\rho_t^{Normal,Normal}$	$\rho_t^{Mixture,Mixture}$	$\rho_t^{Normal,OLS}$	$\rho_t^{Mixture,OLS}$
0.141*	0.116*	0.171*	0.104*	0.159*
(0.006)	(0.005)	(0.005)	(0.004)	(0.005)

Panel B: Differences in Serial Correlations of Estimated Fund Alphas		
${}_{Normal,Normal}\Delta_t^{OLS,OLS}$	${}_{Mixture,Mixture}\Delta_t^{OLS,OLS}$	${}_{Mixture,OLS}\Delta_t^{OLS,OLS}$
-0.025*	0.031*	0.019*
(0.006)	(0.004)	(0.004)

Table 2: In Panel A, serial correlations of estimated fund alphas are computed for rolling, non-overlapping, two-year intervals, 1961-2010. In the last two columns, the hierarchical models are used to predict OLS alphas. In Panel B, differences in these correlations are reported.

Results are shown in Table 2.[32] From the first three columns in Panel A, we see that all alphas are positively serially correlated, consistent with the predictability suggested by the performance of the decile portfolios. The last two columns show the correlations of alphas generated by the hierarchical models in the first period with the OLS alphas from the second period. These are positive as well, and indicate that the positive serial correlations are not artifacts of the Bayesian procedure.

Panel B of Table 2 shows the differences in the correlations. In the second column, we can see that the mixture model's estimates are indeed more strongly serially correlated than the OLS estimates. Most critically, the third column shows that the mixture model predicts the OLS estimates better than OLS itself.

5 Discussion and Implications

The persistent performance of the decile portfolios is impossible to fully reconcile with the rational model of Berk and Green (2004). In their model, managers can possess investing skill but do not exhibit persistent non-zero alphas due to convexities in the costs of managing their funds. As funds grow in size, positive alpha becomes increasingly difficult to generate. In equilibrium all funds have

[32]The same analysis was also performed using rank correlations with essentially identical results.

zero alphas in expectation, but vary in size according to their managers' varying levels of skill. This equilibrium requires investors to evaluate managers in the context of a *known* cross-sectional distribution of alpha, and to dynamically re-allocate capital across funds in each period.

In contrast to the equilibrium predictions of Berk and Green, I find markedly persistent performance in the short-term. At least one of their assumptions must not hold. Costs may not be convex. Reuter and Zitzewitz (2010) use fund flows associated with discrete changes in Morningstar fund ratings to show that their effect on fund performance is inconsistent with the Berk and Green hypothesis. There may be frictions associated with capital re-allocation that prevent investors from responding immediately to new estimates of manager skill.

The most indisputable deviation from the Berk and Green world, however, is that investors do not know the true cross-sectional distribution of alpha. In reality, as in the model presented here, this distribution and the individual fund alphas must be estimated simultaneously. Such an analysis requires MCMC methods—somewhat beyond the capabilities of the typical mutual fund investor, or even the typical financial advisor. The sheer difficulty of the evaluation is sufficient to cause slow learning, even among competent and vigilant investors. Even if all of the other assumptions of Berk and Green hold, this slow learning implies the short-term persistence of alphas that I document.

Flow predictability regressions presented below show that investors do respond to the information contained in the Bayesian mixture alphas. Whatever heuristics they employ to evaluate funds must incorporate the intuition of my model if not its details.

5.1 Convexity of fund flows

Fund flows have a convex relationship with past fund performance. Outflows in response to poor performance are less pronounced than the inflows in response to good performance.[33] Lynch and Musto (2003) offer an explanation: if poor performance makes changes in investment strategy more likely, then poor performance is less informative than good performance regarding future returns. Berk and Green (2004) also predict a convex flow-performance relationship—a consequence of the convex costs of management in their model.[34]

[33] See, e.g., Sirri and Tufano (1998) and Chevalier and Ellison (1997).
[34] See Figure 3 in Berk and Green (2004).

Another possibility is that the optimal updating rule for posterior skill estimates is itself convex. While a normal prior (along with a normal likelihood) gives rise to a linear updating rule, the mixture prior proposed here generates an updating rule that is nonlinear for observed returns that are not extremely far from zero, and convex over the region of greatest interest.[35]

Assume that the true cross-sectional distribution of skill is a mixture-of-normals, as in (2). For comparison, I also consider a normal prior on skill. To obtain the best approximation, the mean and variance of the normal prior are set to match the mean and variance of the mixture distribution:

$$\mu = (1 - w_2) \cdot \mu_1 + w_2 \cdot \mu_2 \tag{5}$$

and

$$V = (1 - w_2) \cdot V_1 + w_2 \cdot V_2 + w_2 \cdot (1 - w_2) \cdot (\mu_1 - \mu_2)^2 . \tag{6}$$

Assume that fund returns are simply managerial skill plus Gaussian disturbances: a simplified version of the model in (1) with no risk factors.

After observing a single return R, investors update their estimates of managerial skill. The posterior expectation of skill under the normal prior is:

$$E_{normal}[\alpha] = \left(\frac{\mu}{V} + \frac{R}{V_\varepsilon}\right) \cdot \left(\frac{1}{V} + \frac{1}{V_\varepsilon}\right)^{-1} . \tag{7}$$

The posterior expectation of skill under the mixture prior is:

$$E_{mixture}[\alpha] = \sum_{i=1}^{N} \widetilde{w}_i \cdot \left(\frac{\mu_i}{V_i} + \frac{R}{V_\varepsilon}\right) \cdot \left(\frac{1}{V_i} + \frac{1}{V_\varepsilon}\right)^{-1} , \tag{8}$$

where

$$\widetilde{w}_i = \frac{w_i}{\sqrt{V_i + V_\varepsilon}} \cdot \exp\left[-\frac{1}{2} \cdot \frac{(R - \mu_i)^2}{V_i + V_\varepsilon}\right] . \tag{9}$$

To make this more concrete, the posterior expectations, (7) and (8), of alpha corresponding to each of these priors after a single return R are shown in Figure 11. Here we can easily see how sensitive the posterior beliefs are to the investor's prior, even when both possible priors have matching first and second moments.

[35] The details depend on the model parameters, but the intuition is that extreme observations are almost certainly drawn from the mixture component with the highest variance. Since this component is normal, the updating rule has a linear asymptote.

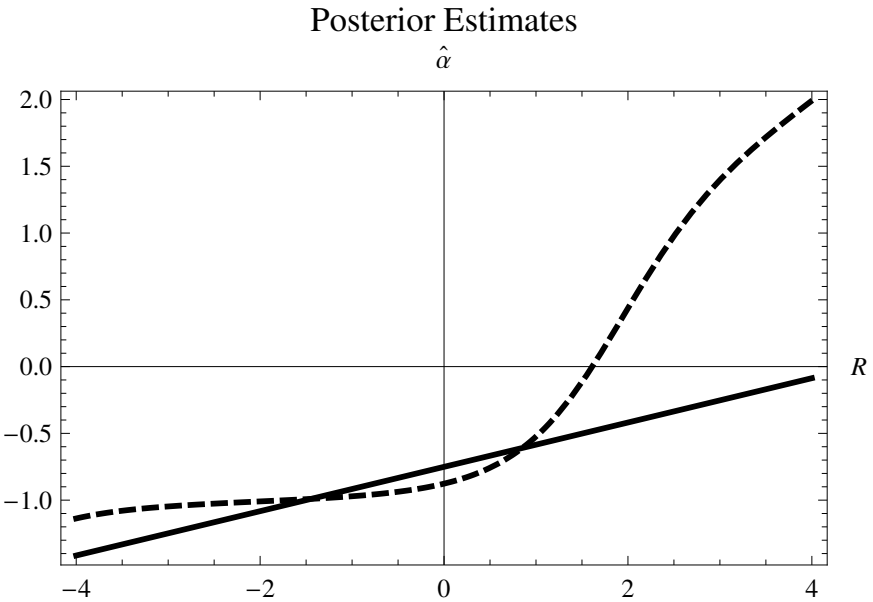

Figure 11: Updating rules for managerial skill with a mixture-of-normals prior (dashed line) and a normal prior with matching first and second moments (solid line). R is a monthly return measured in percent. The parameters of the mixture distribution are: $\mu_1 = -1$, $\mu_2 = 0$, $V_1 = 0.01$, $V_2 = 1$, and $w_2 = 0.1$. Plugging these into (5) and (6), we get the following parameters for the "matching" normal distribution: $\mu = -0.9$ and $V = 0.2$. The updating rule for the normal prior is linear. The updating rule for the mixture prior has a linear asymptote but is nonlinear for non-extreme returns, and is generally convex for monthly returns between -4 and +4 percent.

The updating rule for the mixture prior is generally convex for non-extreme monthly returns. If the relationship between skill estimates and observed returns is convex, and the relationship between fund flows and skill estimates is not too concave, we should expect to observe convex flows.

The mixture-of-normals prior is not special in this regard. Any non-normal prior will generate a nonlinear updating rule. A more flexible prior than the two-component mixture should be able to match the observed flow-performance relationship even more closely.

5.2 Predictability of fund flows

This paper describes a method of accurately estimating mutual fund alphas that fully exploits the information contained in the cross-section of funds. It is natural to ask what resemblance this bears to the methods investors actually use to evaluate funds. Although their methods are not observable, their conclusions are revealed by the fund flow data.

Results for a series of fund flow predictability regressions are shown in Table 3. The dependent variable is the flow ratio, defined as the net fund flow divided by the fund's total assets under management in the previous month. Regression (1) is patterned after the baseline regression model of Sirri and Tufano (1998). This shows that investors are highly responsive to past returns, although they are more responsive to high returns than to low returns—the effect of being in the top quintile of past returns is positive and significant even after controlling for average past returns, $\overline{R_t - R_{mkt}}$. This convexity is a robust feature of mutual fund flows.

Regression (2) adds the lagged OLS alpha as a predictor, while regression (3) adds the lagged Bayesian mixture alpha. Regression (4) includes them both. Regression (2) shows that investors are indeed responsive to the information in OLS alphas. This information does not subsume the effect of past returns, but the coefficient on $\overline{R_t - R_{mkt}}$ is somewhat closer to zero. Regression (3) shows that investors are also responsive to the information in Bayesian mixture alphas.

In regression (4), we see that both the OLS alphas and the Bayesian mixture alphas predict fund flows in a statistically significant manner. Although the point estimate of the coefficient on the Bayesian mixture alpha is higher than that of the OLS alpha, 0.33 versus 0.11, this difference is not significant. The key finding is that the information contained in the Bayesian mixture alphas, incorporating knowledge of the skewness and excess kurtosis in the cross-sectional

distribution of alpha, influences investors' capital allocations, beyond the information contained in OLS alphas and directly in past returns. Although not statistically significant, it also appears that including the Bayesian mixture alphas, in regressions (3) and (4), decreases the convexity of fund flows, consistent with the theoretical predictions of section 5.1.

5.3 Smart money effect

A general finding in the literature is that mutual fund flows help to predict mutual fund performance.[36] The "smart money" effect refers to the *interpretation* that investors are actually able to anticipate future fund performance and allocate capital accordingly, but this is not the only interpretation.[37] One way to assess the "smartness" of investors is to examine the incremental impact that estimated alphas have for fund return prediction, controlling for fund flows.

Table 4 presents estimates from a series of regressions seeking to explain fund returns in terms of past flows and returns. Regression (3) shows that past flows do predict fund returns; adding past returns in regression (4) weakens the predictive power of flows. As expected, regressions (1) and (2) show that including lagged OLS and Bayesian mixture alphas does improve fund return prediction.

The most interesting results are from regression (5), which includes past flows and both OLS and Bayesian mixture alphas. The coefficient on lagged flow ratio remains significant—flows are predictive beyond the information contained in estimated alphas. Had the coefficients on lagged flows all been insignificant, this would have implied that alphas predict fund returns because both are driven by managerial skill: flows are truly smart money. The failure of this test does not imply that investors are not smart, however, since this is a joint test of the accuracy of the estimated alphas and the smart money effect. But the strength of the return predictability documented in section 4.4 and the relative weakness of the flow predictability documented in section 5.2 suggest a mechanism other than "smart money" for the predictability of fund returns by fund flows.

Also noteworthy is the coefficient on lagged OLS alphas in regression (5): it is negative and significant. Controlling for Bayesian mixture alphas, future fund returns are higher when OLS alphas are lower, and vice versa. This is a clear

[36] See, e.g., Gruber (1996) and Zheng (1999).

[37] For example, Lou (2011) presents evidence that fund return predictability is the consequence of flow-induced price pressure on underlying stock holdings.

Fund Flow Predictability, Fama-MacBeth Regressions				
Flow Ratio	(1)	(2)	(3)	(4)
Top Return Quintile	0.62*	0.50*	0.38*	0.43*
	(0.17)	(0.16)	(0.16)	(0.16)
Bottom Return Quintile	0.10	0.21	0.26	0.39
	(0.17)	(0.18)	(0.17)	(0.27)
$\overline{R_t - R_{mkt}}$ (%)	2.45*	1.99*	2.14*	2.16*
	(0.24)	(0.27)	(0.24)	(0.35)
$\sigma(R_t - R_{mkt})$ (%)	-0.04	-0.04	-0.04	-0.06
	(0.03)	(0.04)	(0.03)	(0.04)
Lagged Expense Ratio (%)	-0.15	-0.13	-0.11	-0.09
	(0.18)	(0.17)	(0.18)	(0.18)
Total Flow into Fund Category ($M)	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Lagged Log(Assets Under Mgmt. ($M))	0.36*	-0.38*	-0.38*	-0.37*
	(0.05)	(0.05)	(0.05)	(0.05)
Lagged OLS $\widehat{\alpha}$ (%)		0.13*		0.11*
		(0.02)		(0.05)
Lagged Bayesian Mixture $\widehat{\alpha}$ (%)			0.65*	0.33*
			(0.06)	(0.14)
Constant	0.95	1.35*	1.67*	1.40*
	(0.56)	(0.58)	(0.60)	(0.64)
R^2	0.09	0.09	0.09	0.10

Table 3: Results for Fama-MacBeth regressions predicting flow ratio, defined as this month's net flows divided by last month's total assets under management; 1963-2010. Regression (1) is based directly on the baseline regression in Sirri and Tufano (1998), and shows both the responsiveness of flows to raw past performance, as well as the convexity of fund flows. (2) adds OLS alphas as a predictor, while (3) adds the Bayesian mixture alphas. (4) includes both alphas as predictors. Although overall R^2 does not change dramatically, the strongly significant coefficients on the Bayesian mixture alphas indicate that they contain information not contained in the other predictors, and that investors are sensitive to this information.

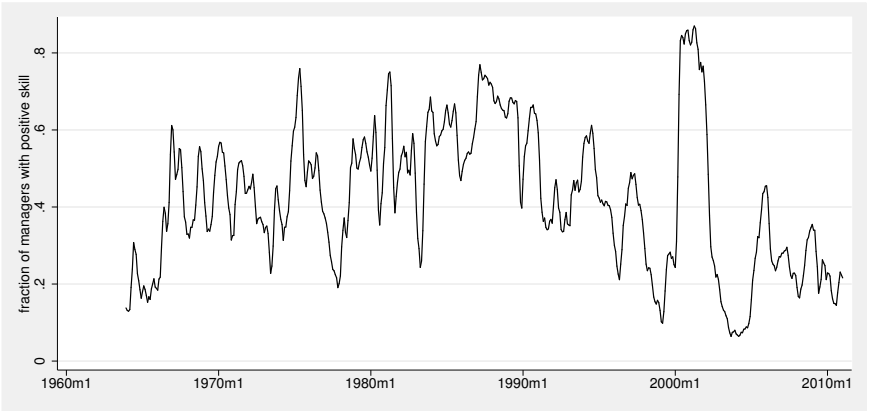

Figure 12: Estimated fraction of managers with positive alpha using the Bayesian mixture model.

illustration of how poorly OLS distinguishes signal from noise in the context of investment performance evaluation.

5.4 Value of active management

Even if some fund managers possess skill, investors do not benefit unless managers generate positive alpha *after* management fees and trading costs. Since all alphas in this study are estimated using net fund returns to the investor, we can directly evaluate the prevalence of skill over time. Panel A of Figure 6 shows the overall mean of the prevailing distribution of skill—it does not diverge substantially from the implied OLS mean—which seems to lend support to the Berk and Green hypothesis: the typical fund has an alpha quite close to zero.

Looking at the mean alone is misleading, however. Figure 7 shows that the prevailing distribution of skill also has fat tails and positive skewness. Instead of simply examining the mean, it is more informative to compute the fraction of managers with positive alpha directly from the estimated parameters of the prevailing cross-sectional distribution. The results are shown in Figure 12.

Barras et al. (2010) classify funds as either unskilled, zero-alpha, or skilled, and find that 75% of funds are zero-alpha while less than 5% are skilled.[38] This paper uses fund-level data to estimate the underlying distribution—sharply separating the negative and positive regions of skill. There is no need for a

[38]See Figure 4 in Barras et al. (2010) for a comparison to Figure 12. Their estimated fraction of skill is both much lower and smoother than the estimates presented here.

"Smart Money" Effect, Fama-MacBeth Regressions					
Monthly Fund Return (%)	(1)	(2)	(3)	(4)	(5)
Lagged OLS $\widehat{\alpha}$ (%)	0.19*				-0.15*
	(0.01)				(0.01)
Lagged Bayesian Mixture $\widehat{\alpha}$ (%)		0.11*			0.11*
		(0.01)			(0.01)
$\overline{R_t - R_{mkt}}$ (%)				0.24*	0.17*
				(0.01)	(0.01)
Flow Ratio in month t_{-1}			0.002*	0.001*	0.001*
			(0.00)	(0.00)	(0.00)
Flow Ratio in month t_{-2}			0.001*	0.00	0.00
			(0.00)	(0.00)	(0.00)
Flow Ratio in month t_{-3}			0.00	0.00	0.00
			(0.00)	(0.00)	(0.00)
Constant	0.45*	0.53*	0.49*	0.38*	0.49*
	(0.09)	(0.09)	(0.01)	(0.01)	(0.01)
R^2	0.02	0.02	0.01	0.02	0.02

Table 4: Results for Fama-MacBeth regressions predicting monthly fund returns; 1963-2010. (1) and (2) show the predictive power of OLS and Bayesian mixture alphas, respectively. (3) shows that flows do help to predict fund performance. They continue to predict performance when (4) past returns are included, and (5) when OLS and Bayesian mixture alphas are included. This evidence, combined with the results from sections 4.4 and 5.2, suggests that "smart money" is *not* the principal mechanism explaining why fund flows predict fund returns.

zero-alpha category unless there is a point mass at zero in the underlying skill distribution. Kosowski et al. (2006) use a bootstrap approach to compute a continuous measure of skill, and find that approximately 5% of managers show evidence of skill net of costs.

The most recent estimates in Figure 12 suggest that approximately 20% of managers have skill, although the fraction has been much higher in the past. The volatility of the estimates over time and the magnitude of the maximum estimates rightly suggest that the distribution has not been perfectly estimated. Nevertheless, the decile portfolios performed well: an imperfectly estimated distribution still greatly assists in the proper ordering of funds.

Even if sizeable fractions of managers possess skill after costs and fees, we would like to know if the active mutual fund industry is providing value in aggregate. Fama and French (2010) employ a bootstrap methodology and find that the aggregate portfolio of mutual funds has a negative net alpha, although there is evidence of superior skill in the extreme upper tail of the cross-sectional distribution of skill.

This study used only one representative share class from each fund. To construct a comparable net industry-wide alpha, we would need to take into account all share classes of each fund, their expense ratios, and assets under management, following French (2008). This has not been done, so the important question of whether the industry provides value to investors remains open. Even still, it is safe to speculate that such an analysis would have conclusions somewhat more favorable to the mutual fund industry than French, given the higher fraction of skilled managers found here.

6 Conclusion

The model explored in this paper is deliberately kept simple to focus attention on the sensitivity of alpha estimates to the assumptions made regarding the cross-section of skill. Existing methodologies ignore the conclusions of the "superstars" literature: the returns to investing skill are likely to be positively skewed and fat-tailed. An optimal approach to performance evaluation should take this into account and explicitly model the cross-sectional distribution of alpha accordingly.

The hierarchical model I propose is such an approach. While it does not impose skewness or excess kurtosis, it "learns" from the data that such features

are indeed present. Despite the simplicity of the mixture-of-normals distribution, the resulting estimates for fund alphas exhibit strong predictability across the entire population of funds—it is not merely the very best and the very worst funds whose performance persists.

In addition to accurately measuring fund alphas, this model also provides a possible explanation for the convexity of fund flows. Any non-normal prior on the distribution of alpha will lead to a nonlinear updating rule: high observed returns can exert much greater influence than low observed returns on the posterior mean of a given fund's alpha. The convexity of fund flows could thus be a rational response to the non-normality of alpha.

A MCMC Algorithm

This section provides details about the Bayesian inference procedure, including the priors on all parameters and a step-by-step summary of the Markov chain Monte Carlo (MCMC) algorithm used to generate samples from the joint posterior distribution of the parameters.

A.1 Parameters to be estimated

$\left(\mu_i, V_i, w_i, \alpha^j, \beta_k^j, V_\varepsilon^j\right)$ for $i = 1, ..., N$; $j = 1, ..., J$; $k = 1, ..., K$; where N is the number of mixture components in (2), J is the number of funds, and K is the number of risk factors in (1).

A.2 Latent indicator

To simplify the estimation procedure, I introduce a latent $J \times 1$ vector z, where each element is Bernoulli distributed,[39]

$$z_j \sim Bernoulli\left(w_2\right). \tag{10}$$

The *conditional* (on z_j) prior on skill is now normal, instead of the original mixture of normals in (2):

$$\alpha^j \sim N\left(\mu_{z_j+1}, V_{z_j+1}\right). \tag{11}$$

A.3 Priors

All priors are proper but diffuse. They are scaled to correspond to the monthly percent returns used in the analysis. For example, an annual alpha of 12% would appear here as 1.

$$\mu_1 \sim N\left(a_\mu^{prior} = 0, A_\mu^{prior} = 10\right) \tag{12}$$

$$\mu_2 \sim N\left(a_\mu^{prior} = 0, A_\mu^{prior} = 10\right) \cdot \mathbf{1}_{\mu_2 \geq \mu_1} \tag{13}$$

$$V_i \sim Gamma\left(1, 10\right) \tag{14}$$

[39] In the general case with more than two mixture components, z_j is categorically distributed with vector parameter w.

$$w \sim Dirichlet\left(\pi^{prior} = \begin{pmatrix}1\\1\end{pmatrix}\right) \tag{15}$$

$$\beta^j_{mkt} \sim N\left(a^{prior}_{\beta_{mkt}} = 1, A^{prior}_{\beta} = 5\right) \tag{16}$$

$$\beta^j_k \sim N\left(a^{prior}_{\beta} = 0, A^{prior}_{\beta} = 5\right) \tag{17}$$

$$V^j_{\varepsilon} \sim Gamma\,(1, 5) \tag{18}$$

A.4 Sampling algorithm

1. Draw z conditional on w, μ_i, V_i, α.

   - Compute (vector) probabilities ω conditional on each mixture component:

   $$\omega_i = \frac{1}{\sqrt{V_i}} \cdot \phi\left(\frac{\alpha - \mu_i}{\sqrt{V_i}}\right). \tag{19}$$

   - Compute (vector) unconditional posterior probability:

   $$p_2 = \frac{\omega_2}{\omega_1 + \omega_2}. \tag{20}$$

   - Draw $z \sim Bernoulli\,(p_2)$.

2. Partition α conditional on z.

   - Partition α into two vectors, α_1 and α_2, corresponding to those elements of α from each of the mixture components, and let $N_i = \text{rows}(\alpha_i)$.

3. Draw V_i conditional on z, μ_i, α_i.

   - Because of the gamma prior, the posterior on V_i is not a known distribution and is sampled using a random walk Metropolis-Hastings draw.

4. Draw μ_i conditional on V_i, α_i.

- Compute posterior hyper-parameters for each mixture component:

$$A_{\mu,i}^{posterior} = \left(\frac{1}{A_{\mu}^{prior}} + \frac{N_i}{V_i} \right)^{-1} \tag{21}$$

$$\frac{a_{\mu,i}^{posterior}}{A_{\mu,i}^{posterior}} = \frac{a_{\mu}^{prior}}{A_{\mu}^{prior}} + \frac{\sum_{n=1}^{N_i} \alpha_{i,n}}{V_i} \tag{22}$$

- Draw $\mu_i \sim N\left(a_{\mu,i}^{posterior}, A_{\mu,i}^{posterior}\right)$.

5. Impose identifying constraint $\mu_2 \geq \mu_1$.

   - If the most recent draws of μ_i violate this constraint, revert to the previous values.

6. Draw w conditional on z.

   - Compute posterior vector hyper-parameter:

$$\pi^{posterior} = \pi^{prior} + \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}. \tag{23}$$

   - Draw $w \sim Dirichlet\left(\pi^{posterior}\right)$.

7. Draw V_{ε}^{j} conditional on $R^j, F, \alpha^j, \beta^j$.

   - Because of the gamma prior, the posterior on V_{ε}^{j} is not a known distribution and is sampled using a random walk Metropolis-Hastings draw.

8. Draw β^j conditional on $R^j, F, \alpha^j, V_{\varepsilon}^{j}$

   - Compute posterior hyper-parameters for each β_k^j:

$$A_{\beta,j,k}^{posterior} = \left(\frac{1}{A_{\beta,k}^{prior}} + \frac{F_k^{'} \cdot F_k}{V_{\varepsilon}^{j}} \right)^{-1} \tag{24}$$

$$\frac{a_{\beta,j,k}^{posterior}}{A_{\beta,j,k}^{posterior}} = \frac{a_{\beta,k}^{prior}}{A_{\beta,k}^{prior}} + \frac{F_k^{'} \cdot R^j - \alpha^j \cdot \sum_{t=1}^{T} F_{k,t} - F_k^{'} \cdot F_{-k} \cdot \beta_{-k}}{V_{\varepsilon}^{j}} \tag{25}$$

- Draw $\beta_k^j \sim N\left(a_{\beta,j,k}^{posterior}, A_{\beta,j,k}^{posterior}\right)$.

9. Draw α^j conditional on $z^j, \mu_{z^j+1}, V_{z^j+1}, \beta^j, V_\varepsilon^j, F, R^j$.

   - Compute posterior hyper-parameters for each α^j:

$$A_{\alpha,j}^{posterior} = \left(\frac{1}{V_{z^j+1}} + \frac{T}{V_\varepsilon^j}\right)^{-1} \tag{26}$$

$$\frac{a_{\alpha,j}^{posterior}}{A_{\alpha,j}^{posterior}} = \frac{\mu_{z^j+1}}{V_{z^j+1}} + \frac{\sum_{t=1}^{T}\left(R_t^j - F_t \cdot \beta^j\right)}{V_\varepsilon^j} \tag{27}$$

   - Draw $\alpha^j \sim N\left(a_{\alpha,j}^{posterior}, A_{\alpha,j}^{posterior}\right)$.

References

[1] 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.

[2] Barras, Laurent, Oliver Scaillet, and Russ Wermers, 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, *Journal of Finance* 65, 179-216.

[3] Berk, Jonathan B. and Richard C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, *Journal of Political Economy* 112, 1269-1295.

[4] Bollen, Nicolas P. B. and Jeffrey A. Busse, 2004, Short-Term Persistence in Mutual Fund Performance, *Review of Financial Studies* 18, 569-597.

[5] Brown, David P. and Youchang Wu, Mutual Fund Families and Performance Evaluation (February 15, 2011). Available at SSRN: http://ssrn.com/abstract=1690539

[6] Busse, Jeffrey A. and Paul J. Irvine, 2006, Bayesian Alphas and Mutual Fund Persistence, *Journal of Finance* 61, 2251-2288.

[7] Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.

[8] Chevalier, Judith and Glenn Ellison, 1997, Risk Taking by Mutual Funds as a Response to Incentives, *Journal of Political Economy* 105, 1167-1200.

[9] 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.

[10] Chung, Kee H. and Ramond A. K. Cox, 1990, Patterns of Productivity in the Finance Literature: A Study of the Bibliometric Distributions, *Journal of Finance* 45, 301-309.

[11] Cohen, Lauren, Andrea Frazzini, and Christopher Malloy, 2008, The Small World of Investing: Board Connections and Mutual Fund Returns, *Journal of Political Economy* 116, 951-979.

[12] Coval, Joshua D. and Tobias J. Moskowitz, 2001, The Geography of Investment: Informed Trading and Asset Prices, *Journal of Political Economy* 109, 811-841.

[13] Efron, Bradley and Carl Morris, 1977, Stein's Paradox in Statistics, *Scientific American* 236, 119-127.

[14] Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns on bonds and stocks, *Journal of Financial Economics* 33, 3-53.

[15] Fama, Eugene F. and Kenneth R. French, 2010, Luck versus Skill in the Cross-Section of Mutual Fund Returns, *Journal of Finance* 65, 1915-1947.

[16] Fernández, Carmen and Mark F. J. Steel, 1998, On Bayesian Modeling of Fat Tails and Skewness, *Journal of the American Statistical Association* 93, 359-371.

[17] French, Kenneth R., 2008, Presidential Address: The Cost of Active Investing, *Journal of Finance* 63, 1537-1573.

[18] Gabaix, Xavier and Augustin Landier, 2008, Why Has CEO Pay Increased So Much?, *Quarterly Journal of Economics* 123, 49-100.

[19] Gelman, Andrew, 2006, Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis* 1, 515-533.

[20] Gruber, Martin J., 1996, The Growth in Actively Managed Funds, *Journal of Finance* 51, 783-810.

[21] Hausman, Jerry A. and Gregory K. Leonard, 1997, Superstars in the National Basketball Association: Economic Value and Policy, *Journal of Labor Economics* 15, 586-624.

[22] Jones, Christopher S. and Jay Shanken, 2005, Mutual fund performance with learning across funds, *Journal of Financial Economics* 78, 507-552.

[23] Kosowski, Robert, Allan Timmerman, Russ Wermers, and Hal White, 2006, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, *Journal of Finance* 61, 2551-2595.

[24] Krueger, Alan B., 2005, The Economics of Real Superstars: The Market for Rock Concerts in the Material World, *Journal of Labor Economics* 23, 1-30.

[25] Leamer, Edward E., 1978, *Specification Searches.*

[26] Levenshtein, V. I., 1966, Binary Codes Capable of Correcting Deletions, Insertions, and Reversals, *Soviet Physics Doklady* 10, 707-710.

[27] Lindley, D. V., 1972, *Bayesian Statistics, A Review.*

[28] Lou, Dong, 2011, A Flow-Based Explanation for Return Predictability, working paper.

[29] Pástor, Ľuboš and Robert F. Stambaugh, 2002, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63, 315-349.

[30] Lynch, Anthony W. and David K. Musto, 2003, How Investors Interpret Past Fund Returns, *Journal of Finance* 58, 2033-2058.

[31] Mamaysky, Harry, Matthew Spiegel, and Hong Zhang, 2007, Improved Forecasting of Mutual Fund Alphas and Betas, *Review of Finance* 11, 359-400.

[32] Massa, Massimo, Jonathan Reuter, and Eric Zitzewitz, 2010, When should firms share credit with employees? Evidence from anonymously managed mutual funds, *Journal of Financial Economics* 95, 400-424.

[33] Reuter, Jonathan and Eric Zitzewitz, 2010, How Much Does Size Erode Mutual Fund Performance? A Regression Discontinuity Approach, NBER Working Paper 16329.

[34] Robert, Christian P. and George Casella, 2004, *Monte Carlo Statistical Methods*.

[35] Rosen, Sherwin, 1981, The Economics of Superstars, *American Economic Review* 71, 845-858.

[36] Sirri, Erik R. and Peter Tufano, 1998, Costly Search and Mutual Fund Flows, *Journal of Finance* 53, 1589-1622.

[37] Stephens, Matthew, 2000, Dealing with label switching in mixture models, *Journal of the Royal Statistical Society B* 62, 795-809.

[38] Wermers, Russ, 2000, Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses, *Journal of Finance* 55, 1655-1695.

[39] Zheng, Lu, 1999, Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability, *Journal of Finance* 54, 901-933.