Cross-Sectional Mutual Fund Performance *

Mark Fisher
Federal Reserve Bank of Atlanta
mark@markfisher.net

Mark J. Jensen
Federal Reserve Bank of Atlanta
Mark.Jensen@atl.frb.org

Paula Tkac
Federal Reserve Bank of Atlanta
Paula.Tkac@atl.frb.org

Abstract. We contribute to the mutual fund performance literature by letting investors learn how skill is distributed over the entire universe of mutual funds. Instead of assuming a known normal distribution whose average and standard deviation is unknown for the population, in this paper, a investor learns how skill is distributed over funds by observing the performance history of all past and present mutual funds. As the panel of returns is observed, the investor reduces his level of ignorance about the population by grouping together similarly skilled funds into sub-populations where the funds belonging to a particular sub-population all have the same average level of ability and variation in this ability. Applying this to the gross returns of a panel of 5,136 domestic equity funds, our investor finds the skill level among mutual funds to be distributed across three different sub-populations. The sub-population with the lowest probability of membership is a group of highly skilled funds whose average performance exceeds a passive four-factor portfolio by approximately six percent per annum. The next sub-population whose proability of membership is in between the other two groups is a group of unskilled funds who on average underperform the passive portfolio by 0.4 percent a year. Lastly, the sub-population with the highest likelihood of membership is a group of break-even fund that beat the passive portfolio by roughly the average fee charge by funds. On average the returns from this break-even group exceeds the passive portofolio by 1.8 percent a year. Knowledge about how skill is distributed across mutual funds leads to our investor predicting that any fund lacking a track record of performance is likely to possess just enough skill to on average cover the fees it charges. He knows there is a very small chance the fund will be highly skilled and capable of beating the market on average by six percent a year. However, he also knows there is a greater chance that the fund will underperform the market and not cover the fees it charges.
Keywords: Mutual fund, performance, Bayesian learning, nonparametric analysis

September 1, 2017

*We would like to thank Vikas Agarwal, Wayne Ferson, Siva Nathan, Jay Shanken and the conference participants at the 10th Annual All Georgia Finance Conference, the 2014 Southern Finance Meetings, the Economics, Finance and Business Section of Bayes250, the 2014 Bayesian Workshop of the Remini Centre of Economic Analysis, the 2015 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics, and the department members at the Institution of Statistics and Mathematics at Vienna University of Business and Economics, the DeGroote School of Business at McMaster University, the Department of Economics at the University of Montreal for their helpful comments and suggestions. The views expressed here are ours and are not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Evaluating the performance of actively managed mutual fund has enjoyed a long and productive research agenda, but only a few of these studies investigate how stock-picking ability is distributed over the population of mutual funds. Kosowski et al. (2006), Barras et al. (2010), Fama & French (2010) and Ferson & Chen (2015) take a classical approach and estimate the population distribution by bootstrapping every fund's estimated skill level. Jones & Shanken (2005) address the distribution question from a Bayesian perspective by estimating the mean and variance of a normal, cross-sectional, distribution of skill. Estimation of the population mean and variance of skill by Jones & Shanken (2005) within a cross-sectional, normal distribution is advantageous because it overcomes the overfitting and imprecision found in classical estimates of fund performance. They do this by tying together every fund's stock-picking skill to the population mean and variance. However, no mutual fund research has asked how an investor can dispense with the normality assumption and use a panel of mutual fund returns to learn in its entirety how skill is distributed across the universe of mutual funds. That is our goal here along with determining each fund's ability under the unknown population distribution of skill. To do this we propose a Bayesian learning approach where the key contribution is estimating how skill is distributed over the entire population of mutual funds so that the cross-sectional performance is consistent with the performance of the individual funds. We refer to a fund's skill and/or stock-picking ability in terms of the alpha a mutual fund adds to an investment before incurring any costs during the production of returns.

Most estimates of alpha either assume something about the nature of the cross-sectional distribution of skill or completely ignore it. In his seminal paper on mutual fund performance, Jensen (1968) implicitly assumes the cross-sectional distribution of alpha does not exist by estimating each funds alpha individually with the ordinary least squares (OLS) estimator. Pástor & Stambaugh (2002*b*) explicitly assume the alphas do not come from a common cross-sectional distribution when they choose a uniform distribution over the entire real line for the prior of alpha. Baks et al. (2001) and Pástor & Stambaugh (2002*a*) both investigate mutual fund performance using a cross-sectional distribution that is set a priori and find a fund's alpha to be sensitive to this choice for the cross-sectional distribution. Avramov & Wermers (2006) investigate fund performance and their predictability using three different beliefs about the cross-sectional distribution of skill; a degenerative distribution where funds only cover their costs; a normal population distribution with a unit variance, centered at the breakeven point; and those who do not believe the alphas come from a population distribution. More recently, both Berk & Green (2004) and Jones & Shanken (2005) assume the cross-sectional distribution of skill is normally distributed and estimate the unknown population mean and variance.[1]

In this paper a investor knows nothing about the distribution family the cross-sectional distribution of skill comes from, hence, in addition to not knowing the population mean and variance, the investor is initially ignorant about the population moments and the multimodality of the distribution. Instead, the cross-sectional distribution of mutual fund performance will be inferred by the investor as he observes a cross-section of mutual fund returns. The investor relies on a Bayesian learning scheme where he assigns each mutual fund to a sub-population of similarly skilled funds. If a funds performance is truly extraordinary the investor assigns the fund to a new sub-population consisting of only the extraordinary fund. By assigning a fund's performance level to a new or existing sub-population the investor flexibly models the number of unknown sub-populations that exist within the universe of mutual funds and also allows for unforeseen sub-populations created by future abnormal mutual funds performance.[2]

After learning how skill is distributed over the population of mutual funds, the investor borrows information from the other funds belonging to same sub-population to gain a more precise estimate about a particular fund's alpha. As in Jones & Shanken (2005), our investor "learns across funds" about each mutual funds ability by linking together alphas through the underlying population distribution, except our investor is more refine. Here the investor only learns from funds belonging to the same sub-population. In other words, it is common sense for our investor to ignore the history of returns from poorly managed funds like the Potomac OTC/Short fund when evaluating the skill of an highly skilled fund like the Schroeder Ultra fund. Estimates of a fund's alpha partially depends on the performance of the other funds in the same sub-population through the average skill level and the variability of the sub-population the particular fund belongs to. This refined learning by the investor reduces the uncertainty around his understanding about a fund's stock-picking ability, especially for a fund with a short performance history who belongs to a large sub-population of funds. It also polices against biasing a fund's alpha towards the average skill level of the entire mutual fund population when unskilled and high-skilled sub-populations exist within the population.[3] Unfortunately for our investor, these sub-populations provide no economic

[1] Although our interest is centered on estimating the population distribution of alpha, in order to better understand the stock-picking ability of a particular mutual funds, our approach could also be used to compare asset pricing models and to test the predictability of stock returns as in Pástor & Stambaugh (2000) and Kandel & Stambaugh (1996), respectively.

[2] Learning the cross-sectional distribution of skill through the assignment of funds to sub-populations is similar to the idea of judging a fund by the company it keeps (see Cohen et al. (2005)). However, unlike Cohen et al. (2005) definition of funds holding a common set of stocks, the company a fund keeps here is a subgroup of funds having the same average level of skill and variability or uncertainty in its skill.

[3] For an example of this shrinkage towards the population average see the performance estimates in Jones

meaning to him about the individual mutual fund. Each sub-population is non-parametric in nature, and hence, they cannot be labeled with any economic meaning for the type of funds belonging to them.

Applying this learning approach to the monthly gross return histories of the 5,136 domestic equity, mutual funds that existed for at least twelve month between January of 1961 and June of 2001, our investor concludes that the performance level of mutual funds is best described by a cross-sectional distribution with three distinct sub-populations. Our investor finds an extremely rare, highly skilled, sub-population of funds whose average performance exceeds a passive four-factor portfolio by approximately six percent per annum. The next most likely sub-population is a group of unskilled funds who on average underperform the passive portfolio by 0.4 percent a year. The third and most probable sub-population beats the passive portfolio by an average of 1.8 percent a year, which is roughly the average fee charged by an actively managed mutual fund. Hence, our investor's understanding of the cross-sectional performance of mutual funds lends support to the theoretical findings of Berk & Green (2004), that an arbitrary mutual fund with no performance history is most likely to possess just enough skill to on average cover the fees it charges its investors. The investor's cross-sectional distribution also supports Barras et al. (2010) and Ferson & Chen (2015) ex ante conjecture that the population of mutual funds consists of three sub-populations, those that are skilled, those that are unskilled, and those that just cover their fees. Since there is only a very slime chance a arbitrary fund possesses the extraordinary skill level needed to beat the market on average by six percent a year, it is most likely that, if the arbitrary fund is not a Berk & Green (2004) breakeven fund, it will be an unskilled fund that reduces the return the investor could have earned on a passive portfolio and charged the investor for this "expertise".

1 Investors decision

We analyze mutual fund performance from the perspective of the investment decision of an investor who can choose between a risk-free asset, a set of benchmark assets, and a number of actively managed mutual funds. This investment choice is slightly different from that made in Baks et al. (2001) (here after BMW) where the investors decision to invest in a particular fund is treated separately from their investment decision in other funds. Instead, our investor follows Jones & Shanken (2005) (here after JS) and chooses from across a number of actively managed mutual funds, drawing on each funds return performance to

& Shanken (2005) and Cohen et al. (2005).

learn about the stock picking ability and skill of past, present, and future mutual funds.

Like Jensen (1968), excess gross returns for J mutual funds are assumed to follow the linear factor model

$$r_{i,t} = \alpha_i + \beta_i' F_t + \sigma_i \epsilon_{i,t}, \tag{1}$$

where $r_{i,t}$, $i = 1, \ldots, J$ and $t = 1, \ldots, T_i$, are the excess returns before expenses and fees for the ith fund in month t.[4] The innovations $\epsilon_{i,t}$ are independent and identically distributed normal with mean zero and a variance of one. Independence here means each fund has its own stock-picking strategy and does not mimic or borrow from the other funds.[5]

The vector of passive benchmark returns F_t are observed by the investor at the end of each month t. Later, in our empirical analysis, F_t will consist of four passive factor returns; the three-factor model of Fama & French (1993) and the momentum portfolio of Carhart (1997). Hence, in our empirical analysis the risk-factor equation will be

$$r_{i,t} = \alpha_i + \beta_{i,R} \cdot \text{RMRF}_t + \beta_{i,S} \cdot \text{SMB}_t + \beta_{i,H} \cdot \text{HML}_t + \beta_{i,M} \cdot \text{MOM}_t + \sigma_i \epsilon_{i,t}, \tag{2}$$

where RMRF_t is the excess market return in the tth month, SMB_t and HML_t are the size and book-to-market factors, and MOM_t is the monthly momentum return.

In Eq. (1) and (2), the magnitude of alpha is assumed to reflect the ith fund's ability to select superior stocks and is the only parameter of stock-picking skill in the model. BMW show that a mean-variance investor will invest in an existing fund if and only if the expected value of the fund's posterior distribution of alpha is greater than zero and covers the fees; i.e., the mutual fund is expected to outperform a costless portfolio comprised of the benchmark returns, F. Investing in an arbitrary fund with no return history will be economical if and only if the expectation over the the population distribution of past and present mutual fund skill is positive and exceeds the average fee charged by funds. How much the investor chooses to invest in a particular mutual fund will depend on the investors level of uncertainty around the expected value of alpha as measured by the respective fund's posterior standard deviation of alpha.

Because alpha is not observed by the investor, he must formulate an initial set of beliefs about the potential level of skill held by each fund by choosing a prior distribution $\pi(\alpha_i)$.

[4]Expenses and fees vary across funds and over time and are generally set by the management company, not the fund manager. For example, the economic model of fund behavior by Berk & Green (2004) predicts the economic rents generated by a skilled fund will be captured by the fund's management company through higher fees.

[5]This assumption could be relaxed as in JS, which would help increase the precision of our estimate of a funds alpha, but it shouldn't affect our findings for the cross-sectional distribution of skill because of the size of our panel.

The investor believes that if the alphas were observable each fund's alpha would be a random realization from the cross-sectional distribution of alpha. Both JS and Berk & Green (2004) assume the alphas from past and present mutual funds are realizations from this cross-sectional distribution. However, the alphas and their cross-sectional distributions are both initially unknown to our investor, but as posterior views on the value of the alphas are developed the investor forms his posterior beliefs about the cross-sectional distribution and vice-a-versa. In this manner, each mutual funds alpha is linked to the other mutual fund alphas because the investor borrows information from the posterior population distribution when forming his posterior understanding about the skill of a particular mutual fund.

Before observing the empirical performance of any fund, the investor's best guess about how skill is distributed across the universe of mutual funds is his prior $\pi(\alpha_i)$. Even though we have assumed each funds stock-picking ability to be independent from the other funds, the investor's knowledge about the ith fund's ability and how skill is distributed across the population of mutual funds increases as the investor observes the risk adjusted gross returns of any fund. If the returns are for the ith fund, the investor directly updates his prior for α_i to the posterior $\pi(\alpha_i|r_i) \propto \pi(\alpha_i)f(r_i|\alpha_i)$, where $f(r_i|\alpha_i)$ is the likelihood the value of α_i leads to the realized history of returns $r_i = (r_{i,1}, \dots, r_{i,T_i})'$. Alternatively, if the returns are for any other fund or funds, the investor first updates his beliefs about the population distribution, in other words, the prior $\pi(\alpha)$, to the updated prior, $\pi(\alpha|r_{-i})$, where r_{-i} are the returns histories of any fund other than the ith fund. This updated prior is the investors new guess for the cross-sectional distribution of skill and it becomes the investors updated prior for the alpha of any fund whose return history has not been observed. In our example here, that fund is the ith mutual fund. After observing r_i, our investors understanding about the ith funds level of skill becomes $\pi(\alpha_i|r_i, r_{-i}) \propto \pi(\alpha_i|r_{-i})f(r_i|\alpha_i)$. Hence, the investor has borrowed information from the performance of other funds and incorporated that information into his posterior cross-sectional distribution, $\pi(\alpha|r_{-i})$, and then used this guess of the cross-sectional distribution to increase his knowledge about the potential skill of any fund lacking a track record of performance.[6]

JS investors assume the population distribution of alpha is normally distributed with an unknown population mean and variance.[7] The symmetrical normal distribution does not allow for sub-populations of extraordinary skilled or unskilled funds. Instead, every fund

[6]Investors could use the cross-section of return histories to update their prior beliefs for β_i and σ_i. However, since our focus is on mutual fund performance we let the priors for β_i and σ_i be ex ante noninformative priors; i.e., we assume there is no population distribution for the risk loadings, β, nor the risk level, σ. Future research calls for investigating how investors learn about these distributions.

[7]It is well known that the normal distribution is a poor population distribution for latent variables like alpha. Normality often result in misleading estimates of the latent variables (see Verbeke & Lesaffre (1996)).

is assumed to come from a population where the average ability and variance is the same for every fund. For example, suppose as Barras et al. (2010) and Ferson & Chen (2015) do, that there are three distinct sub-populations within the universe of mutual funds; a skilled, unskilled and zero-alpha group of funds, where the stock picking abilities of those funds in a particular sub-population vary but on average have a similar level of skill. If the cross-sectional distribution of skill is assumed to be normal by the investor, the three distinct sub-populations get pooled into one overall population. Posterior beliefs by the investor about either a skilled or unskilled fund's alpha will end up being biased towards the average level of skill of the entire population of mutual funds and away from the average ability of its sub-population. In the end the investor posterior beliefs about a truly rare, highly, skilled mutual fund gets lost among the "average" mutual funds.

Here we let the investor's beliefs about the cross-sectional distribution of skill be completely flexible and allow him to learn about the distribution from the information found in a panel of returns. We now make the investors initial beliefs and his learning of the unknown cross-sectional distribution of skill concrete.

2 Investors initial beliefs

We assume the investor's prior beliefs for the distribution of alpha is independent from the funds risk-factors and return variance by letting $\pi(\alpha_i, \beta_i, \sigma_i^2) = \pi(\alpha_i)\pi(\beta_i, \sigma_i^2)$.[8] The ith funds likelihood is $N(r_i|\alpha_i\iota_{T_i} + F_i\beta_i, \sigma_i^2 I_{T_i})$, $i = 1, \ldots, J$, where r_i is the T_i length return history vector for fund i, ι_{T_i} is a T_i length vector of ones, F_i is a $T_i \times 4$ matrix of factor returns, and I_{T_i} is an $T_i \times T_i$ identity matrix. Each funds likelihood is independent from the other funds implying the investment choices of a particular fund conveys no information about any other fund's investment strategy.

Assume that mutual fund stock picking abilities are normally distributed, $N(\alpha_i|\mu_\alpha, \sigma_\alpha^2)$, $i = 1, \ldots, J$, and are conditional on knowing the population mean skill level, μ_α, and population variance, σ_α^2. Since μ_α and σ_α^2 are unobservable to our investor, he will have an inherent prior belief for them. We denote this initial belief with the prior distribution $G(\mu_\alpha, \sigma_\alpha^2)$.[9]

To ease with the presentation of the investors understanding of the cross-sectional distribution we begin by assuming the α_is are observable. Later we will relax this assumption

[8] One could assume investors have a joint prior for $(\alpha_i, \beta_i, \sigma_i^2)$. However, learning this distribution would require him to assign a fund to a sub-population based on all the unknown parameters and not just alpha. Grouping funds into sub-populations of ability would no longer be the investors objective, so, we assume the investor has a separate prior for beta and sigma.

[9] Jones & Shanken (2005) assume G to be a normal, inverse-gamma distribution.

and allow the investor to form posterior beliefs about the alphas, betas, and sigmas, along with μ_α and σ_α, as he observes mutual fund returns and integrate out the uncertainty underlying each funds unknown parameters. After observing the stock-picking ability of J mutual funds through the values of $\alpha_1, \ldots, \alpha_J$, our investor decides to invest or not in a new mutual fund by computing the posterior cross-sectional distribution of mutual fund performance, $\pi(\alpha|\alpha_1, \ldots, \alpha_J, G)$. This posterior understanding of the cross-sectional mutual fund performance by our investor is the same as the posterior predictive distribution of the alpha for a fund with no performance history. It consists of integrating out the parameter uncertainty found in the prior $N(\alpha|\mu_\alpha, \sigma_\alpha^2)$. The formal definition of the posterior predictive or cross-sectional distribution is

$$\pi(\alpha|\alpha_1, \ldots, \alpha_J, G) = \int N(\alpha|\mu_\alpha, \sigma_\alpha^2)\, dG(\mu_\alpha, \sigma_\alpha^2|\alpha_1, \ldots, \alpha_J). \tag{3}$$

In other words, the predictive distribution of skill consists of a mixture of normal distributions over the population mean and variance, μ_α and σ_α^2, whose mixture weights and locations are determined by the posterior distribution, $G(\mu_\alpha, \sigma_\alpha^2|\alpha_1, \ldots, \alpha_J)$.

Conditional on the investors choice for $G(\mu_\alpha, \sigma_\alpha^2)$, the alpha of a mutual fund having no empirical record will be a realization from the above posterior predictive distribution. In practice, a Gibbs sampler is used to draw the latent alphas conditional on the betas, sigmas and μ_α and σ_α^2, and then given the alphas, sample the betas, sigmas and μ_α, and σ_α^2 conditional of the alphas, all conditional on $r_1, \ldots, r_J$ and the investors choice of G.

BMW, Pástor & Stambaugh (2002*b*), and Busse & Irvine (2006) type investors assume they know in its entirety the population distribution of mutual fund skill, $\pi(\alpha)$, but with each assuming something different about it. Since they assume their investors know how skill is distributed, investors do need to learn, nor borrow from the empirical performance of other funds, to update their beliefs about either the cross-sectional mutual fund performance, the average fund's ability, or the variation around this average. Under these circumstances the investors understanding about a particular funds skill will only depend on the performance history of the fund, but it will be sensitive to what the investor assumed about the population distribution, $\pi(\alpha)$. Different beliefs about the cross-sectional distribution will lead to different beliefs about the skill level of a particular fund. Hence, the reason for the wide variety of findings for skill in the mutual fund performance literature.[10]

For example, past research has found positive alphas (see Kosowski et al. (2006), Pástor & Stambaugh (2002*a*), Fama & French (2010)). Since these findings run counter to some

[10]BMW explicitly show how sensitive the inference about the skill level of a particular fund is to the choice of $\pi(\alpha)$.

peoples belief that actively managed funds should not out perform a passive market based investment strategy, it has been argued that investors should rely less on the return data and more on a prior that is skeptical of stock-picking skill. In other words, employ a more informative cross-sectional distribution. In our representation of the cross-sectional distribution, a investor heeding this advice would choose a G where μ_α and σ_α^2 are influenced less by the return data and more by the investors initial beliefs about skill. At the extreme, the investor would let G equal $\mathbf{1}_{\{m_0,s_0^2\}}(\mu_\alpha,\sigma_\alpha^2)$, where G is the degenerative distribution at the point (m_0,s_0^2). Such an investor is certain m_0 is the average skill level over the entire population of funds and s_0^2 is the variation around this average. Given $G \equiv \mathbf{1}_{\{m_0,s_0^2\}}(\mu_\alpha,\sigma_\alpha^2)$, the posterior population distribution in Eq. (3) is the normal distribution

$$\pi\left(\alpha \,\middle|\, \alpha_1,\ldots,\alpha_J, G \equiv \mathbf{1}_{\{m_0,s_0^2\}}(\mu_\alpha,\sigma_\alpha^2)\right) = N(\alpha|m_0,s_0^2),$$

with mean m_0 and variance σ_α^2. Because this investor assumes he knows that the population mean and variance equals m_0 and s_0^2, respectively, his posterior is his prior when inferring every fund's alpha. This dogmatic belief is implicitly assumed in BMW, Pástor & Stambaugh (2002*a*) and Pástor & Stambaugh (2002*b*) and by any Bayesian investor whose prior for alpha has no unknown hyperparameters; e.g., any time values for μ_α and σ_α^2 have been set a priori. For instance, a ordinary least square estimate of a mutual funds alpha implicitly sets $\mu_\alpha \equiv 0$ and $\sigma_\alpha^2 \equiv \infty$ so that $\pi(\alpha)$ is uniform over the entire real line; i.e., the investor strongly believes that the alphas do not come from a common population distribution of skill, but instead, skill is completely idiosyncratic.

In JS seminal article on mutual fund performance an investor learns about the skill of a particular fund by observing the performance of other funds. They do this by learning about the values of μ_α and σ_α^2 from the aggregate performance of every fund and using this population information to make inference about a particular funds stock-picking ability. Because the JS investor learns about the population mean and variance of skill he is initially uncertain about the value of μ_α and σ_α^2. This uncertainty is represented by his prior for σ_α^2, which is an Inverse-Gamma distribution, and his prior for μ_α, which is normally distributed. In our context, a JS type investor sets $G(\mu_\alpha,\sigma_\alpha^2) \equiv NIG(m_0,\sigma_\alpha^2/\kappa_0,\nu_0/2,s_0^2,\nu_0/2)$, where m_0 and σ_α^2/κ_0 are the mean and variance to the conditional normal distribution of μ_α, and $\nu_0/2$ and $s_0^2\nu_0/2$ are respectively the scale and shape of the Inverse-Gamma distribution of σ_α^2. Applying the definition of the prior for the cross-sectional distribution of alpha in Eq. (3), a JS type investor initially believes skill is distributed over the entire universe of mutual funds as a Student-t since

$$\pi(\alpha) \quad = \quad \int N(\alpha|\mu_\alpha,\sigma_\alpha^2)\, NIG(\mu_\alpha,\sigma_\alpha^2|m_0,\sigma_\alpha^2/\kappa_0,\nu_0/2,s_0^2\nu_0/2)\, d(\mu_\alpha,\sigma_\alpha^2), \qquad (4)$$

$$= t_{\nu_0}\left(\alpha \middle| m_0, \left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2\right). \quad (5)$$

Unique to our investigation of mutual fund cross-sectional performance, our investor does not condition his posterior understanding about alpha on a fixed known G. Instead, our investor assumes G is an unknown, random, distribution function, whose uncertainty will be taken into consideration by the investor in his posterior beliefs about mutual fund performance. Being a unknown distribution, a prior is needed for G. Investors will then learn about G and reduce their uncertainty about G as they observe the performance of mutual fund belonging to the population. Because the predictive distribution in Eq. (3) is conditional on G, an unknown G means the cross-sectional distribution $\pi(\alpha)$ is a unknown, random, population distribution function.

We propose and investigate a Bayesian nonparametric predictive distribution for mutual fund skill and performance. Initial beliefs about G are modeled with a Dirichlet process prior distribution (see Ferguson (1973)). To be specific, we model the investors prior beliefs about alpha with Lo (1984) and Escobar & West (1995) Dirichlet process mixture prior (DPM) where the prior for G is a Dirichlet process distribution denoted by $DP(B, G_0)$. The DP distribution is defined in terms of the base distribution, G_0, which is the investor's best guess for G. Here G_0 will be the bivariate, base distribution, $G_0(\mu_\alpha, \sigma_\alpha^2)$. Being the investor's best guess for G it follows that $E[G] = G_0$.[11] An investors level of confidence in his guess G_0 is the value of the positive scalar, B, known as the concentration parameter.[12]

Before our investor observes any return-based information concerning mutual fund performance, he forms his expectations about the stock-picking ability of mutual funds. If he is certain about the average skill and variance of mutual funds he might choose $G_0 \equiv \mathbf{1}_{\{m_0, s_0^2\}}(\mu_\alpha, \sigma_\alpha^2)$, for the Dirichlet process base distribution. As pointed out earlier, degenerative distribution represent strong prior beliefs. Given the degenerative nature of the base distribution, $\mathbf{1}_{\{m_0, s_0^2\}}(\mu_\alpha, \sigma_\alpha^2)$, the investors initial cross-sectional distribution of alpha is

$$E_G[\pi(\alpha)] \equiv E_G\left[\int N(\alpha|\mu_\alpha, \sigma_\alpha^2) dG(\mu_\alpha, \sigma_\alpha^2)\right], \quad (6)$$

$$= \int N(\alpha|\mu_\alpha, \sigma_\alpha^2) dG_0(\mu_\alpha, \sigma_\alpha^2), \quad (7)$$

$$= \int N(\alpha|\mu_\alpha, \sigma_\alpha^2) \mathbf{1}_{\{m_0, s_0^2\}}(\mu_\alpha, \sigma_\alpha^2) d(\mu_\alpha, \sigma_\alpha^2), \quad (8)$$

[11] See Kleinman & Ibrahim (1998), Burr & Doss (2005), Ohlssen et al. (2007), Dunson (2010) and Chapter 23 of Gelman et al. (2013) and references therein for the mathematical details of the Dirichlet process.

[12] B will play an important role in the number of sub-populations, which we explain when we discuss the clustering properties of the DP prior.

$$= \quad N(\alpha|m_0, s_0^2), \tag{9}$$

For brevity, we let $\widehat{\pi}_{\mathbf{1}_{\{m_0, s_0^2\}}}(\alpha) \equiv E_G[\pi(\alpha)]$ when $G \sim DP(B, \mathbf{1}_{\{m_0, s_0^2\}})$.

At the other extreme are investors who have no preconceived notion about the distribution of skill over the population of mutual funds. These investors have a diffuse, uninformative G_0 and rely on return-based information of mutual funds to help them form their posterior beliefs about the distribution of alpha.

We choose to endow our investor with initial beliefs that are flexible enough to answer the question: How do prior beliefs about the population distribution of mutual fund skill impact the posterior performance of past, present and future mutual funds? To establish the Bayesian learning that occurs, we assume the conjugate base distribution, $G_0 \equiv NIG(m_0, \sigma_\alpha^2/\kappa_0, \nu_0/2, s_0^2\nu_0/2)$, where μ_α is normally distribution with mean m_0 and variance, σ_α^2/κ_0, and σ_α^2 is distributed as a Inverse-Gamma distribution with shape, $\nu_0/2$, and scale, $s_0^2\nu_0/2$. Before observing any mutual fund data, our investors thoughtfully determines values for the mean, variance, shape and scale of G_0; i.e., our investor selects values for the hyperparameters, m_0, κ_0, ν_0, s_0^2, based on his subjective beliefs. Equipped with G_0 the investors belief about skill can range from dogmatic when $\kappa_0 \to \infty$, $\nu_0 \to \infty$, and $s_0^2 \to \infty$,[13] to uninformative when $\kappa_0 \to 0$, $\nu_0 \to 0$ and $s_0^2 \to 0$.

When an investors initial belief for G is represented by a DP distribution with a base distribution, G_0, equal to the Normal, Inverse-Gamma distribution, his best guess for the population distribution of mutual fund skill is the Student-t distribution

$$E_G[\pi(\alpha)] \quad = \quad \int N(\alpha|\mu_\alpha, \sigma_\alpha^2) NIG(\mu_\alpha, \sigma_\alpha^2|m_0, \sigma_\alpha^2/\kappa_0, \nu_0/2, s_0^2\nu_0/2)\; d(\mu_\alpha, \sigma_\alpha^2), \tag{10}$$

$$= \quad t_{\nu_0}\left(\alpha \,\middle|\, m_0, \left(\frac{\kappa_0 + 1}{\kappa_0}\right)\nu_0 s_0^2\right), \tag{11}$$

with ν_0 degrees of freedom, mean m_0 and scale $\left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2$. In the absence of any information investors will initially invest in a actively managed fund if their choice for m_0 exceeds fees and expenses. Going forward we will use the notation $\widehat{\pi}_{NIG}(\alpha)$ to represent $E_G[\pi(\alpha)]$ when $G \sim DP(B, NIG)$. Just to be clear $\widehat{\pi}_{NIG}(\alpha) \equiv t_{\nu_0}(\alpha|m_0, (\kappa_0 + 1)\nu_0 s_0^2/\kappa_0)$ is the initial guess about how skill is distributed over the population of mutual funds.

Later in the empirical section of the paper we set the hyperparameters for $G_0 \equiv NIG(m_0, \sigma_\alpha^2/\kappa_0, \nu_0/2, s_0^2\nu_0/2)$ equal to $m_0 = 0$, $\kappa_0 = 0.1$, $\nu_0 = 0.01$ and $s_0^2 = 0.01$. With this G_0 our investor initially believes a typical fund is unable on average of beating the

[13] Under this dogmatic prior the expected value of σ_α^2 is 1 with zero variance. The prior mean for μ_α is m_0 also with zero variance.

passive portfolio ($m_0 = 0$), but is also so uncertain about the level of skill the variance in the expected performance does not exist.

3 Cross-sectional learning

The next step in our investors mutual fund investment decision is to learn about the distribution of the population mean and variance, $G(\mu_\alpha, \sigma_\alpha^2)$. Learning about G leads to the investor increasing his understanding on how skill is distributed over mutual funds. Suppose hypothetically that upon observing the return-based information from a panel of actively managed mutual funds our investor knows the average skill level of the panel and the variance or uncertainty around this average ability and uses them to form an opinion on how they are distributed. For instance, after observing the performance of the first mutual fund the investor knows μ_1 is the funds average level of skill. He also knows σ_1^2 is how the fund's alpha varies around μ_1.[14] Because the investors initial knowledge about the unknown distribution G is the conjugate prior, $DP(B, G_0)$, after seeing μ_1 and σ_1 our investors understanding about this distribution is the updated conditional DP distribution, $G|\mu_1, \sigma_1^2 \sim DP(1 + B, G_1)$. Our investor refines his guess for G by using μ_1 and σ_1^2 to update his base distribution to $G_1 \equiv (BG_0 + \mathbf{1}_{\{\mu_1, \sigma_1^2\}})/(1 + B)$ (see Blackwell & MacQueen (1973) for this conjugate property of the Dirichlet Process distribution). He is also more confident in his guess since the concentration parameter of the conditional DP distribution has increased to $1 + B$.

In G_1 our investor reduces his uncertainty about G by taking his initial guess of G_0 and combining it with the empirical degenerative distribution $\mathbf{1}_{\{\mu_1, \sigma_1^2\}}$. After marginalizing out the uncertainty associated with $G|\mu_1, \sigma_1^2$, our investors new guess for the population distribution is the posterior predictive distribution of alpha, or in other words, the updated prior

$$\begin{aligned} E_{G|\mu_1,\sigma_1^2}[\pi(\alpha)] &= \int N(\alpha|\mu_\alpha, \sigma_\alpha^2)\, dG_1(\mu_\alpha, \sigma_\alpha^2), \\ &= \frac{B}{1+B} t_{\nu_0}\left(\alpha \middle| m_0, \left(\frac{\kappa_0 + 1}{\kappa_0}\right)\nu_0 s_0^2\right) + \frac{1}{1+B} N\left(\alpha|\mu_1, \sigma_1^2\right), \end{aligned} \tag{12}$$

where $1/(1 + B)$ in the second right hand side term is the probability the next mutual fund will belong to the same sub-population as the first fund and have skills that are normally distributed with mean μ_1 and variance σ_1^2. The first right hand side term's $B/(1 + B)$ is

[14] For the time being, one can think of μ_1 and σ_1^2 as being the sample mean and sample variance of the observed return vector r_1. Later we provide the explicit relationship between the mean and variance of a mutual funds skill level and the funds return-based information.

the probability the next fund will belong to a sub-population different from the first fund. In this case the investors level of uncertainty about the funds ability does not benefit from knowing μ_1 and σ_1^2. Hence, the investor relies on his initial beliefs about the cross-sectional distribution of skill $\widehat{\pi}_{NIG}(\alpha) \equiv t_{\nu_0}\left(\alpha \middle| m_0, \left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2\right)$ to predict the new fund's future ability. Writing the updated guess of the population performance of mutual funds in terms of $\widehat{\pi}$ we have the mixture $\widehat{\pi}_{G_1}(\alpha) = \frac{B}{1+B}\widehat{\pi}_{NIG}(\alpha) + \frac{1}{1+B}\widehat{\pi}_{\mathbf{1}_{\{\mu_1,\sigma_1^2\}}}(\alpha)$.

Our investor continues learning in this same sequential manner as he observes a cross-section of μ_is and σ_i^2s. Following the observation of μ_J and σ_J from the last of the J funds, the investors knowledge of G has grown to $G|\mu_1, \sigma_1^2, \ldots, \mu_J, \sigma_J^2 \sim DP(J+B, G_J)$, where the investor's guess for G is $G_J \equiv (BG_0 + \sum_{i=1}^J \mathbf{1}_{\{\mu_j,\sigma_j^2\}})/(J+B)$. Note our investor is now even more confident in his guess for G since his concentration parameter has increased to $J+B$. In other words, our investor puts less and less weight on his initial beliefs about G and relies more and more on the empirical distribution of the μ_is and σ_i^2s. The investor's guess for how skill is distributed across the universe of mutual funds is now

$$E_{G|\mu_1,\sigma_1^2,\ldots,\mu_J,\sigma_J^2}\left[\pi(\alpha)\right] = \frac{B}{J+B}t_{\nu_0}\left(\alpha \middle| m_0, \left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2\right) + \frac{1}{J+B}\sum_{i=1}^{J} N(\alpha|\mu_i, \sigma_i^2). \tag{13}$$

There is now a one-in-$(J+B)$ chance a new funds skill will be distributed like one of the J observed mutual funds. As long as the investor has some doubt about his initial guess G_0, in other words, as long as B is finite, there will be funds with the same mean μ_i and variance σ_i^2 as other funds. As a result some of the existing funds will have the same distribution of idiosyncratic ability. In other words, given our investor's initial DP beliefs for G he will logically group together similarly skilled funds into $K \leq J$ sub-populations as he observes funds having the same average ability and variance. Let μ_k^* and σ_k^*, $k = 1, \ldots, K$, denote the means and variances of the unique sub-populations. Counting up the number of funds belonging to a particular sub-population, the investor finds there are n_k funds belonging to the kth sub-population such that $\sum_{k=1}^K n_k = J$.

Grouping the J funds into their respective sub-populations our investor's posterior cross-sectional distribution of alpha can be written in terms of the different sub-populations means and variances as

$$\begin{aligned}\widehat{\pi}_{G_J}(\alpha) &= E_{G|n_1,\mu_1^*,\sigma_1^{*2},\ldots,n_K,\mu_K^*,\sigma_K^{*2}}\left[\pi(\alpha)\right], \\ &= \frac{B}{J+B}t_{\nu_0}\left(\alpha \middle| m_0, \left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2\right)\end{aligned}$$

$$+\sum_{k=1}^{K}\frac{n_k}{J+B}N(\alpha|\mu_k^*,\sigma_k^{2*}). \tag{14}$$

The probability of a new fund being assigned to the kth sub-population depends on n_k, hence, our investor believes larger sub-populations have a greater chance of attracting new funds. The Dirichlet process prior for G causes our investor to reward larger sub-populations with new funds at the expense of smaller sub-populations. However, we need to remember the degree of grouping also depends on how confident the investor is in his initial guess G_0. This confidence depends on his choice for the concentration parameter B. Because of the important role B plays in the number of sub-populations, in our empirical analysis we let B be unknown and have the investor infer it from the panel of observed mutual funds returns.

Our investor is thus learning across funds as in JS. However, our investor is dramatically different from those in JS who implicitly assume all funds come from a normal population distribution. Our investor assumes nothing about the cross-sectional distribution of skill. Instead, he learns it by flexibly allowing for different sub-populations whose funds have similar ability in terms of their average skill and variance. He does not force every fund's alpha to be a draw from a distribution whose average level of skill is equal to the entire population of both skilled and unskilled funds. Instead, each fund's alpha is a draw from a distribution whose average skill is equal to that of other relevant funds who are similarly skilled. This difference leads to more accurate inference about each funds alpha.

4 Inference

To resolve the uncertainty around the parameters and the unknown distribution of alpha, our investor combines fund-level return data with his initial beliefs to form a posterior view of the unknowns. Our investors decision to invest in a actively managed fund involves a large number of unknowns where the joint posterior distribution of all the unknowns is not a standard distribution. This requires the joint posterior distribution to be judiciously broken into conditional posteriors whose distribution can be drawn from. Cross-sectional mutual fund returns lend themselves to sampling from the conditional posteriors and enables the investor to reduce some of their uncertainty around the unknowns.

Our blocking design of the joint posterior distribution is structured by the hierarchical form of the investors decision. Random draws are made from the conditional posterior distributions which upon convergence generate random draws from the joint posterior. The conditional posterior distribution draws are

1. Draw β_i and σ_i conditional on r_i, and α_i, for $i = 1, \ldots, J$.

2. Draw α_i conditional on r_i, β_i, σ_i and $(\mu^*_{s_i}, \sigma^*_{s_i})$, for $i = 1, \ldots, J$.

3. Draw (μ^*_k, σ^*_k), s, K and B conditional $\alpha_1, \ldots, \alpha_J$, for $k = 1, \ldots, K$.

In Step 2, s_i is the assignment variable where $s_i = k$ when $(\mu_i, \sigma_i^2) = (\mu^*_k, \sigma_k^{2*})$; i.e., s_i is the sub-population assignment variable for the ith fund. In Step 3, s is the J length assignment vector comprised of s_i where $i = 1, \ldots, J$.

After an initial burnin where the draws from the conditionals are thrown away allowing the draws to converge to the posterior distribution, subsequent draws from the conditionals generate a random sample from the joint posterior distribution. These draws of the alphas represent the uncertainty the investor has about the skill level among actively managed funds. Later in Section 5 we make 40,000 draws and keep the last 30,000 for making inference.

In Step 1, our investor has no information on, nor prior beliefs about, the factor loading vector, β_i, or the return variance, σ_i^2. The Jeffreys prior

$$\pi(\beta_i, \sigma_i^2) \propto 1/\sigma_i^2, \tag{15}$$

accurately captures this ignorance and the investors uncertainty will be captured by a normally distributed conditional posterior for the factor loadings with mean and variance equal to the least squares regression estimator of the dependent variable $r_{it} - \alpha_i$ being projected onto the explanatory variables F_{it}, $t = 1, \ldots, T_i$. The marginal conditional posterior distribution for the return variance, σ_i^2, is an Inverse-Gamma with scale, $T_i - 4$, and shape equal to the above linear regression sum of squared error divided by the scale. These conditional posteriors depend only on the return-based information the investor has observed.

To draw the alphas, in Step 2 the investor uses the cross-sectional distribution of the s_ith sub-population, $N(\mu^*_{s_i}, \sigma^{*2}_{s_i})$, as the prior for α_i. Because the prior for the cross-sectional distribution of alpha is a Dirichlet process mixture, the mean, $\mu^*_{s_i}$, and variance, $\sigma^{2*}_{s_i}$, are determined by the returns of funds belonging to the s_i sub-population. Our investor then reduces his uncertainty about α_i by observing the return data, $r_i = (r_{i1}, \ldots, r_{iT_i})'$, and conditional on β_i and σ_i, updates his beliefs about the ith funds ability through the normal posterior distribution whose mean is

$$\left(\frac{\mu^*_{s_i}}{\sigma^{2*}_{s_i}} + \sum_{t=1}^{T_i} r^*_{i,t} \right) \Big/ \left(\frac{1}{\sigma^{2*}_{s_i}} + T_i \right), \tag{16}$$

and whose variance is $(1/\sigma^{2*}_{s_i} + T_i)^{-1}$, where

$$r^*_{i,t} \equiv (r_{i,t} - \beta_i' F_{i,t})/\sigma_i = \alpha_i + \epsilon_{i,t}, \tag{17}$$

is the risk adjusted return.

We point out here that the conditional posterior distribution of α_i does not depend on the number of funds belonging to the s_ith sub-population. Each fund's skill level is independent from the other. However, the information found in the sub-population's $\mu^*_{s_i}$ and $\sigma^*_{s_i}$, is valuable to the investor especially for funds with a short performance window. For example, if a fund with a short history belongs to a sub-population whose membership has a small $\sigma^{2*}_{s_i}$, in Eq. (16) our investor performance measure will be pulled towards the average skill of sub-population, $\mu^*_{s_i}$. Since traditional performance measures for short history funds are noisy and uncertain (see Kothari & Warner (2001)), a small sub-population variance also helps reduce the investors level of uncertainty around his guess about the ability of a short history fund.

This is different from how JS borrows information from other funds. While insightful, JS shrinks every funds alpha toward the average skill level of the entire population. For our investor this amounts to assigning every fund to the same sub-population; i.e., a priori saying $K = 1$. Such an assumption implies that a surviving fund has the same ability and talent as funds that have failed, or a extraordinary skilled fund having the same ability as an average fund.

The above normal distribution for the alphas also shows how our investor's performance measure for an extremely exceptional fund is no different from the investor who does not believe in a cross-sectional performance distribution. To our investor a truly extraordinary fund has no other peers and hence belongs to its own sub-population. As a result, $\sigma^*_{s_i}$ is infinite. Under such circumstances our investor knows not to borrow from the performance of the other ordinary funds. He simply applies Eq. (16), which equals the OLS estimate when $\sigma^*_{s_i} = \infty$, as his guess for the exceptional funds performance measure. The level of certainty in his guess for alpha is also the variance of the OLS alpha. Hence, our investor's performance measure spans the world of Jensen (1968), where every firm is unique, and Jones & Shanken (2005), where every firm has the same average skill level, μ_α, and same standard deviation, σ_α.

Sampling from Step 3 can be thought of as answering the question proposed by JS but adapted to our case, when would the investor discard the information contained in the average skill and variability of the K sub-populations, μ^*_k and σ^*_k, $k = 1, \ldots, K$? Answering this question for each fund amounts to drawing the vector s by sequentially drawing each fund's s_i according to the probabilities

$$P(s_i = k) = \frac{n_k^{(-i)}}{B + J - 1} f_N(\alpha_i | \mu^*_k, \sigma^{*2}_k), \qquad k = 1, \ldots, K^{(-i)}, \tag{18}$$

$$P\left(s_i = K^{(-i)} + 1\right) \;\; = \;\; \frac{B}{B+J-1} f_t\left(\alpha_i | m_0, \left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2\right), \tag{19}$$

where $n_k^{(-i)}$ is the number of funds belonging to the kth pool and $K^{(-i)}$ is the total number of sub-populations, after the ith fund has been excluded from the sample. So $K^{(-i)}$ will equal $K-1$ when the ith funds is the only member of its sub-population. Otherwise, $K^{(-i)}$ equals K.

Eq. (19) is the probability our investor discards the information contain in the performance of the other funds and rely on the fund's ordinary least squares information when determining the funds alpha. The odds of this occurring increase when the α_i drawn in Step 2 is "different" from the average skill of each sub-population. In other words, when the likelihoods, $f_N(\alpha|\mu_k^*, \sigma_k^{2*})$, $k = 1, \ldots, K^{(i)}$, in Eq. (18) are small.

After the investor assigns every fund to a particular sub-population and in the process determines the total number of sub-populations, K, he pools together the alphas of those funds belonging to a particular sub-population and forms his posterior beliefs about the average skill level and variability of the sub-populations skill and draws μ_k^* and σ_k^*. Given the DP base distribution $G_0 \equiv NIG(m_0, \sigma_\alpha^2/\kappa_0, \nu_0/2, s_0^2\nu_0/2)$ from Section 2, the draws of σ_k^{2*}, for $k = 1, \ldots, K$, are from the Inverse-Gamma distribution with shape $(\nu_0 + n_k)/2$ and scale

$$\frac{\nu_0 + n_k}{2}\left[\nu_0 s_0^2 + \sum_{i:s_i=k}(\alpha_i - \overline{\alpha}_k)^2 + \frac{n_k\kappa_0}{\kappa_0 + n_k}(m_0 - \overline{\alpha}_k)^2\right] \Big/ (\nu_0 + n_k)\,,$$

where $\overline{\alpha}_k = n_k^{-1}\sum_{i:s_i=k}\alpha_i$. The μ_k^*s are drawn from a Normal distribution with mean $(\kappa_0 m_0 + n_k\overline{\alpha}_k)/(\kappa_0 + n_k)$ and variance $\sigma_k^{2*}/(\kappa_0 + n_k)$.

Lastly, the investor draws the concentration parameter B from $\pi(B|K)$ using the sampler in Appendix A.5 of Escobar & West (1995).

4.1 Posterior cross-sectional distribution

In Eq. (14) our investors best guess for the cross-sectional distribution of alpha depends on having observed the means and variances of sub-population skill. After observing the return-based information our investor best informed guess at how skill is distributed over mutual funds is the posterior predictive distribution

$$\begin{aligned}\widehat{\pi}_{r_1,\ldots,r_J}(\alpha) \;\; = \;\; & \int\cdots\int E_{G|s,\mu_1^*,\sigma_1^{*2},\ldots,\mu_K^*,\sigma_K^{*2}}\left[\pi(\alpha)\right] \\ & \times\pi\left(s, \mu_1^*, \sigma_1^{*2}, \ldots, \mu_K^*, \sigma_K^{*2}|\alpha_1, \ldots, \alpha_J\right) \\ & \times\pi(\alpha_1|r_1, \beta_1, \sigma_1)\cdots\pi(\alpha_J|r_J, \beta_J, \sigma_J)\end{aligned}$$

$$
\begin{aligned}
& \times\pi(\beta_1, \sigma_1|r_1)\cdots, \pi(\beta_J, \sigma_J|r_J) \\
& \times ds d\mu_1^* \cdots d\mu_K^* d\sigma_1^* \cdots d\sigma_K^* \\
& \times d\alpha_1 \cdots d\alpha_J \; d\beta_1 \cdots d\beta_J d\sigma_1 \cdots d\sigma_J, \qquad (20) \\
\approx \; & M^{-1} \sum_{l=1}^{M} \left[\frac{B^{(l)}}{J + B^{(l)}} t_{\nu_0} \left(\alpha \middle| m_0, \left(\frac{\kappa_0 + 1}{\kappa_0} \right) \nu_0 s_0^2 \right) \right. \\
& \left. + \sum_{k=1}^{K^{(l)}} \frac{n_k^{(l)}}{J + B^{(l)}} N \left(\alpha \middle| \mu_k^{*(l)}, \sigma_k^{2*(l)} \right) \right], \qquad (21)
\end{aligned}
$$

where $\mu_1^{*(l)}, \sigma_1^{*2(l)}, \ldots, \mu_K^{*(l)}, \sigma_K^{*2(l)}$ is the lth draw from the conditional posterior distribution in Step 3 of the above sampling algorithm, and the $n_k^{(l)}$s come from the information contained in the lth draw of $s^{(l)}$, and $B^{(l)}$ is the lth draw from the sampler of Escobar & West (1995). This posterior cross-sectional distribution calculation takes into consideration all the uncertainty about the unknowns, including the distribution of G, by averaging over their posterior distributions.

5 Empirical analysis

Our empirical application consists of the same mutual fund data set used by Jones & Shanken (2005).[15] This data set consists of annual mutual fund gross returns computed monthly from January 1961 to June 2001. It is a panel with a total of 396,820 monthly observations from 5,136 domestic equity funds.[16] Like Baks et al. (2001), Jones & Shanken (2005) and Cohen et al. (2005), we are interested in before cost performance unaffected by idiosyncratic fee schedules so fees and expenses have been added back into the net returns reported in CRSP Mutual Funds data files. Each fund has at least a years worth of return data and all the funds have on average 77.3 monthly returns. Survivorship bias is eliminated by including all domestic equity funds even the 1293 funds that no longer existed at the end of the sample.

5.1 Shrinkage in fund performance

Figure 1 plots each of the 5,136 mutual funds highest 95 percent posterior probability density interval (HPD) of alpha along with the fund's posterior median (represented as dots when visible) beginning at the top with those funds having the shortest return history and ending at the bottom with the longest performing funds. Each funds HPD interval are those values

[15] We would like to thank Chris Jones for graciously providing us with their data.

[16] Funds were eliminated that made substantial investments in other asset classes.

of alpha where there is a 95 percent chance the fund will select stocks that result in an alpha within the HPD interval. Investors minimizing their expected mean absolute loss from investing in a fund chooses a fund based on its posterior median alpha.

Each funds posterior HPD interval is calculated under three different investor beliefs about the underlying cross-sectional distribution of skill. Panel (a) plots the posterior beliefs of the investor who believes skill to be idiosyncratic to the mutual fund and hence, does not believe there is any advantage to be gained by learning from the performance of other funds. These investors prior for the cross-sectional distribution of alpha is $N(0, s_0^2)$, where they assume they know an average fund can do no better than the market and that the cross-sectional variance is $s_0^2 = \infty$.[17]

Panel (b) in Figure 1 plots the posterior beliefs for each fund's alpha of an investor who believes skill is normally distributed over the universe of mutual funds but do not know the population mean or variance. This investor learns about the population mean and variance through the past performance histories of every fund both those in and out of business. He believes aggregating together the performance data of all funds yields important information on the stock-picking ability of the entire population of funds. The investor is endowed with the noninformative Jeffreys prior $\pi(\mu_\alpha, \sigma_\alpha^2) \propto 1/\sigma_\alpha^2$ for the cross-sectional mean and variance; i.e., he has no prior knowledge about the population mean or variance.

Lastly, Panel (c) of Figure 1 plots each funds posterior probability interval for the investor who learns the entire cross-sectional distribution from the history of mutual fund returns. His initial guess for the cross-sectional distribution of alpha is a diffuse Student-t distribution with mean zero, scale, 0.0011, and 0.1 degrees of freedom.

[17] Because this is the Jeffreys prior for each alpha, the intervals in Figure 1(a) are equivalent to the 95 percent confidence intervals from the ordinary least squares (OLS) estimate of each alpha.

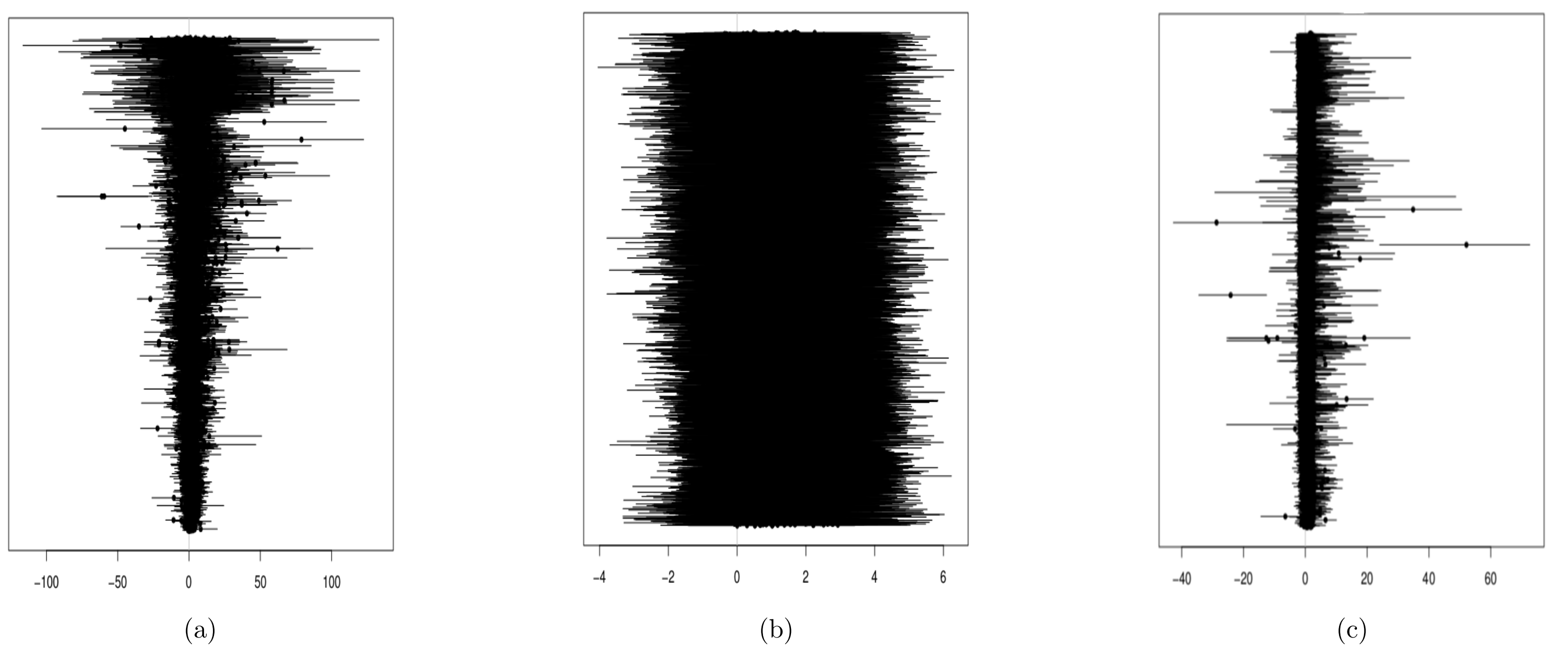


Figure 1: The 95% highest probability density interval of each mutual fund's posterior alpha sorted from shortest (top) to longest (bottom) fund history using (a) no learning occurs and investors prior for α is uniformly distributed on the real line, (b) learning by investors occurs about the cross-sectional mean and variance of skill where investors priors are α is $N(\mu_\alpha, \sigma_\alpha^2)$ and $\pi(\mu_\alpha, \sigma_\alpha^2) \propto 1/\sigma_\alpha^2$ (c) learning occurs by investors about the cross-sectional distribution of skill where investors prior for α is unknown but modeled with the Dirichlet Process Mixture with $(\mu_\alpha, \sigma_\alpha^2) \sim DP(B, G_0)$ and $G_0 \equiv NIG(0, \sigma_\alpha^2/0.1, 0.01/2, 0.01/2 * 0.01)$. A circular dot, where observable, indicates the posterior median alpha for the particular fund.

Comparing the posterior HPD intervals from the three panels of Figure 1, it is clear that an investors assumptions about the population distribution of skill affects their posterior beliefs about a particular mutual fund skill level. Investors in Panels (b) and (c) draw on the performance of other funds when making an informed guess about the stock-picking ability of a particular fund. By borrowing information from other funds the HPD intervals in these two panels are much tighter than those found in Panel (a). As a result the investors of Panel (b) and (c) are more certain about a funds future ability to produce excess returns than Panel (a).

Panel (a) investors are those who do not believe skill is distributed over the population of mutual funds. As a result they are limited to the idiosyncratic performance of a fund when inferring a funds potential skill. Hence, their posterior understanding about a particular fund's level of skill is heavily influenced by the length of the funds performance window. This is evident in the large and noisy HPD intervals of the short lived funds found at the top of Figure 1(a), but it also appears in the noisy HPD of the long lived funds.

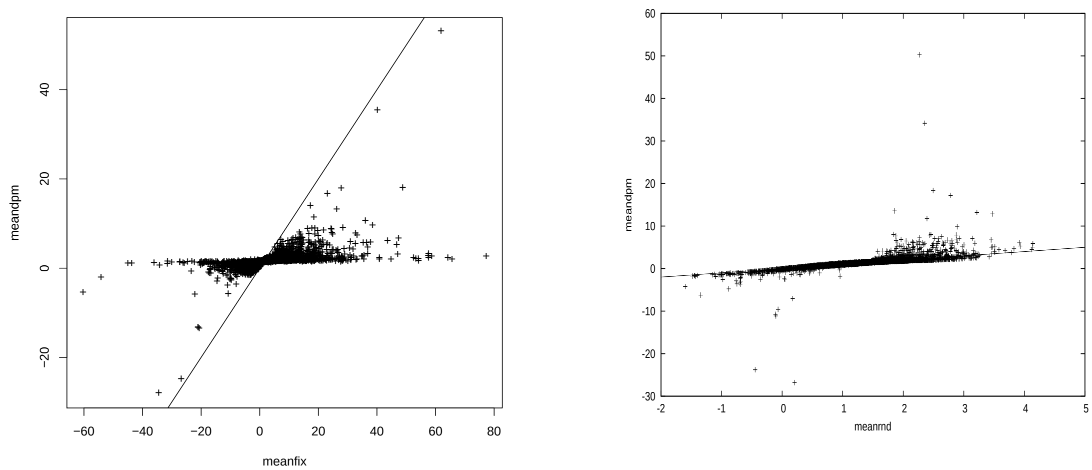
At the other end of the spectrum are the tight and uniform posterior HPDs found in Panel (b) of Figure 1. Believing a fund's performance comes from a normal, cross-sectional, distribution with an unknown population mean and variance, investors in Panel (b) learn about the average stock-picking ability of the population across both old and new funds, surviving and extinct funds, and from both exceptionally skilled and unskilled funds. Posterior beliefs with an assumed symmetrically distributed population distribution having only one mode shrinks every funds alpha towards the population mean. The larger the number of different types of funds the more the investor mistakenly learns from dissimilar funds. Hence, it is not surprising that the HPD of a short lived fund in Panel (b) looks so similar to that of a longer lived fund. By treating all the alpha as draws from a distribution with the same average ability, poor performing funds will do better than they should while highly skilled funds do worse.

In contrast to Panel (b), the investors of Panel (c) borrow information from only those funds belonging to the same similarly skilled group of funds. An extreme example of this is the Schroder Ultra Fund whose alpha has the largest posterior mean of all the funds at 50 percent per annum. Since the next closest skilled fund is the Turner Funds Micro Cap Growth fund with an alpha of 33 percent, the Schroder Ultra fund is likely the member of a small subpopulation. Our investor relies on the performance of a few other funds when making inference about the potential ability of the Schroeder fund. Nor does the performance history of the Schroder Ultra funds have much bearing on our investors belief in the ability of lesser skilled funds. Essentially, our investor treats the Schroder Ultra fund as a truly unique skilled fund, similar to the investor in Panel (a) who views each fund's ability in isolation from other funds. Hence, the Schroder Ultra fund's posterior HPD and median in Panels (a) and (c) look very similar. This contrasts with Panel (b) where the Schroder Ultra fund is not even in the top ten highest performing funds. Panel (b) type investors mistaken the Schroder Ultra fund as belonging to a group of funds consisting of all the other funds, both skilled, unskilled and break-even funds, and thus, fails to discern its exception stock-picking ability from an average fund.

To determine how much a specific mutual fund's alpha is affected by the investor's beliefs about the cross-sectional distribution of mutual fund performance, in Figure 2 we graph two scatter-diagrams containing each of the 5,136 fund's posterior average alpha as determined by our DPM type investor against the other two investor's posterior average alphas. In Panel (a) we plot on the horizontal axis the posterior mean of alpha for the investor who believes skill is idiosyncratic to the fund; i.e. the OLS estimate of a fund's alpha. On the horizontal axis of Panel (b) we graph the posterior mean of each alpha for the investor who believes skill is normally distributed across the universe of mutual funds and has to infer the population mean and variance. The forty-five degree line in both panels helps to locate where the assumptions about the cross-sectional distribution do not affect the posterior inference of a fund's skill level.

In Panel (a) of Figure 2 every mutual fund's expected level of skill has moved, to varying degrees, away from the posterior beliefs of the investor who views each fund in isolation towards zero; i.e., the points have moved vertically away from forty-five degree line towards the x-axis. Hence, the investor who believes there is a unknown cross-sectional distribution of skill underlying each funds performance level, and learns about it, discovers that funds identified by the agnostic investor as being skilled (unskilled) are actually less (more) capable of selecting stocks that beat the market. However, there are a handful of funds so uniquely skilled that treating them in isolation only slightly changes our investor's opinion about the fund's stock-picking ability. These unique funds are those with a posterior mean alpha closest to the forty-five degree line in Panel (a) and include both skilled and unskilled funds. In general, there are fewer extraordinary funds when the investor does not overfit skill but treats each fund's performance as a draw from an unknown distribution. By flexibly learning about the population distribution of skill our investor is able to identify actual fund-specific performance skills while guarding against the overfitting common in the performance measures for short history funds.

In contrast to Panel (a), many of the points in Panel (b) of Figure 2 lie on the forty-five degree line. These points belong to funds having the same average ability and variance



(a) (b)

Figure 2: Scatter plots of the 5,136 points of each mutual fund's posterior mean alpha where along the y-axis of panel (a) and (b) are the posterior mean for the investor who learns across fund performances about the entire cross-sectional distribution of skill. In the panel (a) the x-axis represent the posterior mean of alpha for the uninformed investor who treats each fund's skill independently from the others. The x-axis in panel (b) are the posterior mean of alpha for the investor who also learns across funds but assumes skill is normally distributed over the mutual fund population. To provide a point of reference in each panel a 45-degree line over the x-y space is drawn.

as that in a normal cross-sectional distribution with only one overall mean and variance. Modeling the performance of these funds as draws from a normal cross-sectional distribution does not interfere with the expected value of their posterior alpha. However, identifying this group of funds a priori would be difficult, something our investor does naturally as he learns about the entire cross-sectional performance of mutual funds.

From the off-diagonal points of Figure 2(b) we discover there are alphas being drawn from sub-populations whose means and variances are different from the global moments. Without additional sub-populations many abnormally good and bad stock-picking funds would go undetected. For example, an investor who learns the population distribution arrives at the posterior belief that the Potomac OTC/Short fund on average loses in excess of the market 27 percent a year (this is the worse performing fund in our panel of mutual funds and is the point at the bottom of each panel of Figure 2). In stark contrast, an investor who believes skill is normally distributed over the population finds the Potomac OTC/Short fund producing on average a small excess market return. The Potomac fund is not unique. There are many other extraordinary skilled and unskilled funds that look quite ordinary to the investor who believes skill is normally distributed across funds. Clearly a normally distributed cross-section has a large bearing on an investors beliefs about an abnormal fund's stock-picking ability.[18]

5.2 Cross-sectional performance

Mutual fund skill is thus neither idiosyncratic, nor is it normally distributed across the universe of mutual funds. Instead, cross-sectional mutual fund performance is distributed somewhere in between these two extremes. It is important then to have flexible posterior beliefs about the cross-sectional distribution of mutual fund performance in order to learn about the skill level of a particular fund. Such flexibility gives the investor the room needed to allow for subpopulations of extraordinarily skilled or unskilled funds, and for grouping together ordinary funds into their own sub-population. Better inference about a specific funds ability will also be gained with our nonparametric approach. We now compute the posterior cross-sectional distribution for the two types of investors who believe a funds alpha is a draw from a common cross-sectional distribution.

Figure 3 plots in red the density of the posterior cross-sectional distribution of alpha for the investor who learns how skill is distributed over mutual funds by computing $\widehat{\pi}_{r_1,\ldots,r_J}(\alpha)$ with Eq. (21). Also plotted in Figure 3 is the blue density for the posterior cross-sectional

[18]The performance measure of JS and Cohen et al. (2005) both suffer from this type of shrinkage toward the average of the overall population.

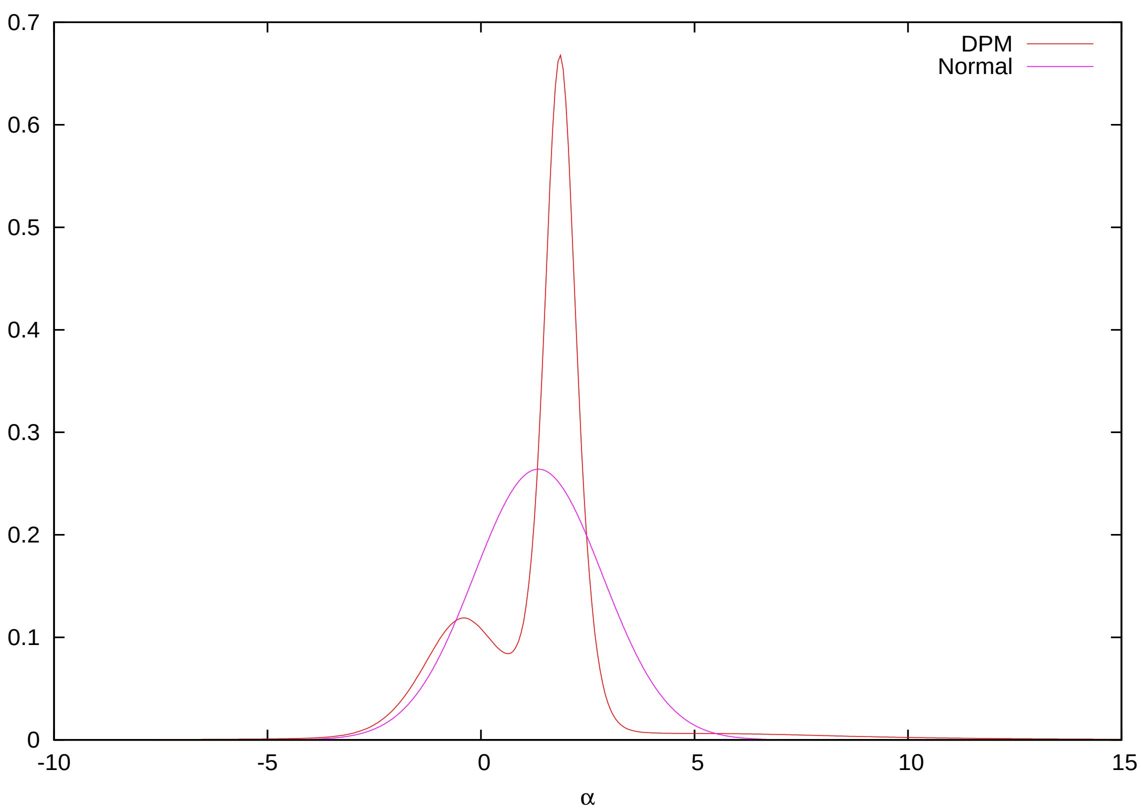


Figure 3: Posterior cross-sectional density of alpha, $\pi(\alpha|r_1, \ldots, r_{5136})$, for investors who believe the underlying distribution is normal (dashed line) and for investors who do not assume a particular distribution for alpha but have placed a Dirichlet Process Mixture prior over the unknown distribution of alpha (solid line).

distribution for those who believe skill is normally distributed but whose unknown population mean and variance are integrated out through

$$\pi(\alpha|r_1,\ldots,r_j) \approx M^{-1}\sum_{l=1}^{M} N(\alpha|\mu_\alpha^{(l)},\sigma_\alpha^{2(l)}),$$

where $(\mu^{(l)},\sigma^{2(l)})$, $l = 1,\ldots,M$, are draws from the normal models posterior $\pi(\mu_\alpha,\sigma_\alpha^2|r_1,\ldots,r_J)$.[19]

Very different conclusions about the cross-sectional performance of mutual funds are drawn from these two predictive densities. By assuming no particular form for the cross-sectional distribution, our investor identifies three distinct sub-populations of ability within the existing universe of mutual fund skill. There is a low probability sub-population of unskilled funds whose average ability results in a return slightly below the market excess return. According to this sub-population, a new fund has a twenty-two percent chance of being an unskilled fund that produces an alpha between -5% to 0.4% a year. There is a more probable sub-population whose expected ability results in an alpha of 1.8 percent, which just covers the fees of the average mutual fund.[20] A new fund has a seventy-three percent chance of belonging to this group and producing an alpha between 0.4% and 4.0%.

The last of the three sub-populations is the very low probability group of highly skilled funds diffusely distributed over the five to ten percent range. This high skilled sub-population is initially hard to see but it becomes clearer in the log predictive density plot of Figure 4. According to the posterior predictive density an arbitrary fund selected by our investor, whose performance either doesn't exist or cannot be observed, has a three percent chance of returning the investor an excess risk factor adjusted return of between four to ten percent a year. This is large relative to the less than a half a percent chance the fund's alpha will be between -4% and -10%.

This multimodal, positively skewed, population distribution stands in stark contrast to the unimodal, symmetrical, normal, cross-sectional, distribution. As a result we find that a fund is more likely to cover its fees, be extraordinarily skilled, or less likely to be a zero-alpha fund than suggested by the normal, cross-sectional, performance, distribution.

In Table 1, the quantiles, standard deviation, skewness and kurtosis are listed for both

[19]The third type of investor from Section 5.1 does not believe the alphas come from a common cross-sectional distribution. Instead, each funds alpha is viewed in isolation and, hence, such investors have nothing to say about the cross-sectional distribution of mutual fund performance.

[20]Chen & Pennacchi (2009) report the average mutual fund's expense fee is 1.14 percent, whereas Berk & Green (2004) choose a slightly higher management fee of 1.5 percent to account for costs not included in the fee when parameterizing their mutual fund model. We perform our analysis with the larger fee of 1.5 percent to compensate for missing trading costs. Wermers (2011) from ICI estimates actively managed mandates expenses and transactions costs of mutual funds and hedge funds amount to at least two percent a year.

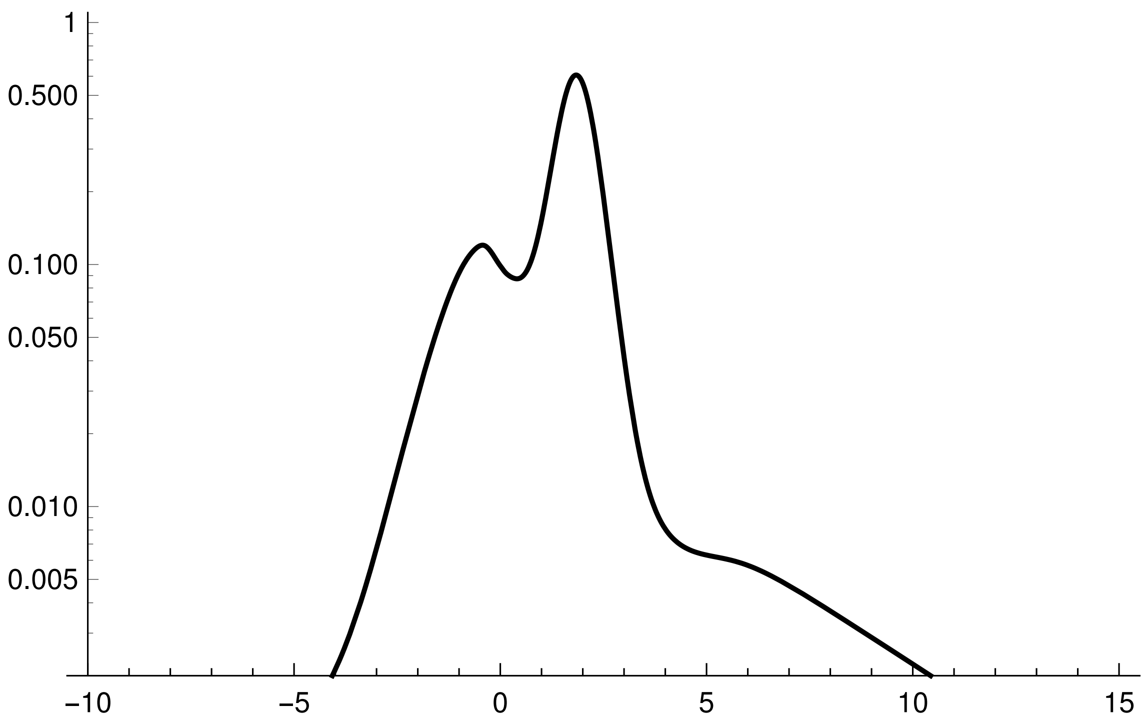


Figure 4: The log, posterior, cross-sectional, density of alpha for the investor who learns how skill is distributed across mutual funds. This density represents the log likelihood the alpha of a fund randomly picked from the population, with no return history, will take on.

	Percentiles									
	0.01	0.05	0.1	0.5	0.9	0.95	0.99	SD	Skew	Kurtosis
DPM	−2.90	−1.40	−0.83	1.43	2.46	3.19	11.48	2.36	4.31	101.49
Normal	−2.19	−1.15	−0.60	1.34	3.28	3.83	4.87	1.51	−0.0002	3.02

Table 1: Posterior cross-sectional percentiles, standard deviation (SD), skewness, and kurtosis for the investor who believes the underlying distribution of skill is normally distributed (Normal) and the investor who learns how skill is distributed across mutual funds (DPM).

posterior cross-sectional distributions. Each distribution's overall mean is slightly less than 1.5 percent a year. Having approximately the same population mean is expected since pooling together the average skill level from the three sub-populations is the average ability of the entire population.

Because one of our distribution's three modes is above five percent, the cross-section is positively skewed, and thus, puts more probability on a fund being skilled than does the posterior population distribution under the assumption of normality. Hence, the likelihood of an arbitrary fund being skilled is higher than previously thought. Our flexible population distribution is also more fat tailed, with a kurtosis of 101.49, than under a normal cross-section. So there is a greater chance that an arbitrary fund will be extraordinarily skilled, or unskilled, than one would expect under a normally distributed population. Distributional properties like positive skewness and excess kurtosis are just a couple of reasons for flexibly modeling the cross-sectional distribution of mutual fund performance.

A natural question to ask of the two investor's cross-sectional distribution is, what is the probability an arbitrary fund possesses enough stock-picking talent to generate an alpha in excess of its fees, and how different will this probability be for the two different type of investors? Both investors believe there is a 81 percent chance an arbitrary fund will generate an alpha greater than zero. Such a high probability is not surprising since JS also found strong evidence of skill with the same panel of mutual funds. However, when fees and expenses are considered our investor believes there is a 57 percent chance the returns from an arbitrary fund will be large enough to cover the 1.5 percent average fee and expense charged by mutual funds. This compares to the 46 percent chance the investor who believes skill is normally distributed gives to the arbitrary fund. So even though the overall potential of an arbitrary fund generating a return in excess of the market is approximately the same, where this potential is located differs for the two investors.

An investor who believes the population distribution of skill is normally distributed assigns a 0.8 percent chance to a arbitrary fund being skilled enough to enhance returns by five percent or more. Because our investor believes there is a non-zero chance the arbitrary fund belongs to the group of highly skilled funds, to him the chances are greater than 0.8 percent. In fact he believes the probability of the fund's alpha exceeding five percent is four times higher at 3.2 percent. Hence, investors who believe mutual fund performance is normally distributed are less incline to invest in a mutual fund than the investor who assumes nothing about the cross-sectional distribution and learns about it.

Both posterior, cross-sectional, densities in Figure 3 are consistent with Berk & Green (2004) theoretical model where in the long run a successful fund breaks even and earns a return that just covers its expenses. This is evident in the overall means both being approximately equal to the average fee of 1.5 percent. The flexible cross-sectional distribution's primary mode is centered near this average fee, meaning a fund is most likely going to generate a market excess return large enough to cover its fees. But there are still the skilled and unskilled subpopulations. One possible explanation for the highly skilled sub-population is that it represents successful funds that have not yet experienced the increasing costs that come from attracting more assets. This short-run versus long-run success argues for having a flexible, time-varying, alpha model. This is the subject of ongoing research.

The flexible posterior cross-sectional distribution's three sub-populations also supports Barras et al. (2010) and Ferson & Chen (2015) claim that funds can be separated into unskilled, zero-alpha, and skilled groups of funds. However, our flexible posterior cross-sectional distribution has a lower probability of a fund being unskilled and a greater probability a fund will break-even fund than this earlier research infers. Our cross-sectional distribution's sub-population of superior performing funds also goes against the empirical results of Ferson & Chen (2015) and Fama & French (2010), who could not find any evidence of skilled funds.

5.3 Robustness to the base distribution

This posterior view of the population distribution of mutual fund skill is for the investor whose initial level of understanding about the distribution of mutual fund skill is represented by a Student-t distribution with $\nu_0 = 0.01$ degrees of freedom, a mean, $m_0 = 0$, and scale, $(\kappa_0 + 1)\nu_0 s_0^2/\kappa_0$, where $s_0^2 = 0.01$ and $\kappa_0 = 0.1$. To validate the robustness of the posterior population distribution of alpha to the investor's choice of G_0, we estimate the posterior cross-section of mutual fund skill using larger values of the degrees of freedom hyperparameter ν_0. Larger values of ν_0 reflect a lower willingness by the investor to learn about the population distribution of alpha. We find that when ν_0 is less than 0.6 the posterior cross-sectional distribution of alpha is no different from Figure 3. However, when ν_0 is greater than or equal to 0.7, the posterior population distribution is no longer multimodal. While the cross-sectional distribution of alpha still consists of a mixture of multiple sub-populations, the posterior distribution is unimodal, centered over an alpha of 1.4 percent, with a skewness of 5.02 and a kurtosis of 110.

One possible explanation for the sensitivity to ν_0 is found in the inter-quartile range of the investor's prior predictive distribution. The inter-quartile range of the investor's initial population distribution, $\widehat{\pi}(\alpha)$, goes from a very diffuse 10^{126} when $\nu_0 = 0.01$, to a range of 3.7 when $\nu_0 = 0.1$, but then to a restrictive range of 0.18 when $\nu_0 = 0.6$. Applying the terminology of JS to the Dirichlet process distribution, a base distribution with a larger ν_0 is a more skeptical prior about skill where learning about different sub-populations by our investor is less pronounced. By increasing the degrees of freedom, our investor's prior predictive distribution of mutual fund skill is more certain about how skill should be distributed over the population of funds. This inflexibility limits the learning by the investor causing him to fail to identify different sub-populations and unique performers. Instead, a wider spectrum of stock picking ability will be grouped together into fewer sub-populations.

A more flexible, but nonconjugate, distribution is to endow our investor with the base distribution $G_0(\mu_\alpha, \sigma_\alpha) \equiv N(\mu_\alpha|0, s_\mu^2)\mathrm{SM}(\sigma_\alpha|1/2, 2, A/\sqrt{3}))$ where

$$f_{SM}\left(\sigma_\alpha \middle| 1/2, 2, A \middle/ \sqrt{3}\right) = \frac{3A\sigma_\alpha}{(A^2 + 3\sigma_\alpha^2)^{3/2}},$$

is the density function to the Singh & Maddala (1976) distribution, $\mathrm{SM}(\sigma_\alpha|1/2, 2, A/\sqrt{3}))$. The Singh & Maddala (1976) distribution is an appealing base distribution for σ_α since it allows for more weight over values of σ_α close to the zero than the Inverse-Gamma distribution. Diffuse priors were then applied to the hyperparameters, s_μ^2 and A, so that their impact on the investor's predictive of the population distribution of alpha could be integrated away. Because of the nonconjugate nature of this base distribution we need to use an alternative algorithm to Section 4 to sample the Dirichlet process distribution unknowns. Using Algorithm 8 of Neal (2000) we find the same multimodal posterior cross-sectional distribution plotted in Figure 3.[21]

5.4 Survivors cross-sectional distribution

Since the cross-sectional densities of Figure 3 are conditional on the past performance of funds that are either still in business or no longer in business, the above result do not suffer

[21] These posterior results are available upon request.

from the surviorship bias pointed out by Elton et al. (1996). In Figure 5 we plot our investors posterior cross-sectional distribution, using only the return histories of the 3,843 funds that survived to the end of the sample, against the above posterior cross-sectional distribution. The two largest modes in the surviving fund's cross-sectional distribution have tighten up around the means of the sub-populations from the original predictive distribution. This helps to clearly delineate the skilled from the unskilled group of funds. Also noticeable in the densities of Figure 5 is the thinner left hand tail when the dead funds are dropped from the analysis. This thinner tail indicates a large share of the extraordinary unskilled funds are no longer in business, but even for the funds that are still in business there is the potential for losing money. Lastly, the probability of the arbitrary fund belonging to the extraordinary group of skilled funds has increased. Hence, removing the dead funds from the investors information set has tightened up the variance of each sub-population and increased the chances the arbitrary fund which is open for business is highly skilled. It did not, however, eliminate the potential for an arbitrary operating fund from being poorly managed and losing money for our investor.

Figure 5 sub-population of poorly managed funds runs counter to the model of Berk & Green (2004). Theoretically rational investors would pull their money from underperforming funds, causing these unskilled funds to either go out of business or experience decreasing returns to scale enabling them to be competitive. This group of unskilled funds who have assets under management leaves open the door that some investors act irrationally when investing in mutual funds (see Gruber (1996)), but that also short term poor performance is tolerated by investors.

5.5 Cross-sectional evolution

Beginning in the year 1993 the number of new mutual funds entering into the asset management business began to accelerate. In the following years more than three-hundred new mutual funds were opening for business each year. This number peaked in 1998 with 659 new funds opening up for business. The large number of new mutual funds opening up for business each year, along with our investor viewing each funds performance as a realization from the unknown population distribution of mutual fund skill, provides us with an opportunity to analyze the evolution in the stock-picking ability of the mutual fund industry and also determine if the new funds were skilled or not.

Starting in 1981 and moving in yearly increments to the year 2000, we estimate the cross-sectional distribution of alpha using the return histories of all the funds ever to have existed up to the specified year. Any new mutual funds were only included if they had at

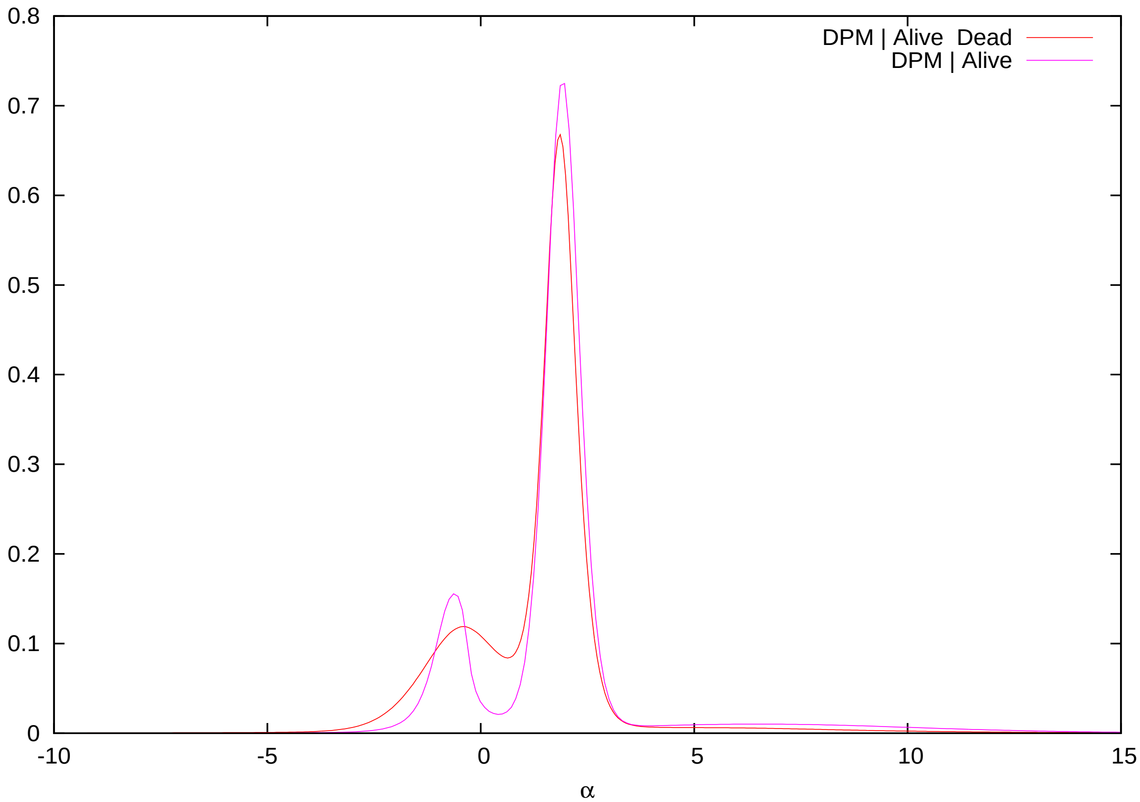


Figure 5: Posterior cross-sectional densities of alpha when investors have a Dirichlet Process Mixture prior for the unknown cross-sectional distribution of alpha and condition their beliefs on the information found in the returns of both existing and extinct funds (red line) and other investors who condition their predictions using only the returns from funds that exist at the end of the sample (blue line). Both densities are calculated from Eq. (21) with $J = 5,136$ for the information set that includes the returns of all funds and $J = 3,843$ when it only includes the returns of those fund existing at the end of the sample.

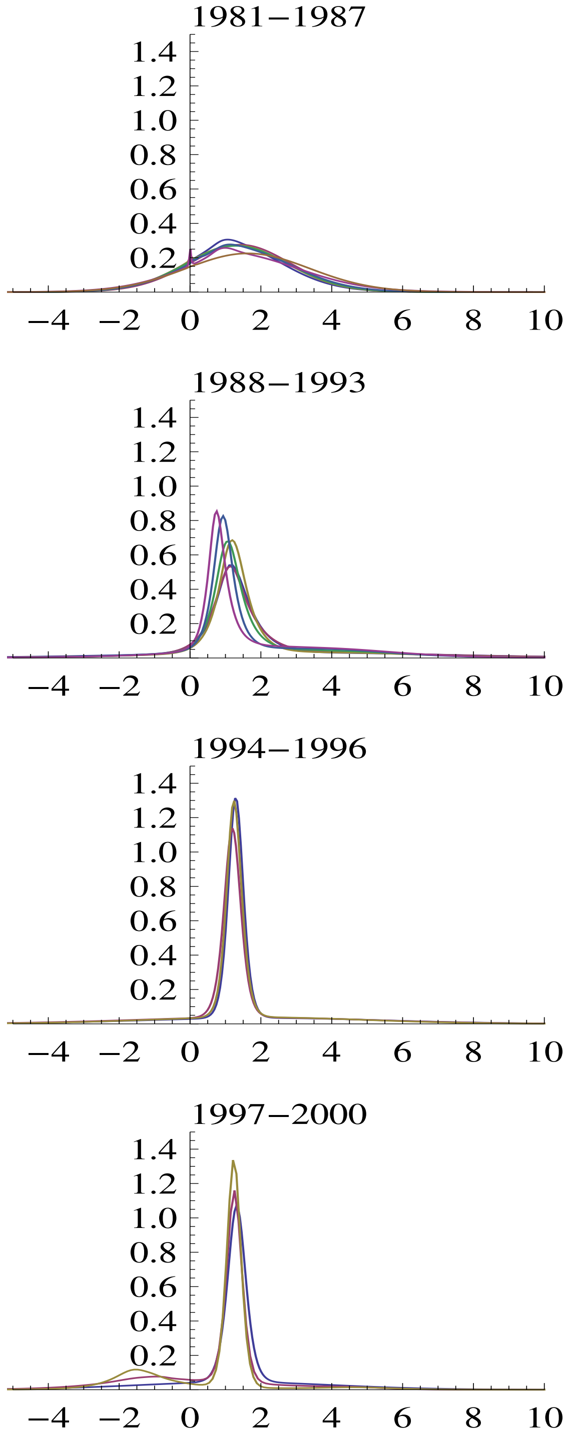


Figure 6: Posterior cross-sectional densities of alpha for, first, a panel of mutual fund that existed during the 1961 to 1981 period, and then increasing the panel by each year's new funds from 1982 to 2000 for the new funds with at least four months of returns. More formally, posterior estimates of the cross-sectional distribution, $\pi(\alpha|R_t)$, $t = 1981, \ldots, 2000$, where R_t is the history of returns from 1961 up until year t of those mutual funds that had existed up to year t. A fund is included if it has at least four months of performance history.

least 4-months worth of returns. What we find is four different episodes or eras of population distributions of mutual fund skill. We plot the four eras population distributions in the four panels of Figure 6. The four eras are i) 1981 to 1897, ii) 1988 to 1993, iii) 1994 to 1996, and iv) 1997 to 2000.

In the first panel of Figure 6 we plot the seven posterior cross-sectional distributions from the growing number of fund histories beginning in 1981 and ending in 1987. Each distribution is symmetrical around the average fund fee of 1.5 percent. This symmetry indicates our investor finds funds to be on average able to cover their costs, but also equally likely to produce abnormally high or low returns. No sub-populations of skill exist in this panel of mutual funds for this time period, so the investor assigns a new fund to the same population as the existing funds.

As entry into the mutual fund industry begins to accelerate from 1988 to 1993, the second panel of Figure 6 shows our investor becoming more and more certain about the level of skill possessed by an arbitrary fund. With each year's inflow of new funds, the primary mode of the population distribution tightens around a value of alpha approximately located at the average mutual fund fee. During this time period the probability a fund selects stocks that will result in an abnormally negative alpha declines relative to the earlier era as the left hand tails of the distribution are now thinner. Highly skilled funds are also on the increase as the right hand tail of the cross-sectional distribution pushes out past six percent to eight percent. Hence, entry of new funds during this time period and performance of existing funds improved the overall performance of mutual funds.

From 1994 to 1996 the population continues to tighten around the average fee charged by funds. However, after 1997 the distribution of stock-picking ability begins to change. In the bottom panel of Figure 6, the population distribution during the years of 1997 to 2000 starts to become skewed to the left. Ultimately a second mode appears in the cross-sectional distribution at abnormally negative values of alpha. There are also more sub-populations among the universe of mutual funds. This era corresponds to the fastest growth period of the mutual fund industry. According to the population distribution the large number of funds entering the business caused some funds to have very poor stock-picking ability, hence, the negative mode.

In Figure 7 we plot the population distributions of alpha from 1995 to 2001 using only the return histories of those funds that opened for business during the 1993 to 2001 time period.[22] Funds that were new to the business were more likely to generate a positive alpha as seen in the positive primary mode. However, over this same time period the probability

[22] A new fund was only included if it had twelve months of return performance.

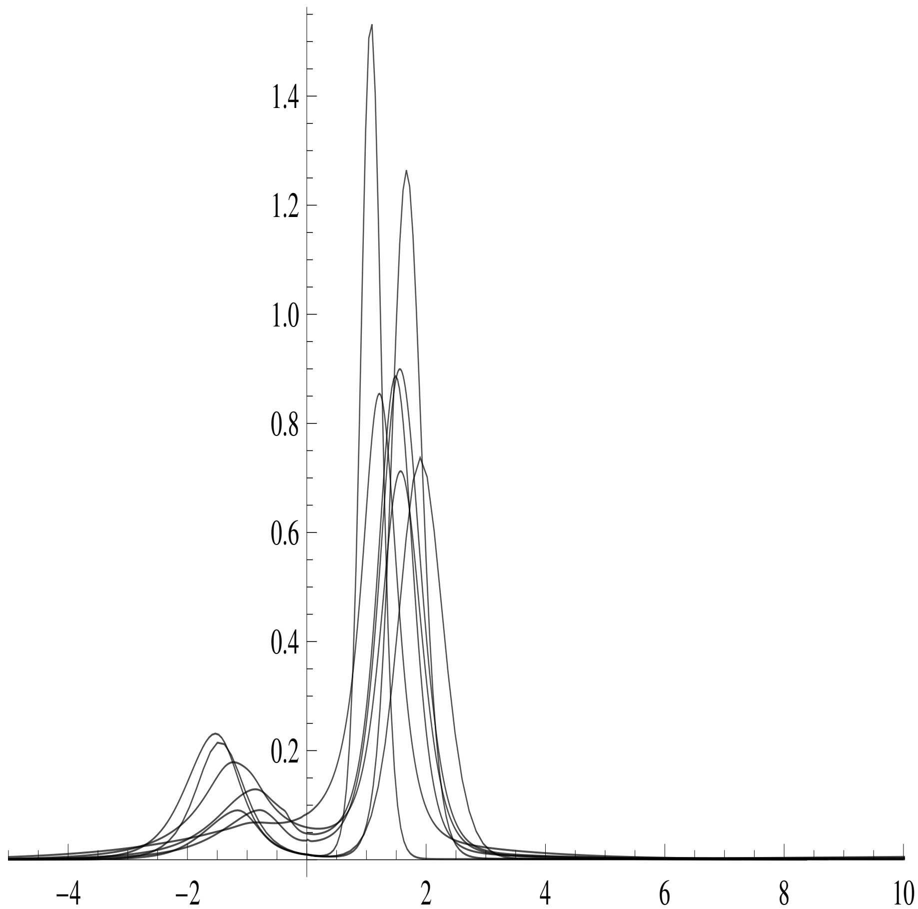


Figure 7: Posterior cross-sectional densities of alpha from 1995 to 2001 using only the performance histories of funds that entered the mutual fund business after 1992 and had a years worth of performance data.

Year	J	K	Median	Skewness	$Q_{0.05}$	$Q_{0.95}$	$P(\alpha > 0)$
1981	328	1	1.220	0.068	−1.139	3.700	0.812
1982	348	1	1.314	0.262	−0.957	4.069	0.853
1983	382	1	1.296	0.042	−1.121	3.774	0.816
1984	432	1	1.318	0.065	−1.122	3.888	0.820
1985	487	1	1.395	0.011	−1.173	3.982	0.816
1986	577	2	1.234	0.379	−1.342	4.987	0.838
1987	665	1	1.595	0.078	−1.288	4.654	0.828
1988	778	2	1.305	1.041	−1.795	7.516	0.892
1989	854	2	1.329	1.016	−1.368	7.703	0.900
1990	902	2	1.279	0.571	−0.951	6.203	0.921
1991	983	2	1.185	1.124	−0.782	6.015	0.920
1992	1073	3	1.036	−0.521	−1.096	5.662	0.914
1993	1258	3	0.963	−0.477	−0.846	5.970	0.918
1994	1599	2	1.303	0.861	−1.151	5.373	0.922
1995	1939	2	1.210	0.544	−1.711	5.194	0.902
1996	2275	2	1.270	0.608	−1.183	4.890	0.918
1997	2704	2	1.314	−0.064	−1.540	4.150	0.900
1998	3364	3	1.164	−4.182	−2.695	3.065	0.803
1999	3977	4	1.160	−1.448	−2.188	2.119	0.778
2000	4539	3	1.444	4.584	−0.766	4.719	0.927

Table 2: Yearly evolution of the cross-sectional mutual fund performance distribution's median, skewness and probability of beating the passive four-factor portfolio, $P(\alpha > 0)$, where J is the number mutual fund having existed up to that year, K is the posterior median number of sub-populations and $Q_{0.05}$ and $Q_{0.95}$ are the 5 percent and 95 percent quantiles.

of a new fund generating a negative alpha is increasing as the negative mode moves further to the left. Thus, we come to the conclusion that during the later half of the 90s when the number of new funds entering into mutual fund industry was accelerating, a new fund was likely to be skilled and capable of covering its fees, but with each year there was an increasing chance the new fund would be unable to earn a high enough return to justify its fees.

Table 2 lists the characteristics and features of each cross-sectional distribution of skill over the years 1981 through 2000. Each line contains the total number of funds, both in, and out of business since 1961 up to that year, J, the median number of sub-populations, K, the median and skewness, and the 5 percent and 95 percent quantiles, $Q_{0.05}$ and $Q_{0.95}$, of the cross-sectional distribution, and the probability an arbitrary mutual fund generates a positive alpha, $P(\alpha > 0)$. Beginning in the 90s an arbitrary fund is exceptionally skilled, as defined by having an alpha in the top 5 percent of the distribution, if it generated an alpha of two to six percent. This was less of a return over the market than a highly skilled fund from the 80s. For example, over the 90s the $Q_{0.95}$ declines from an alpha of approximately 6 percent to 2 percent per annum, where as in the 80s it was never less than 3.7 percent and reached a high of 7.7 percent in 1989.

During the 90s the alpha of an arbitrary bad fund, as defined by the 5 percent quantile, $Q_{0.05}$, also declined but in a noisier fashion. Poor performance in the mutual fund industry went from −0.95 percent in 1990 to a low of −2.7 percent in 1998. With the exception of 1998 and 1999, the probability our investor chooses an arbitrary mutual fund capable of generating a positive alphas stayed right around 90%. Hence, overall mutual fund performance went down during the 90s relative to the 80s.

5.6 Individual predictive distributions

For the purpose of determining the skill level of a specific mutual fund investors form posterior beliefs about the fund's future performance. Called either the posterior predictive distribution or the posterior distribution of the ith fund's alpha, it is proportional to

$$\pi(\alpha_i|r_1,\ldots,r_J) \propto f(r_i^*|\alpha_i)\, \widehat{\pi}_{r_1,\ldots,r_{i-1},r_{i+1},\ldots,r_J}(\alpha_i), \tag{22}$$

for $i = 1,\ldots,J$, and where r_i^* has been defined in Eq. (17) as the risk and factor adjusted return history of the ith fund. After the return histories of the other $J-1$ funds have been observed, but before observing the ith funds returns, our investor's state of knowledge about the ith fund's ability is fully captured by his updated prior distribution, $\widehat{\pi}_{r_1,\ldots,r_{i-1},r_{i+1},\ldots,r_J}(\alpha)$. This updated prior is the posterior population distribution from Eq. (20) but without conditioning on observing the ith fund's empirical returns. Since J equals 5,136 funds, letting any fund be the ith fund and dropping its return history from the posterior analysis has virtually no effect on the posterior cross-sectional distribution $\widehat{\pi}_{r_1,\ldots,r_J}(\alpha)$. Hence, $\widehat{\pi}_{r_1,\ldots,r_{i-1},r_{i+1},\ldots,r_J}(\alpha) \approx \widehat{\pi}_{r_1,\ldots,r_J}(\alpha)$ for $i = 1,\ldots,J$. Applying this approximation to our investors prior beliefs about the ith fund ability in Eq. (22), our investor's prior for the skill level of the ith fund is the posterior cross-sectional density in Figure 3.[23]

As a measure of the investor's state of knowledge about a fund's level of skill, the posterior $\pi(\alpha_i|r_1,\ldots,r_J)$ expresses our investor's knowledge about the ith funds alpha in terms of the probability of its future above market risk factor adjusted returns. It is natural

[23] As written, the likelihood, $f(r_i^*|\alpha_i)$, from Eq. (22) implicitly assumes β_i and σ_i are known. In the sampler of Section 4 we have integrated out the uncertainty behind these parameters through draws from their conditional posterior distribution.

and straight forward for the investor to ask and answer the question, what is the probability the ith mutual fund will pick under priced stocks such that its future returns covers its fees? For our purposes, we choose to denote an exceptional fund as one whose past performance as measured by its likelihood $f(r_i|\alpha_i)$, separates it from the expected future performance of an arbitrary fund as measured by the population distribution, $\widehat{\pi}_{r_1,\ldots,\ldots,r_J}(\alpha)$.

Figure 8 plots the densities of every one of the 5,136 mutual funds posterior predictive distribution of alpha, $\pi(\alpha_i|r_1,\ldots,r_J)$. The predictive densities of the 3,844 mutual funds that are still open for business at the end of the sample are represented by blue lines, and the densities of the 1,293 funds no longer in business are represented by red lines. Transparency has been added to each fund's density so that darker shades of red, blue, or purple (combinations of red and blue) indicates different levels of concentration among the individual densities. To help visually identify the exceptional funds in our panel the posterior cross-sectional distribution's density from Figure 3 is plotted in orange.

Most of the predictive densities for the specific funds in Figure 8, be the fund dead or alive, resemble the multimodal, cross-sectional distribution of alpha (the orange density). This similarity between the specific and cross-sectional posterior predictive densities shows how much information about skill is contained in the updated prior $\widehat{\pi}_{r_1,\ldots,r_{i-1},r_{i+1},\ldots,r_J}(\alpha_i)$ and how much our investor relies on it when forming his beliefs about a particular funds ability. It also indicates what little information is contained in most mutual funds likelihood $f(r_i|\alpha_i,\beta_i,\sigma_i^2)$.[24] For the 'average" fund whose performance is typical and has a likelihood which is relatively flat, the updated cross-sectional distribution resolves most of the uncertainty around the number of modes and their locations for the 'average' funds posterior distributions of skill.

We draw three conclusions from Figure 8 about the ability of the specific funds found in our data set. First, regardless of a fund being in or out of business, its predictive distribution has in general two modes. The smaller of these modes is located at an alpha close to minus one percent, indicating that they have the potential to generate losses, costing the investor not only the active mutual fund's management fee but also the foregone return from investing in the passive risk factor portfolio. These same funds also have a primary predictive mode that is close to an excess market return of two percent. A primary mode of two percent along with a second mode at a negative gross alpha informs the investor that such a fund is likely to cover its fees but there is also a non-trivial chance the fund will fail to cover its costs and could possibly loss money for the investor.

[24]This same point has been made by the least squares literature where the R^2s for the fitted risk-factor models are small for most funds.

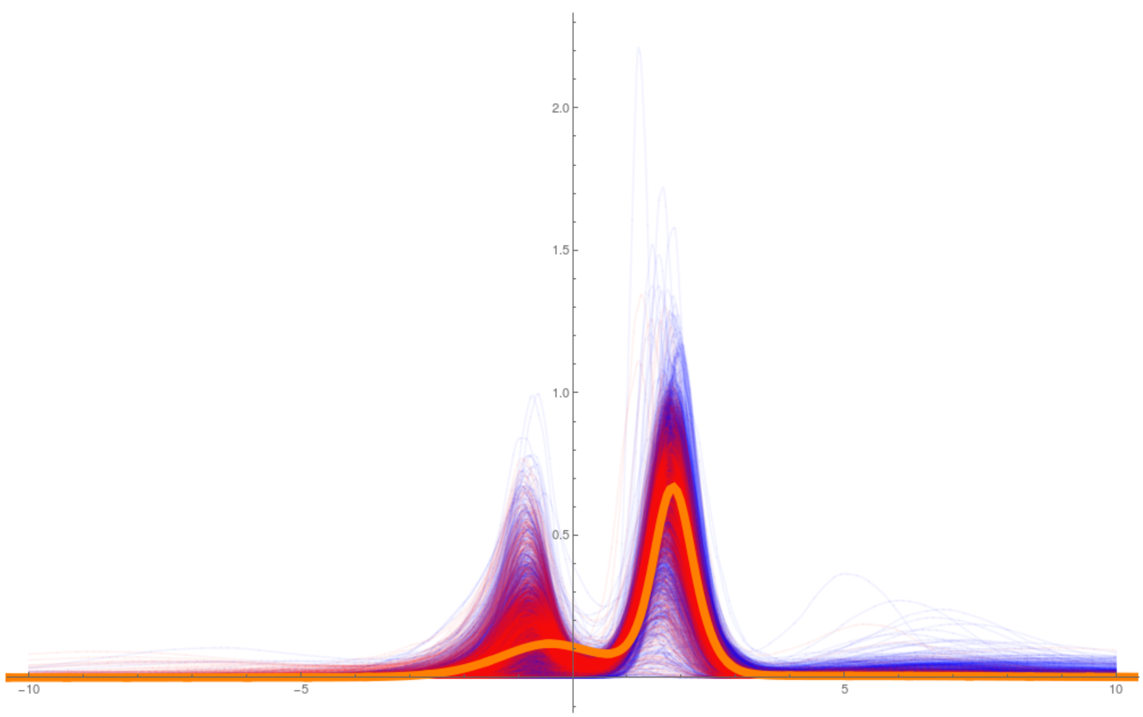


Figure 8: Posterior cross-sectional distribution of alpha in orange plotted against every fund that existed from 1962 to 2001 posterior predictive distribution of alpha, $\pi(\alpha_i|r_1,\ldots,r_{5136})$, $i=1,\ldots,5136$, where the funds that are still in business are plotted in blue, whereas those that are no longer in business are in red. Darker shades of red or blue indicate a higher concentration of funds having similar shaped densities.

Second, in the individual predictive densities of Figure 8 there is credible evidence of exceptional funds, either skilled and unskilled, existing in the population. Later we point out that these funds are truly exceptional since they are so few in number – twenty-one skilled and fifty unskilled. Finding so few exceptional funds runs counter to earlier empirical mutual fund results where there is a larger presence of skilled and/or unskilled funds (see Kosowski et al. (2006), Fama & French (2010), and Barras et al. (2010), and Ferson & Chen (2015)). However, as we have already pointed out in Section 5.1, these earlier findings on extraordinary fund performance suffer from noisy alpha estimates. Hence, most mutual funds are not extraordinarily talented or unskilled. Instead, a fund has a greater chance of being just talented enough to select stocks that on average result in a return that justifies its costs, expenses and fees.

In Figure 8, the funds with superior stock picking skills are those whose predictive densities have sizable probability over values of alphas between five to ten percent. A few of these exceptional funds have modes located at alphas larger than the cross-sectional distribution. For instance, in Panel (b) of Figure 9 we plot the individual densities of the posterior distribution of alpha for each of the twenty-one mutual funds who have at least a 95% chance of its alpha being greater than the average fund fee of 1.5%. Half of the twenty-one skilled funds has a primary mode near 1.5%, and only one is no longer in business. In Table 3 we list from shortest to longest return histories each of the twenty-one skilled funds, its years of operation, and the fund's posterior mean alpha. Four of the funds possess the extraordinary ability to pick stocks such that they are expected to result in an alpha that is greater than 15% per year. However, there is sizable uncertainty around this potential as seen in the four funds flat and diffuse predictive distributions.

The apparent success of these twenty-one highly skilled mutual funds helps answer the question asked by Kosowski et al. (2006) and Fama & French (2010) whether such funds are genuinely skilled or just lucky. We find luck playing no role in the success of these twenty-one skilled funds. Because the posterior cross-sectional distribution of skill is so informative about the typical fund's ability when the investor has a large panel of mutual funds performance, his knowledge about the skill level of an exceptional fund is distinctly different from what he believes about the cross-section. Referring back to Panel (a) of Figure 2 in Section 5.1, we see the posterior means of $\pi(\alpha_i|r_1,\ldots,r_J) \propto f(r_i|\alpha_i)\widehat{\pi}_{r_1,\ldots,r_{i-1},r_{i+1},\ldots,r_J}(\alpha_i)$, (on the vertical axis) shrinking the means of $\pi(\alpha_i) \propto f(r_i|\alpha_i)$ (the horizontal axis) of the highly skilled funds back towards the the cross-sectional distribution's primary mode of 1.5 percent. However, because of the highly skilled fund's superior performance histories, their likelihoods, $f(r_i|\alpha_i)$, enables them to escape the cross-sectional distribution. Each of these

Fund	Years of Operation	Alpha
Turner Funds: Micro Cap Growth	1998-2001	33.07
Schroeder Ultra Fund	1997-2001	49.77
Artisan Mid Cap Fund	1997-2001	16.77
Needham Growth Fund	1996-2001	17.40
Olstein Financial Alert Fund/C	1996-2001	13.84
Fremont Mutual Fds:US Micro Cap Fund	1994-2001	13.23
PIMCO Funds:Stocks Plus Fund/Instl	1993-2001	2.46
Fidelity Dividend Growth	1993-2001	4.87
Managers Funds:US Stock Market Plus	1992-2001	2.44
Fidelity Low Priced Stock	1990-2001	5.93
Victory Funds:Diversified Stock Fund/A	1989-2001	2.88
T Rowe Price Capital Appreciation Fund	1986-1999	3.75
JP Morgan Growth & Income Fund/A	1987-2001	6.30
Gabelli Growth Fund	1987-2001	5.29
Weitz Series Fund:Value Portfolio	1986-2001	4.50
IDEX Janus Growth Fund/A	1986-2001	4.87
Gabelli Asset Fund	1986-2001	4.99
Oppenheimer Growth/A	1973-2001	3.82
AXP Growth Fund/A	1972-2001	4.71
Janus Fund	1970-2001	3.90
Vanguard Morgan Growth/Inv	1968-2001	4.40

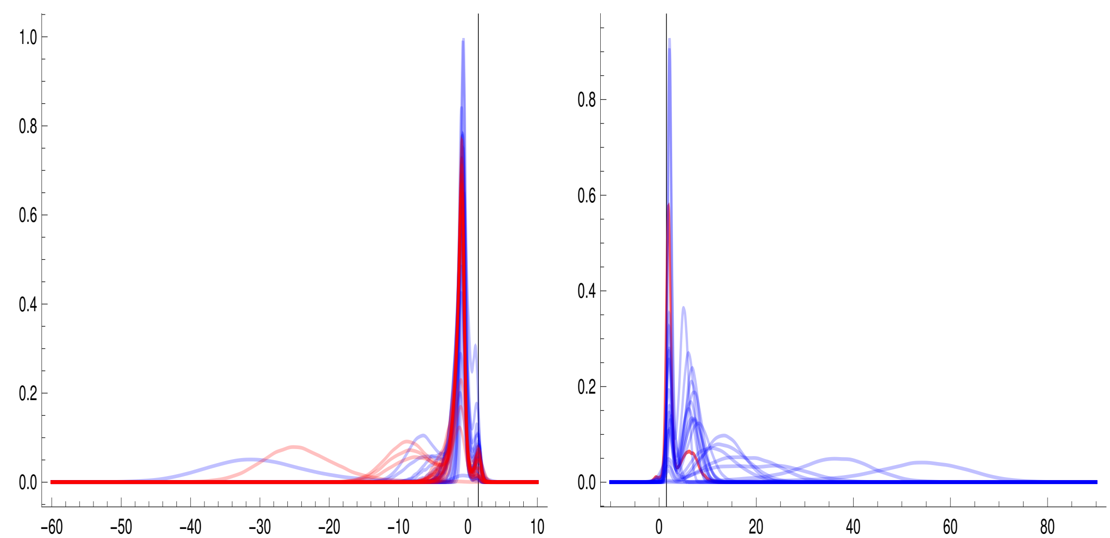
Table 3: Individual mutual funds who have at least a 95 percent chance of returning an investor a market excess return larger than the average mutual fund fee of 1.5 percent, their average posterior alpha, from the newest to oldest.

skilled funds distinguishes itself in Panel (b) of Figure 9 from the population by having a mode of five percent or higher. Hence, there are a few truly skilled funds among our cross-section of mutual funds.

Our third finding concerns the stock-picking ability of funds that are unskilled but are not dead. A number of predictive densities in Figure 8 have fat tails over negative values of alpha, but not all of them are red. There are plenty of blue densities that place a high level of probability on alpha being negative. In Panel (a) of Figure 9 we plot the individual densities of the fifty funds that have a 95% chance or higher of generating an alpha less than the average fee of 1.5%. Twenty-nine of these poor performing funds are still in business at the end of the sample (represented by the blue densities in Panel (a)). This includes the worse performing fund in our cross-section, the Potomac OTC/Short fund, whose expected alpha is -26% and has a 3% chance of losing the investor between twenty to forty percent a year. Except for the two closed funds, Bowser Growth Fund, whose expected alpha is -24%, and Ameritor Industry Fund, with an expected alpha of -7%, the other forty-seven unskilled funds engage in trades that are expected to result in gross losses of between zero to five percent a year. Nine of these are funds have been in business since the early 1960s and are still in business. So, poor performance by a fund does not necessarily lead to investors divesting their money from a unskilled fund. Perhaps there are restrictions placed on the fund investors prohibiting withdrawals or limiting redemptions. In other cases, investors may not be paying attention to the losses on their mutual fund investment because they were not the original investor.

6 Conclusion

By allowing investors to learn the cross-sectional distribution of mutual fund skill and not just its unknown mean and variance, we find the population cross-sectional distribution of mutual fund performance consists of three sub-populations. The most likely sub-population has a mean gross alpha of 1.8 percent per year, followed by an unskilled group of funds whose average alpha is -0.5 percent, and lastly, a low probability, but high performing, sub-population whose average performance is greater than five percent a year. Since the cross-sectional distribution is approximately the prior for a particular mutual fund's posterior predictive distribution of skill, we find 22 (50) funds out of our panel of 5,136 mutual funds have a 95% chance or higher of (not) producing a gross alpha that exceeds the average fee charged by a fund. Hence, the performance history of most funds is not more informative than the cross-sectional distribution, but there are few skilled and unskilled funds.



(a) Unskilled funds

(b) Skilled funds

Figure 9: Specific fund's densities for the posterior predictive distribution of future skill where Panel (a) are the fifty unskilled fund's who have at least a 95% chance its alpha will be less than 1.5 percent, and Panel (b) are the twenty-one skilled funds who have at least a 95% chance its alpha will be greater than 1.5 percent (the vertical line in both plots is at the average fund's fee of 1.5 percent). Densities for mutual funds that are in business at the end of the sample are plotted blue, whereas the densities for funds that have closed are in red. In Panel (a) there are a total of twenty-one dead unskilled funds, and in Panel (b) there is only one dead skilled fund (the T. Rowe Price Capital Appreciation Fund).

The approach we have taken here to learn how skill is distributed over the cross-section of mutual funds can also be applied to other finance related questions. Bayesian updating of an unknown distribution is applicable to any finance problem where one is interested in making inference about a population distribution. For example, one could estimate the unknown distribution of a time varying CAPM beta in order to forecasts beta into the future by sampling from its updated posterior for future time periods. We are currently investigating this and other similar types of research ideas from the standpoint of an investor who learns about an unknown distribution.

References

Avramov, D. & Wermers, R. (2006), 'Investing in mutual funds when returns are predictable', *Journal of Financial Economics* **81**, 339–377.

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

Barras, L., Scaillet, O. & Wermers, R. (2010), 'False discoveries in mutual fund performance: Measuring luck in estimated alphas', *The Journal of Finance* **65**(1), 179–216.

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

Blackwell, D. & MacQueen, J. (1973), 'Ferguson distributions via polya urn schemes', *The Annals of Statistics* **1**, 353–355.

Burr, D. & Doss, H. (2005), 'A bayesian semiparametric model for random-effects meta-analysis', *Journal of the American Statistical Association* **100**(469), 242–251.

Busse, J. A. & Irvine, P. J. (2006), 'Bayesian alphas and mutual fund persistence', *The Journal of Finance* **61**(5), 2251–2288.

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *The Journal of Finance* **52**(1), pp. 57–82.

Chen, H.-L. & Pennacchi, G. G. (2009), 'Does prior performance affect a mutual funds choice of risk? theory and further empirical evidence', *Journal of Financial and Quantitative Analysis* **44**, 745–775.

Cohen, R. B., Coval, J. D. & Pástor, L. (2005), 'Judging fund managers by the company they keep', *The Journal of Finance* **60**(3).

Dunson, D. B. (2010), Nonparametric bayes application to biostatistics, *in* N. L. Hjort, C. Holmes, P. Mueller & S. G. Walker, eds, 'Bayesian Nonparametrics', Cambridge University Press, pp. 223–268.

Elton, E., Gruber, M. & Blake, C. (1996), 'Survivor bias and mutual fund performance', *Review of Financial Studies* **9**(4), 1097–1120.

Escobar, M. D. & West, M. (1995), 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association* **90**(430), 577–588.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3 – 56.

Fama, E. F. & French, K. R. (2010), 'Luck versus skill in the cross-section of mutual fund returns', *The Journal of Finance* **65**(5), 1915–1947.

Ferguson, T. (1973), 'A Bayesian analysis of some nonparametric problems', *The Annals of Statistics* **1**(2), 209–230.

Ferson, W. & Chen, Y. (2015), How many good and bad fund managers are there, really? Working Paper.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian Data Analysis*, Chapman and Hall/CRC.

Gruber, M. J. (1996), 'Another puzzle: The growth in actively managed mutual funds', *The Journal of Finance* **51**(3), pp. 783–810.

Jensen, M. C. (1968), 'The performance of mutual funds in the period 1945-1964', *The Journal of Finance* **23**(2), pp. 389–416.

Jones, C. S. & Shanken, J. (2005), 'Mutual fund performance with learning across funds', *Journal of Financial Economics* **78**, 507–552.

Kandel, S. & Stambaugh, R. F. (1996), 'On the predictability of stock returns: An asset-allocation prespective', *Journal of Finance* **51**, 385–424.

Kleinman, K. P. & Ibrahim, J. G. (1998), 'A semiparametric bayesian approach to the random effects mode', *Biometrics* **54**, 921–938.

Kosowski, R., Timmermann, A., Wermers, R. & White, H. (2006), 'Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis', *The Journal of Finance* **61**(6), pp. 2551–2595.

Kothari, S. & Warner, J. B. (2001), 'Evaluating mutual fund performance', *The Journal of Finance* **56**(5), 1985–2010.

Lo, A. Y. (1984), 'On a class of Bayesian nonparametric estimates. i. density estimates', *The Annals of Statistics* **12**, 351–357.

Neal, R. (2000), 'Markov chain sampling methods for Dirichlet process mixture models', *Journal of Computational and Graphical Statistics* **9**, 249–265.

Ohlssen, D. I., Sharples, L. D. & Spiegelhalter, D. J. (2007), 'Flexible random-effects models using bayesian semi-parametric models: applications to institutional comparisons', *Statistics in Medicine* **26**(9), 2088–2112.

Pástor, L. & Stambaugh, R. F. (2000), 'Comparing asset pricing models: An investment perspective', *Journal of Financial Economics* **56**, 335–381.

Pástor, L. & Stambaugh, R. F. (2002*a*), 'Investing in equity mutual funds', *Journal of Financial Economics* **63**(3), 351 – 380.

Pástor, L. & Stambaugh, R. F. (2002*b*), 'Mutual fund performance and seemingly unrelated assets', *Journal of Financial Economics* **63**(3), 315 – 349.

Singh, S. & Maddala, G. (1976), 'A function for size distribution of incomes', *Econometrica* **44**, 963–970.

Verbeke, G. & Lesaffre, E. (1996), 'A linear mixed-effects model with heterogeneity in the random-effects population', *Journal of the American Statistical Association* **91**(433), 217–221.

Wermers, R. (2011), 'Performance measrement of mutua funds, hedge funds, and institutional accounts', *Annual Review of Finanacial Economics* **3**, 537–574.