Bayesian Nonparametric Learning of How Skill is Distributed Across the Mutual Fund Industry *

Mark Fisher
Federal Reserve Bank of Atlanta
1000 Peachtree Street, N.E.
Atlanta, GA 30309-4470
United States
mark@markfisher.net

Mark J. Jensen†
Federal Reserve Bank of Atlanta
1000 Peachtree Street, N.E.
Atlanta, GA 30309-4470
United States
Mark.Jensen@atl.frb.org

Abstract. In this paper, we use Bayesian nonparametric learning to estimate the skill of actively managed mutual funds and also to estimate the population distribution of skill. A nonparametric hierarchical prior, where the hyperprior distribution is unknown and modeled with a Dirichlet Process prior, is used to model the skill parameter, with its posterior predictive distribution being an estimate of the population distribution. Our nonparametric approach is equivalent to an infinitely ordered mixture of normals where we resolve the uncertainty in the number of mixture components by learning how to partition the funds into groups according to the average ability and the variability in the skill of a group. By resolving the mixture's uncertainty, our nonparametric prior avoids having to sequentially estimate and test an array of pre-specified, finite ordered, mixture priors. Applying our Bayesian nonparametric learning approach to a panel of actively managed, domestic equity funds, we find the population distribution of skill to be fat-tailed, skewed towards higher levels of performance, with two distinct modes – a primary mode where the average ability covers the average fees charged by funds, and a secondary mode at a performance level where a fund loses money for its investors.
Keywords: Bayesian nonparametrics, mutual funds, unsupervised learning
JEL: C14, C23, C58

April 23, 2020

*We would like to thank Vikas Agarwal, Jerry Dwyer, Wayne Ferson, Kris Gerardi, Siva Nathan, Jay Shanken, Paula Tkac, and the guest editors of this special issue, Nick Polson and Jun Yu, for their feedback and suggestions. We especially thank the referees for improving the paper through their comments and input. Thanks also go to the conference participants at the 10th Annual All Georgia Finance Conference, the 2014 Southern Finance Meetings, the Economics, Finance, and Business Section of ISBA Bayes 250, the 2014 Bayesian Workshop of the Remini Centre of Economic Analysis, the 2015 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics, the 2015 European Seminar on Bayesian Econometrics, the 2017 ISBA Conference on Bayesian Nonparametric, the 2018 Workshop on Bayesian Methods in Finance held at the ESSEC Business School, and the department members at the Institution of Statistics and Mathematics at the Vienna University of Business and Economics, the DeGroote School of Business at McMaster University, Clemson University, the University of North Carolina–Charlotte, Queen Mary–University of London, the University of Montreal, and Michigan State University. The views expressed here are ours and are not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

†Corresponding Author. Tel.: +1 404 498 8019.

1 Introduction

Since the seminal article by Jensen (1968), estimating the skill level of actively managed mutual funds has been widely researched and debated.[1] In addition to estimating the skill of a fund, others have investigated how skill is distributed across the industry. For instance, Kosowski et al. (2006), Barras et al. (2010), Fama & French (2010), and Ferson & Chen (2015) take a frequentist approach and estimate the population distribution by bootstrapping the estimated skill of the funds. Chen et al. (2017) and Harvey & Liu (2018) both model the population distribution with a finite mixture of normals and estimate the mixture parameters with an EM algorithm. Barras et al. (2018) use a nonparametric method to estimate the population distribution but do not use the information from the population when estimating a fund's level of skill.

Jones & Shanken (2005) estimate the population distribution with a Bayesian model and a parametric hierarchical normal prior for skill. Others, like Pástor & Stambaugh (2002*b*), assume a parametric form for the population distribution. Baks et al. (2001), Pástor & Stambaugh (2002*a*), and Avramov & Wermers (2006) also assume a known cross-sectional distribution for skill. Each finds the estimate of a fund's level of skill to be sensitive to the choice of the population distribution.

To our knowledge, no one has estimated mutual fund skill by letting the population distribution be entirely unknown, estimating it, and using it to infer the skill level of the funds. We do this here by modeling the unknown population distribution with a Bayesian, nonparametric, hierarchical prior. Our nonparametric prior is an infinite mixture of normals whose mixture weights, locations, scales, and order are unknown. We infer these unknown mixture parameters using an unsupervised learning approach where we partition the panel of mutual funds into a finite number of groups (mixture clusters).[2] Each member of a specific group has the same average stock-picking ability and variability (mixture location and scale).[3]

We leverage these random partitions to resolve the uncertainty in the skill level of a fund belonging to a particular group by pooling the information from the group's other funds. Sharing information within the group is especially important when resolving the uncertainty in the skill level of newer funds whose performance histories are short. Partitioning the funds into different groups also eliminates the global shrinkage issues that plague

[1]See Elton & Gruber (2013) for a review of this literature.

[2]See Murphy (2012) for an introduction to unsupervised learning.

[3]Learning about the cross-sectional distribution of skill by partitioning funds into groups is similar to Cohen et al.'s (2005) idea of judging a fund by the company it keeps. However, our approach is unsupervised and, hence, does not use any information about a fund beyond its return history.

parametric hierarchical priors, and also finite mixture models, whose number of clusters is determined by testing nested mixture models. Under our nonparametric hierarchical prior, an extraordinarily (un)skilled fund is allowed to be the only member of its group and, hence, not have its estimate of skill shrunk towards the global industry average as in Jones & Shanken (2005), or towards the average skill of a (below)above-average group of funds as in Chen et al. (2017) and Harvey & Liu (2018).

Using return data from the entire actively managed, US domestic equity, fund industry, we find the population distribution of skill to be bi-modal, fat-tailed, and slightly skewed towards better stock-picking ability. Our estimate of the population distribution of skill suggests there is a greater chance of a fund being extraordinarily skilled relative to both a normally distributed population and a Gaussian mixture model with two clusters. We also find our exceptionally skilled and unskilled funds look rather ordinary when skill is assumed to be normally distributed across mutual funds.

We organize the paper in the following manner. In Section 2, we present a mutual fund investor's investment decision and how he applies Bayes rule to learn about the population distribution of skill and the skill level of a particular fund. Section 3 presents our nonparametric, hierarchical, Dirichlet process mixture, prior for the skill level of the funds and the initial population distribution from this nonparametric prior. We then describe in Section 4 the Bayesian unsupervised learning that comes from the Dirichlet process, followed in Section 5 by the model's Markov Chain Monte Carlo (MCMC) sampler. In Section 6, we apply our Bayesian nonparametric approach, a Bayesian parametric hierarchical model, and a idiosyncratic Bayesian parametric model, to a panel of 5,136 actively managed mutual funds. Section 7 summarizes our findings and provides our conclusions.

2 Investors decision

In this section, we analyze the population distribution of mutual fund performance from the perspective of rational Bayesian investors who choose between a risk-free asset, a set of benchmark assets, and an array of actively managed mutual funds. Our investors' decision differs from that in Baks et al. (2001) (BMW) where the decision to invest in a specific fund is treated independently from the deliberations around investing in the other funds. Instead, we follow Jones & Shanken (2005) (JS) and assume the investors choose to invest in a current mutual fund by analyzing the return performance of past and present mutual funds when determining skill and the population distribution.

Following Jensen (1968), the risk-free adjusted, gross returns[4] for J, past, and present, mutual funds, are assumed to follow the linear risk factor model

$$r_{i,t} = \alpha_i + \beta_i' F_t + \sigma_i \epsilon_{i,t}, \qquad i = 1, \ldots, J, \text{ and } t = \tau_i, \ldots, T_i, \tag{1}$$

where $1 \leq \tau_i$. The length of each series is $\mathcal{T}_i = T_i - \tau_i + 1$ such that $\mathcal{T}_i$ and $\mathcal{T}_{i'}$ do not have to be equal. As an unbalanced panel, the starting points, τ_i, do not need to be same for different is.

We first assume the innovations are homoskedastic, Gaussian white noise, $\epsilon_{i,t} \overset{iid}{\sim} N(0, 1)$. Later, in Section 6.4, we allow for time-varying heteroskedastic sampling distributions by modeling each fund's returns with autoregressive conditional heteroskedasticity (ARCH) variances.

We let each fund have its own stock-picking strategy. Hence, the return innovations are conditionally uncorrelated across funds such that $\text{Cov}(\epsilon_{i,t}, \epsilon_{i',t}) = 0$.[5] Under these assumptions, the sampling distribution representation of Eq. (1) is

$$f(r_1, \ldots, r_J | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \prod_{i=1}^{J} \prod_{t=\tau_i}^{T_i} N(r_{i,t} | \alpha_i + \beta_i' F_t, \sigma_i^2),$$

where $r_i = (r_{i,\tau_i}, \ldots, r_{i,T_i})'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)$, and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_J^2)'$.

The vector of passive benchmark returns, F_t, are observed by our investors at the end of each month t. Later, in the empirical application, these benchmark returns will consist of the four passive risk factors; the three-factor model of Fama & French (1993) and the momentum portfolio factor of Carhart (1997). Under these risk factors, Eq. (1) becomes

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

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 both Eq. (1) and (2), the parameter α_i is assumed to measure fund i's ability to identify under-valued stocks and is the only parameter measuring this skill. BMW show that Bayesian, mean-variance, investors invest in an existing fund if and only if the expected posterior value of α_i is greater than zero and at least as large as the fund's fees; i.e., the mutual fund is expected to outperform a cost-less portfolio comprised of the benchmark

[4] We analyze gross returns because expenses and fees vary across funds and over time, and because the management company generally sets the fees. For example, in the economic model of fund behavior by Berk & Green (2004), the model predicts the economic rents generated by a skilled fund will be captured by management through higher fees.

[5] JS relax the cross-sectional independence assumption and use a hidden factor model, which improved the precision but did not affect the location of the skill estimates.

returns, F, and cover the fund's fees. According to BMW, Bayesian investors choose to invest in a new fund if and only if the expected value of alpha over the posterior population distribution is positive and exceeds the average fee charged by the industry. How much investors invest in either a new or existing fund depends on the level of uncertainty in the investment as measured by the standard deviation of the relevant posterior.

In our Bayesian nonparametric approach, an investor's initial guess about the population distribution of skill or a particular fund's level of expertise is represented by the hierarchical prior $\pi(\alpha|\theta)$ with the hyperparameter θ. Knowledge about the ith fund's ability, along with an understanding of the population, increases as we observe the risk-adjusted gross returns of any actively managed fund. For instance, if we only see the returns for the ith fund, we directly update our guess about the skill of the ith fund by applying Bayes rule

$$\pi(\alpha_i|r_i, \theta) \propto \pi(\alpha_i|\theta) f(r_i|\alpha_i), \tag{3}$$

where

$$f(r_i|\alpha_i) = \prod_{t=\tau_i}^{T_i} N\left(r_{i,t} - \beta_i' F_t, \middle| \alpha_i, \sigma_i^2\right), \tag{4}$$

is the likelihood conditional on β_i and σ_i^2.[6]

Alternatively, if the returns are from the $J-1$ other funds, we update our guess about the population distribution with the posterior predictive distribution

$$\pi(\alpha_i|r_{-i}) \quad = \quad \int \pi(\alpha_i|\theta) dG(\theta|r_{-i}), \tag{5}$$

where α_i is the level of skill for the ith fund and r_{-i} are the return histories of the $J-1$ funds besides the ith fund. Note that α_i can generically represent the skill level of any fund whose performance history is not found in r_{-i}.

In Eq. (5),

$$G(\theta|r_{-i}) \propto G(\theta) p(r_{-i}|\theta), \tag{6}$$

is the posterior for the hyperparameter where the connection between θ and r_{-i} is made via $p(r_{-i}|\theta) = \int f(r_{-i}|\alpha_{-i})\, \pi(\alpha_{-i}|\theta)\, d\alpha_{-i}$.

The posterior predictive distribution in Eq. (5) has updated our initial guess for the population distribution, the prior, $\pi(\alpha)$, to the posterior population distribution, $\pi(\alpha|r_{-i})$. This "updated prior" gets augmented with the additional information observed in the ith

[6]Note we have suppressed β_i and σ_i^2 from conditioning argument of the likelihood to simplify the notation.

fund's likelihood. Now our assessment of the ith fund's level of skill is found in the fund's posterior

$$\pi(\alpha_i|r_i, r_{-i}) \propto \pi(\alpha_i|r_{-i})f(r_i|\alpha_i). \tag{7}$$

Given the cross-sectional information in Eq. (7) of past and present mutual fund performance, the posterior for fund i's alpha is at least as well informed as $\pi(\alpha_i|r_i)$ and better if r_i is short or non-existent.

Thus far we have been learning only about the alphas and θ conditional on the particular prior distribution $\pi(\alpha, \theta) = \pi(\alpha|\theta)G(\theta)$ where the distributions $\pi(\alpha|\theta)$ and $G(\theta)$ are assumed to be known. In the following section, we let our beliefs about the cross-sectional distribution of skill be completely flexible, in other words, nonparametric, by letting G be unknown. We then learn about the population distribution of skill by using the information from the panel of returns to update G, the alphas, and θ. We now show how one learns about the unknown cross-sectional distribution of skill as one learns about G.[7]

3 Initial beliefs about the population

We assume the prior beliefs about the distribution of alpha is independent from the 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] We follow Müller & Rosner (1997) and let the prior for the alphas be the nonparametric, Dirichlet Process mixture, prior (DPM)

$$\alpha_i|\mu_{\alpha,i}, \sigma_{\alpha,i}^2 \sim N(\mu_{\alpha,i}, \sigma_{\alpha,i}^2), \tag{8}$$

$$\mu_{\alpha,i}, \sigma_{\alpha,i}^2|G \sim G, \tag{9}$$

$$G|G_0 \sim DP(B, G_0), \tag{10}$$

where the hyperprior distribution, G, is unknown and modeled in terms of Ferguson's (1973) Dirichlet process, $DP(B, G_0)$. The DP arguments are the concentration parameter, $B > 0$, and the base distribution G_0.[9]

[7]We could apply our nonparametric approach to the population distributions of the betas and sigmas. However, since our focus is on mutual fund skill we let the priors for β_i and σ_i be ex ante uninformative priors; i.e., we assume the betas and sigmas are idiosyncratic over the cross-section of funds. Investigating how to learn about the betas and sigmas would be a worthy research project and would showcase how straight forward our nonparametric approach is relative to other approaches.

[8]One could assume investors have a joint prior for $(\alpha_i, \beta_i, \sigma_i^2)$. However, learning this distribution would require assigning a fund to a cluster based on all the unknown parameters and not just alpha. Grouping funds by ability would no longer be the objective, so, we assume a separate prior for the alphas, betas and sigmas.

[9]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.

The DPM has been used extensively in econometrics to model the unknown distributions of observable data (Chib & Hamilton 2002, Hirano 2002, Jensen 2004, Jensen & Maheu 2010, Bassetti et al. 2014). A primary reason for this is the DP's almost sure discrete representation of the unknown hyperprior distribution

$$G \overset{as}{=} \sum_{k=1}^{\infty} \omega_k \mathbf{1}_{\{\mu^*_{\alpha,k},\sigma^{2*}_{\alpha,k}\}}, \tag{11}$$

with $(\mu^*_{\alpha,k}, \sigma^{2*}_{\alpha,k}) \overset{iid}{\sim} G_0$, $\omega_k = w_k \prod_{k'<k}(1-w_{k'})$, and where $w_k \sim \text{Beta}(1, B)$, and $\mathbf{1}_{\{\mu^*_{\alpha,k},\sigma^{2*}_{\alpha,k}\}}$ is a point mass at $(\mu^*_{\alpha,k}, \sigma^{2*}_{\alpha,k})$. This almost sure discreteness of G leads to the partitioning of the unknown alphas into groups.

Another important reason for using the DP to model G is its ease of use. As a conjugate distribution the DP lends itself to the simple, and efficient, sampling algorithm of West et al. (1994).[10] Draws from this sampler are made using conditional posterior distributions that are known, and the draws quickly converge to realizations from the posterior distribution of the nonparametric hierarchical prior.

The expectation of $DP(B, G_0)$ is $E[G] = G_0$, so the base distribution, G_0, represents our initial guess of G, and indirectly, our initial guess of the population distribution. Since $\text{Var}[G] \equiv [G_0(1-G_0)]/(1+B)$, the concentration parameter B can be thought of as the inverse variance of G. The larger B is the more confident we are about G_0 being the hyperprior G. In the limit, $G \to G_0$ and $\omega_k \to 0$ as $B \to \infty$.[11] In our empirical application B is unknown and estimated.

One needs to be thoughtful about choosing G_0 since it plays an important role in how open-minded we are about mutual fund skill. For example, if we were certain about the average skill level and the variance of the population we might choose the base distribution $G_0 \equiv \mathbf{1}_{\{m_0,s_0^2\}}(\mu_\alpha, \sigma^2_\alpha)$ where m_0 and s_0^2 are set to prespecified values. Given the degenerative nature of this base distribution, our initial guess for the cross-sectional distribution of the alphas would be

$$\widehat{\pi}_{\mathbf{1}_{\{m_0,s_0^2\}}}(\alpha) \equiv E_G\left[\int N(\alpha|\mu_\alpha, \sigma^2_\alpha) dG(\mu_\alpha, \sigma^2_\alpha)\right] \tag{12}$$

$$= \int N(\alpha|\mu_\alpha, \sigma^2_\alpha) dG_0(\mu_\alpha, \sigma^2_\alpha) \tag{13}$$

$$= \int N(\alpha|\mu_\alpha, \sigma^2_\alpha) \mathbf{1}_{\{m_0,s_0^2\}}(\mu_\alpha, \sigma^2_\alpha) d(\mu_\alpha, \sigma^2_\alpha) \tag{14}$$

[10] If we were concerned about the computing time involved in re-estimating the nonparametric, posterior, population distribution as new return data becomes available we could compute in real time the posterior population distribution using the particle learning, sequential sampler of Carvalho et al. (2010).

[11] B plays an important role in the creation of clusters as the number of funds grows. We will explain this when we present the clustering properties of the DP prior.

$$= N(\alpha|m_0, s_0^2). \quad (15)$$

For those whose prior is $\widehat{\pi}_{\mathbf{1}_{\{m_0,s_0^2\}}}(\alpha)$, they believe they know the population to be normally distributed with a mean and variance equal to m_0 and s_0^2, respectively. BMW, Pástor & Stambaugh (2002*a*), and Pástor & Stambaugh (2002*b*), either implicitly or explicitly assume such strong prior beliefs about the population. For instance, any empirical study of mutual fund skill where ordinary least square (OLS) estimates of the alphas are used implicitly sets $m_0 = 0$ and $1/s_0^2 = 0$. As a result, the prior predictive distribution $\widehat{\pi}_{\mathbf{1}_{\{0,\infty\}}}(\alpha) \propto C$ says there is no information to be found in the cross-section. Instead, each fund's level of skill is idiosyncratic to the fund.

If we are sure about G_0 being the unknown hyperprior, G, then $B \to \infty$, and we would only need to learn about μ_α and σ_α^2. Suppose we are confident the normal, inverse-gamma, base distribution is the hyperprior, then

$$G \to NIG(m_0, \sigma_\alpha^2/\kappa_0, \nu_0/2, s_0^2, \nu_0/2), \text{ as } B \to \infty,$$

where m_0 and σ_α^2/κ_0 are the mean and variance to the conditional normal distribution for μ_α, and $\nu_0/2$ and $s_0^2\nu_0/2$, are, respectively, the scale and shape of the inverse-gamma distribution for σ_α^2.

According to Bernardo & Smith (2000, Appendix A2), such prior beliefs about the hyperprior are those where the prior predictive distribution is the Student-t distribution

$$\pi_{NIG}(\alpha) = \int N(\alpha|\mu_\alpha, \sigma_\alpha^2)\, NIG\left(\mu_\alpha, \sigma_\alpha^2 \left| m_0, \sigma_\alpha^2/\kappa_0, \nu_0/2, s_0^2\nu_0/2\right.\right) d(\mu_\alpha, \sigma_\alpha^2) \quad (16)$$

$$= t_{\nu_0}\left(\alpha \left| m_0, \left(\frac{\kappa_0 + 1}{\kappa_0}\right)\nu_0 s_0^2\right.\right), \quad (17)$$

with ν_0 degrees of freedom, a mean of m_0, and scale $\sqrt{[(\kappa_0 + 1)/\kappa_0]\nu_0 s_0^2}$.

When $B \to \infty$ in the DP, we are so sure G is equal to G_0, $\pi_{NIG}(\alpha)$ will be our posterior predictive distribution. For this value of the concentration parameter value we are so confident in our initial guess for the population distribution we choose to learn nothing about skill from the cross-section of fund performance data.

In this paper, we choose to learn everything about the population distribution from the cross-section of fund returns. To do this our prior knowledge about the hyperprior is captured by

$$G_0 \equiv NIG(m_0, \sigma_\alpha^2/\kappa_0, \nu_0/2, s_0^2, \nu_0/2), \quad (18)$$

with $m_0 = 0$, $\kappa_0 = 0.1$, $\nu_0 = 0.01$ and $s_0^2 = 0.01$. The Student-t prior predictive

$$\widehat{\pi}_{NIG}(\alpha) = \int N(\alpha|\mu_\alpha, \sigma_\alpha^2)E[dG(\mu_\alpha, \sigma_\alpha^2)]$$

$$= t_{\nu_0}\left(\alpha \middle| m_0, \left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2\right), \tag{19}$$

is then proper, but extremely diffuse.

From the arguments to Eq. (19), our initial beliefs are that the average fund does not have the skill to beat a passive portfolio ($m_0 = 0$). Furthermore, with $\nu_0 = 0.01$ our level of uncertainty about skill is so high our initial guess for the population variance does not even exist. We now explain how we learn about G, and hence, learn about the population distribution $\pi(\alpha)$.

4 Bayesian nonparametric learning

The next step is learning about the hyperprior distribution, G. Suppose, hypothetically, that we are able to see a realization of the hyperparameters, $(\mu_{\alpha,1}, \sigma^2_{\alpha,1})$, from the hyperprior, G, where $\mu_{\alpha,1}$ is the average skill level of a fund and $\sigma^2_{\alpha,1}$ is the variance in the fund's skill level. Being "data" from G, we use $(\mu_{\alpha,1}, \sigma^2_{\alpha,1})$ to increase our understanding about G with the updated posterior DP

$$G|\mu_{\alpha,1}, \sigma^2_{\alpha,1} \sim DP(1+B, G_1), \tag{20}$$

where the updated base distribution is[12]

$$G_1 \equiv \frac{B}{1+B} G_0 + \frac{1}{1+B}\mathbf{1}_{\{\mu_{\alpha,1}, \sigma^2_{\alpha,1}\}}. \tag{21}$$

In Eq. (20), the concentration parameter has increased to $1+B$. As a result, we are a bit more confident after observing $\mu_{\alpha,1}$ and $\sigma^2_{\alpha,1}$ in the updated base distribution, G_1, representing G. This new estimate of G consists of a mixture of our original guess, G_0, and the information found in the empirical distribution, $\mathbf{1}_{\{\mu_{\alpha,1}, \sigma^2_{\alpha,1}\}}$.

Given G_1, our estimate for the population distribution is now the posterior predictive distribution

$$\begin{aligned} \widehat{\pi}_{NIG}(\alpha|\mu_{\alpha,1}, \sigma^2_{\alpha,1}) &= \int N(\alpha|\mu_\alpha, \sigma^2_\alpha)\, dG_1(\mu_\alpha, \sigma^2_\alpha) \\ &= \frac{B}{1+B} t_{\nu_0}\left(\alpha \middle| m_0, \left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2\right) + \frac{1}{1+B} N\left(\alpha|\mu_{\alpha,1}, \sigma^2_{\alpha,1}\right). \end{aligned} \tag{22}$$

Equation (22) captures our unsupervised probabilistic approach to learning how skill is distributed across mutual funds by clustering funds into groups where members all have the same average level of skill and the same level of variability in their skill. For example,

[12]See Blackwell & MacQueen (1973).

$1/(1+B)$ is the probability that a mutual fund we know nothing about is assigned to the group whose average level of skill is $\mu_{\alpha,1}$ and whose variability is $\sigma^2_{\alpha,1}$. $B/(1+B)$ is probability this same fund gets assigned to a new group whose average skill and variance is different from $\mu_{\alpha,1}$ and $\sigma^2_{\alpha,1}$. Assigning funds into new groups provides us with the flexibility to continue to increase the number of mixture clusters as the number of mutual funds in our panel grows; i.e., the DP has an infinite number of mixture clusters available for new funds to be assigned to.

We continue to apply this unsupervised probabilistic approach to assigning funds as we hypothetically observe realizations from G for the J mutual funds. After "seeing" $\mu_{\alpha,i}$ and $\sigma^2_{\alpha,i}$, $i = 1, \dots, J$, the posterior DP for G is

$$G|\mu_{\alpha,1}, \sigma^2_{\alpha,1}, \dots, \mu_{\alpha,J}, \sigma^2_{\alpha,J} \quad \sim \quad DP(J+B, G_J), \tag{23}$$

where

$$G_J \equiv \frac{B}{J+B} G_0 + \sum_{k=1}^{K} \frac{n_k}{J+B} \mathbf{1}_{\{\mu^*_{\alpha,k}, \sigma^{2*}_{\alpha,k}\}}, \tag{24}$$

is our guess for G.

In Eq. (24), the updated base distribution, G_J, shows how our Bayesian nonparametric method of learning has uncovered $K \leq J$ groups each with its own unique mean, $\mu^*_{\alpha,k}$, and variance, $\sigma^{2*}_{\alpha,k}$, $k = 1, \dots, K$. The kth group contain n_k funds and so it follows that $\sum_{k=1}^{K} n_k = J$.

The concentration parameter in Eq. (23) has increased to $J+B$, so our confidence in the guess for G has grown as has our confidence in the estimate for the population distribution

$$\begin{aligned} \widehat{\pi}_{NIG}(\alpha|\mu_{\alpha,1}, \sigma^2_{\alpha,1}, \dots, \mu_{\alpha,J}, \sigma^2_{\alpha,J}) \quad &= \quad \frac{B}{J+B} t_{\nu_0} \left(\alpha \,\middle|\, m_0, \left(\frac{\kappa_0 + 1}{\kappa_0} \right) \nu_0 s_0^2 \right) \\ &\quad + \sum_{k=1}^{K} \frac{n_k}{J+B} N(\alpha|\mu^*_{\alpha,k}, \sigma^{2*}_{\alpha,k}). \end{aligned} \tag{25}$$

According to Eq. (25), when we know nothing about a fund, the probability of assigning the fund to one of the K groups depends on the group's size n_k. In our mind, larger groups have a greater chance of having a new fund assigned to it. However, the number of groups also depends on how confident we are in our initial guess G_0; i.e., the concentration parameter B. The larger B is the more groups we will partition the cross-section of mutual funds in to.

We are thus learning about the population distribution of mutual fund skill by assuming very little about the cross-sectional distribution of skill, but then learning about it by flexibly forming a mixture of normals where the number of clusters are identified by partitioning the funds into groups having the same average skill level and variability in skill. Because our approach is based on unsupervised learning, it allows the number of groups to grow with the size of the cross-section, introduces new groups when a new extraordinarily skilled or unskilled fund opens for business, and includes such extraordinary occurrences in the probability of future mutual fund performances.

4.1 Finite mixture approach

In contrast to our Bayesian nonparametric learning approach, the frequentist approaches of Harvey & Liu (2018) and Chen et al. (2017) require pre-specifying the number of mixture clusters, K. To determine the number of clusters, they estimate and test a K-ordered mixture model against the alternative $(K+1)$-ordered mixture model.[13] There are drawbacks to the frequentist approach. First, it requires incrementally estimating and testing a number of different ordered mixture models. This takes time and increases the complexity. For instance, our estimate of the yearly evolution of the population distribution of skill in Section 6.6 would be very tedious with the finite mixture approach.

Second, even if the correct number of clusters is chosen, predictions with finite-ordered mixture models are less flexible. Confining the mixture to a finite number of clusters lacks the DP's infinite number of clusters that are available when a new fund's skill level does not align with one of the existing group's average level of skill.

Lastly, frequentist tests for the correct number of clusters require the null and alternative mixture models to be nested. This is the reason a K-ordered mixture model is tested against a $(K+1)$-ordered alternative. Because of this nesting, the frequentist's finite-ordered mixture approach does not generalize to the estimation of the population distributions of two or more parameters. For instance, one cannot extend the frequentist approach to estimating the population distribution of one of the risk factor parameters, β_i, while simultaneously modeling the population distribution of the alphas. Whereas our nonparametric DPM prior can be applied separately to as many of the parameters as one desires. Given the limitations of the finite mixture approach, we strongly argue in favor of applying our nonparametric DPM prior to the estimation and prediction of mutual fund skill.

[13] Kasahara & Shimotsu (2015) propose a likelihood-ratio test for determining the number of clusters in a finite-ordered Gaussian mixture regression model of observable data. Because the alphas are latent, their test is not applicable here. As a result, Harvey & Liu (2018) have to simulate the critical values for the test.

5 Inference

To resolve the uncertainty around the alphas, betas, and unknown mixture parameters, we combine fund-level return data with our initial beliefs to form a joint posterior for the unknowns. However, the joint posterior distribution for these unknowns is very complex and does not have a known analytical distribution. Analysis of the complex joint posteriors requires judiciously breaking it up into its conditional posteriors and using a Markov Chain Monte Carlo sampler to make joint posterior draws by sequentially sampling from the conditional posteriors.

The conditionals we sample from are structured by the hierarchical form of our non-parametric model. Let $s = (s_1, \ldots, s_J)$ be a J length vector containing all the fund's group assignments s_i, where $s_i = k$ when $(\mu_{\alpha,i}, \sigma^2_{\alpha,i}) = (\mu^*_{\alpha,k}, \sigma^{2*}_{\alpha,k})$. The sequence of conditional posterior distributions are sampled by:

1. Drawing β_i and σ_i^2 conditional on r_i and α_i for $i = 1, \ldots, J$.
2. Drawing α_i conditional on r_i, β_i, σ_i^2 and $(\mu^*_{\alpha,s_i}, \sigma^{2*}_{\alpha,s_i})$ for $i = 1, \ldots, J$.
3. Drawing s, K, and $(\mu^*_{\alpha,k}, \sigma^{2*}_{\alpha,k})$, $k = 1, \ldots, K$, conditional on $\alpha_1, \ldots, \alpha_J$.
4. Drawing B conditional on K.

In Step 1, our prior knowledge for the factor loading vector, β_i, and the return variance, σ_i^2, is represented by the Jeffreys prior

$$\pi(\beta_i, \sigma_i^2) \propto 1/\sigma_i^2. \tag{26}$$

Under this prior, the conditional posterior for β_i in Step 1 depends only on the return-based information, r_i. The conditional $p(\beta_i | r_i, \alpha_i, \sigma_i^2)$ is a normally distributed conditional posterior with mean and covariance equal to the least squares regression estimator of the dependent variable, $r_{it} - \alpha_i$, projected onto the explanatory variables F_{it}, $t = \tau_i, \ldots, T_i$. The marginal conditional posterior distribution $p(\sigma_i^2 | r_i, \alpha_i, \beta_i)$ is a inverse-gamma distribution with scale, $\mathcal{T}_i - 4$, and shape equal to the sum of squared errors from the above linear least squares regression divided by the scale $\mathcal{T}_i - 4$.

In Step 2, the prior for α_i is the cross-sectional distribution of skill for the s_ith group

$$\alpha_i | \mu^*_{\alpha,s_i}, \sigma^{2*}_{\alpha,s_i} \sim N(\mu^*_{\alpha,s_i}, \sigma^{*2}_{\alpha,s_i}).$$

Combining this cross-sectional information with the likelihood from the ith fund's return history, r_i, the conditional posterior in Step 2 is

$$\alpha_i | r_i, \beta_i, \sigma_i^2, \mu^*_{\alpha,s_i}, \sigma^{2*}_{\alpha,s_i} \sim N(a_i, b_i), \tag{27}$$

where the posterior mean is

$$a_i = \left(\frac{\mu^*_{\alpha,s_i}}{\sigma^{2*}_{\alpha,s_i}} + \sum_{t=\tau_i}^{T_i} r^*_{i,t}\right) \Bigg/ \left(\frac{1}{\sigma^{2*}_{\alpha,s_i}} + \mathcal{T}_i\right), \tag{28}$$

with

$$r^*_{i,t} \equiv (r_{i,t} - \beta_i' F_{i,t})/\sigma_i, \tag{29}$$

being the risk adjusted return, and the posterior variance is

$$b_i = (1/\sigma^{2*}_{\alpha,s_i} + \mathcal{T}_i)^{-1}. \tag{30}$$

In Step 3, we can think of the sampler answering the question asked by JS and adapted to our case – when would investors discard the information found in the average skill and variability of the K sub-populations, $\mu^*_{\alpha,k}$, and $\sigma^*_{\alpha,k}$, $k = 1, \ldots, K$? Answering this question for each fund amounts to drawing the assignment vector s by sequentially drawing each fund's s_i according to the probabilities

$$P(s_i = k) = \frac{n_k^{(-i)}}{B + J - 1} f_N(\alpha_i | \mu^*_{\alpha,k}, \sigma^{*2}_{\alpha,k}), \qquad k = 1, \ldots, K^{(-i)}, \tag{31}$$

$$P\left(s_i = K^{(-i)} + 1\right) = \frac{B}{B + J - 1} f_t\left(\alpha_i \middle| m_0, \left(\frac{\kappa_0 + 1}{\kappa_0}\right) \nu_0 s_0^2\right), \tag{32}$$

where $n_k^{(-i)}$ is the number of funds belonging to the kth group after the ith fund has been removed from the sample, and $K^{(-i)}$ is the number of groups. $K^{(-i)}$ equals $K - 1$ if the ith funds is the only member of its group. Otherwise, $K^{(-i)}$ equals K.

Equation (32) is the probability the ith fund is uniquely skilled with a mean alpha and variance different from all the other funds. This probability is a function of the concentration parameter, B, and the value of the prior predictive distribution evaluated at the draw of α_i. If s_i were assigned $K^{(-i)} + 1$ a new cluster would be created along with a unique mean and variance such that $(\mu^*_{\alpha,K^{(-i)}+1}, \sigma^{*2}_{\alpha,K^{(-i)}+1}) = (\mu_{\alpha,s_i}, \sigma^2_{\alpha,s_i})$.

From sweep to sweep of the sampler, the elements of the assignment vector, s, experience label switching (Richardson & Green 1997, Frühwirth-Schnatter 2006, Geweke 2007). Unless more structure is added to the means and variances, $(\mu^*_{\alpha,k}, \sigma^{*2}_{\alpha,k})$, $k = 1, \ldots, K$, such as strictly ordering the mixture means, k can be the label of the low skilled group of funds for one sweep and, in the next sweep, be the label for the high skilled group. Because of label switching we are unable to estimate which group each fund belongs to.[14]

[14] Malsiner-Walli et al. (2017) identify cluster assignment within a finite mixture model by employing an informative prior based on assumed shapes present in the data. As far as we know this approach has not been applied to latent data such as skill since the assumed shapes in the latent data are not known a priori.

After we assign s_i for every fund, and by doing so determining the total number of mixture clusters, K, we pool together the alphas from those funds belonging to the same group and form our posterior beliefs about the average skill level and variability by drawing $\mu^*_{\alpha,k}$ and $\sigma^{2*}_{\alpha,k}$ for each group $k = 1, \ldots, K$. Given the DP base distribution, $G_0 \equiv NIG(m_0, \sigma^2_\alpha/\kappa_0, \nu_0/2, s_0^2\nu_0/2)$, from Section 3, draws of $\sigma^{2*}_{\alpha,k}$ are made from

$$\sigma^{2*}_{\alpha,k}|\{\alpha_i\}_{i:s_i=k} \sim NIG\left(\frac{\nu_0+n_k}{2}, \frac{1}{2}\left[\nu_0 s_0^2 + \sum_{i:s_i=k}(\alpha_i-\overline{\alpha}_k)^2 + \frac{n_k\kappa_0}{\kappa_0+n_k}(m_0-\overline{\alpha}_k)^2\right]\right), \quad (33)$$

where $\overline{\alpha}_k = n_k^{-1}\sum_{i:s_i=k}\alpha_i$. Draws of $\mu^*_{\alpha,k}$ are then sampled from

$$\mu^*_{\alpha,k}|\{\alpha_i\}_{i:s_i=k}, \sigma^{2*}_{\alpha,k} \sim N\left(\frac{\kappa_0 m_0 + n_k\overline{\alpha}_k}{\kappa_0+n_k}, \frac{\sigma^{2*}_k}{\kappa_0+n_k}\right). \quad (34)$$

Lastly, in Step 4 we draw the concentration parameter B from $\pi(B|K)$ using the sampler described in Appendix A.5 of Escobar & West (1995).

Later in the empirical application, we initialize our sampler by setting all the funds' alphas equal to zero. The concentration parameter, B, is initialized with a random draw from its prior, $\pi(B) \equiv \text{Gamma}(2.0, 30.0)$. This draw of B, along with the normal, inverse-gamma, base distribution, G_0, in Eq. (18), are used to initialize s, K, and $\{\mu^*_{\alpha,k}, \sigma^{2*}_{\alpha,k}\}_{k=1,\ldots,K}$, by making J random draws from $DP(B, G_0)$. Given these initial values we then begin to iterate over the sampler by using Step 1 to draw the β_is and σ_i^2s.

After a burn-in of the sampler where the draws from the conditional posteriors are thrown away to allow the sampler to converge to the posterior distribution, subsequent draws are kept and treated as random realizations from the joint posterior distribution. This randomness in the sampled alphas represents our beliefs about the skill level for each of the funds. We choose to iterate the sampler 40,000 times keeping the last 30,000 draws of the unknowns to infer all the posteriors.

5.1 Flexibility

We point out that the conditional posterior distribution draw of α_i in Eq. (27) does not depend on the performance history of the other funds. In other words, our sampler's draw of each fund's alpha is independent of the other funds. However, the skill level of the other funds does influence our guess for α_i through the s_ith group's average level of skill, μ^*_{α,s_i}, and variance, σ^*_{α,s_i}.

Cluster information is especially important for a fund with a short performance history, in other words, when $\mathcal{T}_i$ is small, or when a fund's performance history is noisy such that σ_i^2 is large. Traditional measures of alpha for funds with limited histories are noisy and uncertain (Kothari & Warner 2001). But from Eq. (28), we see that average conditional draw of α_i is also determined by the average skill of the s_ith group, μ^*_{α,s_i}. Hence, our nonparametric estimator of a fund's alpha will depend more on the average performance of a fund's group, and less on the fund's own performance history, when the fund has a noisy or short performance history.[15]

The conditional mean of alpha in Eq. (28) also shows how our Bayesian nonparametric learning approach uses the cross-sectional information differently from JS. In JS there is but one cluster ($K = 1$); i.e., all the funds belong to the same group whose average is the industry-wide average, μ^*_α, with variability, σ^{2*}_α. While insightful in their use of cross-sectional information, we will see in Section 6 that by not allowing for an unknown number of clusters the parametric approach of JS under-predict the alphas of skilled funds and over-predict the alphas of unskilled funds. Over and under-prediction of skill occurs because the JS estimate for α_i shrinks towards the average ability of the entire population.[16] Such bias is also prone to exist in pre-specified, finite ordered, mixture models.

In terms of our Bayesian nonparametric approach, shrinkage towards the population average by JS is equivalent to the econometrician thinking there is only one group among the mutual funds. One group will be found if the concentration parameter of the DP prior is $B = 0$. When $B = 0$, the mixture weight, $\omega_1 = 1$, in Eq. (11)'s almost sure, discrete representation of G. Hence, the base distribution, G_1, in the updated DP of Eq. (20), consists of only one cluster, $(\mu^*_\alpha, \sigma^{2*}_\alpha) = (\mu_{\alpha,1}, \sigma^2_{\alpha,1})$, where $(\mu_{\alpha,1}, \sigma^2_{\alpha,1}) \sim G_0$. Each $(\mu_{\alpha,i}, \sigma^2_{\alpha,i})$, $i = 2, \ldots, J$, is a realization from the degenerative base distribution $\mathbf{1}_{\{\mu_{\alpha,1},\sigma^2_{\alpha,1}\}}$, so that the posterior base distribution for $G|\mu_{\alpha,1}, \sigma^2_{\alpha,1}, \ldots, \mu_{\alpha,J}, \sigma^2_{\alpha,J}$ is

$$G_J = \frac{J}{J+B}\mathbf{1}_{\{\mu^*_\alpha,\sigma^{2*}_\alpha\}}.$$

Step 1 & 2 of our sampler remain the same, but Step 3 now only involves drawing σ^{2*}_α and μ^*_α from Eq. (33) and (34), respectively.[17]

At the other extreme is when $B \to \infty$. According to the updated base distribution G_J in Eq. (24), when $B \to \infty$ every fund's hyperparameter $(\mu_{\alpha,i}, \sigma^2_{\alpha,i})$, $i = 1, \ldots, J$, is seen

[15] This feature of our Bayesian nonparametric approach would be very helpful measuring the skill level of self-reporting hedge funds where there is no regulation requiring them to report their performance.

[16] It is well known that the normal hierarchical prior can lead to poor estimates of the population distribution and the unknown parameter (Verbeke & Lesaffre 1996).

[17] This is equivalent to the sampler found in JS.

as a new independent draw from the base distribution, G_0. Since each realization of the hyperparameter is idiosyncratic, fund's are not partitioned into groups, so we learn nothing about the value of the hyperparameters from other fund's in the group. Instead, we would ignore what we learn about the skill of a fund when looking at other funds. Step 2 of our sampler would then consist of J independent draws of the alphas where the priors are the idiosyncratic prior distributions, $N(\mu_{\alpha,i}, \sigma^2_{\alpha,i})$, $i = 1, \ldots, J$.

Equation (27)–(30) also shows how our guess of an extraordinary fund's alpha is no different from the opinion of an econometrician who chooses to treat such highly skilled funds idiosyncratically. With our Bayesian nonparametric learning approach, highly skilled funds have few peers, and, hence, likely belong to small groups. In the extreme, a highly skilled fund is so talented it has no peers; i.e., $n_{s_i} = 1$ and σ^*_{α,s_i} is infinite. In this hypothetical situation, our Bayesian nonparametric approach does not borrow information from the cross-section when learning about the fund's alpha. Instead, each extraordinary fund is treated separately from the other funds, and, according to Eq. (28), draws of the highly skilled fund's conditional posterior alpha come from a normal distribution whose first and second moments are those of an OLS estimator of alpha. We will see evidence of this in the empirical investigation of Section 6.7 and the similarity between our nonparametric estimate of a highly skilled fund's alpha and the least squares estimate of its alpha.

5.2 Posterior cross-sectional distribution

In Eq. (25), our best guess for the cross-sectional distribution of the alphas depends on having hypothetically observed the means and variances of the clusters. After observing the return histories from a cross-section of funds, we can account for the uncertainty in these unknown mixture means and variances by Rao-Blackwellizing the conditional posterior predictive distribution over the M posterior draws of the unknown parameters

$$\begin{aligned}\widehat{\pi}_{DPM}(\alpha|r_1,\ldots,r_J) \quad \approx \quad & M^{-1}\sum_{l=1}^{M}\left[\frac{B^{(l)}}{J+B^{(l)}}t_{\nu_0}\left(\alpha\,\middle|\,m_0,\left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 s_0^2\right)\right. \\ & \left.+\sum_{k=1}^{K^{(l)}}\frac{n_k^{(l)}}{J+B^{(l)}}N\left(\alpha\,\middle|\,\mu^{*(l)}_{\alpha,k},\sigma^{2*(l)}_{\alpha,k}\right)\right], \qquad (35)\end{aligned}$$

where $(\mu^{*(l)}_{\alpha,k}, \sigma^{*2(l)}_{\alpha,k})$, $k = 1, \ldots, K^{(l)}$, $l = 1, \ldots, M$, is the lth sweep's draw from the conditional posterior distribution in Step 3, and $n_k^{(l)}$ comes from the information found in $s^{(l)}$. Lastly, $B^{(l)}$ is the lth draw from Step 4 of the sampler. This posterior predictive distribution for the alphas takes into consideration all the uncertainty about the unknowns, including the unknown hyperprior distribution, G, by averaging over the posteriors of all the unknowns.

6 Empirical investigation

Our empirical application consists of applying our Bayesian nonparametric learning approach to the estimation of the alphas of the actively managed mutual funds found in the data set of Jones & Shanken (2005).[18] This data set is comprised of the monthly gross returns over the period from January 1961 to June 2001. It consists of a unbalanced panel of 396,820 monthly observations from 5,136 domestic equity funds.[19] Like Baks et al. (2001), Jones & Shanken (2005), and Cohen et al. (2005), we are interested in before cost performance unaffected by the dynamics of the funds' 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 the funds have on average 77.3 monthly returns. We include all actively managed, domestic, equity funds in our panel, even the 1,292 funds that were no longer open for business at the end of the sample in order to avoid any survivorship bias.

6.1 Sampler convergence

As mentioned earlier, we throw away the first 10,000 draws of the alphas, α_i, risk-factor vectors, β_i, and nonparametric mixture parameters, before keeping the subsequent 30,000 draws for posterior inference. To determine if our MCMC sampler has converged to the unknown posterior distributions, we compute Geweke's (1992) z-score convergence diagnostic statistic for the posterior draws of each fund's alpha. For the 5,136 funds' alphas, the average z-score is 0.0275, with 4,419 of the fund's z-scores being less than two in absolute value. Such z-score values are evidence that our draws of the individual fund's alpha have converged to a sample from their posterior distribution.

We also compute Geweke's (1992) numerical measure of correlation to gauge the degree of independence between each draw of the alphas. Using this measure of sampling efficiency, we find that the draws of the alphas are nearly independent, with efficiency levels exceeding 100 and, on average, equaling 3,123. Thus, our sample of the alphas represents near independent realizations from the posterior distributions of skill.

To test if the posterior predictive distribution of alpha in Eq. (35) has converged after a burn-in of 10,000 draws, we compare it to one computed after a burn-in of 100,000 draws. From this test, we find the two posterior predictive distribution's densities to be nearly indistinguishable. Both posterior predictive densities of alpha have exactly the same

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

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

properties seen in the DPM plot of Figure 1.[20] Hence, after a burn-in of 10,000 draws our MCMC sampler has converged to the unknown posterior population distribution.

6.2 Posterior number of clusters

For our Bayesian nonparametric approach where the population distribution of the alphas are modeled with the nonparametric hierarchical prior of Eq. (8)–(10), we find the posterior median number of clusters, K, to equal four, with a minimum posterior draw of three and a maximum of six clusters. The 95% highest posterior probability density (HPD) interval for K is three to five clusters. Hence, the 5,136 funds are randomly partitioned over a handful of the infinite possible mixture clusters.

The posterior mean for the concentration parameter, B, is 0.1245, with a 95% HPD interval of $(0.004, 0.261)$. A concentration parameter this close to zero supports our earlier claims of mutual fund skill not being idiosyncratic to a fund, nor being normally distributed over the population. Instead, the population distribution of skill is represented by a mixture over a small number of normal distributions. 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.

6.3 Cross-sectional distribution

In Figure 1, the red line is the density for the posterior cross-sectional distribution of alpha where we learn how skill is distributed over mutual funds by computing $\widehat{\pi}_{DPM}(\alpha|r_1,\ldots,r_J)$ with Eq. (35). The blue density line is the posterior cross-sectional distribution where skill is normally distributed, with its unknown mean, μ_α, and variance, σ^2_α, modeled with the uninformative Jeffreys prior, $\pi(\mu_\alpha,\sigma^2_\alpha)\propto 1/\sigma^2_\alpha$. The uncertainty around the mean and variance is integrated out of the normal population distribution with

$$\widehat{\pi}_{JS}(\alpha|r_1,\ldots,r_J)\approx M^{-1}\sum_{l=1}^{M}N\left(\alpha\left|\mu_\alpha^{(l)},\sigma_\alpha^{2(l)}\right.\right),$$

where the $(\mu^{(l)},\sigma^{2(l)})$ are draws from $\pi(\mu_\alpha,\sigma^2_\alpha|r_1,\ldots,r_J)$.

The green density in Figure 1 is the density of the empirical distribution for the fund-by-fund OLS estimates of the alphas. It is estimated with the posterior predictive distribution of the DPM applied directly to the 5,123 OLS alphas, $\widehat{\alpha}_i$.[21]

[20] Plots of the two posterior predictive densities are available upon request from the authors.

[21] Please note that the OLS alphas are idiosyncratic to the fund and, hence, the actual population distribution under the OLS estimator is a uniform distribution over the entire real line. As a result, the green density in Figure 1 is a smoothed "histogram" of the OLS alphas and is not an actual estimate of the population distribution.

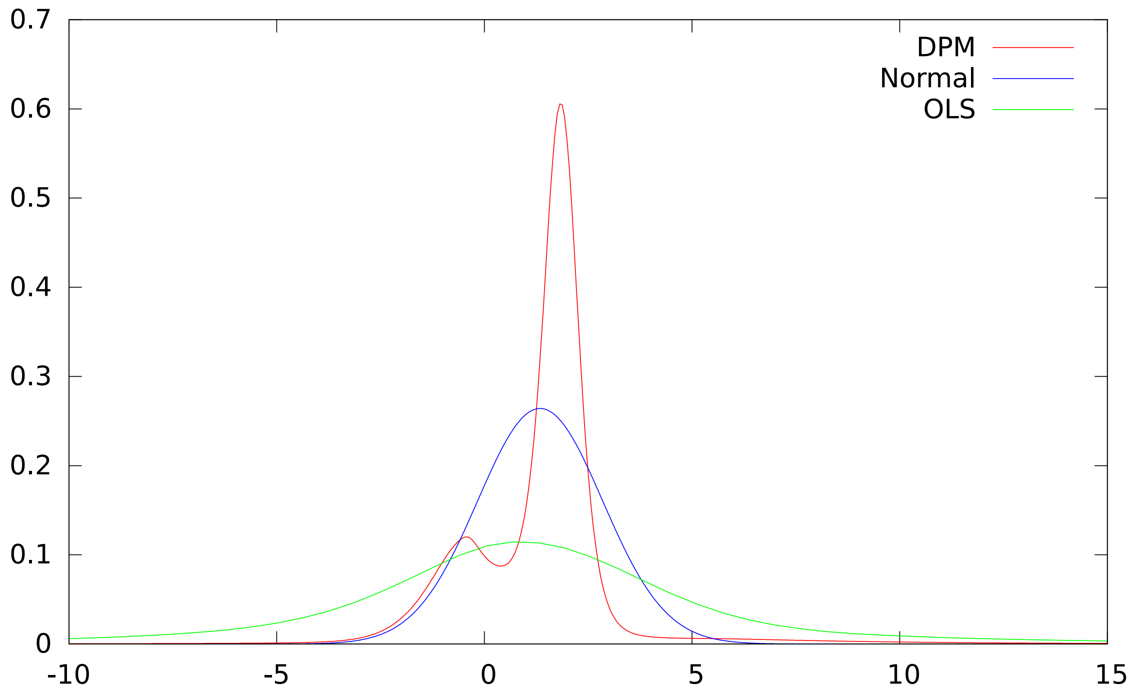


Figure 1: Posterior population distribution of alpha for the JS type investors, $\widehat{\pi}_{JS}(\alpha|r_1,\ldots,r_J)$, who believe the underlying population distribution is normal but with unknown mean and variance (blue line), the posterior population distribution for our investors, $\widehat{\pi}_{DPM}(\alpha|r1,\ldots,r_J)$, who do not assume a particular distribution for alpha but have placed a DPM prior on the unknown distribution of alpha (red line), and the empirical distribution of the fund-by-fund OLS estimates of the alphas estimated by applying the DPM to the OLS alphas (green).

Very different conclusions about the cross-sectional performance of mutual funds are drawn from the estimated population densities in Figure 1. Our Bayesian nonparametric estimator of the population distribution finds two modes and a flat density section around 6%. In contrast, the JS population distribution by definition has only one mode. The primary mode for the nonparametric population distribution is 1.8% and the JS mode is 1.5%. Such values for alpha just offset the fees charged by the average mutual fund.[22] An investor who applied the JS approach, thus, believes that, on average, a fund, they know nothing about, will just break even. Whereas with our Bayesian nonparametric approach we find that there is a seventy-three percent chance the unknown fund will have an alpha between 0.4% and 4.0%.

The secondary mode of the nonparametric population distribution in Figure 1 is located at -0.65%. This negative mode suggests there are actively managed funds whose average stock-picking ability is detrimental to what an investor could earn on a passively managed fund.[23] Given the information from this second mode, we find that there is a 22% chance that a fund, for which we have no information about, would produce an alpha between -5% to 0.4% a year.

The flat section of the nonparametric population distribution at 6% is a diffuse, low probability, area of the predictive density. Such probability is only possible with a normal mixture model with more than three clusters. There is a 3% chance a fund we know nothing about will have the skill level to produce an alpha of four to ten percent. On the other hand, such a fund has less than a half a percent chance of its alpha being between -4% and -10%. Hence, it is more likely we will find a highly skilled fund than an unskilled fund.

The empirical distribution of the OLS alphas in Figure 1 identifies the average skill level of the population, but it fails to uncover the bi-modality, skewness, or tightness of the population distribution. The inability of the empirical distribution to find important characteristic of the population distribution can be understood in terms of signal-noise extraction where the signal is the population distribution, $\pi(\alpha)$. Conditional on knowing the variances and betas of the funds, the OLS estimate's predictive distribution can be

[22] 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.

[23] Theoretically, rational investors would pull their money from under-performing funds, causing them to go out of business. Negative alphas leave open the door that some investors act irrationally (Gruber 1996), or that investors tolerate short term poor performance.

written as

$$\pi(\widehat{\alpha}|\alpha) = N(\widehat{\alpha}|\alpha, \sigma^2),$$

where σ^2 is the level of noise in the OLS estimates of the alphas.[24] Assuming we know the cross-sectional distribution of skill, $\pi(\alpha)$, the marginal distribution of the OLS alphas is the weighted average

$$\pi(\widehat{\alpha}) = \int \pi(\widehat{\alpha}|\alpha)\pi(\alpha)d\alpha.$$

Hence, the empirical distribution of the OLS alphas, $\pi(\widehat{\alpha})$, is a "blurred" version of the population distribution, $\pi(\alpha)$. If the noise in the OLS alphas, σ^2, were smaller, $\pi(\widehat{\alpha})$ would be closer to $\pi(\alpha)$. However, many funds have noisy or short performance histories, so the empirical distribution smears skilled and unskilled funds into a unimodal population.

In Table 1, we list the percentiles, standard deviation, skewness, and kurtosis of the parametric, and nonparametric, posterior cross-sectional distributions. Medians of both distributions are close to the average fee of 1.5% a year. At first glance, these medians support the theoretical findings of Berk & Green (2004) where, in the long run, successful funds break-even with an alpha that matches their fees. However, we have also uncovered multiple modes along with a skewness towards highly skilled funds. These characteristics are only possible when the Gaussian mixture model has four or more mixture clusters.

One explanation for having multiple modes is that funds whose alpha is close to 6% are newer funds with fewer assets under management and have not yet experienced the diminishing returns to scale assumed in the model of Berk & Green (2004). As these young skilled funds attract assets, grow, and mature, we expect that their alphas will move towards the population median; i.e., towards the break-even alpha.

	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
JS	−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 when the underlying distribution of skill is believed to be normally distributed (JS) and the nonparametric, hierarchical, prior (DPM).

According to the percentiles in Table 1, the probability of a new, or unknown fund, being extraordinarily skilled is higher than previously thought since the nonparametric population distribution's 99th-percentile is an alpha of 11.48%, and the 99th-percentile for the

[24] In the case where we do not condition on the return variances and betas a Student-t distribution would replace the Gaussian.

parametric distribution is only 4.87%. Our flexible, nonparametric, population distribution is also more fat-tailed, with a kurtosis of 101.49, and more skewed towards finding skill in the industry with a skewness of 4.31 than the parametric population distribution. Our nonparametric population distribution also has a slightly more negative 1%-percentile of -2.90% compared to -2.19% for the parametric distribution. So, compared to the parametric population distribution, our nonparametric approach finds that there is a greater chance a fund will be either extraordinarily skilled or unskilled.

A similar tail relationship also exists between our nonparametric population distribution and a finite ordered mixture model. Given the bi-modality of our nonparametric posterior predictive distribution, one might expect a finite ordered mixture model with two Gaussians fits the population best. In fact, after testing several orders of Gaussian mixture models, Harvey & Liu (2018) concluded a mixture of two produced the best model of mutual fund skill.

A two-cluster, Gaussian, mixture model, however, suffers shrinkage just like the normal model of JS. All the extraordinary skilled funds get shrunk back towards the average alpha of the skilled group, while all the extremely unskilled funds get pulled up to the average alpha of the lower skilled group. As a result, Harvey & Liu (2018) find fewer exceptionally skilled and unskilled funds relative to our nonparametric approach. Only with more mixture components will Harvey & Liu's (2018) finite-ordered mixture model accurately capture the tail behavior of the population and unveil the highly skilled funds of the industry.

6.4 Robustness to ARCH

A well known empirical regularity of daily market and corporate equity returns is their time series dynamics of time-varying variance. One might think that a skilled fund manager would select stocks that help minimize this time-varying variance in their portfolio. However, to our knowledge time-varying variance in the monthly performance histories of actively managed mutual funds has not been documented.

To investigate how robust our nonparametric estimates of the population distribution and the alphas for each of the funds are to time-varying variance, and to see if the risk level in a fund's portfolios is time-varying, we relax the homoskedasticity condition assumed in the sampling distribution of Eq. (1). We now allow the sampling distribution to have time-varying variances in the form of autoregressive conditional heteroskedasticity (ARCH)

$$\sigma_{i,t}^2 = \gamma_{i,0} + \gamma_{i,1}\epsilon_{i,t-1}^2, \tag{36}$$

where $\sigma_{i,t}^2$ is guaranteed to be positive by restricting $\gamma_0 > 0$ and $0 < \gamma_1 < 1$.

Since the variance is a function of $\epsilon_{i,t-1} = r_{i,t-1} - \alpha_i - \beta_i' F_{t-1}$, the posterior Gibbs draws of α_i and β_i found in Steps 1 & 2 of our sampler are no longer valid. In place of these Gibbs draws, we substitute independent, random-walk, Metropolis draws of α_i and β_i where the step-sizes to the normal candidate distributions are set equal to the standard deviations from the OLS estimate of Eq. (1).

Random-walk Metropolis draws are also made for the conditional posterior draws of $\gamma_{i,0}$ and $\gamma_{i,1}$. Normally distributed candidate draws are made for the two ARCH parameters using the standard deviations from the linear regression of the OLS variances on lagged squared residuals as the step-size. Since Step 3 & 4 of our sampler conditions on a particular sweep's draw of the alphas, we continue to use these steps to draw the DPM parameters.

To improve the Metropolis acceptance rates, we keep the draw from every twentieth sweep from a total of 200,000 draws after an initial burn-in of 30,000 sweeps. For all 5,136 funds the Metropolis acceptance rates for the alphas, betas, and gammas, are all 100%. The posteriors distributions for the ARCH coefficients, $\gamma_{1,i}$, are heavily concentrated near zero, and are on average 0.4. Still there are a few funds whose ARCH parameters are distributed close to one.[25]

In Figure 2, we plot the log of the nonparametric posterior predictive distribution of the alphas from Figure 1 and the posterior predictive distribution when the sampling distributions have ARCH variances. One can see that both the left and right hand tails of the predictive distribution are not as fat under ARCH as they were when the variances were homoskedastic. Hence, under ARCH the population of funds is not expected to be as extraordinarily skilled or unskilled as before.

Bi-modality is still a prominent feature of the population distribution under ARCH. Investors would still expect a fund they know nothing about to likely have enough skill to break even, but with a non-negligible possibility the fund would lose money.

When we regress the posterior means of the alphas from the ARCH version of the model, $\bar{\alpha}_{ARCH,i}$, on our original estimates of the alphas, $\bar{\alpha}_i$, we find

$$\bar{\alpha}_{ARCH,i} = 0.1322 + 0.7865\, \bar{\alpha}_i.$$

The two estimates of the alphas are also highly correlated at 0.91. In general, each fund's posterior mean of alpha is slightly closer to zero under ARCH innovations. Our nonparametric estimates of skill are thus robust to the presence of ARCH in the sampling distribution

[25]Plots of the posterior densities for the $\gamma_{1,i}$s are available from the authors upon request. We should point out that if testing for ARCH were the focus of this paper then the correct way to do this would be to apply the nonparametric learning approach of this paper to the population distribution of the $\gamma_{1,i}$s and use this cross-sectional information in the estimation of the specific ARCH parameters.

of the performance histories.

6.5 Robustness to the base distribution

To test the robustness of our nonparametric approach to the choice of G_0, or, in other words, to the prior predictive, $\widehat{\pi}_{NIG}(\alpha)$, we estimate the population distribution using a larger scale parameter, ν_0, for the normal, inverse-gamma base distribution. Since ν_0 is the degrees of freedom of the prior predictive, increasing ν_0 leads to a more informed prior predictive distribution. For $\nu_0 \leq 0.6$, the posterior cross-sectional distributions of alpha are no different from the nonparametric population distribution plotted in Figure 1. However, when $\nu_0 \geq 0.7$ the posterior population distributions are no longer multi-modal, instead, they are uni-modal with a mode near 1.4%. Hence, the nonparametric distribution has fewer clusters as the degrees of freedom of its prior predictive distribution increases.

When $\nu_0 = 0.7$, the skewness of the nonparametric population distribution increases to 5.02 from the original 4.31. So when the secondary mode at the large value of alpha is not identified, we find skill to be more probable. We also find the population distribution is more fat-tailed than before with a kurtosis of 110. Hence, under the Bayesian nonparametric estimator of the population distribution, there is a greater chance of a fund, for which we have no information about, being highly skilled, regardless of the value of ν_0.

We can use the inter-quartile range of the prior predictive distribution to help explain how the choice of ν_0 affects the number of modes of the population. The inter-quartile range for $\widehat{\pi}_{NIG}(\alpha)$ goes from a very diffuse, 10^{126}, when $\nu_0 = 0.01$, to a relatively tight 0.18 when $\nu_0 = 0.6$. The tighter range of the prior predictive limits us from learning about the different skilled groups. Instead, a wider spectrum of stock picking ability gets blurred together into larger groups.

An alternative class to the normal, inverse-gamma, base distribution is the flexible, but non-conjugate, normal-SM, base distribution

$$G_0(\mu_\alpha, \sigma_\alpha) \equiv N\left(\mu_\alpha | 0, s_\mu^2\right) \text{SM}\left(\sigma_\alpha \left| 1/2, 2, A/\sqrt{3}\right.\right), \tag{37}$$

where $\text{SM}\left(\sigma_\alpha \left| 1/2, 2, A/\sqrt{3}\right.\right)$ is the base distribution for the standard deviations of the mixture.

The SM distribution is defined in Singh & Maddala (1976) and has the density function

$$f_{SM}\left(\sigma_\alpha \left| 1/2, 2, A\middle/\sqrt{3}\right.\right) = \frac{3A\sigma_\alpha}{(A^2 + 3\sigma_\alpha^2)^{3/2}}. \tag{38}$$

The SM distribution is appealing since it allows for more weight than the inverse-gamma distribution does for σ_αs close to zero. The independence between μ_α and σ_α in Eq. (37)

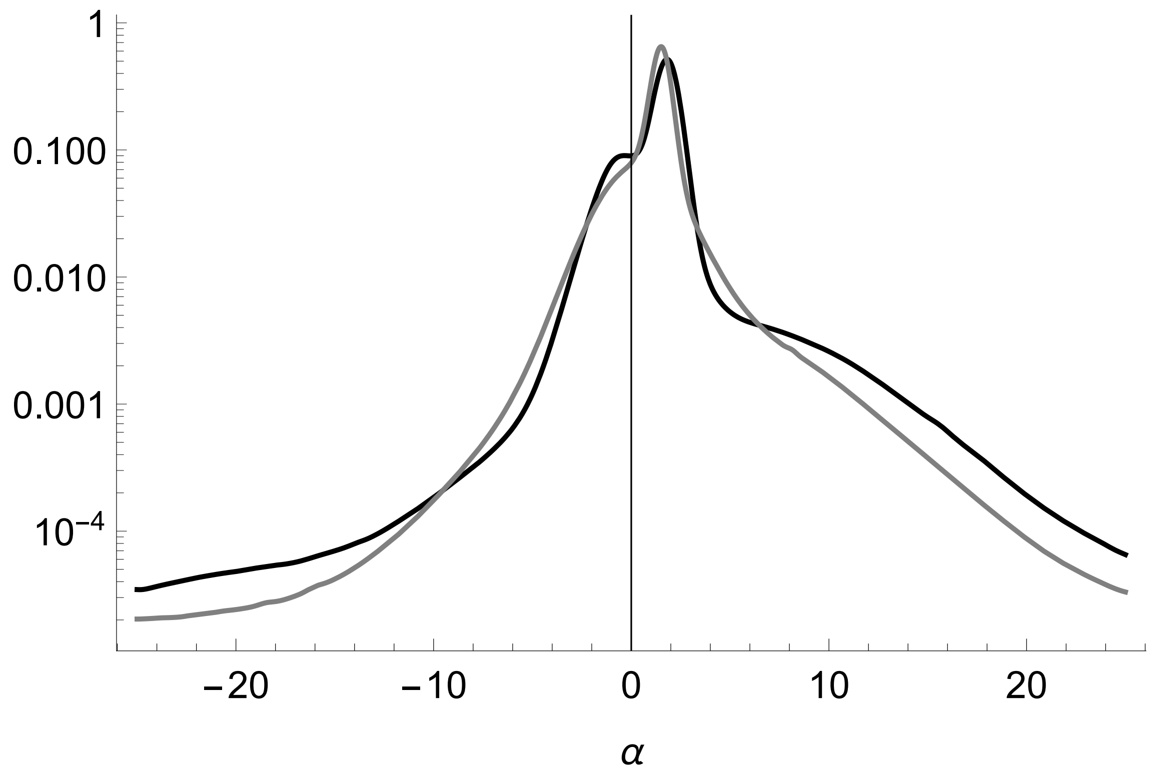


Figure 2: Log transformation of the posterior population distribution of alpha for the Bayesian nonparametric investors, $\ln \widehat{\pi}_{DPM}(\alpha|r_1, \ldots, r_J)$, under homoskedastic returns (black), and autoregressive conditional heteroskedastic returns (gray).

also proves advantageous since it allows the mixture location to move separately from the mixture scales. We set $s_{\mu}^2 = 2$ and $A = 25$, and, because the SM is not conjugate, apply Algorithm 8 of Neal (2000) to make the draws in Step 3 of our sampler.

Under the normal-SM base distribution, we again find the population distribution of skill to be multi-modal with modes located near those of the nonparametric density in Figure 1.[26] The posterior draws of K range from four to fourteen clusters with a median of six clusters. Hence, the number of mixture clusters is marginally larger under the normal-SM base distribution than the normal, inverse-gamma.

The posterior mean of the concentration parameter, B, is also slightly larger under the normal-SM base distribution at 0.59 compared to the normal, inverse-gamma's 0.125. Both of these posterior estimates of B, along with K, support our earlier conclusion that the population distribution of skill is neither normally distributed, nor is skill idiosyncratic to a fund. Instead, the population distribution of skill requires the flexibility that our Bayesian nonparametric approach provides.

6.6 Evolution of the population

Beginning in 1993, the number of new mutual funds entering the actively managed fund industry accelerated. Following 1993 more than 300 mutual funds opened each year. Entry peaked in 1998 with 659 funds opening up for business. This history of funds opening for business allows us to analyze how the population of skill evolved over this time period and also investigate if these new funds were more skilled than the old ones.[27]

Starting in 1981, we move forward in one-year increments up to the year 2000 and iteratively re-estimate the cross-sectional distribution of alpha using the return histories of all the funds to have ever existed up to the specified year.[28] We find that there are four episodes or eras for the population distribution of skill; i) 1981 to 1897, ii) 1988 to 1993, iii) 1994 to 1996, and iv) 1997 to 2000. In Figure 3, we plot in the four panels the population distributions from each of these eras.

In the first panel of Figure 3, 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 fee of 1.5%. This symmetry indicates funds

[26]The posterior results using the normal-SM distribution are available upon request.

[27]To our knowledge a fully time-varying, nonparametric, population distribution has not been done. MacEachern (1999) has come closest with a dependent Dirichlet process.

[28]New mutual funds were included when they had 4-months worth of returns. In contrast with how straight-forward it is to re-apply our nonparametric approach to each year's data set, re-estimating and testing an array of different ordered, finite mixture, models would be very tedious.

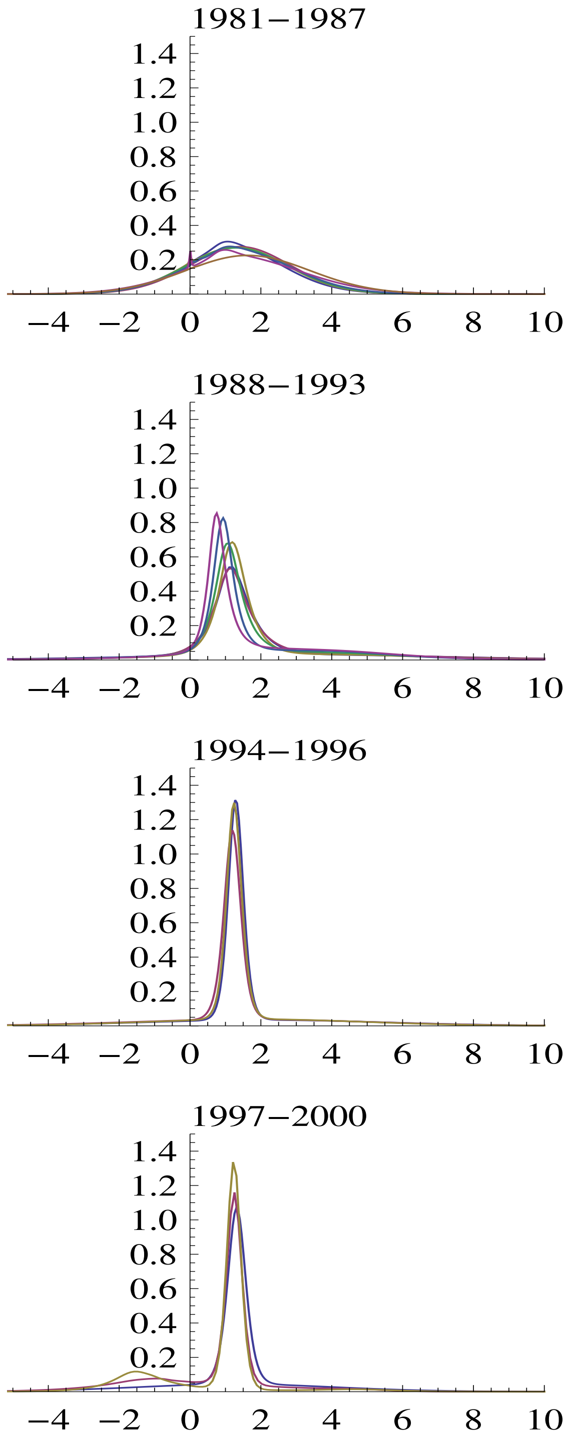


Figure 3: Posterior cross-sectional distributions of alpha starting with the return histories of funds up to 1981 and then incriminating forward one year at a time from 1982 to 2000. More formally, plots of $\widehat{\pi}_{DPM}(\alpha|R_t)$, $t = 1981, \ldots, 2000$, where R_t are the return histories of all the mutual funds ever in business during the 1961 to year t time period. A new fund is only included if it has a history of least four months. Each panel contains the population distributions from that era; i) 1981–1987, ii) 1988–1993, iii) 1994–1996, iv) 1997–2000.

Year	J_t	K_t	Median	Skew	$p_{0.05}$	$p_{0.95}$	$P(\alpha > 0)$
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 median, skewness, and probability of beating the passive four-factor portfolio, $P(\alpha > 0)$, where J_t is the number of mutual fund, both alive and dead, at year t, K_t is the posterior median number of clusters, and $p_{0.05}$, and $p_{0.95}$, are the 5th and 95th percentiles of the cross-sectional mutual fund performance distribution.

are skilled enough on average to cover their costs and equally likely to cover or not cover their fees.

As the entry into the mutual fund industry accelerated during the 1988 to 1993 period, the population distribution in the second panel of Figure 3 tightens around the mean. During this era, the probability of a fund selecting stocks that will result in an abnormally negative alpha declines relative to the earlier episode as the left-hand tails for these distributions are now thinner. Probability of finding highly skilled funds are also on the increase as the right-hand tail of the population pushes out past six percent to eight percent. Hence, the entry of new funds during this era and the performance of existing funds improved the overall performance of the population.

During the 1994 to 1996 era, the population distribution continues to tighten around the mean. However, after 1997 the population begins to change. In the bottom panel of Figure 3, the population distribution starts to skew to the left. Ultimately a second mode appears at a negative alpha. This last era corresponds to the fastest growing period of the mutual fund industry and, according to the population distribution, poorer stock-picking ability.

In Table 2, we list the characteristics and features of the cross-sectional distributions from Figure 3. Each line contains the total number of funds, both in and out of business, from 1961 up to that year, J_t, the posterior median of the number of clusters, K_t, the median and skewness, and the 5th-percentile, $p_{0.05}$, and 95th-percentile, $p_{0.95}$, of the cross-sectional distribution, and the probability of an unknown mutual fund generating 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% of the distribution, if it generated an alpha of two to six percent. This value of alpha is smaller than a highly skilled fund from the 80s. For example, over the 90s the 95th-percentile declined from approximately 6% to 2% per annum. In contrast, these percentiles were never less than 3.7% in the 80s and reached a high of 7.7% in 1989.

During the 90s, the alpha of an unskilled fund, as defined by the 5th-percentile, also declined but in a more noisy fashion than did the alpha for a skilled fund. Poor performance in the mutual fund industry went from -0.95% in 1990 to a low of -2.7% in 1998. Except for 1998 and 1999, the probability of an unknown mutual fund being capable of generating positive alphas stayed right around 90%. Thus, overall mutual fund performance went down during the 90s relative to the 80s.

In Figure 4, 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 period.[29] Funds that were new to the industry were more likely to generate a positive alpha as seen in the positive primary mode. However, over this same period the probability of a new fund generating a negative alpha is also increasing as the negative mode moves further to the left. Thus, we conclude that during the latter 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. However, with each year there was an increasing chance the new fund would be unable to earn a high enough return to justify its fees.

6.7 Comparison of the individual alphas

In Figure 5, we plot each of the 5,136 mutual fund's 95% HPD interval for its alpha along with the posterior median (represented as dots when visible). At the top of the figure are the posterior results for the funds with the shortest histories. At the bottom of the figure

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

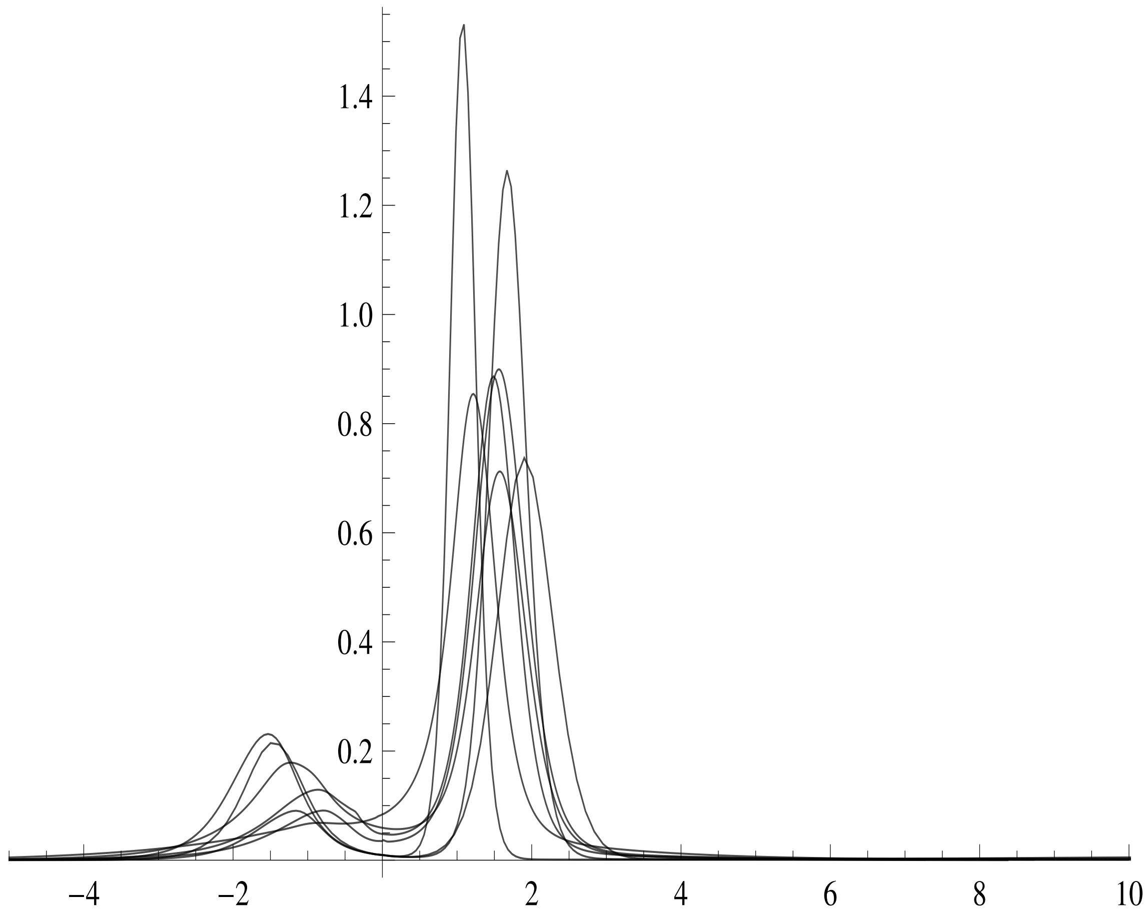
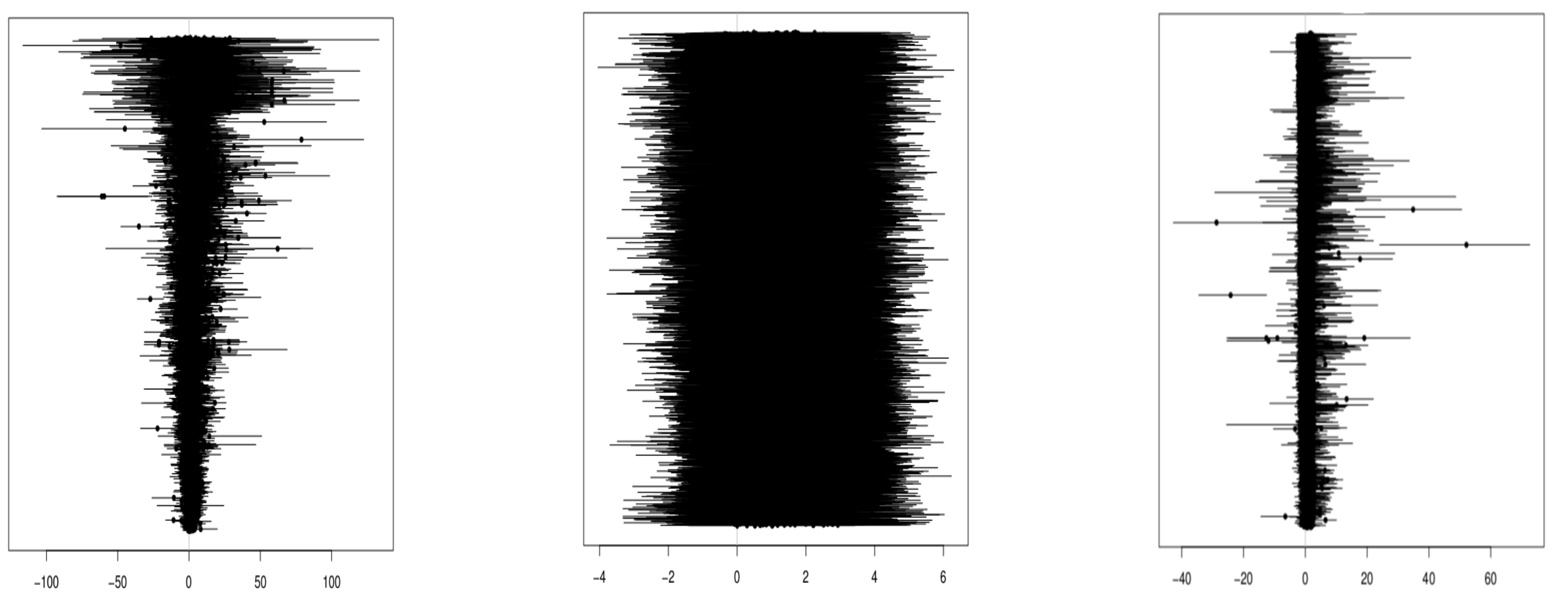


Figure 4: Posterior cross-sectional distribution of alpha from 1995 to 2001 using only the performance histories of the funds that entered the business after 1992 and had a years worth of performance data.

are the results for the funds with the longest histories.

We calculate the 95% HPD intervals for three different models of the population. Panel (a) of Figure 5 plots the HPD intervals where ability is believed to be idiosyncratic to the mutual fund and the prior for alpha is $N(0, s_0^2)$ with $1/s_0^2 = 0$.[30] Panel (b) plots the HPD intervals using the approach of JS. The population is normally distributed but the mean and variance are unknown and modeled with the uninformative Jeffreys prior, $\pi(\mu_\alpha, \sigma_\alpha^2) \propto 1/\sigma_\alpha^2$. Panel (c) of Figure 5 plots the 95% HPD interval of each fund's alpha using our Bayesian nonparametric approach. Our initial guess for the cross-sectional distribution of alpha is again the diffuse Student-t distribution found in Eq. (19) whose mean is zero, scale 0.0011, and 0.1 degrees of freedom.

[30] Because this is the Jeffreys prior the intervals in Panel (a) are equivalent to the 95% confidence intervals of the OLS estimate of alpha.



(a) (b) (c)

Figure 5: Plots of each funds posterior 95% highest probability density interval for alpha sorted from shortest (top) to longest (bottom) fund return history. Panel (a) assumes skill is idiosyncratic to the fund with the prior, $\pi(\alpha) = N(0, s_0^2)$, where $1/s_0^2 = 0$. Panel (b) assumes skill is normally distributed where the unknown population mean, μ_α, and variance, σ_α^2, have the prior, $\pi(\mu_\alpha, \sigma_\alpha^2) \propto 1/\sigma_\alpha^2$. Panel (c) assumes the population distribution is unknown and modeled with the hierarchical, nonparametric, DPM, prior, where $\alpha|\mu_\alpha, \sigma_\alpha^2 \sim N(\mu_\alpha, \sigma_\alpha^2)$, $(\mu_\alpha, \sigma_\alpha^2)|G \sim G$, and, $G \sim DP(B, G_0)$, with $G_0 \equiv NIG(0, \sigma_\alpha^2/0.1, 0.01/2, 0.01/2 * 0.01)$. Where visible the dots in the plots are the posterior median alpha for the fund.

Comparing the funds' posterior HPD intervals in the three panels of Figure 5, it is clear that what one assumes about the nature of the population distribution affects the estimate of a particular mutual fund's alpha. The posteriors in Panels (b) and (c) draw on the performance of other funds to make 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 tighter than those in Panel (a). As a result, the posteriors in Panel (b) and (c) use information from the population to be more precise about a fund's future ability to produce excess returns than the idiosyncratic approach used in Panel (a).

Because the posteriors in Panel (a) of Figure 5 view skill idiosyncratically, the length of a fund's performance window influences the posteriors. Short-lived funds found at the top of Panel (a) have larger and noisier HPD intervals than do the long-lived funds located at the bottom. Although there are also funds with long performance histories that have wide HPD intervals due to noisy and erratic performance histories.

At the other end of the spectrum are the tight and homogeneous HPD intervals in Panel (b) of Figure 5. Believing a fund's performance comes from a normal, cross-sectional, distribution with an unknown population mean and variance shrinks a fund's estimated alpha towards the overall average alpha of the industry. Exceptionally skilled funds get pooled together with average funds and funds with short or noisy performance histories take on the skill characteristics of the population. Hence, the homogeneity of the HPD intervals in Panel (b).

By treating all the alphas as draws from a normal population distribution, unskilled funds, like the one near the top of Panel (a) of Figure 5 where the median alpha is close to -50%, look better in Panel (b) than maybe they should. Furthermore, a highly skilled fund like those in Panel (a) with posterior medians greater than 50% do not look so extraordinary in Panel (b). Our Bayesian nonparametric learning approach automatically determines if such funds should be pooled together or treated separately. In contrast to Panel (b), where K is set, a priori, equal to one, the intervals calculated with our nonparametric approach in Panel (c) randomly group together similarly skilled funds and integrate away the uncertainty of K. Conditional on the random group, information is borrowed from the other funds in the group and used to infer each of the fund's alphas.

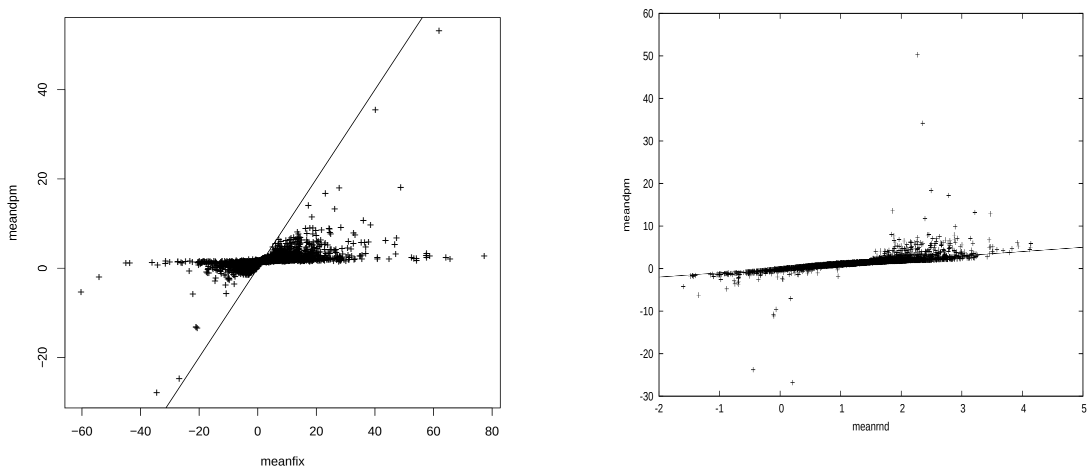
The benefits from letting K be unknown is found in the alpha for the Schroder Ultra Fund. In Panel (c) of Figure 5, Schroder Ultra has the highest posterior median alpha of all the funds at 50% per annum. The next closest fund is the Turner Funds Micro Cap Growth fund at 33%. Our nonparametric sampler in Section 5 randomly groups Schroder Ultra with other funds. Given the likely small size of this random, but highly skilled group, the variance of the group, $\sigma^{2*}_{\alpha,k}$, is likely large. According to Eq. (28), this large variance causes Step 2 of the sampler to make posterior draws of the Schroder fund's alpha that are on average weighted more towards the sample average of the risk-factor adjusted returns of the fund, $\mathcal{T}_i^{-1}\sum_{t=\tau_i}^{T_i} r^*_{i,t}$, and less toward the average of the group, $\mu^*_{\alpha,k}$. In the extreme case where a fund has no peers, our Bayesian nonparametric approach essentially treats the fund idiosyncratically as in Panel (a). As a result, the Schroder Ultra fund's HPD intervals and medians in Panels (a) and (c) are very similar. This similarity stands in stark contrast to Panel (b) where, because of the shrinkage towards the population average, the posterior alpha under a normal population does not even put Schroder Ultra among the top ten performing funds.

6.8 Shrinkage

To determine how much a mutual fund's alpha is affected by one's beliefs about the cross-sectional distribution of mutual fund performance, in Figure 6 we graph two scatter-plots. In each scatter-plot we plot on the y-axis the posterior mean alpha for each of the 5,136 funds from our Bayesian nonparametric method. Panel (a) of Figure 6 plots these nonparametric posterior mean alphas against the posterior mean alpha where skill is believed to be idiosyncratic. In Panel (b), we graph on the x-axis the posterior mean of the alphas where skill is believed to be normally distributed. The forty-five degree line in both plots shows where the assumption about the cross-sectional distribution has no affect on the estimate of a fund's alpha relative to our nonparametric approach.

In Panel (a) of Figure 6, every mutual fund's alpha has moved, to varying degrees, away from the posterior beliefs of someone who believes the skill is idiosyncratic, and towards zero; i.e., the points have moved vertically away from the forty-five degree line towards zero. Hence, those who believe there is an unknown cross-sectional distribution of skill underlying each fund's performance level, and learns about it, discovers that funds identified by those who view skill idiosyncratically as being skilled (unskilled) are less (more) capable of selecting stocks that beat the market.

There are a handful of points in Panel (a) of Figure 6 where skill is so unique and the return histories so informative that treating the fund idiosyncratically is only slightly different from our nonparametric estimates. These funds are those whose posterior mean alpha are 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 we do not treat skill idiosyncratically, but, instead, treat each fund's performance as a draw from an unknown population distribution. By flexibly learning how skill is distributed over the mutual fund



(a) (b)

Figure 6: Scatter plots of the 5,136 fund's posterior mean alphas where in both Panel (a) and Panel (b) the y-axis are the posterior mean under our nonparametric hierarchical prior. The x-axis of Panel (a) are the posterior mean alphas when funds are treated idiosyncratically and the alphas' priors is the Jeffreys prior. In Panel (b) the x-axis are the posterior mean alphas under a normal prior whose mean and variance are unknown and modeled with the Jeffreys prior, $\pi(\mu_\alpha, \sigma_\alpha^2) \propto 1/\sigma_\alpha^2$. To provide a point of reference a 45-degree line has been added to both panels to represent those posterior means where the two relative approaches have the same guess for alpha.

industry, we identify actual fund-specific performance skills while guarding against the noisy performance measures the idiosyncratic approach finds for short lived funds.

Panel (a) of Figure 6 also helps us answer the question posed by Kosowski et al. (2006) and Fama & French (2010) as to whether skilled funds are genuinely talented or just lucky. We find luck playing little to no role in the success of the extraordinary funds whose posterior mean alpha is greater than 4% on both the vertical and horizontal axis. The posterior cross-sectional distribution of skill is very informative about the typical fund's ability in the sense that such funds have performance histories that are not exceptional enough for their posterior distributions to differ from the population. In contrast, highly skilled funds have posterior distributions that are distinctly different from the posterior population distribution. In Panel (a) of Figure 6, these exceptionally skilled funds were not lucky but were truly skilled since their performance lead to posterior mean alphas greater than 4% under both the nonparametric hierarchical prior (y-axis) and the idiosyncratic Jeffreys prior (x-axis). A lucky fund is one of the many "skilled" funds whose posterior mean on the horizontal axis is larger than 5%, for instance, the extreme fund whose posterior mean is nearly 80%, but then shrinks close to zero under the nonparametric hierarchical prior. Later, in Table 3, we list the names of these highly skilled funds.

In contrast to Panel (a), many of the points in Panel (b) of Figure 6 lie on the forty-five degree line. These alphas belong to funds having the same average ability and variance as that of the normal cross-sectional distribution. Modeling the performance of this particular group of funds as draws from a normally distributed cross-section would not interfere with the expected value of the posterior alphas. However, identifying this group of funds from the cross-section of funds, a priori, would be impossible. Our nonparametric approach on its own identifies these ordinary funds as it learns about the population distribution.

From the off-diagonal points in Panel (b) of Figure 6 we discover some funds' alphas are drawn from groups whose means and variances are different from the global average level of skill and its variance. If the population is thought to be normally distributed, many extraordinarily skilled and unskilled funds would go undetected. For example, by learning the population distribution we find that the posterior alpha for Potomac OTC/Short fund under-performs the market by an average of 27% a year, and does not have a positive alpha.[31] In stark contrast, investors who think skill is normally distributed over the population believe the Potomac OTC/Short fund can produce, on average, an excess market return of less than 1% per year (located on the x-axis of Panel (b)).

[31] Potomac is the worse performing fund in our panel of mutual funds whose posterior mean alpha in both panels of Figure 6 is the point located at the very bottom.

The Potomac fund is not unique in this respect. Many other extraordinarily skilled and unskilled funds look quite ordinary to those who believe skill is normally distributed across funds. Clearly, the assumption of a normal population distribution adversely affects the estimate of these funds alphas.[32]

6.9 Posterior distribution of each fund's alpha

To evaluate and compare the skill level of the funds in our panel, we plot in Figure 7 the posterior distribution of alpha for all 5,136 mutual funds. The blue lines are the posteriors of the 3,844 mutual funds that were still open for business at the end of our sample. The red lines are the posterior distributions for the 1,292 funds that were no longer in business. Both types of colored lines have a degree of transparency so that darker shades of red, blue, or purple (combinations of red and blue) show where the posterior distributions are concentrated.

As we showed in Eq. (7), a fund's posterior is proportional to

$$\pi(\alpha_i|r_1,\dots,r_J) \propto \pi(\alpha_i|r_{-i})f(r_i^*|\alpha_i), \ \ i=1,\dots,J,$$

where r_i^* is the risk and factor adjusted return history of the ith fund.[33] In other words, when the return histories of the other $J-1$ funds, r_{-i}, have been observed, but before the ith funds returns, r_i, are seen, our understanding of the ith fund's skill level is best described by the "updated prior" distribution, $\pi(\alpha|r_{-i})$; i.e., the posterior population distribution or posterior predictive distribution.

How different a fund's posterior, $\pi(\alpha_i|r_1,\dots,r_J)$, is from the posterior predictive distribution, $\pi(\alpha_i|r_{-i})$, is how exceptional the fund is from a fund we have no information about. The more uniquely shaped a fund's posterior the more its likelihood, $f(r_i^*|\alpha_i)$, plays into our understanding of its ability. Because our panel consists of 5,136 funds, the skewness, kurtosis, and multi-modality of the nonparametric posterior population distribution in Figure 1 is unaffected when we drop any fund from the panel. Given this robustness of the posterior population distribution, we define a fund as being exceptional when its posterior for alpha does not have the same bi-modal shape as $\widehat{\pi}_{DPM}(\alpha|r_1,\dots,r_J) \approx \pi(\alpha_i|r_{-i})$. To identify these exceptional funds in Figure 7 we plot in orange the posterior population distribution, $\widehat{\pi}_{DPM}(\alpha|r_1,\dots,r_J)$.

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

[33] We could integrate out the betas and the return variance in which case a fund's likelihood function, $f(r_i|\alpha_i)$, is a Student-t distribution but we would require a different sampler for the DP unknowns.

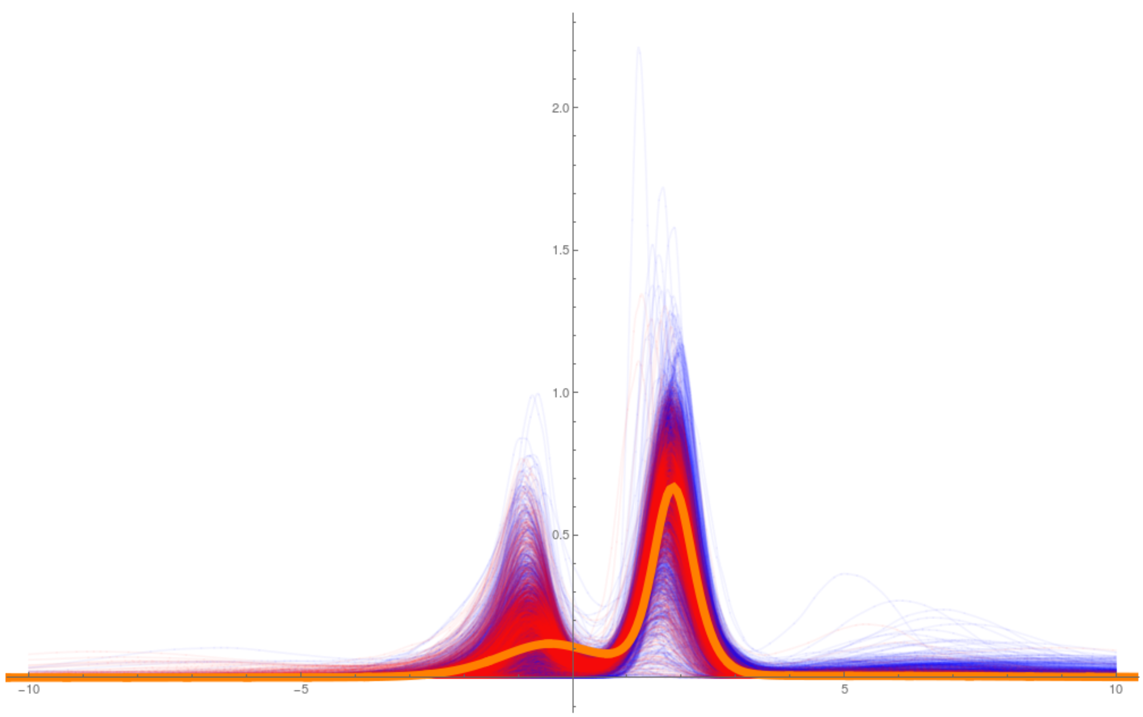


Figure 7: Posterior cross-sectional distribution of alpha in orange and every funds posterior distribution of alpha, $\pi(\alpha_i|r_1, \ldots, r_{5136})$, $i = 1, \ldots, 5136$, where the funds that were still in business at the end of the sample are plotted in blue, whereas those funds that were not in business are in red. Darker shades of red or blue indicate a higher concentration of funds having similar shaped posteriors.

Many of the posterior distributions in Figure 7, be the fund dead or alive, resemble the bi-modal, cross-sectional distribution of alpha. The large number of funds whose posterior is similar to the population shows how important the posterior population distribution is in determining the skill level of a fund, and, hence, how vital it is to modeling the population. It also shows how little information there is about skill in most mutual funds return history.[34] For the typical fund with average performance, meaning its likelihood is relatively flat, the posterior cross-sectional distribution resolves most of the uncertainty in its alpha and, hence, its posterior distribution shrinks towards the population distribution. Also, their is not enough information in the fund's performance history for it to be classified in a particular group. Instead, it is likely the fund finds itself belonging to the different clusters just like a fund we know nothing about.

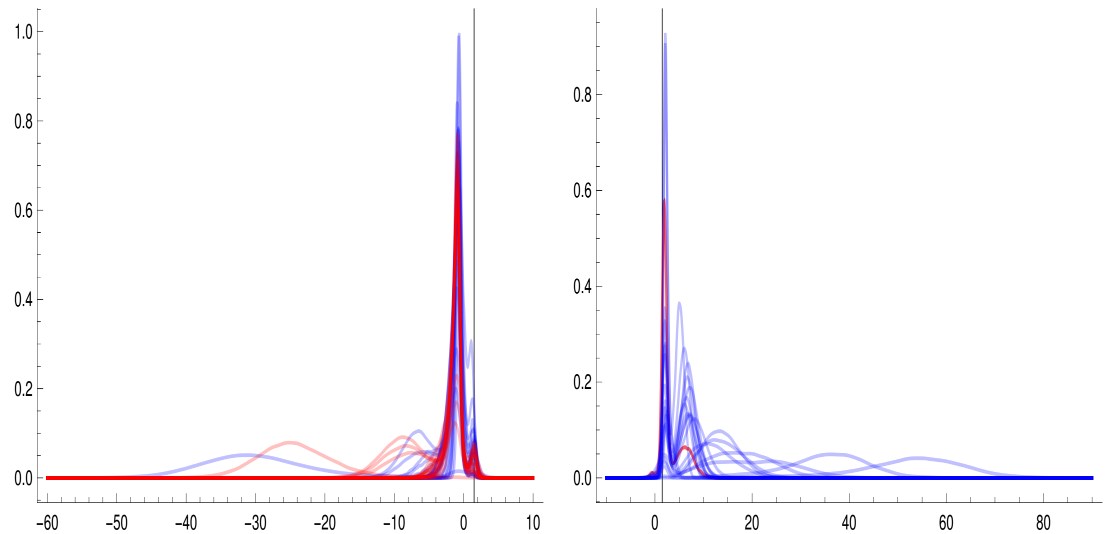
From Figure 7, we draw three conclusions about the skill level of the funds. First, regardless of a fund being in or out of business, its posterior in general has two modes. The secondary modes are close to -1%, indicating that the funds have the potential to generate losses. The primary modes are close to 2%. Thus, many of the 5,136 funds are likely to cover their fees, but there is a non-trivial chance the fund will fail to earn a return that covers its costs.

The second finding from the posteriors in Figure 7 is the credible evidence of there being a few exceptional funds. These funds are extraordinary either because they are unskilled or because they are highly skilled. There are 21 exceptionally skilled funds where there is a 95% probability of its alpha being greater than 1.5%, and 50 exceptionally unskilled funds that have the same probability of being less than 1.5%.

Finding so few exceptional funds runs counter to earlier empirical findings where there is a larger presence of skilled and/or unskilled funds (Kosowski et al. 2006, Fama & French 2010, Barras et al. 2010, Ferson & Chen 2015). However, as we have already pointed out, these earlier findings suffer from noisy estimates of alpha. Hence, we conclude that most mutual funds are not extraordinarily talented or unskilled. Instead, most funds have a higher chance of being just gifted enough to select stocks that, on average, result in a return that justifies their costs, expenses, and fees.

Our third finding concerns those funds that are exceptionally unskilled but, for some reason, are still in business. Several predictive densities in Figure 7 have fat tails over negative values of alpha and were still not out of business. In Panel (a) of Figure 8, we plot in blue the alive fund's posterior and in red, the dead fund's posterior distributions for the

[34]Not being able to accurately estimate a funds alpha with only its return history has been a well known problem with OLS estimates of the alpha. The mutual fund skill literature is full of results where the R^2s are small.



(a) Unskilled funds (b) Skilled funds

Figure 8: The posterior distribution of alpha where Panel (a) plots the posteriors of 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 posteriors for 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). Posteriors for the funds that are in business at the end of the sample are plotted in blue, whereas those that have closed are in red. In Panel (a) there are a total of twenty-one dead unskilled funds. In Panel (b) there is only one dead skilled fund (T. Rowe Price Capital Appreciation Fund).

50 exceptionally unskilled funds. Twenty-nine of these poor performing funds were still in business at the end of our sample. This includes the worse performing fund from our panel – the Potomac OTC/Short fund, whose expected alpha is -26% and has a 3% chance of losing between 20% to 40% percent a year.

The posterior mean alphas for the extinct funds Bowser Growth Fund and Ameritor Industry Fund are the most negative in the panel at -24% and -7%, respectively. The other forty-seven unskilled funds earn their investors, on average, a gross alpha of between zero to five percent a year. Nine of these funds had been in business since the early 1960s and were still in business in 2001. Thus, we conclude that poor performance does not necessarily lead investors to divest their money from the fund. Perhaps there are restrictions prohibiting investors from withdrawing their investment or limiting redemptions.

In Panel (b) of Figure 8, we plot the posterior distribution for the 21 exceptionally skilled funds. Notice how different these funds' posteriors are from the nonparametric posterior population distribution in Figure 1. These skilled funds have densities where the

Fund	Years of Operation	Alpha	95% HPD
Turner Funds: Micro Cap Growth	1998-2001	33.07	[13.81, 51.01]
Schroeder Ultra Fund	1997-2001	49.77	[24.07, 71.86]
Artisan Mid Cap Fund	1997-2001	16.77	[1.26, 27.92
Needham Growth Fund	1996-2001	17.40	[0.51, 33.22]
Olstein Financial Alert Fund/C	1996-2001	13.84	[5.86, 21.87]
Fremont Mutual Fds:US Micro Cap Fund	1994-2001	13.23	[1.35, 21.94]
PIMCO Funds:Stocks Plus Fund/Instl	1993-2001	2.46	[1.33, 3.96]
Fidelity Dividend Growth	1993-2001	4.87	[1.08, 10.25]
Managers Funds:US Stock Market Plus	1992-2001	2.44	[1.44, 3.72]
Fidelity Low Priced Stock	1990-2001	5.93	[1.59, 9.34]
Victory Funds:Diversified Stock Fund/A	1989-2001	2.88	[1.16, 6.01]
T Rowe Price Capital Appreciation Fund	1986-1999	3.75	[1.49, 6.95]
JP Morgan Growth & Income Fund/A	1987-2001	6.30	[1.08, 12.24]
Gabelli Growth Fund	1987-2001	5.29	[1.40, 9.10]
Weitz Series Fund:Value Portfolio	1986-2001	4.50	[1.11, 9.07]
IDEX Janus Growth Fund/A	1986-2001	4.87	[1.15, 10.04]
Gabelli Asset Fund	1986-2001	4.99	[1.62, 8.01]
Oppenheimer Growth/A	1973-2001	3.82	[1.10, 8.10]
AXP Growth Fund/A	1972-2001	4.71	[1.24, 8.73]
Janus Fund	1970-2001	3.90	[1.21, 7.93]
Vanguard Morgan Growth/Inv	1968-2001	4.40	[1.94, 6.50]

Table 3: Individual mutual funds' posterior mean of alpha and 95% highest posterior density (HPD) interval from the newest to oldest, and who have at least a 95% chance of returning a market excess return greater than the average mutual fund fee of 1.5%.

weight assigned to the alphas between 5% to 10% is noticeably greater than the population distribution. A few of these exceptionally skilled funds have a primary mode at an alpha larger than the primary mode of the population. Half of these funds have a primary mode near 1.5%. Only one of the funds was no longer in business.

In Table 3, we list from shortest to longest return history, the 21 skilled funds posterior mean of alpha, the number of years the fund had been in business, and the 95% HPD interval. Four of the funds possess the extraordinary ability to pick stocks such that their expected alpha is greater than 15% per year. The top performers are newer funds with less than seven years of experience. There are also a few highly skilled funds, like the Janus Fund, that have been in business since the 1970s. Hence, not all skilled funds suffer from Berk & Green's (2004) theory of decreasing returns to scale.

7 Conclusion

With our Bayesian nonparametric learning approach, we find the posterior population distribution of the alphas for the actively managed mutual fund industry to be bi-modal, fat-tailed, and skewed towards higher levels of skill. The primary mode is an alpha of 1.8% per year and the secondary mode is -0.65%.

These modes and the skewness of the population means there is a higher probability than previously thought that a fund, we know nothing about, will be skilled. In contrast, under either a parametric, Gaussian, hierarchical, prior for alpha, or a finite ordered mixture population model comprised of two Gaussian distributions, the alpha of an extraordinary fund gets shrunk towards the modes of these parametric distributions. Hence, we also find more exceptionally skilled funds than these parametric approaches.

Compared to an idiosyncratically skilled population, our nonparametric approach finds fewer extraordinarily skilled and unskilled funds. Out of 5,136 funds, we find 21 (50) funds have a 95% chance of its alpha (not) exceeding the average fee charged by a fund of 1.5%. However, these extraordinary funds have very informative likelihoods that drives a wedge between their alpha's posterior and the population distribution. Hence, these skilled funds are genuinely talented at picking under-valued stocks and are not just lucky. These skilled fund's posterior mean alphas were similar under both our nonparametric prior and the OLS's idiosyncratic prior. Ordinary funds that had larger posterior mean alphas under the idiosyncratic prior lack the empirical performance in their likelihood to drive a wedge between their posterior and the nonparametric population distribution of skill.

Learning how skill is distributed across the mutual fund industry with our Bayesian nonparametric approach and using this information to better infer the stock picking skill of a fund applies to other problems. For example, one can nonparametrically estimate the population distribution of the multiple risk-factors' beta coefficients. Currently, we are investigating these and other questions, along with extending our Bayesian nonparametric learning approach to a time-varying setting.

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., Gagliardini, P. & Scaillet, O. (2018), The cross-section distribution of fund skill measures, Technical report, Université De Genéve.

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.

Bassetti, F., Casarin, R. & Leisen, F. (2014), 'Beta-product dependent Pitman–Yor processes for Bayesian inference', *Journal of Econometrics* **180**(1), 49–72.

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

Bernardo, J. M. & Smith, A. F. M. (2000), *Bayesian Theory*, Wiley.

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.

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

Carvalho, C. M., Lopes, H. F., Polson, N. G. & Taddy, M. A. (2010), 'Particle learning for general mixtures', *Bayesian Analysis* **5**(4), 709–740.

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

Chen, Y., Cliff, M. & Zhao, H. (2017), 'Hedge funds: The good, the bad, and the lucky', *Journal of Financial and Quantitative Analysis* **52**(3), 1081–1109.

Chib, S. & Hamilton, B. (2002), 'Semiparametric Bayes analysis of longitudinal data treatment models', *Journal of Econometrics* **110**, 67–89.

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, p. 223–268.

Elton, E. J. & Gruber, M. J. (2013), Mutual Funds, Vol. 2 of *Handbook of the Economics of Finance*, Elsevier, p. 1011–1061.

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.

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, Springer Series in Statistics. New York/Berlin/Heidelburg.

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

Geweke, J. (1992), Evaluating the accuracy of sampling based approaches to the calculation of posterior moments, *in* Bernardo, Berger, Dawid & Smith, eds, 'Bayesian Statistics', Vol. 4, Oxford: Clarendon Press.

Geweke, J. (2007), 'Interpretation and inference in mixture models: Simple MCMC works', *Computational Statistics & Data Analysis* **51**(7), 3529–3550.

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

Harvey, C. R. & Liu, Y. (2018), 'Detecting repeatable performance', *The Review of Financial Studies* **31**(7), 2499–2552.

Hirano, K. (2002), 'Semiparametric Bayesian inference in autoregressive panel data models', *Econometrica* **70**, 781–799.

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

Jensen, M. J. (2004), 'Semiparametric Bayesian inference of long-memory stochastic volatility models', *Journal of Time Series Analysis* **25**(6), 895–922.

Jensen, M. J. & Maheu, J. M. (2010), 'Bayesian semiparametric stochastic volatility modeling', *Journal of Econometrics* **157**(2), 306–316.

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

Kasahara, H. & Shimotsu, K. (2015), 'Testing the number of components in normal mixture regression models', *Journal of the American Statistical Association* **110**(512), 1632–1645.

Kleinman, K. P. & Ibrahim, J. G. (1998), 'A semiparametric Bayesian approach to the random effects model', *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.

MacEachern, S. N. (1999), 'Dependent nonparametric processes', *American Statistical Association: Proceeding of the Section on Bayesian Statistics* .

Malsiner-Walli, G., Frühwirth-Schnatter, S. & Grün, B. (2017), 'Identifying mixtures of mixtures using Bayesian estimation', *Journal of Computational and Graphical Statistics* **26**(2), 285–295.

Murphy, K. P. (2012), *Machine Learning: A Probabilistic Perspective*, MIT Press, Cambridge, Mass.

Müller, P. & Rosner, G. L. (1997), 'A Bayesian population model with hierarchical mixture priors applied to blood count data', *Journal of the American Statistical Association* **92**(440), 1279–1292.

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. (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.

Richardson, S. & Green, P. J. (1997), 'On Bayesian analysis of mixtures with an unknown number of components', *Journal of the Royal Statistical Society, Series B* **B,59**, 731–92.

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 measurerement of mutual funds, hedge funds, and institutional accounts', *Annual Review of Finanacial Economics* **3**, 537–574.

West, M., Muller, P. & Escobar, M. (1994), Hierarchical priors and mixture models with applications in regression and density estimation, *in* P. R. Freeman & A. F. Smith, eds, 'Aspects of Uncertainty', John Wiley.