ESTIMATING MUTUAL FUND SKILL: A NEW APPROACH*

Angie Andrikogiannopoulou† Filippos Papakonstantinou‡
London School of Economics Imperial College London

August 2016

Abstract

We propose a novel methodology that jointly estimates the proportions of skilled/unskilled funds and the cross-sectional distribution of skill in the mutual fund industry. We model this distribution as a three-component mixture of a point mass at zero and two components — one negative, one positive — that we estimate semi-parametrically. This generalizes previous approaches and enables information-sharing across funds in a data-driven manner. We find that the skill distribution is non-normal (asymmetric and fat-tailed). Furthermore, while the majority of funds have negative alpha, a substantial 13% generate positive alpha. Our approach improves out-of-sample portfolio performance and significantly alters asset allocation decisions.

Keywords: Mutual Funds, Skill, Performance, Specification Uncertainty, Asset Allocation

JEL Classification: G11, G23, C11, C52, C58

---

*For helpful comments, we would like to thank Enrico Biffis, Markus Brunnermeier, Pasquale Della Corte, Michael Dempster, Robert Kosowski, Olivier Scaillet, Fabio Trojani, and Dimitri Vayanos, seminar participants at Imperial College London and the University of Geneva, and conference participants at the 2012 Cambridge-Princeton conference, the 2014 European Finance Association meeting, and the 2014 European Seminar on Bayesian Econometrics. This manuscript was previously circulated under the title "A Direct and Full-information Estimation of the Distribution of Skill in the Mutual Fund Industry."

†London School of Economics, Department of Finance, Houghton Street, London WC2A 2AE, UK, e-mail: a.andrikogiannopoulou@lse.ac.uk.

‡Imperial College Business School, Tanaka Building, South Kensington Campus, London SW7 2AZ, UK, e-mail: fpapakon@imperial.ac.uk.

1 Introduction

The evaluation of mutual funds' ability to generate excess returns has great importance both for the study of financial markets and for investors. For example, estimating the population distribution of fund skill is useful for assessing the industry's performance and the rationality of investing in funds, while estimating the skill of individual funds is useful in asset allocation. The literature has followed two approaches for estimating skill in the *population*: One focuses on distinguishing between funds with zero and nonzero alpha and estimates the proportions of these types using fund-level hypothesis tests, while the other estimates the alpha distribution assuming it has a simple shape, usually normal. The literature has also followed two approaches for estimating the skill of *individual* funds: One estimates each fund's alpha in isolation using fund-level factor regressions, while the other combines fund-level information with information from the population distribution, which is presumed to be known a priori or is estimated under specific parametric assumptions (normality).

In this paper, we propose a novel methodology that generalizes previous approaches in several important respects, and leads to improved inference both at the population and at the individual level. First, it introduces the proportions of skilled/unskilled funds as model parameters that are jointly estimated with the entire distribution of skill in a unified framework. Second, it estimates the skill distribution using a flexible semi-parametric model that is able to fit a wide range of distributions. Third, treating the population distribution as unknown and estimating it using a flexible model enables us to share information across funds in a manner dictated by the data, hence to improve our estimates of individual fund alphas, as whatever we learn about the population feeds back into the estimation of each fund's alpha (and vice versa).

Applying our methodology to the monthly returns of 3,497 actively managed open-end US equity funds for the period January 1975 through December 2011, we obtain three main results. First, even though the majority of funds have negative skill, there is a substantial proportion that generates positive excess returns, net of fees and expenses. Second, the distribution of alpha is highly non-normal, with fat tails, negative skewness, and a mode slightly below zero. Third, our methodology has significant effects on asset allocation decisions and leads to improved portfolio performance compared to alternative methodologies that do not incorporate information sharing across funds and/or do so under restrictive parametric assumptions.

Prior practice and economic sense (e.g., Jensen, 1968; Ferson and Schadt, 1996; Barras, Scaillet and Wermers, 2010) suggest that funds can be separated into three skill groups: i) those that generate negative excess returns because, e.g., they consistently trade on misinformation, suffer from exploitable behavioral biases, or have high transaction costs or fees, ii) those that generate zero excess returns, consistent with the long-run equilibrium of Berk and Green (2004), and iii) those that generate positive excess returns, e.g., because they possess superior information or trading skill. Furthermore, it has been suggested that the alpha distribution of mutual funds may be non-normal, due to the heterogeneous investment strategies and risk-taking of funds (Kosowski et al., 2006). We accommodate all these features by modeling the cross-sectional distribution of alpha with a three-component specification consisting of a point mass at zero and two components — one with negative and one with positive support — that we estimate semi-parametrically. Specifically, we represent each of the nonzero components of the skill distribution as a mixture of log-normals, where we treat the number of components in each mixture as unknown parameters to be estimated. We estimate our model using Bayesian techniques.

Our three-component specification introduces the proportions of zero-, negative-, and positive-alpha funds as model parameters that contribute to the likelihood. As we show in simulations, this significantly improves inference about the proportions relative to the extant fund-by-fund hypothesis-testing approach. The first reason for this is that it uses *more* information: it uses all the return information for all funds instead of summary scores (p-values or t-statistics) derived separately for each fund from fund-by-fund tests. Second, it uses this information in a *direct* manner: the proportions can be estimated directly using the information in the likelihood instead of being calculated in a separate step from the number of hypothesis rejections. Third, it estimates these proportions *simultaneously* with the entire distribution, hence information about the shape of the distribution feeds into the estimation of the proportions. Notably, the two approaches yield estimates that have quite distinct economic implications. For example, while Barras, Scaillet and Wermers (2010) find that positive-alpha funds are almost non-existent (at 0.6% of the population) and conclude that the industry is at the long-run equilibrium of Berk and Green (2004) as most funds (75%) have zero alpha, we find a rather sizable proportion (13%) of positive-alpha funds and that most funds (78%) have negative alpha, while only a small mass of funds, if any, have zero alpha. These estimated proportions can also be useful for simple models of investor behavior examining, e.g., the rationality of investing in mutual funds, or models of simplistic investor behavior (e.g., of categorical thinking as in Mullainathan, 2000 or Barberis and Shleifer, 2003).

The use of a mixture model with an unknown number of components is appealing because, as is well known in the literature, such a representation can approximate a wide range of distributions very well (e.g., Miller and Harrison, 2015). Our model is a substantial generalization of previous approaches, which assume a specific skill distribution (e.g., Baks, Metrick and Wachter, 2001) or a specific form for it (e.g., a normal in Jones and Shanken, 2005), as it does not *impose* but rather *allows* for a variety of features, such as skewness, fat tails, multiple modes, and the existence of a point mass at zero alpha, and lets the data determine whether they are present in the true distribution. Indeed, our simulations demonstrate the flexibility of our model compared to simpler alternatives: simulating skill from data generating processes ranging from discrete values, to distributions with and without skewness and kurtosis, and with and without a point mass at zero alpha, we find that our model performs well in all cases. Applying our methodology to the real data, we find that the estimated distribution of alpha is significantly non-normal, with fat tails and negative skewness. Specifically, we find that about 99% of the funds in the population have alpha between −6% and 3% per year, while concentrating on the right tail, we find that about 5% of funds yield more than 1% in excess of the benchmark return per year, and a little more than 1% of funds yield an excess return above 2%. In contrast, the normal model estimates fewer funds with alphas far from zero: for example, it places about 0.5% of fund alphas outside the range (−4%, +2%), while we estimate about 3% (about 6 times as many) outside this range. Measures of skewness and tail weight significantly differ from those for the normal, and distance measures between our estimated skill distribution and the normal suggest that they differ about as much as a standard normal from a t distribution with 2 degrees of freedom. Notably, these non-normality measures are robust, i.e., they ignore the extreme tails which cannot be estimated precisely.

Our approach should also improve the estimation of fund-level alphas, as it enables us to share information across funds 'objectively,' i.e., in a way dictated by the data rather than imposed by restrictive distributional assumptions. To be specific, we share information across funds in two ways. One, we use information from all funds' allocations to skill types and their alphas to learn in a flexible manner about the population, i.e., the proportion of each type and the entire distribution of alpha. And two, we use information from this population distribution together with the information in each fund's returns to learn about its allocation to skill types and its alpha.[1] Out-of-sample tests confirm our methodology provides an advantage over simpler alternatives in identifying funds with high alphas. In particular, we conduct 5-year rolling estimations for the period 1975–2011 to compare the performance of monthly-rebalanced portfolios formed using our methodology with that of portfolios formed using i) the normal model, and ii) the False Discovery

[1] We note that this estimation solves for an equilibrium density — the joint posterior of all model parameters — which is a stochastic analogue of a fixed point. As we explain in Section 2.4, we implement it using a Gibbs sampler, which is an iterative procedure that is itself the stochastic analogue of an algorithm that solves for a fixed point.

Rate (FDR) methodology of selecting funds with high probability of having a positive alpha, which Barras, Scaillet and Wermers (2010) show to be superior to a simple ranking of portfolios by their Ordinary Least Squares (OLS) alpha or t-statistic.[2] Estimating, for each portfolio, rolling 4-factor OLS alphas at each month in the period 1990–2011, we find that the portfolio constructed using our methodology delivers a mean annualized alpha of 3.2% over the years, versus 2.3% for the FDR portfolio and 1.7% for the portfolio from the normal model.

An application of our improved fund-level inference is in asset allocation. Most previous studies in Bayesian portfolio analysis approach the investment decision from the perspective of investors with specific prior beliefs with respect to the distribution of fund skill. Instead, like Jones and Shanken (2005), we do not take this distribution as given, but we learn about it from the returns of all funds. However, we generalize the approach of Jones and Shanken (2005), who allow for *restricted* learning under the assumption that skill is normal, and we allow for almost *unrestricted* learning under a flexible specification for the skill distribution. Thus, our approach is more suitable for agnostic investors who do not hold specific prior beliefs about the skill distribution or its shape. Studying the simple asset allocation problem of an investor who allocates wealth between the risk-free asset, the benchmark portfolios, and a single mutual fund, we show that unrestricted learning across funds under our flexible model has an economically significant impact on the optimal allocation decision.

A well-founded concern with our (as with any) Bayesian estimation is that the results may be sensitive to the priors. In the absence of concrete prior information, we utilize hierarchical priors, which model the lack of prior information about model parameters through a distribution over priors. More specifically, we have two layers of priors, where the first layer — the population distribution of skill — is itself part of the model and estimated from the data, and the second layer — the hyperpriors for the population parameters — contains very little information. Thus, the hierarchical approach contains more data-driven information and yields more objective and robust inference (Gelman, 2006; Robert, 2007). Indeed, we perform an extensive sensitivity analysis and show that our results are robust to alternative prior specifications.

We also perform several robustness checks. We allow for dependence between skill, the factor loadings, and error precision, since they may all be related to a fund's investment strategy, but find the correlations to be small. Furthermore, we use a conditional version of the asset pricing model that allows for time-varying market risk exposures, and we relax the assumption that errors are cross-sectionally independent; in both cases, a small probability mass is transferred from positive

[2]The FDR methodology (Benjamini and Hochberg, 1995; Storey, 2002) corrects for the expected proportion of funds for which the null hypothesis of zero alpha is incorrectly rejected in multiple hypothesis testing.

to negative values of alpha. We also show that our estimated distribution of skill is not driven by our distributional assumption for the factor loadings and the errors. Finally, we estimate two variations of our model: one in which we blur the distinction between zero- and nonzero-alpha funds and replace the point mass with a narrow normal centered at zero, and one in which we ignore the distinction between negative- and positive-alpha funds and assume they are drawn from a common rather than from two non-overlapping distributions.

This paper contributes to a growing literature on mutual fund performance evaluation. Within the frequentist framework, studies have focused on fund-by-fund hypothesis tests. Controlling for false discoveries in multiple testing, Barras, Scaillet and Wermers (2010) estimate the proportion of funds with zero, negative, and positive alpha. Combining hypothesis tests with a simulation approach, Kosowski et al. (2006) examine whether it is likely that some funds have skill, by comparing the distribution of alpha t-statistics derived from simulated returns of zero-alpha funds with that derived from actual returns. Fama and French (2010) use a similar approach to estimate the distribution of skill under the assumption that it is normal with mean zero. These studies use some cross-sectional information — the t-statistic (or p-value) distribution — to produce partial population-level inference, but this information is not useful for fund-level inference. Within the Bayesian framework, studies have combined information about each fund with information from the population distribution of skill. Typically, they assume that this population distribution is known a priori with certainty (e.g., Baks, Metrick and Wachter, 2001; Pastor and Stambaugh, 2002*a*,*b*; Busse and Irvine, 2006; Avramov and Wermers, 2006), therefore they do not share any information across funds. Our paper is closer to the hierarchical model of Jones and Shanken (2005), who propose that funds' skill is drawn from a normal distribution whose mean and variance they estimate from the returns of all funds, hence they share information across funds to draw inference both at the fund- and at the population-level. We extend and generalize their approach, by replacing the assumption that fund skill is normal with a flexible representation of the cross-sectional skill distribution.[3] As a result, we are also able to estimate the proportions of skilled/unskilled funds simultaneously with the entire distribution, and to share information across funds in a manner that is not restricted by the normal parametric assumption.

[3]Also see Sastry (2015), who proposes that skill is drawn from two normals; this makes it possible to estimate a distribution with fat tails and skewness, but has limited flexibility to estimate these features in an unconstrained, data-driven manner. Also, Cohen, Coval and Pastor (2005) informally incorporate information sharing across funds, since they calculate a fund's skill as a weighted average alpha across funds with correlated holdings, the intuition being that high correlation between a fund's holdings and those of funds with high/low performance should be informative about the former's skill. This approach improves fund-level inference, but is not relevant for population-level inference.

The remainder of the paper is structured as follows. In Section 2, we present our model of fund skill and our estimation algorithm. In Section 3, we use simulated data to compare our approach with extant alternatives. Subsequently, we focus our analysis on real data from the US mutual fund industry; we present these data in Section 4. In Section 5, we present our results on the proportions of skilled/unskilled funds and the entire distribution of mutual fund skill. In Section 6, we study the implications of our methodology for fund-level inference by solving a simple asset allocation problem and conducting out-of-sample tests of the performance of fund portfolios constructed using our methodology and alternatives. In Section 7, we present additional results using subgroups of funds with different investment objectives, funds that are active during different subperiods, as well as fund returns *before* fees and expenses. In Section 8 we present robustness checks and in Section 9 we conclude. An appendix contains additional details on the estimation, and an online appendix contains additional tables and figures.

2 Econometric model and estimation

Here, we present the model of returns, our model of fund skill, a hierarchical version of our model which enables us to perform Bayesian estimation, and our estimation algorithm.

2.1 Model of returns

We evaluate fund performance using the linear factor model

$$r_{it} = \alpha_i + F_t' \beta_i + \varepsilon_{it}, \tag{1}$$

with r_{it} the month t return of fund i in excess of the risk-free return, α_i the fund-specific alpha or skill, F_t the month t factor returns, β_i the fund-specific factor loadings, and $\varepsilon_{it} \sim \mathcal{N}\left(0, h_i^{-1}\right)$ the fund- and time-specific errors, with h_i the fund-specific precision.[4]

In the baseline specification, we use the Carhart (1997) 4-factor model, i.e., F_t contains the excess return of the market portfolio (*M*) and the returns of zero-investment factor-mimicking portfolios for size (*SMB*), book-to-market (*HML*), and momentum (*UMD*). We use alternatives (the 3-factor model of Fama and French, 1993 and the conditional 4-factor model of Ferson and Schadt, 1996) in robustness checks. As is standard in this literature, we do not take a position in the debate about whether these portfolios are proxies for (all) risk factors, and we define skill/ability

[4]As argued in Fama and French (2010), assuming serial independence in errors should be innocuous. In robustness checks in Section 8, we introduce cross-sectional dependence and we also show that our results are not driven by the normality assumption.

as the alpha in Equation 1, i.e., the average return in excess of a comparable passive portfolio.[5]

Skill α_i, factor loadings β_i, and error precision h_i are fund-specific parameters that are assumed to vary randomly in the population according to the model presented below.

2.2 Model of skill under specification uncertainty

We propose that fund skill α_i is an i.i.d. draw from a mixture population distribution, i.e.,

$$p\left(\alpha_i \left|\left\{\pi^q, \theta_\alpha^q\right\}\right.\right) := \sum_{q \in Q} \pi^q f^q\left(\alpha_i \left|\theta_\alpha^q\right.\right), \tag{2}$$

where $p(\cdot)$ and $p(\cdot|\cdot)$ denote generic marginal and conditional densities, respectively, Q is the set of mixture components, f^q is the density of the q-th component and θ_α^q its parameter vector, and π^q is the weight of the q-th component, with $\sum_{q\in Q}\pi^q = 1$.[6]

It is common in the literature to classify funds into three groups based on their skill: good funds whose alpha is positive, bad funds whose alpha is negative, and funds with zero alpha. Furthermore, it has been suggested that the cross-sectional distribution of mutual fund skill is non-normal, potentially exhibiting skewness, fat tails, and/or multimodality. To accommodate these features, we propose that the mixture density in Equation 2 has three components $Q = \{0, -, +\}$ corresponding to funds with zero, negative, and positive alpha, and we incorporate model specification uncertainty, that is we adopt a flexible semi-parametric specification for the alpha distribution of the nonzero components. In particular, for funds with negative (positive) alpha, we assume that α_i is drawn from a mixture of log-normals with K^- (K^+) components, where we treat the number of components as model parameters to be estimated. Notably, this can be viewed as a model selection analysis, in which we select the 'best' model across mixture models with different numbers of components. Thus, Equation 2 becomes

$$p\left(\alpha_i \left|\left\{\pi^q, \theta_\alpha^q\right\}\right.\right) := \pi^0 f^0\left(\alpha_i \left|\theta_\alpha^0\right.\right) + \pi^- f^-\left(\alpha_i \left|\theta_\alpha^-\right.\right) + \pi^+ f^+\left(\alpha_i \left|\theta_\alpha^+\right.\right),$$

[5]This definition of skill is used by most of the related literature (e.g., Baks, Metrick and Wachter, 2001; Jones and Shanken, 2005; Kosowski et al., 2006; Barras, Scaillet and Wermers, 2010; Fama and French, 2010). Berk and Green (2004) define skill as alpha before costs (including, importantly, information acquisition), Pastor, Stambaugh and Taylor (2015) define it as alpha adjusted for fund and industry size (i.e., the alpha on the first dollar invested in the fund and industry), while Koijen (2014) defines it, under market efficiency, as the price of the active-portfolio risk (i.e., the compensation for holding assets that earn a risk premium).

[6]Identifying profitable opportunities is a zero-sum game, so if mutual funds hold a significant proportion of the market their alphas should be negatively correlated. Actively managed US equity funds only held 19% of outstanding US stocks at the end of 2013 (Investment Company Fact Book), so our assumption that alphas are independent is unlikely to be restrictive. Using simulations, Jones and Shanken (2005) show that incorporating negative correlation among fund alphas has a trivial effect on posteriors.

where π^0, π^-, π^+ are the proportions of funds with zero, negative, and positive alpha, f^0 is a point mass at 0, and f^-, f^+ have negative and positive support. In detail,

$$f^0\left(\alpha \left|\theta_\alpha^0\right.\right) := \delta_0$$

$$f^-\left(\alpha \left|\theta_\alpha^-\right.\right) := \sum_{1\le k\le K^-} \frac{\pi_k^-}{\pi^-} f_{\ln\mathcal{N}}\left(|\alpha| \left|\theta_{\alpha,k}^-\right.\right), \text{ for } \alpha<0$$

$$f^+\left(\alpha \left|\theta_\alpha^+\right.\right) := \sum_{1\le k\le K^+} \frac{\pi_k^+}{\pi^+} f_{\ln\mathcal{N}}\left(|\alpha| \left|\theta_{\alpha,k}^+\right.\right), \text{ for } \alpha>0,$$

where δ_0 is the Dirac mass at 0; $f_{\ln\mathcal{N}}$ is the density of the log-normal distribution; π_k^- and π_k^+ are the proportions of the k-th negative and positive component, with $\sum_k \pi_k^- = \pi^-$ and $\sum_k \pi_k^+ = \pi^+$; $\theta_{\alpha,k}^q := \left(\mu_{\alpha,k}^q, V_{\alpha,k}^q\right)$ contains the mean and variance of the k-th component; and $\theta_\alpha^q := \left\{\pi_k^q, \theta_{\alpha,k}^q\right\}$ for $q \in \{-,+\}$ and $\theta_\alpha^0 := 0$ arbitrarily.

Intuitively, each fund is drawn from one of the three skill types with probability π^0, π^-, π^+, respectively, and conditional on belonging to the negative- or positive-alpha type, it is drawn from the type's k-th component with probability $\frac{\pi_k^-}{\pi^-}$ or $\frac{\pi_k^+}{\pi^+}$, respectively. That is, there exist unobserved 'allocation' variables e_i^0, $e_{i,k}^-$, $e_{i,k}^+ \in \{0,1\}$ that sum to 1 and indicate which type and mixture component fund i belongs to. For example, a fund belonging to the k-th component of the positive-alpha type has $e_{i,k}^+ = 1$ and its α is drawn from the log-normal distribution whose underlying normal has mean $\mu_{\alpha,k}^+$ and variance $V_{\alpha,k}^+$. Letting $\pi := \left(\pi^0, \{\pi_k^-\}, \{\pi_k^+\}\right)$ be the probability vector, the allocation vector $e_i := \left(e_i^0, \left\{e_{i,k}^-\right\}, \left\{e_{i,k}^+\right\}\right)$ follows the multinomial $M(1,\pi)$ distribution.

As we have explained in the introduction, this model for skill is appealing for the following reasons. First, it provides flexibility and can fit a wide range of distributions for fund skill. Indeed, density estimation through mixtures of finite mixtures — such as the one we employ — is similar to non-parametric density estimation, both in terms of mathematical properties and in terms of flexibility (e.g., Miller and Harrison, 2015); see also our simulations in Section 3. Second, this model enables us to estimate the proportions of fund skill types *directly* and *simultaneously* with the entire distribution, using *all* the return information for *all* funds. This approach leads to improved inference over indirect alternatives that use fund-by-fund hypothesis tests (see, e.g., our simulations in Section 3). An additional benefit of our approach is that it incorporates in a single unified framework two strands of literature that have been disparate so far: the approach of estimating the proportions of skilled/unskilled funds and that of modeling the skill distribution. Finally, using a flexible model for the population distribution leads to improved fund-level inference, as the sharing of information across funds is driven to a larger extent by the data rather than by the priors (see, e.g., our portfolio performance analysis in Section 6.1).

To complete the model, we need to specify the distributions of the fund-specific factor loadings, β_i, and error precisions, h_i. In the baseline case, we assume β_i and h_i are independent from each other and from α_i, with β_i drawn from a normal $\mathcal{N}\left(\mu_\beta, V_\beta\right)$ with mean μ_β and variance V_β, and h_i drawn from a Gamma $\mathcal{G}\left(\kappa_h, \lambda_h\right)$ with shape κ_h and scale λ_h. We note that, since $\varepsilon_{it} \sim \mathcal{N}\left(0, h_i^{-1}\right)$, we compound a normal distribution with a Gamma precision, so the errors pooled across funds have the t distribution, whose fat tails make it suitable for robust analysis. In Section 8.2, we relax the assumption that α_i, β_i, and h_i are independent and allow for a full correlation matrix between all model parameters, and in Section 8.4 we show that our results are not driven by our normality assumption about the factor loadings and the errors.[7]

To summarize our model, in Figure 1 we show its directed acyclic graph representation, where squares represent observed quantities and circles represent parameters to be estimated.

2.3 Hierarchical priors

To render the estimation of our model feasible, we utilize Bayesian methods, which produce posterior inference by combining the information contained in the data with priors. In the absence of concrete prior information about the model parameters, it is important to choose weak priors in order to let the data determine the posteriors, as well as to conduct a sensitivity analysis to check that the posteriors are robust to alternative prior specifications.

We utilize hierarchical priors, which model the lack of prior information about model parameters according to the Bayesian paradigm, i.e., through a distribution over priors. In this approach, we have two layers of priors. In the first layer, which is essentially part of the model presented above, we propose that fund-level alphas are drawn from a flexible mixture distribution with unknown population parameters which are estimated from the data using information from all funds. In the second layer, we propose a weak 'hyperprior' distribution for these population parameters. Thus,

[7] It is interesting to note that, in their model of skill for private equity funds, Korteweg and Sorensen (2016) assume errors are homogeneous across funds and drawn from a mixture of normals. This assumption is both more and less general than ours, as it can capture potential asymmetries — not just fat tails — but does not allow for fund-specific precision heterogeneity. Our analysis in Section 8.4 shows that, in our application, the assumption of symmetric errors is not restrictive. In contrast, allowing for fund-specific heterogeneity is important since i) ex ante, the wide range of possible strategies followed by mutual funds implies that their error distributions are likely to be different, and ii) ex post, indeed we estimate very different precision parameters across funds.

the hierarchical approach contains more data-driven information, hence allows a more "objective approach to inference" (Gelman, 2006) and has been shown to improve the robustness of the resulting posteriors (Robert, 2007).[8]

For the population proportions π, we use the symmetric Dirichlet prior $\mathcal{D}(1)$, which implies an uninformative uniform distribution over all possible values of π. For the population means $\left\{\mu_{\alpha,k}^{q}\right\}$, μ_{β} and variances $\left\{V_{\alpha,k}^{q}\right\}$, V_{β} of the nonzero components of the skill distribution and of the factor loadings, we use the independent Normal-inverse-Wishart prior. To make priors even weaker, we introduce an additional hierarchical layer such that the parameters of the hyperpriors for the population mean and variance of skill are themselves random parameters to be estimated by combining the data with even deeper hyperparameters. For the population shape κ_h and scale λ_h of the error precision distribution, we use the Miller (1980) conjugate prior with baseline parameters that correspond to a 'prior sample' with a handful of funds, so it contains very little information relative to the data. For the numbers K^- and K^+ of the mixture components in the alpha distribution of negative- and positive-alpha funds, we employ the commonly used truncated Poisson distribution, which penalizes mixtures with more components and is truncated for convenience of computation and presentation.

For a detailed presentation of the prior distributions and the baseline hyperparameter values, see Section A of the appendix. Also see Section 8.1 for an extensive sensitivity analysis that shows that our posteriors are robust to alternative prior specifications.

2.4 Estimation

To estimate our model, we need to derive the joint posterior distribution of the model parameters conditional on the factor returns and the fund returns. This posterior is proportional to the likelihood times the joint prior but cannot be calculated analytically. Instead, we use a Markov chain Monte Carlo (MCMC) algorithm to generate draws that form a Markov chain with stationary distribution equal to the posterior. In particular, we employ the Reversible Jump MCMC algorithm with Metropolis-Hastings-within-Gibbs sampling (see Metropolis et al., 1953; Hastings, 1970; Geman and Geman, 1984; Green, 1995; and Richardson and Green, 1997). We briefly discuss our algorithm below, and we present details in the appendix.

Since the joint posterior is not from a known distribution, we use a Gibbs sampler which parti-

[8]We note that using a hierarchical prior is different from using a marginal prior that compounds the two hierarchical layers, since the latter does not estimate the first layer — the population distribution — from the data, and therefore contains more prior-driven instead of data-driven information.

tions the parameters into blocks such that we can sequentially draw from the conditional posterior of each block given the data and the other blocks. The basic intuition of the Gibbs sampler is the following. Consider, for simplicity, that we only have two blocks of parameters, A and B. From Bayes' rule, we know that we can draw from their joint posterior by drawing from the marginal posterior of A and then from the conditional posterior of B given A. So, assuming we have an initial draw for A, we can use it to generate a valid draw for B from the joint posterior, which we can in turn use to generate a valid draw for A from the joint posterior. Continuing like this, we make draws forming a Markov chain which, under weak conditions (see Geweke, 1999), converges to the joint posterior regardless of the initial draw. To provide more concrete intuition about how this sampler works, we succinctly describe some of the steps we use. First, funds' allocations $\{e_i\}$ to skill types constitute a sample from the multinomial $M(1, \pi)$ distribution, and are used to learn about the population proportions π. Second, funds' alphas $\{\alpha_i\}$ and their allocations $\{e_i\}$ provide information about the population distribution for each type. For example, with one negative component, the alphas of negative-alpha funds constitute a sample drawn from $\ln \mathcal{N}\left(\mu_\alpha^-, V_\alpha^-\right)$, and are used to learn about μ_α^- and V_α^-.

A complication arises from the fact that α_is are drawn from three sub-populations with no common support. To be able to explore the whole support of $\{\alpha_i\}$, we need to simultaneously learn about $\{e_i\}$ and $\{\alpha_i\}$ by combining information from the data with information from the proportions and the entire alpha distribution. The idea is that, for each fund, we combine information in its likelihood with the population distribution to learn about the probability that the fund belongs to each type and about the distribution of its alpha conditional on belonging to each type. Intuitively, the posterior probability that a fund has zero alpha is higher when the proportion of zero-alpha funds in the population is high, the fund's likelihood at zero is high, and the overlap between its likelihood and the population density over nonzero alphas is low. Furthermore, the density of a fund's alpha conditional on it being nonzero is determined by the overlap between its likelihood and the population density over nonzero alphas. One of our technical innovations is to design a Metropolis-Hastings step that enables us to draw from the joint conditional posterior of $\{e_i\}, \{\alpha_i\}$. In this step, we make candidate draws using a generating density similar to that in Gottardo and Raftery (2008) for drawing from mixtures of mutually singular distributions. These candidates are accepted or rejected based on an acceptance probability that generally moves the chain toward areas of the parameter space where posterior probability is high.

A second complication arises from the fact that, to estimate the unknown numbers of components and their distribution parameters, we need to construct a Markov chain that jumps between models with different dimensions. For this purpose, we use the Reversible Jump MCMC algorithm, which introduces two dimension-changing steps: one that splits or merges existing components, and one that creates a new or deletes an existing empty component. We extend the standard algorithm to account for the fact that we have two mixture distributions with an unknown number of parameters, as well as a point mass.

We make 5 million MCMC draws, discard the first 10% as burn-in, and retain every 50$^{\text{th}}$ to mitigate serial correlation. Convergence of the draws to the posterior is monitored by inspecting their trace plots and density plots and using standard diagnostic tests.[9] Using these draws, in our results we present posterior information, e.g., the mean, median, standard deviation, and density plots of the posterior distributions for all parameters of interest.

3 Simulation analysis

In this section, we conduct a simulation analysis: We combine various data generating processes for skill with the factor model in Equation 1 to generate samples of fund returns, and we use our methodology and alternatives to estimate the proportion of skilled funds and the skill distribution. In Section 3.1, we compare our approach to estimating the proportions of skilled/unskilled funds with the fund-by-fund hypothesis-testing approach, and in Section 3.2 we compare our approach to estimating the skill distribution with a model that assumes skill is normal.

The parameters we employ in our simulations resemble the corresponding parameters in the real data we use subsequently (see Section 4). In each sample, the number of funds (3,500) equals the number of funds in the data, and the number of observations per fund is drawn from its empirical distribution in the data. The factor returns (F_t) are drawn from a normal with mean and covariance equal to their sample counterparts in the data. To derive the distribution of the other simulation parameters, we conduct fund-level OLS estimations of Equation 1 using the real data: The factor loadings (β_i) are drawn from a normal with mean and covariance equal to their sample counterparts in the fund-level OLS estimates, and the errors (ε_{it}) are drawn from normals with mean zero and fund-specific precisions drawn from a Gamma with parameters fitted by maximum likelihood to the fund-level OLS error precision estimates. To mitigate the effect of simulation noise in the creation of the data, each methodology is applied to 50 simulated data sets.

[9] The algorithm was coded in Matlab and C++. Making 5 million draws takes 30 hours on the c4.8xlarge instance on Amazon's EC2 service. We note that, both in the simulations and in the estimation with the real data that we discuss below, convergence is relatively fast and 500,000 draws are more than sufficient to obtain accurate estimates, so run times can be quite short at the expense of a slight increase in numerical standard errors.

3.1 Comparison with fund-by-fund hypothesis testing

The extant approach to estimating the proportions of fund skill types is to i) perform multiple fund-level OLS regressions of fund returns on factor returns, and ii) count the proportion of funds for which the null hypothesis $H_{0,i} : \alpha_i = 0$ versus the alternative $H_{A,i} : \alpha_i \neq 0$ is rejected. To adjust for the fact that the probability of incorrectly rejecting the null hypothesis is inevitably increased when performing multiple tests, Barras, Scaillet and Wermers (2010) apply the FDR approach. The main idea of the FDR approach is the following: Since the distribution of p-values arising from null hypotheses (funds with zero alpha) is uniform on [0, 1], then *assuming* that all p-values above some threshold arise from funds with zero alpha, one can extrapolate the proportion of incorrect null rejections (i.e., those corresponding to 'lucky' or 'unlucky' funds with zero alpha) and adjust the calculated proportion of null rejections to estimate the true proportion of funds with nonzero alpha.[10] However, the assumption that all p-values above a threshold correspond to zero-alpha funds is equivalent to the assumption that p-values under the alternative hypothesis of nonzero skill are concentrated near zero, hence that the alphas of nonzero funds are far from 0. This is unlikely to be true, since the distribution of skill, hence the density of p-values, is likely to have more complex features. In this case, the FDR methodology overestimates the proportion of true null hypotheses because the power of the test stays below 1 even for high values of the aforementioned threshold. Our approach deals with this issue by introducing the proportions of the skill groups as population parameters that contribute to the likelihood. As a result, we do not rely on low-power fund-by-fund hypothesis tests, nor do we deterministically allocate all funds with p-value above some threshold to the zero-alpha type to extrapolate the proportion of zero-alpha funds. Instead, we estimate the proportions directly and simultaneously with the entire distribution, using more of the information in the data (all the returns for all funds). In particular, as we have explained previously, we use an iterative procedure. In one step, we use information about the population (i.e., the proportions and the distribution of alpha) together with the information in each fund's returns to determine the posterior probability that the fund belongs to each skill type as well as the posterior distribution of its alpha. In another step, we use this information for all funds to determine the posterior about the population (i.e., the proportions and the parameters of the alpha distribution).[11]

[10] For a detailed description of the FDR approach, applied in the context of mutual fund performance, see Barras, Scaillet and Wermers (2010).

[11] Certainly, our methodology is not immune to the problem caused by data that have (almost) the same distribution under the null and under the alternative, in which case there may not be enough information to differentiate between the two, but it is less prone to mis-estimation for the reasons we have just explained.

In Table 1, we compare our estimated proportions of fund skill types with those estimated by fund-by-fund hypothesis tests, with and without the FDR correction. For this comparison, we employ data generating processes (DGPs) in which alphas are drawn from a mixed distribution with a point mass at zero and with nonzero alphas drawn from a variety of distributions. Our starting point is the DGP that Barras, Scaillet and Wermers (2010) utilize in their simulation analysis to demonstrate the strength of the FDR methodology. In this DGP (henceforth DGP D-1), alphas are generated from a discrete distribution that places probability 75%, 23%, and 2%, respectively, on values 0%, -3.2%, and 3.8% per year; that is, the proportion of zero-alpha funds is large, and nonzero alphas are discrete and far from zero. For the other DGPs, we want to vary the proportions of funds with zero and nonzero alpha, the magnitude of nonzero alphas, as well as their distribution, to make it harder to differentiate between zero- and nonzero-alpha funds. Succinctly, we draw nonzero alphas from i) discrete distributions with large values (DGP D-1) and small values (DGP D-2), ii) normals with high variance (DGP N-1) and small variance (DGP N-2), and iii) log-normals far from zero (DGP L-1), close to zero (DGP L-2), and similar to the ones we estimate from the real data using our model (DGP L-3); the proportions of zero-, and nonzero-alpha funds also vary across DGPs. In Table 1, we provide detailed descriptions of these DGPs, and we report the true and the average (across 50 simulations) estimated proportions of funds with zero, negative, and positive alpha from the various estimation methods. On the one hand, we see that our model estimates the true proportions accurately, even when, e.g., nonzero alphas are discrete or normal. On the other hand, the alternative approach performs well when the values of nonzero-alpha funds are far from zero (i.e., for DGPs D-1, N-1, and L-1), but it becomes overly conservative as these values get closer to zero (i.e., for DGPs D-2, N-2, and L-2). Indeed, the FDR correction does not significantly improve performance in the latter cases because, as we explained earlier, its assumption that nonzero alphas are far from zero fails. It is also interesting to note that, for the case in which alpha is drawn from a distribution close to the one our methodology estimates from the *real* data (i.e., for DGP L-3), the FDR approach estimates the proportions of zero-, negative-, and positive-alpha funds to be 73%, 26%, and 1%, respectively; these are not only far from the true proportions in the simulated data (9%, 78%, and 13%), but also similar to the proportions that Barras, Scaillet and Wermers (2010) estimate using the FDR methodology on the real data (75%, 24%, and 1%). In Section C.2 of the online appendix, we also report the percentiles of the simulated distributions and the distributions estimated using our methodology. We see that, as with the proportions of skilled/unskilled funds, the percentiles of the simulated and estimated distributions are also very similar.

3.2 Comparison with the hierarchical normal model

Jones and Shanken (2005) are the first to estimate the entire distribution of fund skill; they do so under the assumption that alpha is drawn from a normal. However, the skill distribution may exhibit complex features such as skewness and fat tails, which a normal distribution would fail to adequately fit. Our model deals with this issue by replacing the normality assumption with a semi-parametric representation which yields considerable flexibility and can fit a wide range of distributions.

In Figure 2, we compare our estimated skill distribution with that estimated by the hierarchical normal model. For this comparison, we employ plausible DGPs that also provide good tests for the flexibility of the two alternative models. More specifically, we draw alpha from i) a normal distribution (DGP C-1) and ii) a continuous distribution with negative skewness and fat tails (DGP C-2), where in both DGPs the quantiles of alpha are similar to those estimated from the real data. The normal distribution is a particularly interesting case, not only because it is a good test of our model's flexibility as it has no point mass at zero, no skewness, and no fat tails, but also because most previous studies have assumed that the alpha distribution is normal. In Figure 2, we provide detailed descriptions of these DGPs, and we present Quantile-Quantile (Q-Q) plots of the alpha distributions estimated using our model (in Panels *a* and *b*) and the hierarchical normal model (in Panels *c* and *d*) against the simulated alpha distributions. On the one hand, in Panels (a) and (b), we see that the Q-Q plots with our estimated posteriors lie very close to the 45° line, demonstrating that our model fits the simulated distributions well. In particular, we see that our model is flexible enough to perform well even when alphas are simulated from a single normal distribution (see Panel *a*). Furthermore, there is no point mass in either simulated alpha distribution; the fact that our model estimates the point mass to be very small in both cases (4% for DGP C-1 and 6% for DGP C-2) and, even more importantly, fits the overall distributions well suggests that allowing for a point mass in our model is not particularly restrictive. On the other hand, the hierarchical normal model performs well when the true alpha distribution is normal (see Panel *c*) but is not robust to mis-specification, as it largely mis-estimates the non-normal distribution (see Panel *d*); clearly, using in Panel (d) a DGP that is more skewed and/or fat-tailed than DGP C-2 would yield a Q-Q plot that deviates even more strongly from the 45° line. In Section C.2 of the online appendix, we also present the percentiles of the simulated and the estimated alpha distributions using our model and the hierarchical normal model, which verify the results we discuss here.[12]

[12] The priors we use for the hierarchical normal model are weak and similar to the ones we use for our model (see Section A in the appendix). This, together with the prior sensitivity analysis we conduct, indicates that the results from comparing our model to the hierarchical normal are not driven by the priors.

4 Data description

In our subsequent analyses, we use monthly return data and accompanying fund information from the Center for Research in Security Prices (CRSP) Survivorship-Bias-Free US Mutual Fund Database, for the period January 1975 to December 2011. We focus on actively managed open-end US equity funds, so we exclude index, fixed income, international/global, accrual, money market, and sector funds. To exclude these, we use the reported fund classifications and information obtained from keyword searches in funds' names.

One issue with the CRSP database is that it treats share classes of the same fund as separate funds. Since these share classes hold the same portfolio, their performance is identical, so we need to identify and merge them rather than treat them as independent observations. To identify funds' share classes, we use the MFLINKS database from Wharton Research Data Services. Since MFLINKS only covers part of the CRSP database, in both the time and the cross-sectional dimension, we complement it with our own algorithm designed to detect patterns in fund names that indicate different share classes belonging to the same fund. For the subset of our data covered by MFLINKS, our algorithm generally agrees with it, but we also identify 850 cases of disagreement. We manually check these and find that they mostly correspond to share classes that MFLINKS omits even though it contains other share classes of the same fund, and to share classes of the same fund that are treated by MFLINKS as separate funds. After we identify funds' share classes, we compute the monthly return for each fund as the weighted average of the returns of its classes, with weights equal to the beginning-of-month total net asset value of each class.

We use fund returns as reported in the CRSP database, which are net of fees, expenses, and transaction costs. To improve the accuracy of the returns information in the data, we omit any monthly return that directly follows a missing return, as it may compound multiple months' returns. We also keep funds with at least 60 monthly return observations, not necessarily contiguous but with no gaps greater than a year.[13] Our final data set consists of 3,497 funds. In the online appendix, we present summary statistics for the characteristics of the funds in our sample.

To construct the benchmarks against which we measure fund performance, we use the CRSP NYSE/Amex/NASDAQ value-weighted index as the market factor and the one-month Treasury bill rate as the risk-free rate, while monthly returns for the *SMB*, *HML*, and *UMD* factor-mimicking portfolios are downloaded from Kenneth French's website.

[13] All studies in the literature use such a filter, though the number of monthly observations required ranges from 8 (e.g., Fama and French, 2010) to 36 (e.g., Koijen, 2014) to 60 (e.g., Kosowski et al., 2006). Using a filter of 36 months has little impact on our main results, so any survivorship bias due to the 60-month requirement should be small.

5 The prevalence of skill

In this section, we present our estimation results for the baseline specification presented in Section 2. In Section 5.1 we discuss our estimated proportions of funds with zero, negative, and positive skill, and in Section 5.2 we turn to our estimates of the entire distribution of alpha, to obtain a complete picture of the prevalence of skill in the mutual fund industry.

5.1 The proportions of skilled and unskilled funds

In Table 2, we present our estimated proportions of zero-, negative-, and positive-alpha funds. We see that most funds (78%) have negative skill, but a substantial proportion (13%) generate positive excess returns. These are very different from the estimates obtained from fund-level hypothesis tests: for example, Barras, Scaillet and Wermers (2010) classify 75.4% of funds as zero-alpha, 24% as negative-alpha, and only 0.6% as positive-alpha. Thus, while the extant estimates would imply that investing in a randomly chosen mutual fund is relatively safe but not worth it, we find that doing so has a large downside risk but also considerable upside.[14] Indeed, it has long been considered a puzzle that people invest in active mutual funds despite the evidence that they underperform (e.g., Gruber, 1996), leading some to question the rationality of this behavior. Our finding that there is a substantial proportion of positive-alpha funds suggests that investing in mutual funds could actually be rational for (at least some) investors.[15]

Our findings also relate to the equilibrium theory of Berk and Green (2004), in which fund alphas equal zero in the long run because funds face decreasing returns to scale and competitive investors direct their capital to alpha-generating funds. The earlier finding that a large majority of funds have zero alpha indicates that the industry is at this long-run equilibrium, but our finding that most funds have nonzero alpha suggests the opposite. This could be the case, for example,

[14] In a concurrent study, Ferson and Chen (2015) use a simulation approach to improve on the FDR methodology. They assume alphas can take three values — zero, a specific negative, or a specific positive value — and utilize tests of all three point null hypotheses to estimate the negative and positive alpha value and their proportions. This approach shares a similarity with ours as it explicitly models nonzero alphas, but its assumption that negative and positive alphas take a single value each is very different from our flexible model for nonzero alphas. Furthermore, our approach uses more information and more directly as it does not rely on fund-by-fund tests and simulations. They still find four times as many zero-alpha funds as we do (36% vs. 9%) and almost zero positive-alpha funds versus the 13% we find, so their conclusions are also different. Though their approach improves on the FDR approach, it may still be conservative. This is supported by their simulations: they write that "when the true fraction of zero-alpha funds is 10%, the [FDR] estimator . . . produces an average estimate [of zero-alpha funds] of 75% [. . . while their] estimate is 27%; still biased but much more accurate."

[15] Baks, Metrick and Wachter (2001) show that investing in mutual funds is rational for individuals holding certain specific prior beliefs about the distribution of skill. Since we estimate the skill distribution from the data, we do not make an assumption about individuals' prior beliefs. Therefore, our results could be useful in an analysis of the rationality of mutual fund investing that would apply more broadly to investors who do not hold specific prior beliefs.

because frictions such as taxes and agency costs prevent investors from quickly moving their money out of (into) under-performing (over-performing) funds or because behavioral biases such as the disposition effect make investors reluctant to sell their shares in under-performing funds. We note that this conclusion is not driven by a strict interpretation of the theory that alpha equals *exactly* zero, since it continues to hold if we replace the point mass at zero with a narrow normal centered at zero (see Section 8.6), nor is it driven by newer funds since it continues to hold if we exclude the returns observations for the first two years of each fund's life (see Section C.18 in the online appendix).

Importantly, the differences between our estimates and those in the literature are not due to a disagreement between methodologies on the level of alpha that is *economically* different from zero. Looking at our estimate of the entire skill distribution in the next section, we see that the differences between the two sets of estimates remain substantial for reasonable subjective assessments of the level of alpha that is economically significant. For example, if we classify as zero-alpha funds those with $|\alpha| < 0.5\%$ per year, we still find that 63% have negative and 8% have positive alpha. As mentioned above, the difference arises from the fact that our approach uses more information and estimates the proportions directly and simultaneously with the alpha distribution, while the FDR approach becomes quite conservative when there is limited information at the fund level, hence the power of the fund-level tests is low. A back-of-the-envelope calculation helps to illustrate that the FDR methodology fails to correctly classify funds, even when alphas are economically very large. In the CRSP data, the median fund has about 150 monthly observations, its fund return volatility is about 5% per month, and about 90% of this volatility is explained by the 4-factor model. Therefore, for such a fund, the standard error of the OLS coefficient estimate for α is approximately 0.13% per month ($\sqrt{\frac{0.05^2 \cdot 0.1}{150}}$). Assuming normality and using $\lambda = 0.5$ which as noted by Barras, Scaillet and Wermers (2010) yields close to "optimal" results for the FDR methodology, a fund with an economically significant alpha of, e.g., $\pm 2\%$ ($\pm 1\%$) per year has p-value greater than λ with probability 25% (42%). This means that $\frac{25\%}{1-\lambda} = 50\%$ ($\frac{42\%}{1-\lambda} = 84\%$) of funds with $\alpha = \pm 2\%$ ($\alpha = \pm 1\%$) are misclassified as zero-alpha funds by the FDR approach.

5.2 The distribution of fund skill

The proportions of skilled/unskilled funds in the population can be useful, e.g., for investors who think in terms of categories or process coarse information (and for models of such behavior), but they only provide partial information about the prevalence of skill in the mutual fund industry. For a more complete picture, we turn to our estimates of the skill distribution. In Table 3, we present the posterior for the unknown number of components in the skill distribution for negative- and positive-alpha funds. We see that the model with the highest posterior probability has two negative components and one positive, so this could be considered the 'selected' model.[16]

In Figure 3 and Table 4, we present information about our estimated population distribution of alpha. Formally, this is the posterior predictive distribution, i.e., the distribution of an unobserved fund's alpha after observing all our data. In Figure 3b, we plot the estimated density; since the point mass at zero is a Dirac mass which can be thought of as an infinitely high and infinitely thin spike, we adopt the convention to represent it with a vertical arrow of arbitrary length and to indicate its mass in the legend. In Figure 3c, we plot a histogram of values simulated from this distribution; for our mixed distribution, this helps to visualize how probability mass is divided over the real line, particularly around zero. In Table 4, we present point and interval estimates for various percentiles of this distribution.[17]

We find that 99% of the funds in the population have alpha between −6.60% and 3.24% per year, and half have alpha smaller than −0.72%. Concentrating on the right tail, we find that 10% of funds yield a return of more than 0.30% in excess of the benchmark per year, 5% of funds yield more than 0.87% in excess return, and 1% of funds yield an excess return above 2.35%. We also observe that the distribution is asymmetric and has a mode between −0.5% and −1% per year. This mode is due to the fact that a substantial proportion of funds have *before*-fees/expenses alphas close to zero (see Table 15) and the distribution of fees and expenses has a mode at 0.95% (see Section C.4 in the online appendix). Importantly, in Figure 3a we plot the *prior* predictive densities that correspond to the baseline and to an alternative prior specification, both of which yield almost the same posterior predictive density, plotted in Figure 3b. We see a wide deviation between the two prior distributions as well as between each of the priors and the posterior, indicating that the posterior is robust and driven by the data rather than by the priors.[18]

[16]In most of our subsequent analyses, we present results on the skill distribution that are *not* conditional on the most likely model but rather incorporate the model uncertainty we allow for in our baseline specification, since these estimates provide better inference and have higher efficiency properties (Richardson and Green, 1997; Marin, Mengersen and Robert, 2005). Results conditional on the selected model are not very different (see Section C.19 in the online appendix), so for computational convenience, we employ it in some of our subsequent analyses (in Sections 6.1, 7.1, 7.2, and 8.2).

[17]For clarity, we note that we calculate all statistics (here and throughout the paper) for the estimated alpha *distribution*, not for the posterior means of the fund alphas in the *sample*.

[18]Observing Figure 3b, it is interesting to note that alphas are more often equal to exactly zero than just slightly below or above it. This is driven by the fact that our model allows for a point mass at zero, in conjunction with the noise in the data and the law of parsimony embodied in Bayesian inference. That is, the noisy returns of funds with alpha very close to zero are explained equally well by the point mass at zero and by the mixture model that represents nonzero alphas; therefore by the law of parsimony, these funds are allocated with higher probability to the point mass, since it is a simpler 'model.' We also note that, even if one believes a priori that a point mass at zero alpha is implausible, this allocation has little impact in economic terms on our estimated distribution, since the alternative allocation would not affect, e.g., the histogram in Figure 3c, the percentiles in Table 4, and the measures of non-normality in Tables 5 and 6.

Non-normality of the skill distribution Next, we compare our estimated distribution of alpha with the normal, and particularly with the distribution estimated using a hierarchical model that assumes alphas are normal, as in Jones and Shanken (2005). First, in Table 4, we present percentiles of the two distributions. We see that the normal model estimates fewer funds with large (in absolute value) alphas. For example, the normal model places about 0.5% of fund alphas outside the range $(-4\%, +2\%)$, while we estimate about 3% of alphas (i.e., about 6 times as many) outside this range. This discrepancy becomes even larger if we consider larger alphas: The normal places 0.025% of alphas outside the range $(-6\%, +3\%)$, while we estimate more than 1% of alphas (i.e., more than 40 times as many) outside this range. These differences can be seen graphically in Figure 4 which presents a Quantile-Quantile (Q-Q) plot that compares the estimated normal against the distribution estimated from our model, and in Figure 5 which shows zoomed-in density plots to show tail detail. The S-shape of the Q-Q plot indicates that our distribution has fatter tails than the normal and a more prominent peak, while the fact that the left tail deviates farther from the 45° line than the right tail indicates negative skewness.

To see more formally that the distribution we estimate is different from the normal, we calculate measures of skewness and tail weight, we calculate measures of the distance between the two distributions, and we conduct a normality test. To study the distribution's skewness and tail weight, we do not rely on the traditional moment-based measures proposed by Pearson, because their values are greatly influenced by the extreme tails which cannot be estimated precisely since, by definition, there are very few data points there. Instead, we use quantile-based measures, which are robust as they ignore the extreme tails. Groeneveld and Meeden (1984) propose the skewness measure

$$\frac{F^{-1}(1-p)+F^{-1}(p)-2F^{-1}(0.5)}{F^{-1}(1-p)-F^{-1}(p)},$$

where $F^{-1}(x)$ is the x^{th} quantile of the distribution and $0 < p < \frac{1}{2}$. Essentially, this calculates the difference between the deviation of the right and the left tail quantile from the median, and scales it by the inter-quantile range, therefore it takes values in $[-1, 1]$, with negative (positive) values indicating left (right) skewness. Brys, Hubert and Struyf (2006) propose left and right tail weight measures which do the same for the left and right half of the distribution, respectively. As a result, these measures have similar characteristics, but their values are more readily interpreted by calculating differences from the values that correspond to the normal distribution. In Table 5, we see that the estimated measures are all significantly different from those of the normal, at least at the 5% level, indicating that our estimated alpha distribution is significantly non-normal. Specifically, the measure of skewness (-0.20) and the left and right tail weight measures (0.34 and 0.27 in excess of the normal values) indicate that, ignoring the extreme tails, the asymmetry in the estimated alpha distribution is similar to that in the χ^2 (35) distribution and its tails are as fat as those of the t (2) distribution.[19] Another way to get a sense of how strongly our estimated distribution deviates from normality is to compute measures of the distance between our estimated distribution and the estimated normal (see Table 6). We find that the Wasserstein distance between the two distributions is 0.22 and significantly different from 0 at the 1% level; this is comparable to the Wasserstein distance between the standard normal and i) a normal with the same mean but 25% smaller or greater standard deviation, ii) the χ^2 (4) distribution, or iii) the t (3) distribution.[20,21] Finally, the normality null is also rejected with the robust Jarque-Bera test at the 1% level (and even more strongly with non-robust tests). Given that our model is flexible enough to estimate a normal distribution with considerable accuracy (see Section 3), our finding that the estimated distribution deviates substantially from a normal suggests that our more general and flexible specification is necessary to model the true distribution of alpha. As we have noted earlier, our results are not driven by the normality assumption about the factor model errors; see Section 8.4 for more details.

The alpha distribution we estimate is also remarkably different from the distribution of the OLS estimates of α_i in our sample. In Table 4, we present percentiles of the OLS estimates, and in Figure 4 we present a Q-Q plot comparing their distribution against our estimated distribution. We see that the former has much higher dispersion and takes more extreme values than the latter. For example, more than 10% of OLS alpha estimates are outside the range (-6%, $+3\%$) while we estimate that a little more than 1% of fund alphas are outside this range. Thus, ignoring the sampling error inherent in OLS estimates and using their distribution as an estimate of the distribution of alpha in the population can be quite misleading.

Finally, in Figure 6 and Table 7, we see that there is significant heterogeneity across funds in all factor loadings, suggesting that funds follow quite diverse strategies. For example, the mean loading on the market is 0.95 and its standard deviation 0.21. Furthermore, 75% (25%) of funds

[19] We want to use as much of the distribution's range as can be estimated with reasonable accuracy, so we use $p = 0.005$, i.e., 99% of the range. Using a smaller range yields the same results, except that the skewness measure gradually becomes insignificant, indicating that the asymmetry is primarily in the tails (but not just the extreme tails, as all robust measures exclude them).

[20] The Wasserstein distance between densities f_X, f_Y is $\inf_{f_{XY}} \mathrm{E}\left[\|X-Y\|\right]$, where f_{XY} is any joint with marginals f_X, f_Y. Intuitively, it is the minimum work needed to turn f_X into f_Y, where work is probability mass moved times the distance it is moved. This measure is very intuitive, but we also calculate alternatives, with similar results. E.g., for the Hellinger distance — defined as $1 - \int \sqrt{f_X f_Y} dx$, essentially a weighted average of the densities' odds ratio — the distance between our and the estimated normal distribution is 0.08, which is similar to the distance between the standard normal and i) a normal with the same mean but half or twice the standard deviation, ii) a χ^2 (3), or iii) a t (1) distribution.

[21] For additional context, in Section C.6 of the online appendix we present robust measures of skewness and tail weight for various known distributions, as well as distance measures between the standard normal and these distributions.

have a positive (negative) loading on the *SMB* factor, while about half load negatively and half positively on the *HML* and *UMD* factors. Given such disparity in funds' strategies, a natural question is whether a fund's strategy is related to its alpha. We address this in Section 7.1 where we repeat our analysis for subgroups of funds that follow different investment strategies, and in Section 8.2 where we allow for dependence between alpha and the factor loadings.

In the online appendix, we present additional results, e.g., posteriors for the population mean and standard deviation of alpha and the factor loadings, and trace plots of the MCMC draws.

6 Portfolio analysis

In this section, we use our methodology to study the out-of-sample performance of fund portfolios constructed using simple rules (in Section 6.1) and an optimal asset allocation problem (in Section 6.2). The former analysis can be thought of as an out-of-sample test of whether our methodology can successfully identify funds with high alphas. The latter analysis provides a sense of the economic significance of incorporating information sharing and specification uncertainty in portfolio selection.

6.1 Portfolio performance

The mixture model we propose should improve the estimation of individual funds' alphas, for the following reasons. First, it utilizes more information than a fund-by-fund analysis since it uses all the information in each fund's returns instead of a summary score (the p-value) and combines it with information on all funds to derive a posterior. Second, it uses a flexible semi-parametric representation for alpha hence shares information across funds more objectively than a model that imposes restrictive distributional assumptions. Here, we examine out-of-sample whether our methodology can successfully identify funds with high alpha by forming portfolios of funds based on their alpha and calculating their performance over time. Furthermore, we compare the performance of fund portfolios formed using our methodology to that of portfolios formed using i) a hierarchical model in which alphas are drawn from a normal, and ii) the FDR methodology of selecting funds with high probability of having a positive alpha, which Barras, Scaillet and Wermers (2010) show to be superior to a simple ranking of portfolios by their OLS-estimated alpha or t-statistic. Before we proceed, it is important to keep in mind the caveat that (as pointed out by Fama and French, 2010) this type of analysis is subject to noise as it allocates funds to portfolios using only a few years of data. Therefore, the results in this section should be viewed as a complement to our theoretical arguments in Section 2 and our simulation analysis in Section 3.

To construct and evaluate portfolios, we proceed as follows. At the beginning of each month in the period 1980–2011, we use the preceding 5 years of returns to estimate the 4-factor model, and we form and hold until the end of the month an equal-weighted portfolio of funds that are estimated to have a positive alpha with high probability. For the hierarchical models, we simultaneously use the data for all funds to derive the posterior probability that each fund's alpha is *not* positive; we then sort funds by this probability and, starting from the top, we select the group of funds such that the average probability (across this group of funds) is closest to 0.1 and not greater than 0.2. Similarly, for the FDR methodology, we use fund-level OLS regressions and select the group of funds such that the estimated false discovery rate (i.e., the probability of incorrectly allocating a fund to the positive-alpha type) is closest to 0.1 and not greater than 0.2.[22] For months during which these criteria yield an empty group of funds, we either leave the portfolio empty — we call this the 'conservative' portfolio — or we select the funds whose posterior mean alpha (for the hierarchical methodologies) or OLS t-statistic (for the FDR methodology) is in the top 1% among all funds in the data set for the preceding 5 years — we call this the 'aggressive' portfolio. In what follows, we mostly focus on the performance of the aggressive portfolios, and in the online appendix (Section C.7) we show that our results are similar for the conservative portfolios as well as for alternative portfolio construction rules: portfolios formed using a 3-year (instead of a 5-year) rolling estimation window, portfolios that keep the top 2% (instead of the top 1%) of funds in the aggressive portfolio, and portfolios that keep in all months the top 1% of funds sorted by their posterior mean alpha.[23]

Our portfolio is more selective than the alternatives: Across the years it contains, on average, 11 funds, and the FDR ('normal') portfolio contains 19 (27) funds. But it is not merely a subset of the alternatives, since its overlap with the FDR ('normal') portfolio — i.e., the number of common funds divided by the number of all funds in the two portfolios — is, on average, 33% (27%). These differences in portfolio composition reflect the different ways in which each methodology uses the information in the data to derive the fund-level estimates. For example, our approach is more selective than the FDR because it shrinks each fund's posterior toward the estimated population distribution, whose mass is mostly concentrated on negative alphas, and it is more selective than the

[22] We have chosen this as our baseline portfolio formation rule for comparability with Barras, Scaillet and Wermers (2010), who show that the FDR portfolio constructed using this rule performs very well. As they note, given a loss function that penalizes including funds with non-positive α and excluding funds with positive α, this portfolio formation rule roughly corresponds to a Bayes action that minimizes expected loss for some level of the penalties.

[23] In the online appendix, we also present results on performance for portfolios formed using a model in which alphas are drawn from two normals; they are similar to those for the portfolios from the model with one normal.

normal model which estimates thin tails but a higher dispersion over intermediate values and more mass over positive alphas. We also note that there is considerable turnover in the composition of all portfolios over time: For our portfolio, 38% of funds remain in the portfolio after a year, 16% remain after 3 years, and 9% remain after 5 years; these percentages are also very similar for the alternatives.

In Figure 7, we examine the evolution over time of the conservative portfolios (Panel *a*) and the aggressive portfolios (Panel *b*) formed using all methodologies: at each month in the period 1990–2011, we plot OLS 4-factor alphas using the monthly portfolio returns from January 1980 through that month. We see that the portfolios constructed using our methodology consistently deliver positive alpha which, focusing on the aggressive portfolios, varies across the years between 2.50% (in 1998) and 4.54% (in 1991), with a mean value of 3.22%. Furthermore, we see that our portfolios consistently produce higher alpha relative to both the FDR portfolios and the portfolios formed using the hierarchical normal model. Specifically, the alpha of the FDR portfolio ranges over the years from 1.67% (in 1999) to 3.34% (in 2001), with a mean value of 2.31%, while the alpha of the portfolio formed using the hierarchical normal model ranges from 1.06% (in 1998) to 2.76% (in 1991), with a mean value of 1.72%.

In Table 8, we present various performance measures — annualized OLS 4-factor alpha ($\hat{\alpha}$) and $\hat{\alpha}$ t-statistic, information ratio, and Sharpe ratio — for the whole sample as well as for the two halves of our sample period separately (1980–1995 and 1995–2011). We see that the portfolio constructed using our methodology delivers an economically and significantly positive alpha in both sub-periods as well as over the entire sample. In comparison, both alternative methodologies fail to deliver significantly positive alphas in one of the two sub-periods. Our portfolio also delivers superior performance based on all performance measures, again in both sub-periods and over the entire sample. To be specific, over the entire sample from 1980 to 2011, our portfolio has an annualized alpha of 2.80% with information ratio 0.69, while for the FDR portfolio we calculate $\hat{\alpha} = 2.01\%$ with information ratio 0.47, and for the 'normal' portfolio we calculate $\hat{\alpha} = 1.71\%$ with information ratio 0.38. Hence, compounded over the 32 years in the period 1980–2011, our portfolio generates an additional 53% (80%) in return in excess of the benchmarks relative to the FDR portfolio (the normal portfolio).[24]

[24] We note that a classical statistical test of the null that our portfolio's alpha equals that of the alternatives is not informative: assuming a true difference equal to our point estimate of the difference in portfolios' alphas, we calculate that the power of a test with size of 5% would be just 15%, far below any acceptable level to be useful. In Bayesian terms, the posterior probability that our portfolio's alpha is higher than that of the FDR portfolio is 84%, and rises to 92% for the comparison with the normal portfolio.

Finally, we compare the performance of quantile-based portfolios for the methodologies that estimate the distribution of skill, i.e., the hierarchical models, in order to conduct an out-of-sample analysis of how well each methodology estimates the skill distribution. At the beginning of each month in 1980–2011, we use the preceding 5 years of returns to estimate the 4-factor model using each methodology, then we sort funds into ten quantiles based on the posterior mean alpha, and we hold these quantile-based portfolios until the end of the month. In the online appendix, we present various measures of performance for each portfolio. We find that the slope of performance going from the bottom to the top quantile is steeper for the portfolios constructed using our methodology than for those constructed using the alternatives. In particular, the portfolio that buys the funds in the top and sells the funds in the bottom quantile has OLS $\hat{\alpha} = 3.23\%$ per year for the hierarchical normal model and $\hat{\alpha} = 4.29\%$ for our methodology, with the difference in $\hat{\alpha}$ being statistically significant at the 1% level. These results suggest that our methodology can successfully identify funds at the tails (both the right and the left tail) of the skill distribution.

6.2 Optimal asset allocation

An application of our improved fund-level inference is in asset allocation. Here, we study the optimal decision of an investor who allocates wealth between the risk-free asset, the benchmark portfolios, and a single mutual fund. This is not intended as a complete analysis of asset allocation to mutual funds or as investment advice, as it ignores load fees, taxes, leverage, and the potential to invest in multiple funds. Rather, we seek to compare how optimal allocation in this simplified setting varies with the methodology used to estimate fund skill, in order to get a sense of how *economically* different the methodologies can be. In particular, we compare the optimal portfolios derived using our methodology to those derived using i) the methodology which estimates parameters fund-by-fund hence does not incorporate information sharing across funds, and ii) a hierarchical model which assumes alphas are normally distributed hence ignores specification uncertainty.

In the asset allocation problem we consider, an investor solves

$$\begin{aligned} &\max_{w_i, w_F} \mathrm{E}\left[u\left(W_{T+1}\right)\right] \\ &\text{s.t. } W_{T+1} = W_T\left(1 + r_{f,T+1} + w_i r_{i,T+1} + F_{T+1}' w_F\right), \end{aligned} \tag{3}$$

where W is wealth; T is the last month in the sample; w_i and w_F are portfolio weights on fund i and the benchmark portfolios; r_f is the return of the risk-free asset, r_i the net return of fund i in excess of r_f, and F the returns of the benchmark portfolios; and u is the constant relative risk aversion

(CRRA) utility.[25] To solve this problem, we use numerical optimization to find the weights that maximize the mean of $u(W_{T+1})$ across 10 million draws of W_{T+1}. For the fund-by-fund approach, assuming that the returns of the benchmark portfolios and fund i are jointly normal, the mean is taken with respect to the distribution $\mathcal{N}\left(\hat{\mu}_{i,F},\, \hat{\Sigma}_{i,F}\right)$ where $\hat{\mu}_{i,F},\, \hat{\Sigma}_{i,F}$ are the maximum likelihood estimates of the returns' mean and covariance. For the hierarchical methodologies, the mean is taken with respect to the posterior predictive density $p\left(r_{i,T+1}, F_{T+1} \middle| r, F\right)$, where r and F collect all the returns of all funds and the benchmark portfolios.[26]

First, we solve the allocation problem for each fund that is active at the end of our sample period, and we pick the 'best' fund and its corresponding optimal portfolio, which yields the maximal expected utility among all portfolios across all funds. In Table 9 we present, for each methodology, the ticker symbol of the best fund and the optimal weight on it. We also present the certainty equivalent (CE) gain from this portfolio relative to one that invests only in the risk-free asset and the benchmarks. We see that the best funds and their optimal allocations as well as the associated CE gains differ across methodologies for almost all levels of risk aversion ρ. The average gain (across levels of ρ) is 33 basis points per month for the fund-by-fund approach, 18 basis points for the hierarchical normal model, and 26 basis points for our model; these differences are economically very significant. Overall, both in terms of the optimal portfolios and in terms of the economic gain associated with these portfolios, the results from our methodology lie between those from the alternative approaches. This is because, focusing on the best-performing funds, the fund-by-fund methodology estimates fund return distributions with higher means as there is no learning across funds, hence no shrinkage toward the population mean. The hierarchical normal model, on the other hand, yields fund return distributions with lower means as it is not flexible enough to adequately fit the right tail of the alpha distribution.

To see this more clearly, we compare across estimation methodologies the optimal portfolio weights on a specific fund. We choose the Sequoia fund as the subject of the analysis, because it was the top-performing fund in 2011 according to the financial press, and is therefore likely to be relevant to investors.[27] The results in Table 10 show that the optimal weights on this fund differ

---

[25] This problem is similar to that in Kandel and Stambaugh (1996) and Jones and Shanken (2005). Following these studies, we impose $w_i \geq 0$, $w_F \geq \mathbf{0}$, and $w_i + w_F'\mathbf{1} \leq 0.99$ to avoid expected utility equaling $-\infty$. Also, we fix r_f at 6% per year, which is close to its mean over our sample period.

[26] For details on how to make draws from this predictive density, see Section C.8 of the online appendix.

[27] This ranking, from Bloomberg (2011), excludes funds with assets below \$250 million and funds requiring a minimum investment above \$100,000, and is based on performance from January 1 to December 2 of 2011; the Sequoia fund topped the ranking with a total return of 12.1% over the evaluation period.

widely across estimation methodologies. For our methodology, the optimal investment varies from full (for $\rho = 0$) to about 50% (for $\rho = 1$ to 5). For the fund-by-fund methodology, the optimal investment is as high as twice that using our methodology, while for the hierarchical normal model, the optimal investment for $\rho = 0$ is none (versus full for our methodology) while for other ρs it is close to half that using our methodology.[28]

Finally, we calculate the certainty equivalent loss (according to our estimates) if we ignore information sharing and/or specification uncertainty in the portfolio formation, i.e., if we select the best fund and the optimal weights using one of the alternative models instead of our model. The results in Table 11 show that, as expected, the loss is higher when using the fund-by-fund approach than when using the hierarchical normal model, and that the loss is generally decreasing with risk aversion. Importantly, for risk aversions $\rho = 0$ through 2, the loss is economically significant at more than 10 basis points per month for both models. Furthermore, for intermediate levels of risk aversion ($\rho = 1$ and 2), this loss has similar magnitude for both models, indicating that incorporating specification uncertainty, i.e., allowing for unrestricted learning across funds, is of similar importance to allowing for *some* learning under the normality assumption.

7 Additional results

In this section, we repeat our estimation using subgroups of funds with different investment objectives, funds that are active during different subperiods, as well as fund returns before fees and expenses. We also conduct a post-estimation analysis of the relation between capital flows and estimated alphas, to check if capital is directed toward funds with superior performance and how this affects subsequent performance.

[28] It is notable that, with the hierarchical normal model, the optimal allocation to the Sequoia fund for $\rho = 0$ is zero, while for higher ρ it first increases and then decreases. As we have explained, this model is not flexible enough to adequately capture the tails of the alpha distribution, so it substantially shrinks the alphas of top-performing funds, including Sequoia, toward the mean, making them seem less appealing. A risk-neutral investor ($\rho = 0$) fully invests in the risky asset with the highest return; for the hierarchical normal model, this is one of the benchmarks, *not* the Sequoia fund to which he therefore allocates no wealth. A risk-averse investor seeks to diversify hence invests in Sequoia, and an even more risk-averse investor seeks to invest in safer assets hence invests less in Sequoia.

7.1 The distribution of skill by fund investment objective

Here, we estimate the skill distribution for groups of funds that follow different investment strategies. For this purpose, we utilize a restricted data set that contains funds covered by the Thomson database and linked to CRSP through MFLINKS, because as is generally accepted, Thomson provides more consistent investment objective information over time and across funds. The restricted data set contains 1,865 funds classified into three investment categories: Growth & Income (405 funds), Growth (1,230 funds), and Aggressive Growth (230 funds). We present summary statistics for the characteristics of funds in this restricted sample in the online appendix.

The results in Table 12 show that the Aggressive Growth (AG) category has the highest proportion of positive- and the lowest proportion of negative-alpha funds (24% and 40%), followed by the Growth (G) category (14% and 66%) and the Growth & Income (GI) category (10% and 76%). While these proportions are consistent with previous findings that AG funds have superior performance (e.g., Grinblatt and Titman, 1989), the percentiles in Table 13 reveal that the proportions do not tell the whole story. By estimating the entire alpha distribution for each investment group, we find that the one for AG funds has much fatter tails, *both* left and right, than those for the other groups. For example, we find that 10 times more AG funds than GI funds have $\alpha > 2\%$ but also that 10 times more AG funds than GI funds have $\alpha < -4\%$, per year. Thus, we conclude that investments in AG funds have larger upside but also larger downside.

Regarding the distributions of the factor loadings, in Figure 8 we see that there is significant variation both within and across investment objectives. For the posterior mean loading on all factors, there is an ordering from GI, to G, to AG funds, which indicates that the classification scheme broadly captures differences in investment strategies that are reflected in funds' factor loadings. But, as we will show in Section 8.2, we find little correlation between alpha and the factor loadings, so the differences in performance between funds with different investment objectives are not simply due to differences in their factor loadings.

7.2 The prevalence of short-term skill and its evolution over time

In the baseline estimation of our model, we assume that fund skill and the population distribution from which it is drawn are both constant over time. However, skill may change, e.g., because the manager changes, and the population distribution may change, e.g., because the stock market is becoming more efficient or due to regulatory changes. The same is true for the other model parameters, like the factor loadings. Here, we repeat our baseline analysis for sample subperiods, which enables us to relax the assumption that skill and other model parameters are constant throughout the whole sample, though at the cost of an increase in statistical error as we necessarily use less information in each estimation.

We consider skill to be constant in the short term, and we estimate our model using 5-year rolling windows starting from the period 1975–1979 and ending in 2007–2011, advancing one month at a time. In Figure 9, we show the evolution of short-term performance. We see that the proportions of zero-, negative-, and positive-alpha funds are almost equal until the early 1990s, and subsequently the proportion of negative-alpha funds increases at the expense of the other two types. Similarly, the estimated alpha distribution is quite stable until the early 1990s, but afterward it becomes more concentrated and many quantiles switch from positive to negative values.[29]

7.3 Analysis of gross-of-expenses returns

Here, we re-estimate our baseline model using gross returns (before fees and expenses, but after transaction costs), which we compute by adding to each fund's monthly net returns the monthly equivalent of its annual expense ratio, as reported by CRSP. This analysis enables us to examine if the distribution of gross-of-expenses alphas under the assumed asset pricing model is consistent with the hypothesis that the stock market is sufficiently efficient that risk-adjusted returns after transaction costs equal zero.

In Table 14, we present the estimated proportions of funds with zero, negative, and positive alpha, and in Table 15 we present various percentiles of the estimated distribution of alpha (the distribution of the factor loadings is as before). We find that a significant proportion of funds, 22%, have an excess return of zero, 44% of funds generate positive excess returns, while 34% generate negative excess returns. Our results suggest that, conditional on the asset pricing model, there are significant market inefficiencies, which close to half the funds can exploit. These results are very different from the ones obtained from a fund-level regression analysis. For example, Barras,

[29]Pastor, Stambaugh and Taylor (2015) use a fund-by-fund analysis to find a similar pattern in the evolution of the average alpha across funds. Their statement that "funds have become more skilled over time" does not contradict our finding, since they define skill differently, as alpha adjusted for fund and industry size, i.e., the alpha on the first dollar invested in the fund when there are no other funds in the industry. Assuming no other changes in the industry, they show that while funds' ability to generate alpha would have increased over time if competition in the mutual fund industry had remained at 1980 levels, increased competition has actually reduced their ability to generate alpha, which is consistent with our results.

Scaillet and Wermers (2010) find that 85% of funds have zero alpha, 5% have negative alpha, and 10% have positive alpha before expenses, which would imply that the market is generally efficient, with few and not easily exploitable inefficiencies.

Comparing the proportions estimated using returns before and after expenses, we see that, as expected, the proportion of funds with negative (positive) alpha before expenses is lower at 34% vs. 78% (higher at 44% vs. 13%) than that after expenses. Our estimates also imply that half the funds with positive alpha charge fees above the rents they generate, about a fifth (the 9% of funds with zero net-of-expenses alpha) capture the rents they generate, and a third leave money on the table. It is noteworthy that the fund-by-fund approach fails to categorize funds on the basis of returns before and after expenses in a consistent manner. For example, Barras, Scaillet and Wermers (2010) find that 85% of funds have zero alpha before expenses, while 75% have zero alpha after expenses, which is contradictory given that expenses are almost always nonzero. This is a further indication that the fund-by-fund approach can be overly conservative.

7.4 Fund flow analysis

Our finding that most funds have negative alpha and very few have zero alpha net of fees and expenses suggests that the mutual fund industry is not at a zero-alpha long-run equilibrium (see Berk and Green, 2004). Here, we examine whether the *forces* discussed by Berk and Green (2004) are at work, namely whether investors direct capital toward funds with good past performance and whether funds exhibit decreasing returns to scale in deploying their skill.

For each fund in our sample, we compute the annual net flow of capital as the percentage increase in total net assets in excess of the net investment gains. Each year, we sort funds into quintiles based on their annual flows, and we calculate the performance of each quintile in the previous and in the subsequent 5-year period. We measure fund performance using i) the posterior probability of having a positive alpha, ii) the posterior probability of having a negative alpha, and iii) the posterior mean of alpha, all relative to the 4-factor model and estimated using our methodology. In Panel A (B) of Table 16, we present the past (future) performance of each flow quintile, averaged across all 5-year non-overlapping periods in our sample (for the results for each period separately, see Section C.11 in the online appendix). In Panel A, we see strong evidence that investors direct capital toward funds with better past performance, as funds in higher flow quintiles have significantly better performance in the period prior to the flow measurement. For example, the average alpha prior to flow measurement is -0.19% per year for funds in the top flow quintile versus -0.82% for funds in the bottom quintile. Comparing for each quintile its past performance in Panel A with its future performance in Panel B, we also see evidence that funds that experience capital outflows (inflows) exhibit higher (lower) performance subsequently. For example, for funds in the bottom flow quintile there is an average increase of 0.34% in annualized alpha between the periods prior to and subsequent to the flow measurement, while for funds in the top flow quintile there is a decrease of 0.41% in alpha.[30] Interestingly, however, this force seems to push funds toward negative rather than zero alpha, suggesting that convergence to an equilibrium in which most funds have alpha equal or close to zero is unlikely to occur soon. In the online appendix, we present a regression analysis which yields similar results.

8 Robustness checks

In this section, we conduct the following robustness checks: i) we conduct a prior sensitivity analysis, ii) we allow model parameters to be correlated, iii) we allow for time-varying market risk exposures, iv) we examine whether our results are driven by our distributional assumptions about the factor loadings and the factor model errors, v) we relax the assumption of independent factor model errors, and vi) we consider variations for the distributions of zero and nonzero alphas. We briefly discuss our results here, and we present related tables and figures in the online appendix.

8.1 Prior sensitivity analysis

Here, we conduct a prior sensitivity analysis and show that our main results are robust to alternative priors (see Table 17, and also Section C.12 in the online appendix).

Our baseline prior for the number, K^- and K^+, of negative and positive components in the skill distribution is the Poisson, $p(K) \propto \frac{1}{K!}$, truncated at $K^-_{\max} = K^+_{\max} = 4$. We vary both the functional form (from Poisson to uniform, i.e., $p(K) = \frac{1}{K_{\max}}$) and the truncation point (from 4 to 6), and we estimate our model using various combinations of the two. As expected, the posterior probabilities of mixtures with more components are slightly higher with the uniform prior, but overall, the effects of these variations on the posterior estimates of population proportions and percentiles are small.

Our baseline prior for the population proportions, π, is the symmetric Dirichlet $\mathcal{D}\left(\underline{\pi}\right)$ with $\underline{\pi} = 1$, which is an uninformative uniform over all values of π. Fruhwirth-Schnatter (2006) shows

[30] Performance may partially persist even in the presence of decreasing returns to scale, because investors may be slow to direct capital toward (away from) funds with good (poor) performance. This may explain why, in Panel B, we observe a small difference between the bottom and top flow quintile in performance in the 5-year period after the flow measurement.

that the $\mathcal{D}(1)$ prior may lead to a high risk of choosing too many components, and suggests using the $\mathcal{D}(4)$ prior instead. Using the latter, we find that the posterior is concentrated to a smaller number of components but the change is small. The changes in the percentiles of the estimated skill distribution are also small. We also repeat our estimation replacing the symmetric $\mathcal{D}(1)$ prior with the asymmetric $\mathcal{D}(1, 3, 3)$ prior which overweights low values for the proportion of zero-alpha funds; in the former case, the marginal prior on the proportion of zero-alpha funds has the Beta distribution $\mathcal{B}(1, 2)$ with mean 0.33 while in the latter it has distribution $\mathcal{B}(1, 6)$ with mean 0.14. Again, our results are quite robust to this change.

We also vary the deep parameters governing the prior of the population mean and variance of the nonzero components of the alpha distribution. Since these are hyperparameters at a second hierarchical layer, we find that they have a trivial effect on the posteriors.

Finally, we analyze the sensitivity of the estimated skill density, i.e., the posterior predictive density, on the *prior* predictive density. This analysis is different from the above in that, instead of focusing on varying the marginal priors of the population parameters, it focuses on varying the prior predictive density of alpha itself, which compounds all the marginals. We find that a drastic change in the prior predictive density — from one in which the mass is concentrated very close to zero to one in which the mass is concentrated far from zero — has almost no effect on the posterior predictive density.

8.2 Dependence between model parameters

In our baseline specification, we assume that fund-specific skill α_i, factor loadings β_i, and error precision h_i are drawn independently. However, a fund's alpha may be related to some aspect of its investment strategy, hence possibly to its risk exposure. It may also be related to its error precision: As noted by Jones and Shanken (2005), if skilled managers identify mispricings across a wide range of stocks, we might expect a positive correlation between α_i and h_i, while if they generate alpha by focusing on a few stocks and forgo diversification, we might expect a negative correlation; or maybe funds that closely follow their benchmarks have both larger h_i and smaller absolute value of α_i. To study this issue, in this section we allow for a full correlation matrix between all model parameters, that is, we assume that α_i, β_i, and h_i are drawn from a multivariate mixture distribution.

For tractability, we do not tackle model specification uncertainty here, rather we use the model with the highest posterior probability, i.e., with $K^- = 2$ negative and $K^+ = 1$ positive mixture components. We propose that the density of $\varphi_i := \left(\alpha_i, \beta_i', h_i\right)'$ is $p\left(\varphi_i \left|\left\{\pi^q, \theta_\varphi^q\right\}\right.\right) = \sum_{q\in Q} \pi^q f^q\left(\varphi_i \left|\theta_\varphi^q\right.\right)$, where $\left\{\pi^q, \theta_\varphi^q, f^q\right\}$ are defined analogously to $\left\{\pi^q, \theta_\alpha^q, f^q\right\}$ in Equation 2. Specifically, the density of the transformed parameter vector $\tilde{\varphi}_i := \left(\ln|\alpha_i|, \beta_i', \ln h_i\right)'$ is

$$\begin{aligned}
p\left(\left(\beta_i, \ln h_i\right) | \, e_i^0 = 1, \theta_\varphi^0\right) &= f_{\mathcal{N}}\left(\left(\beta_i, \ln h_i\right) | \, \theta_\varphi^0\right) \\
p\left(\tilde{\varphi}_i \, | \, e_{i,k}^- = 1, \theta_{\varphi,k}^-\right) &= f_{\mathcal{N}}\left(\tilde{\varphi}_i \, | \, \theta_{\varphi,k}^-\right) \\
p\left(\tilde{\varphi}_i \, | \, e_{i,k}^+ = 1, \theta_{\varphi,k}^+\right) &= f_{\mathcal{N}}\left(\tilde{\varphi}_i \, | \, \theta_{\varphi,k}^+\right),
\end{aligned}$$

from which we can derive the $\{f^q\}$ using the Jacobian, and where $f_{\mathcal{N}}$ is the normal density.[31]

We find that the proportions of zero-, negative-, and positive-alpha funds (15%, 65%, and 20%) are not very different from those we estimate in the specification with no correlation (14%, 71%, and 15%) in the model with $K^- = 2$ and $K^+ = 1$ negative and positive components. Furthermore, we find that the estimated distribution of α is similar to that in the model without correlation, with some mass having shifted from negative to positive values. While we see that the distributions of the βs and h are different across types, as we explain in Section 8.4, the distributional assumption about the βs (and h) has no material effect on the fund-level posteriors of these parameters and, hence, on the fund- and population-level posteriors of α. Finally, we find that the correlation between α and β and between α and h is small, with posterior mean estimates below 0.1 for most pairwise correlations. Overall, we conclude that the simplifying assumption that alphas, factor loadings, and error precision are uncorrelated does not significantly affect our main results. Detailed results are presented in Section C.13 of the online appendix.

8.3 Conditional asset pricing model

Here, we examine the sensitivity of our results to using the conditional version of the Carhart (1997) 4-factor model, where funds' risk exposures are allowed to vary with the state of the economy as proxied by predetermined public information variables (see Ferson and Schadt, 1996). We allow for time-varying loadings on the market factor M_t, i.e.,

$$r_{it} = \alpha_i + F_t'\beta_i + \left(M_t \cdot C_{t-1}'\right)\gamma_i + \varepsilon_{it},$$

[31] To allow for dependence, we assume multivariate normality, so h_i is log-normal instead of Gamma as in Section 2. This should have little effect on our results since these distributions are similar, and also we have sufficient fund-level information to estimate h_i accurately so the population shape should not affect the posteriors.

where γ_i is the coefficient vector for the conditional term, C_{t-1} is the vector of lagged deviations of the conditioning variables from their time-series means, and the conditioning variables are: i) the one-month T-bill rate; ii) the dividend yield of the CRSP value-weighted stock index; iii) the term spread, proxied by the yield difference between a constant-maturity 10-year Treasury bond and a three-month T-bill; and iv) the default spread, proxied by the yield difference between Moody's Baa- and Aaa-rated corporate bonds.[32]

As expected, using the conditional specification, we find a higher proportion of negative-alpha funds (81% vs. 78%) and a smaller proportion of positive-alpha funds (11% vs. 13%) than with the unconditional one. However, the general shape of the estimated distribution remains unchanged. Detailed results are presented in Section C.14 of the online appendix. In unreported results, we also use the CAPM and the Fama and French (1993) 3-factor model, for which, as expected, we find a slightly lower (higher) proportion of funds with negative (positive) alpha.

8.4 Distributional assumptions about errors and factor loadings

Here, we examine whether our estimated alpha distribution is affected by our distributional assumptions about the fund return errors and the factor loadings in Equation 1.

In the baseline model, we assume that fund-specific errors ε_{it} are normal with fund-specific precision h_i drawn from a Gamma. This implies that the marginal distribution for the pooled errors is a t distribution, which is a symmetric distribution with fat tails. Since it incorporates heterogeneity across funds and fat tails, our assumption is appropriate both for studying the performance of heterogeneous funds and for robust analysis. However, it does not account for possible asymmetries in the distribution of the pooled errors nor for non-normalities in the fund-specific errors, so it could affect our results. To study this issue, first we examine whether we observe significant skewness in the distribution of the estimated residuals, $\hat{\varepsilon}_{it} := r_{it} - \hat{\alpha}_i - F_t\hat{\beta}_i, \forall i, t$, pooled across funds, where $\hat{\alpha}_i$ and $\hat{\beta}_i$ denote the posterior means of our estimates. Various robust skewness measures are close to zero, indicating very little skewness.[33] Second, we use the robust Jarque-Bera test (Brys, Hubert and Struyf, 2004) to simultaneously test for each fund the null that its estimated residuals are normal. Accounting for false discoveries, we reject the normality null for only 3% of funds. Finally,

[32] The T-bill rate is from CRSP; the dividend yields are constructed from returns with and without dividends, also from CRSP; and the term and default spreads are derived from Federal Reserve data.

[33] Sample skewness is 0.80, but it is sensitive to even a single outlier, so we compute robust measures based on the 25/75 and the 0.5/99.5 quantiles and find them to be 0.01 and 0.02. These indicate almost no skewness, as they are smaller than the corresponding measures for, e.g., the χ^2 (1000), which is approximately normal.

we use a forward-search procedure (Coin, 2008) to identify and remove for each fund the residuals that deviate from the values of an equal-sized normal random sample; in total, we remove 1.8% of all observations. Re-estimating our model using this reduced data set for which our distributional assumption for the error should be even more accurate, we find very similar estimates to those using the entire data set, so we conclude that our results are not driven by this assumption.[34]

In the baseline model, we also assume that fund-specific factor loadings β_i are normal. While this necessarily implies that our posterior for β is normal, it has no appreciable effect on the results that mainly interest us. The reason is that the data contain sufficient information to estimate each fund's β precisely, so information sharing across funds has a very small effect on the posteriors of fund-level βs and, therefore, of fund-level αs and their population distribution. To demonstrate this, first we find that our posterior estimates are very similar with the OLS estimates of fund-level βs, which do not assume normality and do not induce information sharing and shrinkage, with most pair-wise correlations at 0.99. Second, we repeat our estimation, this time *imposing* through the priors a large variance in the population distribution for βs, effectively eliminating shrinkage in the posterior estimates of fund-level βs. Comparing our estimation results in this "no shrinkage" case with our baseline results, we observe that they are very similar, suggesting that the normality assumption has a very limited impact. For more details on this analysis, see Section C.15 of the online appendix.

8.5 Cross-sectional error dependence

Here, we relax the assumption of cross-sectional independence in the factor model errors by allowing their covariance to have the linear factor structure proposed by Jones and Shanken (2005). Specifying the error of fund i as $\varepsilon_{it} := G_t'\delta_i + \xi_{it}$, Equation 1 becomes

$$r_{it} = \alpha_i + F_t'\beta_i + G_t'\delta_i + \xi_{it},$$

where G_t is the month t vector of latent error factors which are assumed to be normal with mean 0 and variance 1 (without loss of generality) and orthogonal to each other and to the factors F_t; δ_i are fund-specific error factor loadings; and $\xi_{it} \sim \mathcal{N}\left(0, h_i^{*-1}\right)$ is the cross-sectionally independent part of the error term, with h_i^* a fund-specific precision. The benchmark is still assumed to be correctly specified, so α_i can still be interpreted as a measure of skill. In one specification, we

[34]Specifically, in this reduced data set, the sample skewness of the pooled residuals is 0.03 and the *non-robust* Jarque-Bera test against normality is only rejected for 3% of funds. The alpha distribution estimated from this data set has slightly fatter tails, especially in the left. This is because the excluded observations have large residuals, so the posteriors of fund-level alphas have lower variance hence exhibit less shrinkage toward the mean, and so more alphas are placed in the tails.

assume that there is a single latent factor on which all funds may load. In another specification, we allow for additional dependence across funds that follow the same strategy by introducing 4 latent error factors: one on which all funds may load, and three strategy-specific factors on which only funds that follow a specific investment strategy (Growth & Income, Growth, and Aggressive Growth, respectively) may load.

Our baseline model is quite robust to this cross-sectional error dependence. Using, e.g., the model with one latent factor, we estimate that the proportions of zero-, negative-, and positive-alpha funds are 8%, 79%, and 13%, and the distribution of negative-alpha funds shifts slightly to the left. For detailed results, see Section C.16 of the online appendix.

8.6 Variations to model specification

Finally, we estimate two variations of our model for the alpha distribution. The first retains the distinction between zero- and nonzero-alpha funds, but ignores the distinction between negative- and positive-alpha funds, i.e., it assumes α_i is drawn from a mixture that consists of a point mass at 0 and a normal component. The second retains the distinction between negative- and positive-alpha funds, but blurs the distinction between zero- and nonzero-alpha funds by replacing the point mass at zero with a narrow normal centered at zero. First, we find that our estimate of the point mass at zero is not driven by our assumption that nonzero alphas are drawn from two non-overlapping distributions. That is, our finding that a small mass of funds, if any, have zero alpha is robust to this alternative specification for the distribution of nonzero alpha. Second, we find that our estimated percentiles of the alpha distribution are not driven by our model's feature to allow for a point mass at zero. For detailed results, see Section C.17 of the online appendix.

9 Concluding discussion

In this paper, we propose a novel methodology that allows for a rich and flexible representation of the cross-sectional distribution of skill. We model the skill distribution using a three-component mixture that consists of a point mass at zero and two components, one with negative and one with positive support. To avoid restrictive parametric assumptions, we tackle specification uncertainty by modeling the negative and the positive components of the skill distribution as mixture densities across an unknown, but estimable, total number of components. This approach enables us to jointly estimate the proportions of skilled/unskilled funds and the cross-sectional distribution of skill in a unified framework, as well as to use the full information in the data without imposing restrictive parametric assumptions. Thus, as we show in simulations and out-of-sample tests, we are able to improve inference at the population and at the fund level.

We find that the skill distribution is highly non-normal, exhibiting fat tails and negative skewness, and that while the majority of funds is classified to the negative-alpha group, there is a substantial proportion that generate positive alpha (net of fees and expenses). These results could, for example, be useful in investigating the rationality of investing in mutual funds. They also have significant implications for asset allocation, as the optimal portfolios calculated using our methodology are substantially different from those calculated using the fund-by-fund regression approach or the hierarchical model that assumes alphas are normal. Quantifying economically these implications for asset allocation, we find that allowing for unrestricted learning across funds has similar incremental economic value to allowing for some learning under the normality assumption in the first place.

In closing, we note that our methodology can be applied more widely to a variety of issues in finance. First, it can be used to estimate the proportions of different types in a population, e.g., different types of investors (in terms of risk preferences, sophistication, investment style, etc.). Second, it can be used to flexibly estimate the population distribution of an unobserved characteristic (e.g., firm or bank quality), especially if this distribution is likely to exhibit complex features. Third, it enables the sharing of information across individual units without restrictions, which can be particularly useful in settings in which data as well as prior information is limited; for example, it can be used to obtain improved inference for individual-level trading behavior. Finally, it can be used to conduct complex model selection analyses, e.g., to select the asset pricing model that best explains some test assets.

References

Avramov, D, and R Wermers. 2006. "Investing in mutual funds when returns are predictable." *Journal of Financial Economics*, 81(2): 339–377.

Baks, K, A Metrick, and J Wachter. 2001. "Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation." *Journal of Finance*, 56(1): 45–85.

Barberis, N, and A Shleifer. 2003. "Style investing." *Journal of Financial Economics*, 68(2): 161–199.

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

Benjamini, Y, and Y Hochberg. 1995. "Controlling the false discovery rate: A practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1): 289–300.

Berk, J. B, and R. C Green. 2004. "Mutual fund flows and performance in rational markets." *Journal of Political Economy*, 112(6): 1269–1295.

Brys, G, M Hubert, and A Struyf. 2004. "A robustification of the Jarque-Beta test of normality." In *Compstat 2004 Symposium*, Physica-Verlag, Springer.

Brys, G, M Hubert, and A Struyf. 2006. "Robust measures of tail weight." *Computational Statistics and Data Analysis*, 50: 733–759.

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

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

Celeux, G. 1998. "Bayesian inference for mixture: The label switching problem." In *COMPSTAT*, ed. Roger Payne and Peter Green, 227–232. Physica-Verlag HD.

Chib, S, and E Greenberg. 1995. "Understanding the Metropolis-Hastings algorithm." *The American Statistician*, 49(4): 327–335.

Cohen, R, J Coval, and L Pastor. 2005. "Judging fund managers by the company they keep." *Journal of Finance*, 60(3): 1057–1096.

Coin, D. 2008. "Testing normality in the presence of outliers." *Statistical Methods and Applications*, 17(1): 3–12.

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

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

Ferson, W, and R Schadt. 1996. "Measuring fund strategy and performance in changing economic conditions." *Journal of Finance*, 51(2): 425–461.

Ferson, W, and Y Chen. 2015. "How many good and bad fund managers are there, really?" Unpublished Paper.

Fruhwirth-Schnatter, S. 2006. *Finite mixture and Markov switching models: Modeling and applications to random processes.* New York:Springer.

Gelman, A. 2006. "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis*, 1(3): 513–533.

Geman, S, and D Geman. 1984. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." *Pattern Analysis and Machine Intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6: 721–741.

Geweke, J. 1999. "Using simulation methods for Bayesian econometric models: Inference, development, and communication." *Econometric Reviews*, 18(1): 1–73.

Gottardo, R, and A Raftery. 2008. "Markov chain Monte Carlo with mixtures of mutually singular distributions." *Journal of Computational and Graphical Statistics*, 17(4): 949–975.

Green, P. J. 1995. “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.” *Biometrika*, 82(4): 711–732.

Grinblatt, M, and S Titman. 1989. “Mutual fund performance: An analysis of quarterly portfolio holdings.” *Journal of Business*, 62(3): 393–416.

Groeneveld, R. A, and G Meeden. 1984. “Measure of skewness and kurtosis.” *Journal of Finance*, 33(4): 391–399.

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

Hastings, W. 1970. “Monte Carlo sampling methods using Markov chains and their applications.” *Biometrika*, 57(1): 97–109.

Jensen, M. 1968. “The performance of mutual funds in the period 1945–1964.” *Journal of Finance*, 23(2): 389–416.

Jones, C, and J Shanken. 2005. “Mutual fund performance with learning across funds.” *Journal of Financial Economics*, 78(3): 507–552.

Kandel, S, and R. F Stambaugh. 1996. “On the predictability of stock returns: An asset-allocation perspective.” *Journal of Finance*, 51(2): 385–424.

Koijen, R. S. 2014. “The cross-section of managerial ability, incentives, and risk preferences.” *Journal of Finance*, 69(3): 1051–1098.

Korteweg, A. G, and M Sorensen. 2016. “Skill and luck in private equity performance.” *Journal of Financial Economics*, to appear.

Kosowski, R, A Timmermann, R Wermers, and H White. 2006. “Can mutual fund stars really pick stocks? New evidence from a bootstrap analysis.” *Journal of Finance*, 61(6): 2551–2595.

Marin, J.-M, K. L Mengersen, and C. P Robert. 2005. “Bayesian modelling and inference on mixtures of distributions.” In *Handbook of Statistics*, Vol. 25, ed. D Dey and C R Rao, Chapter 16, 459–507. Amsterdam:Elsevier.

Metropolis, N, A Rosenbluth, M Rosenbluth, A Teller, and E Teller. 1953. “Equation of state calculations by fast computing machines.” *Journal of Chemical Physics*, 21(6): 1087–1092.

Miller, J. W, and M. T Harrison. 2015. “Mixture models with a prior on the number of components.” arXiv:1502.06241.

Miller, R. 1980. “Bayesian analysis of the two-parameter gamma distribution.” *Technometrics*, 22(1): 65–69.

Mullainathan, S. 2000. “Thinking through categories.” Unpublished Paper.

Pastor, L, and R Stambaugh. 2002*a*. “Investing in equity mutual funds.” *Journal of Financial Economics*, 63(3): 351–380.

Pastor, L, and R Stambaugh. 2002*b*. “Mutual fund performance and seemingly unrelated assets.” *Journal of Financial Economics*, 63(3): 315–349.

Pastor, L, R Stambaugh, and L. A Taylor. 2015. “Scale and skill in active management.” *Journal of Financial Economics*, 116(1): 23–45.

Richardson, S, and P. J Green. 1997. “On Bayesian analysis of mixtures with an unknown number of components (with discussion).” *Journal of the Royal Statistical Society: Series B (statistical methodology)*, 59(4): 731–792.

Robert, C. P. 2007. *The Bayesian choice: From decision-theoretic foundations to computational implementation.* . Second ed., New York:Springer Science & Business Media.

Sastry, R. 2015. “The cross-section of investing skill.” Unpublished Paper.

Stephens, M. 1997. “Bayesian methods for mixtures of normal distributions.” PhD diss. University of Oxford.

Stephens, M. 2000. “Bayesian analysis of mixture models with an unknown number of components – An alternative to reversible jump methods.” *Annals of Statistics*, 28(1): 40–74.

Steverman, B, and W Lu. 2011. “Best and worst mutual funds of 2011.” *Bloomberg*, December 17. http://tiny.cc/xxk0nx.

Storey, J. 2002. “A direct approach to false discovery rates.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3): 479–498.

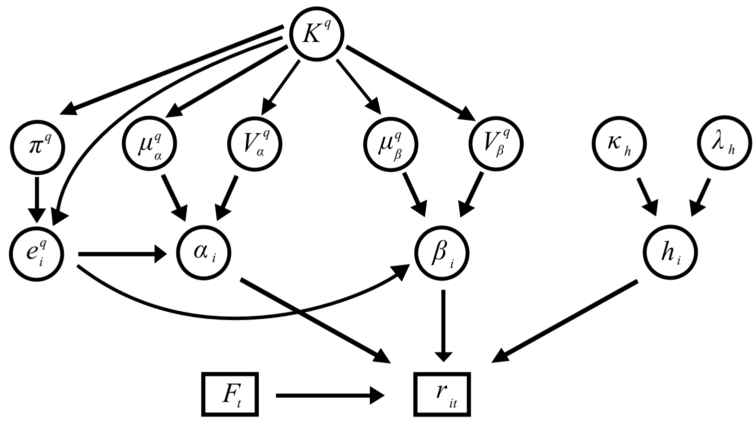


Figure 1: Representation of the baseline mixture model as a directed acyclic graph. Squares represent observed quantities and circles represent unknown model parameters that need to be estimated.

Table 1: Simulations — Estimated Proportions of Fund Skill Groups

Estimation results from simulations in which nonzero alphas (expressed as annualized percentages) are generated from discrete distributions (in Panel A), a normal distribution (in Panel B), and log-normal distributions (in Panel C). The data generating processes (DGPs) within each panel differ in the proportions π^0, π^-, π^+ of funds with zero, negative, and positive alpha, respectively, and/or in the distance of nonzero alphas from zero. For each DGP, we report the true proportions of funds with zero, negative, and positive alpha, and the average (across 50 simulations) estimated proportions from the fund-by-fund methodology that relies on hypothesis tests — with and without the FDR correction — and from our methodology.

Panel A: Discrete nonzero alphas

$$\alpha \sim \pi^0\delta_0 + \pi^-\delta_{x^-} + \pi^+\delta_{x^+}$$

		π^0	π^-	π^+
DGP D-1: *Unequal proportions, large nonzero alphas*	True	75.0	23.0	2.0
$\pi^0 = 0.75, \pi^- = 0.23, \pi^+ = 0.02$	No FDR	75.4	18.9	5.7
$x^- = -3.2, x^+ = 3.8$	FDR	76.5	20.1	3.4
	Our Model	75.2	22.8	2.0
DGP D-2: *Equal proportions, small nonzero alphas*	True	34.0	33.0	33.0
$\pi^0 = 0.34, \pi^- = 0.33, \pi^+ = 0.33$	No FDR	75.6	10.1	14.3
$x^- = -1.2, x^+ = 1.8$	FDR	70.2	12.5	17.3
	Our Model	30.6	35.0	34.4

Panel B: Normal nonzero alphas

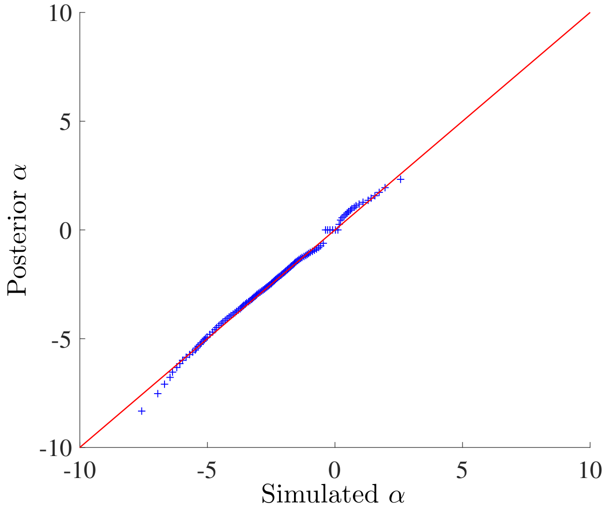
$$\alpha \sim \pi^0\delta_0 + \pi^{-,+} f_{\mathcal{N}}\left(\alpha \,|\mu, \sigma^2\right)$$

		π^0	π^-	π^+
DGP N-1: *Large variance*	True	90.0	5.8	4.2
$\pi^0 = 0.90, \pi^{-,+} = 0.10$	No FDR	82.2	9.7	8.1
$\mu = -1.45, \sigma^2 = 72$	FDR	86.8	7.5	5.7
	Our Model	90.8	5.2	4.0
DGP N-2: *Small variance*	True	35.0	45.8	19.2
$\pi^0 = 0.35, \pi^{-,+} = 0.65$	No FDR	66.8	24.6	8.6
$\mu = -1.45, \sigma^2 = 7.2$	FDR	63.4	28.7	7.9
	Our Model	41.6	41.6	16.8

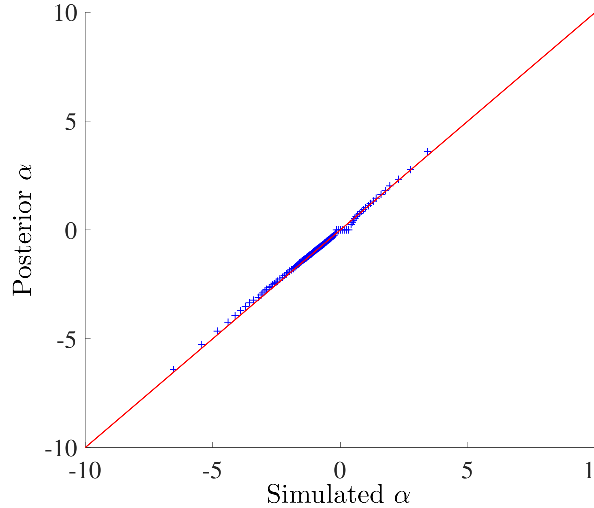
Panel C: Log-normal nonzero alphas

$$\alpha \sim \pi^0\delta_0 + \pi^- f^-\left(\alpha \,|\theta_\alpha^-\right) + \pi^+ f^+\left(\alpha \,|\theta_\alpha^+\right)$$

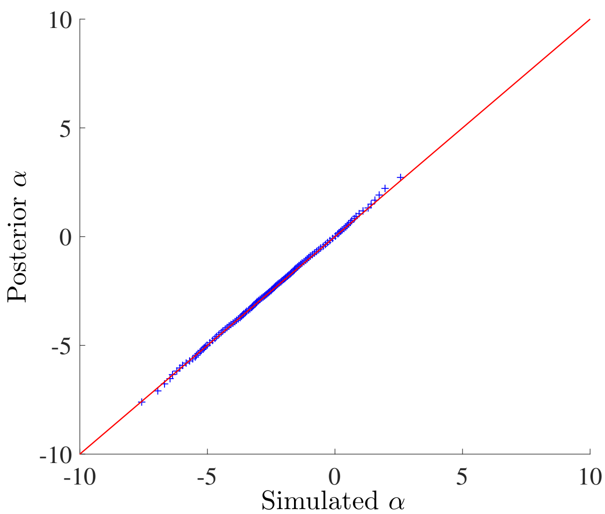
		π^0	π^-	π^+
DGP L-1: *Nonzero alphas far from zero*	True	45.0	28.0	27.0
$\pi^0 = 0.45, \pi^- = 0.28, \pi^+ = 0.27$	No FDR	46.1	27.3	26.6
$f^-(\alpha) = f_{\ln\mathcal{N}}(\lvert\alpha\rvert \,\lvert 2, 0.2)$ for $\alpha < 0$	FDR	45.4	27.6	27.0
$f^+(\alpha) = f_{\ln\mathcal{N}}(\lvert\alpha\rvert \,\lvert 2, 0.2)$ for $\alpha > 0$	Our Model	44.9	28.1	27.0
DGP L-2: *Nonzero alphas close to zero*	True	45.0	28.0	27.0
$\pi^0 = 0.45, \pi^- = 0.28, \pi^+ = 0.27$	No FDR	62.7	19.1	18.2
$f^-(\alpha) = f_{\ln\mathcal{N}}(\lvert\alpha\rvert \,\lvert 1, 0.35)$ for $\alpha < 0$	FDR	59.4	20.6	20.0
$f^+(\alpha) = f_{\ln\mathcal{N}}(\lvert\alpha\rvert \,\lvert 1, 0.35)$ for $\alpha > 0$	Our Model	43.8	28.5	27.7
DGP L-3: *Our posterior*	True	9.0	78.0	13.0
$\pi^0 = 0.09, \pi^- = 0.78, \pi^+ = 0.13$	No FDR	75.5	19.7	4.8
$f^-(\alpha) = 0.74 \cdot f_{\ln\mathcal{N}}(\lvert\alpha\rvert \,\lvert 0.05, 0.23)$	FDR	72.6	26.0	1.4
$+0.26 \cdot f_{\ln\mathcal{N}}(\lvert\alpha\rvert \,\lvert 0.05, 1.25)$ for $\alpha < 0$	Our Model	9.7	79.7	10.6
$f^+(\alpha) = f_{\ln\mathcal{N}}(\lvert\alpha\rvert \,\lvert -0.53, 0.98)$ for $\alpha > 0$				



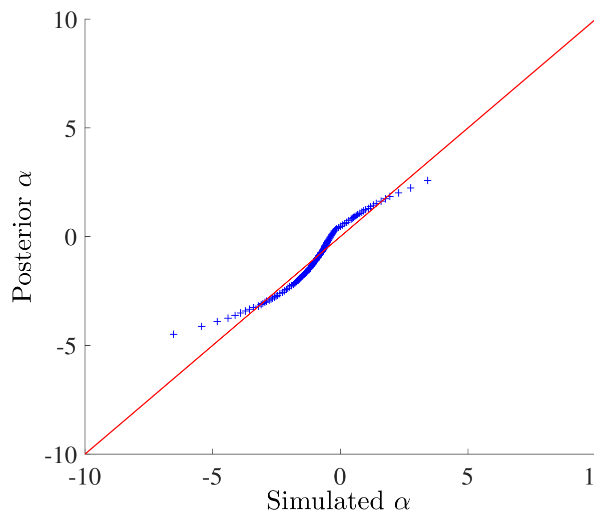
(a) Our posterior vs. simulated normal distribution.



(b) Our posterior vs. simulated skewed/fat-tailed distribution.



(c) Normal posterior vs. simulated normal distribution.



(d) Normal posterior vs. simulated skewed/fat-tailed distribution.

Figure 2: Quantile-Quantile plots of posterior alphas estimated using our methodology and using the hierarchical normal model versus simulated alphas. The blue cross marks plot the quantiles (0.5^{th} through 99.5^{th}), and the solid red line plots the 45° line. In Panels (a) and (b) we plot for DGPs C-1 and C-2, respectively, the quantiles of the posterior alpha distribution estimated using our model against the quantiles of the simulated alpha distribution. In Panels (c) and (d) we plot for DGPs C-1 and C-2, respectively, the quantiles of the posterior alpha distribution estimated using the hierarchical normal model against the quantiles of the simulated alpha distribution. In DGP C-1, alphas are drawn from a normal distribution, in particular $\alpha \sim \mathcal{N}(-2.5, 4)$. In DGP C-2, alphas are drawn from a negatively skewed and fat-tailed distribution, in particular $\alpha \sim \pi^N f^N\left(\alpha \left|\theta_\alpha^N\right.\right) + \pi^- f^-\left(\alpha \left|\theta_\alpha^-\right.\right) + \pi^+ f^+\left(\alpha \left|\theta_\alpha^+\right.\right)$ where $\pi^N = 0.10$, $\pi^- = 0.80$, $\pi^+ = 0.10$, $f^N = f_{\mathcal{N}}(\cdot\,|0, 0.1)$, $f^-(\alpha) = f_{\ln\mathcal{N}}(|\alpha|\,|0.1, 0.5)$ for $\alpha < 0$, and $f^+(\alpha) = f_{\ln\mathcal{N}}(|\alpha|\,|0.1, 0.5)$ for $\alpha > 0$. For clarity, we note that the quantiles plotted on the vertical axis of each plot correspond to the estimated posterior predictive distribution of alpha, not to the posterior means of the alphas for the funds in each simulated sample.

Table 2: Proportions of Fund Types

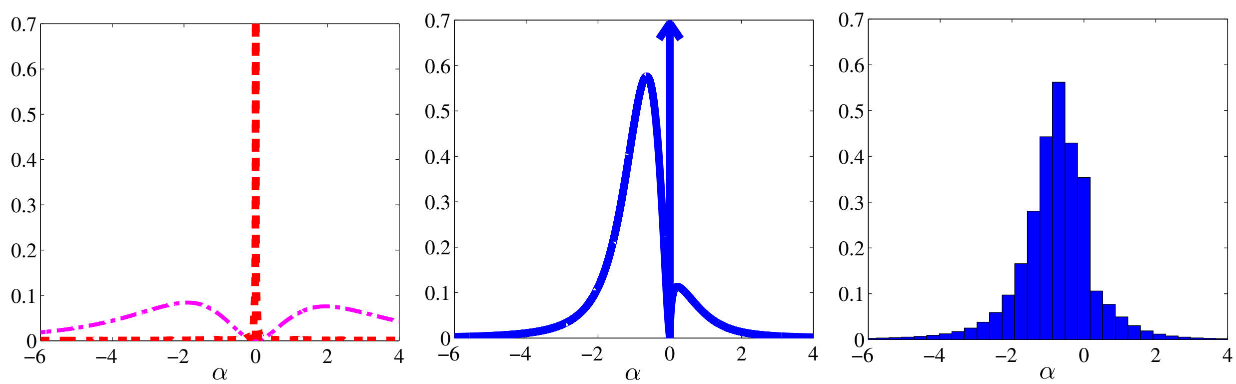
Results on the posterior distributions of the population proportions of funds with zero, negative, and positive alpha, estimated using our baseline model presented in Section 2 with returns net of expenses for 3,497 funds. The 95% Highest Posterior Density Interval (HPDI) is the smallest interval such that the posterior probability that a parameter lies in it is 0.95. NSE stands for autocorrelation-adjusted numerical standard errors for the posterior mean estimate of each parameter.

	Mean	**Median**	**Std.Dev.**	95% **HPDI**	**NSE**
π^0	0.09	0.07	0.07	[0.00 , 0.26]	0.00
π^-	0.78	0.78	0.06	[0.66 , 0.88]	0.00
π^+	0.13	0.12	0.04	[0.06 , 0.22]	0.00

Table 3: Number of Mixture Components

The joint posterior probability for the number of mixture components K^- in the alpha distribution of negative-alpha funds and the number of mixture components K^+ in the alpha distribution of positive-alpha funds, estimated using our baseline model presented in Section 2 with returns net of expenses for 3,497 funds.

	$K^+ = 1$	$K^+ = 2$	$K^+ = 3$	$K^+ = 4$
$K^- = 1$	0.09	0.03	0.00	0.00
$K^- = 2$	0.31	0.23	0.02	0.00
$K^- = 3$	0.13	0.11	0.01	0.00
$K^- = 4$	0.03	0.04	0.00	0.00



(a) Prior density of alpha.

(b) Estimated density of alpha.

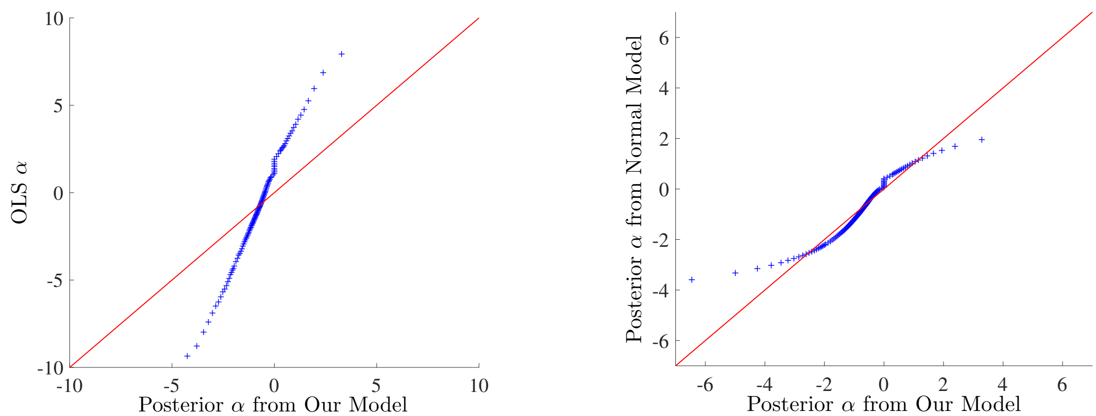
(c) Estimated histogram of alpha.

Figure 3: Visual representation of the prior and posterior predictive distribution of annualized 4-factor alphas (expressed as a percent), from our model in Section 2 estimated using net returns for 3,497 funds. In Panel (a), we plot the prior predictive density corresponding to our baseline prior specification (in dotted red) and to an alternative (in dash-dotted magenta). For clarity, we do not represent the point mass at zero alpha, which has probability 0.33 for the baseline and 0.14 for the alternative. For details on these prior specifications, see Section A in the appendix and Section C.12 in the online appendix. In Panel (b), we plot the posterior predictive density (which is almost exactly the same for both priors); the vertical arrow at zero alpha has arbitrary length and represents a point mass with probability 0.09. In Panel (c), we plot a histogram of values simulated from this estimated distribution, where bar heights are normalized so the area of each bar represents the probability of the corresponding interval.

Table 4: Percentiles of Estimated Skill Distributions — Our Model vs. Alternatives

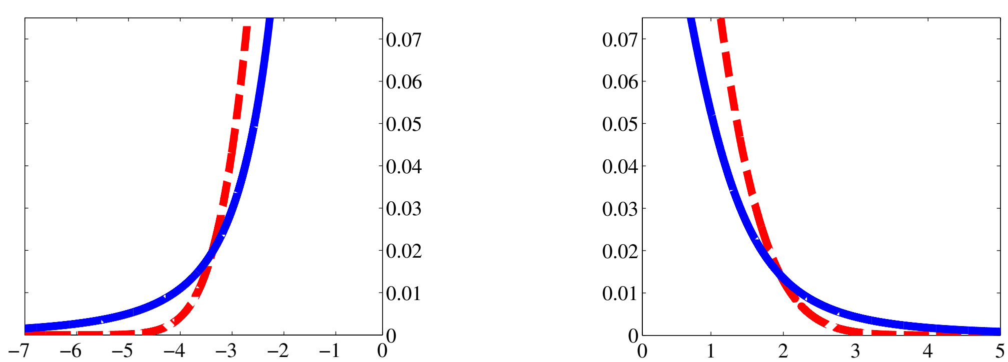
Percentiles of the distribution of annualized alpha (expressed as a %) estimated from our model and the hierarchical normal model, and the empirical distribution of fund-by-fund OLS alphas. For the hierarchical normal and our model, we also present each percentile's 95% HPDI. All distributions are estimated with net returns for 3,497 funds.

	Percentiles														
	0.5^{th}	**1^{st}**	**5^{th}**	**10^{th}**	**20^{th}**	**30^{th}**	**40^{th}**	**50^{th}**	**60^{th}**	**70^{th}**	**80^{th}**	**90^{th}**	**95^{th}**	**99^{th}**	**99.5^{th}**
OLS	-13.28	-10.42	-5.96	-4.20	-2.71	-1.89	-1.31	-0.81	-0.27	0.39	1.20	2.39	3.53	6.87	7.93
Hierarchical Normal Model															
Posterior Mean	-3.55	-3.28	-2.54	-2.16	-1.68	-1.34	-1.05	-0.78	-0.51	-0.22	0.12	0.59	0.98	1.72	1.99
2.5%	-3.76	-3.48	-2.70	-2.28	-1.79	-1.44	-1.13	-0.85	-0.58	-0.30	0.02	0.46	0.83	1.50	1.73
97.5%	-3.32	-3.07	-2.38	-2.02	-1.58	-1.28	-0.98	-0.71	-0.44	-0.14	0.21	0.71	1.12	1.90	2.19
Our Model															
Posterior Mean	-6.60	-5.01	-2.62	-1.96	-1.42	-1.12	-0.96	-0.72	-0.55	-0.36	0.00	0.30	0.87	2.35	3.24
2.5%	-8.97	-6.28	-3.06	-2.22	-1.57	-1.28	-1.08	-0.90	-0.71	-0.52	-0.31	0.00	0.40	1.70	2.28
97.5%	-5.11	-3.91	-2.18	-1.73	-1.28	-0.98	-0.75	-0.58	-0.39	0.00	0.09	0.69	1.24	3.13	4.50



(a) OLS estimates vs. our posterior.
(b) Posterior from hierarchical normal vs. our posterior.

Figure 4: Quantile-Quantile plots of estimated alpha distributions from alternative methodologies. The blue crosses plot the quantiles (0.5^{th} through 99.5^{th}), and the red solid line plots the 45° line. In Panel (a), we plot the quantiles of the OLS alpha estimates versus those of the alpha distribution estimated from our model. In Panel (b), we plot the quantiles of the distribution estimated from the hierarchical normal model versus those of the distribution estimated from our model. We note that the quantiles for our model (horizontal axis, Panels *a* and *b*) and the hierarchical normal model (vertical axis, Panel *b*) correspond to the respective estimated distribution, not the posterior means for the funds in the sample. All estimations use net returns for 3,497 funds. Alpha is annualized and expressed as a %.



(a) Densities of estimated alphas (negative).
(b) Densities of estimated alphas (positive).

Figure 5: Estimated densities of annualized 4-factor alphas (expressed as a percent), for negative (in Panel *a*) and positive (in Panel *b*) alphas, zoomed in to show tail detail. In both panels, we plot the estimated density from our model (in solid blue) and the estimated density from a hierarchical normal model (in dashed red). For all plots, returns net of expenses for 3,497 funds are used.

Table 5: Robust Measures of Skewness and Tail Weight for the Skill Distribution

Resuts on robust quantile-based measures of skewness and tail weight that rely on 99% of the range of the distribution of alpha estimated using our baseline model with returns net of expenses for 3,497 funds. The measure of skewness is as in Groeneveld and Meeden (1984) — $S := {[Q(1-p)+Q(p)-2Q(0.5)]}/{[Q(1-p)-Q(p)]}$ — and the measures of left and right tail weight are as in Brys, Hubert, and Struyf (2006) — $LTW := -\left[Q\left(\frac{1-p}{2}\right)+Q\left(\frac{p}{2}\right)-2Q(0.25)\right]/\left[Q\left(\frac{1-p}{2}\right)-Q\left(\frac{p}{2}\right)\right]$ and $RTW := \left[Q\left(\frac{1+q}{2}\right)+Q\left(1-\frac{q}{2}\right)-2Q(0.75)\right]/\left[Q\left(\frac{1+q}{2}\right)-Q\left(1-\frac{q}{2}\right)\right]$ — where $Q(x)$ is the x^{th} quantile of the distribution, and we use $p=0.005$ and $q=0.995$. Measures are reported as deviations from the corresponding values for the normal distribution. We present point estimates, the 95% HPDI, and the NSE.

	Mean	**Median**	**Std.Dev.**	95% **HPDI**	**NSE**
Skewness	−0.20	−0.20	0.10	[−0.38 , −0.00]	0.00
Left Tail Weight	0.34	0.34	0.03	[0.28 , 0.40]	0.00
Right Tail Weight	0.27	0.29	0.08	[0.08 , 0.37]	0.00

Table 6: Measures of Distance Between the Normal and Our Estimated Skill Distribution

Resuts on robust measures of distance between the normal and our estimated skill distribution estimated using our baseline model with returns net of expenses for 3,497 funds. The Hellinger distance between densities f_X, f_Y is $H^2 := 1 - \int \sqrt{f_X(s)\, f_Y(s)}ds$, and takes values in [0, 1]. The Wasserstein distance between densities f_X, f_Y is $W := \inf_{f_{XY}} E[\|X-Y\|]$ where f_{XY} is any joint density with marginals f_X, f_Y, and takes values in $[0, +\infty)$. For the Wasserstein distance, we present values that rely on 99% of the range of the distribution, i.e., we exclude the extreme tails to make the distance measure robust. For each measure, we present point estimates, the 95% HPDI, and the NSE.

	Mean	**Median**	**Std.Dev.**	95% **HPDI**	**NSE**
Hellinger Distance $\left(H^2\right)$	0.11	0.10	0.04	[0.05 , 0.22]	0.00
Wasserstein Distance (W)	0.22	0.22	0.04	[0.15 , 0.31]	0.00

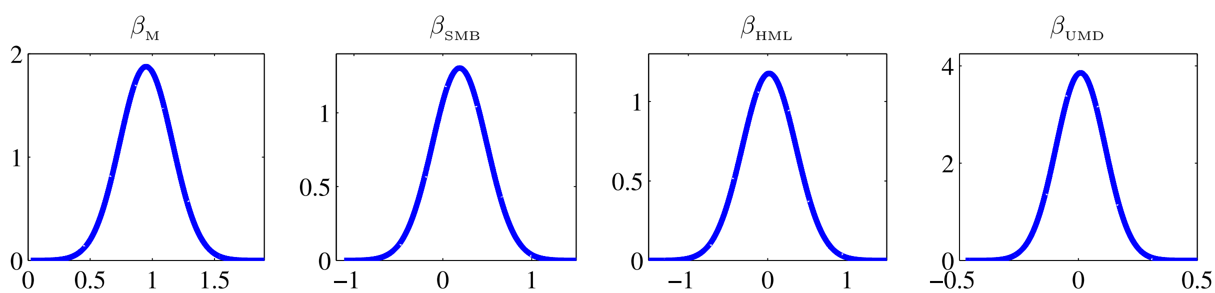


Figure 6: Estimated population densities of factor loadings from our baseline model with net returns for 3,497 funds.

Table 7: Percentiles of Estimated Distributions of Factor Loadings

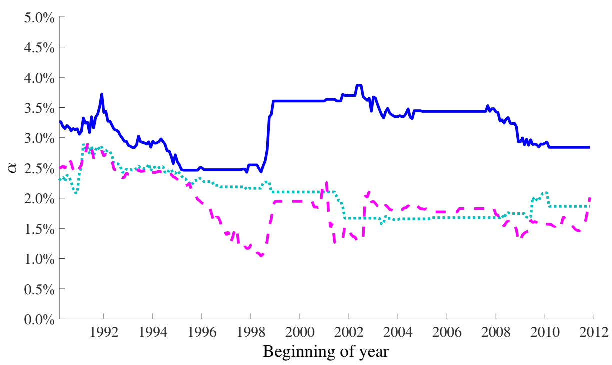
Percentiles of estimated distributions of factor loadings from our model estimated with net returns for 3,497 funds.

	Percentiles														
	0.5th	**1st**	**5th**	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**	**95th**	**99th**	**99.5th**
$\beta_{\mathbf{M}}$	0.40	0.46	0.60	0.68	0.77	0.84	0.90	0.95	1.00	1.06	1.13	1.22	1.30	1.44	1.49
$\beta_{\mathbf{SMB}}$	-0.60	-0.52	-0.31	-0.20	-0.07	0.03	0.11	0.19	0.27	0.35	0.45	0.58	0.69	0.90	0.98
$\beta_{\mathbf{HML}}$	-0.86	-0.77	-0.54	-0.41	-0.27	-0.16	-0.07	0.02	0.11	0.20	0.31	0.46	0.58	0.81	0.89
$\beta_{\mathbf{UMD}}$	-0.26	-0.23	-0.16	-0.12	-0.08	-0.04	-0.02	0.01	0.04	0.06	0.10	0.14	0.18	0.25	0.28

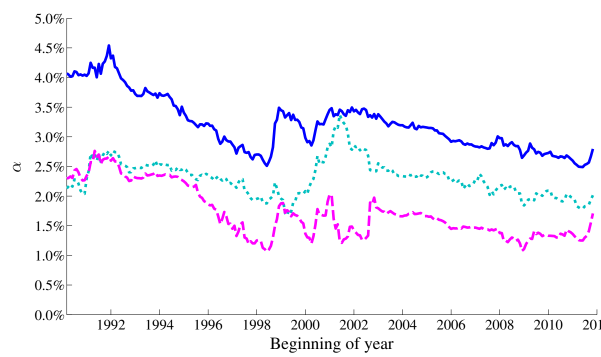
Table 8: Out-of-sample Portfolio Performance

Out-of-sample performance measures for portfolios that select funds using the FDR methodology, a hierarchical model in which fund αs are normal, and our methodology. At the beginning of each month in 1980–2011, we use the preceding 60 months of fund returns to estimate the 4-factor model using each methodology, and we form and hold until the end of the month equal-weighted portfolios of funds that are estimated to have high probability of having a positive α (see Section 6.1 for details). During months in which all funds have a low probability of having a positive α, we select funds whose posterior mean α (for the hierarchical methodologies) or OLS t-statistic (for the FDR methodology) is in the top 1% among all funds in the data set for the preceding 60 months. For each portfolio we construct, we use its monthly portfolio returns from 1980 to 1995 (Panel A), from 1995 to 2011 (Panel B), and from 1980 to 2011 (Panel C) to estimate its annualized OLS 4-factor alpha $\hat{\alpha}$ and residual standard deviation $\hat{\sigma}_\varepsilon$ (both expressed as percents), $\hat{\alpha}$ t-statistic, Information Ratio ($\hat{\alpha}/\hat{\sigma}_\varepsilon$), mean and standard deviation (both expressed as percents) of its return in excess of the risk-free rate, and its Sharpe Ratio (mean/std. dev. of excess return).

	Panel A: 1st–half			**Panel B: 2nd–half**			**Panel C: Full sample**		
	FDR	**Normal**	**Our Model**	**FDR**	**Normal**	**Our Model**	**FDR**	**Normal**	**Our Model**
$\hat{\alpha}$	2.45	2.22	3.36	1.54	1.92	2.24	2.01	1.71	2.80
$\hat{\alpha}$ t-statistic	2.73	2.74	3.51	1.27	1.70	2.28	2.67	1.97	3.68
$\hat{\sigma}_\varepsilon$	3.05	3.03	3.73	4.99	4.20	3.74	4.25	4.45	4.08
Information Ratio	0.80	0.73	0.90	0.31	0.46	0.60	0.47	0.38	0.69
Mean Return	8.52	8.49	10.59	7.04	6.51	7.21	7.74	7.45	8.81
Std. dev. Return	15.90	14.83	15.47	17.48	12.41	14.46	16.73	13.59	14.93
Sharpe Ratio	0.54	0.57	0.68	0.40	0.52	0.50	0.46	0.55	0.59



(a) Alpha of conservative portfolios.



(b) Alpha of aggressive portfolios.

Figure 7: Plots of the evolution over time of the estimated annualized alpha (expressed as a percent) for portfolios that select funds using our estimation methodology (in solid blue lines), a hierarchical model in which fund alphas are normally distributed (in dashed magenta lines), and the FDR methodology (in dotted cyan lines). At the beginning of each month from January 1980 through December 2011, we use the preceding 60 months of fund returns to estimate the 4-factor model using each methodology, and we form portfolios by selecting funds with high estimated probability of having a positive alpha (see Section 6.1 for more details). During months in which all funds have a low probability of having a positive alpha, portfolios in Panel (a) remain empty while portfolios in Panel (b) select funds with high posterior mean alpha (for the two hierarchical methodologies) or high OLS t-statistic for alpha (for the FDR methodology). For each portfolio, we estimate by OLS regression its 4-factor alpha at each month t from January 1990 to December 2011, using the monthly portfolio returns from January 1980 through month t, and we plot the evolution of these alphas over time.

Table 9: Optimal Asset Allocation — Best Fund

Comparison of the optimal portfolio that invests in one mutual fund, the risk-free asset, and the 4 benchmarks, for a CRRA investor with various levels of risk aversion ρ, and for various estimation models: the methodology that performs fund-by-fund OLS regressions (Panel A), the hierarchical normal model (Panel B), and our model (Panel C). We present the ticker symbol of the best mutual fund, the weight w^* on it, and the certainty equivalent (CE) gain (in basis points per month) from the optimal portfolio that invests in the best mutual fund relative to one that invests only in the risk-free asset and the 4 benchmarks.

	Panel A: OLS			Panel B: Normal			Panel C: Our Model		
ρ	**Ticker**	w^*_{OLS}	**CE Gain**	**Ticker**	w^*_{N}	**CE Gain**	**Ticker**	w^*_{our}	**CE Gain**
0	FSCSX	0.99	46.08	PBMBX	0.99	30.50	ARTMX	0.99	36.17
1	VGHCX	0.99	33.57	PBMBX	0.76	17.27	ARTMX	0.99	29.11
2	VGHCX	0.99	30.55	LRSCX	0.62	13.65	ARTMX	0.72	22.60
5	VGHCX	0.74	21.18	FLPSX	0.52	10.14	FLPSX	0.60	16.29

Table 10: Optimal Asset Allocation — Sequoia Fund

For a portfolio that invests in the risk-free asset, the 4 benchmarks, and the best-performing fund of 2011 (the Sequoia fund according to Bloomberg), we present optimal weights on this fund for a CRRA investor with various levels of risk aversion ρ, and various estimation models. Weights w^*_{OLS}, w^*_{N}, and w^*_{our} denote optimal weights according to the methodology that performs fund-by-fund OLS regressions, the hierarchical normal model, and our model.

ρ	w^*_{OLS}	w^*_{N}	w^*_{our}
0	0.99	0.00	0.99
1	0.99	0.32	0.56
2	0.77	0.42	0.55
5	0.63	0.29	0.48

Table 11: Optimal Asset Allocation — Economic Significance

Certainty equivalent (CE) loss, for a CRRA investor with various levels of risk aversion ρ, from ignoring information sharing and/or model specification uncertainty in determining the optimal fund and weights in the portfolio that invests in the risk-free asset, the 4 benchmarks, and a mutual fund. We present CE loss (measured in basis points per month) from using fund-by-fund OLS regressions, i.e., ignoring information sharing, and from using the hierarchical normal model, i.e., ignoring model specification uncertainty.

	Certainty Equivalent Loss	
ρ	**OLS**	**Normal**
0	42.56	6.97
1	16.04	12.78
2	12.59	10.54
5	9.99	0.37

Table 12: Proportions of Fund Types — By Investment Objective

Results on the posteriors of the population proportions of funds with zero, negative, and positive alpha, estimated with returns net of expenses using the $K^- = 2, K^+ = 1$ model separately for funds classified to the 3 investment objectives: Growth & Income (Panel A), Growth (Panel B), and Aggressive Growth (Panel C). The estimation uses data on 1,865 funds for which we have investment objective information from the Thomson database.

	Panel A: Growth & Income Objective			Panel B: Growth Objective			Panel C: Aggressive Growth Objective		
	Mean	**Median**	**Std.Dev.**	**Mean**	**Median**	**Std.Dev.**	**Mean**	**Median**	**Std.Dev.**
π^0	0.14	0.13	0.10	0.20	0.20	0.12	0.36	0.35	0.21
π^-	0.76	0.76	0.09	0.66	0.65	0.09	0.40	0.41	0.16
π^+	0.10	0.08	0.06	0.14	0.13	0.09	0.24	0.22	0.13

Table 13: Percentiles of Estimated Skill Distribution — By Investment Objective

Percentiles of the estimated population distributions of annualized alpha (expressed as a percent), estimated with returns net of expenses using the $K^- = 2$, $K^+ = 1$ model separately for funds classified to the three investment objectives: Growth & Income, Growth, and Aggressive Growth.

	Percentiles														
	0.5th	**1st**	**5th**	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**	**95th**	**99th**	**99.5th**
Growth & Income	-3.75	-3.15	-2.04	-1.66	-1.34	-1.14	-0.96	-0.80	-0.63	-0.40	0.00	0.00	0.47	1.32	1.71
Growth	-4.90	-4.10	-2.56	-1.99	-1.45	-1.12	-0.87	-0.64	-0.40	0.00	0.00	0.10	0.48	2.28	3.62
Aggressive Growth	-17.81	-12.26	-4.35	-2.36	-0.96	-0.36	0.00	0.00	0.00	0.00	0.30	1.12	1.99	4.63	6.25

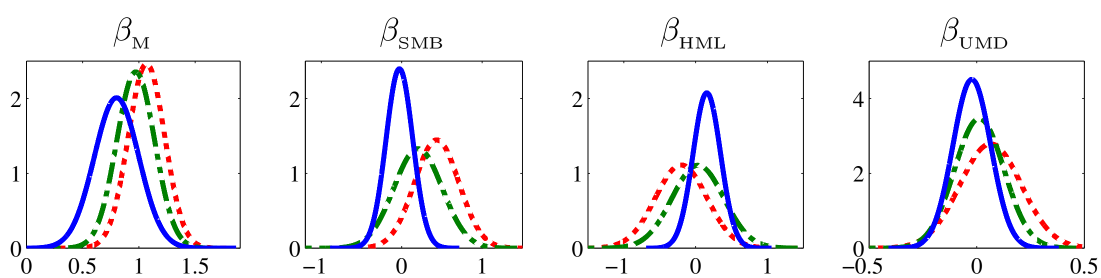
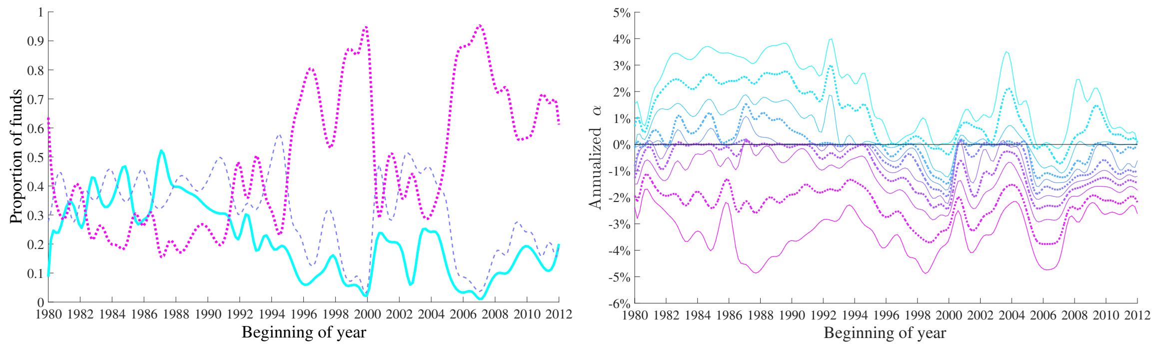


Figure 8: Estimated population densities for the factor loadings, estimated separately for Growth & Income funds (plotted in solid blue lines), Growth funds (plotted in dash-dotted green lines), and Aggressive Growth funds (plotted in dashed red lines). All densities are estimated using the $K^- = 2$, $K^+ = 1$ model with two negative and one positive component for the distribution of alpha, with returns net of expenses, and using data on 1,865 funds for which we have investment objective information from the Thomson database.



(a) Population proportions. (b) Percentiles of α distribution.

Figure 9: Evolution over time of the proportions of fund types and the percentiles of the estimated alpha distribution, with short-term skill. In Panel (a), we plot posterior means of the population proportions of zero-alpha funds (dashed blue line), negative-alpha funds (dotted violet line), and positive-alpha funds (solid cyan line). In Panel (b), we plot various percentiles (5^{th}, 10^{th}, 20^{th}, ..., 80^{th}, 90^{th}, 95^{th}) of the estimated distribution of annualized alpha (expressed as a percent); higher percentiles are plotted above lower percentiles and, for clarity, line style is alternated between solid and dotted. Estimation is performed at the beginning of each month from January 1980 to December 2011, using data from the preceding 60 months. All estimations use returns net of expenses in the $K^- = 2$, $K^+ = 1$ model with two negative and one positive component for the distribution of alpha.

Table 14: Proportions of Fund Types — With Gross Returns

Results on the posterior distributions of the population proportions of funds with zero, negative, and positive alpha, estimated using our baseline model from Section 2 with gross returns for 3,497 funds. The 95% HPDI is the smallest interval such that the posterior probability that a parameter lies in it is 0.95. NSE stands for autocorrelation-adjusted numerical standard errors for the posterior mean estimate of each parameter.

	Mean	**Median**	**Std.Dev.**	95% **HPDI**	**NSE**
π^0	0.22	0.20	0.14	[0.01 , 0.53]	0.00
π^-	0.34	0.34	0.10	[0.16 , 0.53]	0.00
π^+	0.44	0.44	0.08	[0.28 , 0.58]	0.00

Table 15: Percentiles of Estimated Skill Distribution — With Gross Returns

Percentiles of the distribution of annualized alpha (expressed as a percent) estimated using our baseline model with gross returns for 3,497 funds. We present point estimates and the 95% HPDI.

		Percentiles														
		0.5th	**1st**	**5th**	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**	**95th**	**99th**	**99.5th**
Posterior	Mean	-4.28	-3.07	-1.25	-0.76	-0.35	-0.12	0.00	0.00	0.21	0.58	0.96	1.55	2.15	3.84	4.78
	2.5%	-5.76	-4.00	-1.53	-0.99	-0.58	-0.32	-0.15	-0.05	0.00	0.06	0.76	1.32	1.89	3.17	3.74
	97.5%	-2.92	-1.89	-0.94	-0.51	0.00	0.00	0.00	0.25	0.48	0.78	1.16	1.74	2.41	4.57	5.89

Table 16: Fund Flows and Fund Performance

Quintile analysis of the relation between fund flows and past and future fund performance. At the end of each non-overlapping 5-year period from 1975 to 2010, we sort funds into quintiles (Q1 through Q5) based on their flows in the subsequent year, and in Panel A we report the average performance and average flows of each flow quintile across all periods. At the beginning of each non-overlapping 5-year period from 1980 to 2010, we sort funds into quintiles (Q1 through Q5) based on their flows in the previous year, and in Panel B we report the average performance and average flows of each flow quintile across all periods. We measure fund performance using i) the posterior probability (expressed as a percent) of having a positive alpha, ii) the posterior probability (expressed as a percent) of having a negative alpha, and iii) the posterior mean of alpha (expressed as an annualized percent), all estimated from our model with 4 factors; we report results for each of these measures of performance in separate lines labeled accordingly. We also report the average fund flow (expressed as a percent of beginning-of-year total net asset value, per year) for funds in each quintile. In columns labeled 'Q5−Q1', we report the difference between the top and the bottom flow quintile. */**/*** indicate significance of this difference at the 10%/5%/1% levels.

Panel A: Flows and Past Performance

	Flow Quintiles					
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q5−Q1**
Positive-α probability (as a %)	16.88	19.04	20.08	23.69	26.89	10.01***
Negative-α probability (as a %)	54.64	51.26	48.29	44.99	42.62	−12.03***
α (as a %/year)	−0.82	−0.69	−0.50	−0.34	−0.19	0.63***
Flows (as a %/year)	−30.03	−14.15	−6.46	8.25	116.38	146.41***

Panel B: Flows and Future Performance

	Flow Quintiles					
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q5−Q1**
Positive-α probability (as a %)	22.76	21.25	22.48	24.16	21.51	−1.25***
Negative-α probability (as a %)	45.50	47.34	46.75	45.70	48.84	3.34***
α (as a %/year)	−0.48	−0.58	−0.54	−0.42	−0.60	−0.12***
Flows (as a %/year)	−28.04	−12.67	−2.34	17.59	178.43	206.47***

Table 17: Population Proportions and Percentiles of Estimated Alpha Distribution — Prior Sensitivity Analysis

Prior sensitivity analysis of estimated posterior means of population proportions of zero-, negative-, and positive-alpha funds, and of percentiles of the estimated alpha distribution (expressed as an annualized percent). Panel A shows estimations with $\underline{\pi} = 1$ and Panel B shows estimations with $\underline{\pi} = 4$, where $\pi \sim \mathcal{D}\left(\underline{\pi}\right)$ is the symmetric Dirichlet prior for the population proportions π. In each of these panels, we vary the maximum number of negative and positive mixture components in the skill distribution, $K^{-}_{\max}$ and $K^{+}_{\max}$, from 4 to 6, the prior for the number of mixture components from $p(K) \propto \frac{1}{K!}$ to $p(K) = \frac{1}{K_{\max}}$, and the prior hyperparameters $\underline{K}_{\kappa_\alpha}$ from 10 to 100, $\underline{\Delta}_{K_\alpha}$ from 1 to 100, $\underline{\lambda}_{\Delta_\alpha}$ from 1 to 0.2, and $\underline{\Delta}_{\Delta_\alpha}$ from 1 to 0.02. In the leftmost column, we indicate how each prior specification differs from the baseline prior specification presented in Section A of the appendix. Panel C shows two estimations with the asymmetric prior $\pi \sim \mathcal{D}(1, 3, 3)$ for the population proportions: one uses the baseline values for all other prior parameters, and one uses alternative values ($\underline{\kappa}_{\kappa_\alpha} = -6$, $\underline{K}_{\kappa_\alpha} = 0.5$, and $\underline{\lambda}_{K_\alpha} = 4$) which effect a large change to the prior predictive density for α. All estimations use returns net of expenses for 3,497 funds.

	Proportions			Percentiles														
	π^0	π^-	π^+	**0.5th**	**1st**	**5th**	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**	**95th**	**99th**	**99.5th**
Panel A: Estimations with $\pi \sim \mathcal{D}(1)$																		
Baseline	0.09	0.78	0.13	-6.60	-5.01	-2.62	-1.96	-1.42	-1.12	-0.91	-0.72	-0.58	-0.36	0.00	0.30	0.87	2.35	3.24
large $K^{-}_{\max}$, $K^{+}_{\max}$, $p(K) = \frac{1}{K_{\max}}$	0.10	0.77	0.13	-6.68	-5.05	-2.62	-1.95	-1.41	-1.12	-0.91	-0.73	-0.56	-0.37	0.00	0.31	0.89	2.32	3.16
large $K^{-}_{\max}$, $K^{+}_{\max}$, $p(K) \propto \frac{1}{K!}$	0.07	0.80	0.13	-6.55	-5.03	-2.67	-1.98	-1.41	-1.11	-0.89	-0.71	-0.54	-0.37	0.00	0.30	0.87	2.39	3.30
$p(K) = \frac{1}{K_{\max}}$	0.09	0.78	0.13	-6.61	-5.03	-2.62	-1.96	-1.41	-1.12	-0.91	-0.73	-0.56	-0.36	0.00	0.32	0.88	2.33	3.21
small $\underline{K}_{\kappa_\alpha}$	0.09	0.79	0.12	-6.50	-4.97	-2.63	-1.97	-1.42	-1.12	-0.90	-0.72	-0.54	-0.35	0.00	0.26	0.84	2.36	3.27
large $\underline{\Delta}_{K_\alpha}$	0.09	0.77	0.14	-6.52	-4.87	-2.56	-1.95	-1.44	-1.16	-0.94	-0.75	-0.56	-0.33	0.00	0.29	0.80	2.35	3.27
small $\underline{\lambda}_{\Delta_\alpha}$	0.08	0.78	0.14	-6.48	-5.01	-2.67	-1.98	-1.42	-1.11	-0.90	-0.71	-0.54	-0.36	0.00	0.30	0.85	2.36	3.27
small $\underline{\Delta}_{\Delta_\alpha}$	0.07	0.80	0.13	-6.52	-5.03	-2.68	-1.99	-1.41	-1.10	-0.88	-0.70	-0.53	-0.36	0.00	0.27	0.82	2.42	3.38
Panel B: Estimations with $\pi \sim \mathcal{D}(4)$																		
Baseline	0.13	0.75	0.12	-6.41	-5.03	-2.71	-2.01	-1.43	-1.11	-0.88	-0.70	-0.52	-0.31	0.00	0.23	0.77	2.42	3.40
large $K^{-}_{\max}$, $K^{+}_{\max}$, $p(K) = \frac{1}{K_{\max}}$	0.12	0.76	0.12	-6.50	-5.03	-2.67	-1.98	-1.41	-1.10	-0.89	-0.70	-0.53	-0.34	0.00	0.24	0.81	2.43	3.40
large $K^{-}_{\max}$, $K^{+}_{\max}$, $p(K) \propto \frac{1}{K!}$	0.13	0.75	0.12	-6.44	-5.04	-2.71	-2.01	-1.43	-1.11	-0.88	-0.70	-0.52	-0.32	0.00	0.22	0.77	2.41	3.39
$p(K) = \frac{1}{K_{\max}}$	0.13	0.75	0.12	-6.50	-5.01	-2.67	-1.99	-1.43	-1.12	-0.90	-0.71	-0.53	-0.32	0.00	0.23	0.78	2.43	3.42
small $\underline{K}_{\kappa_\alpha}$	0.13	0.75	0.12	-6.42	-5.03	-2.72	-2.02	-1.43	-1.11	-0.88	-0.70	-0.52	-0.31	0.00	0.21	0.77	2.43	3.41
large $\underline{\Delta}_{K_\alpha}$	0.13	0.74	0.13	-6.51	-5.03	-2.66	-2.00	-1.45	-1.15	-0.92	-0.72	-0.52	-0.26	0.00	0.22	0.71	2.32	3.30
small $\underline{\lambda}_{\Delta_\alpha}$	0.13	0.75	0.12	-6.43	-5.04	-2.72	-2.02	-1.43	-1.11	-0.88	-0.70	-0.52	-0.31	0.00	0.22	0.77	2.40	3.37
small $\underline{\Delta}_{\Delta_\alpha}$	0.12	0.76	0.12	-6.52	-5.09	-2.73	-2.02	-1.42	-1.10	-0.87	-0.69	-0.51	-0.32	0.00	0.21	0.74	2.38	3.37
Panel C: Estimations with $\pi \sim \mathcal{D}(1, 3, 3)$																		
Baseline	0.07	0.80	0.13	-6.47	-5.03	-2.69	-1.99	-1.41	-1.10	-0.88	-0.70	-0.53	-0.36	0.00	0.29	0.82	2.40	3.37
Alternative	0.08	0.80	0.12	-6.55	-5.05	-2.67	-1.98	-1.41	-1.10	-0.89	-0.71	-0.54	-0.37	0.00	0.29	0.92	2.39	3.26

Appendix

A Prior specification

Here, we present the priors for the population parameters, whose joint prior density is[35]

$$
\begin{aligned}
& p\left(K^{-}\right)\times p\left(K^{+}\right)\times p\left(\pi\right)\times\left\{\prod_{i=1}^{I} p\left(e_i \,|\, \pi\right)\right\}\times\left\{\prod_{i=1}^{I} p\left(h_i \,|\kappa_h,\lambda_h\right)\right\}\times p\left(\kappa_h,\lambda_h\right) \\
& \times\left\{\prod_{i=1}^{I} p\Big(\left(\alpha_i,\beta_i\right)\Big| e_i,\left\{\left(\mu^q_{\alpha,k},V^q_{\alpha,k}\right)\right\},\left(\mu_\beta,V_\beta\right)\Big)\right\}\times\left\{\prod_{q\in\{-,+\}}\left[\prod_{1\le k\le K^q} p\left(\mu^q_{\alpha,k},V^q_{\alpha,k}\right)\right]\right\}\times p\left(\mu_\beta,V_\beta\right). \qquad \text{(A.1)}
\end{aligned}
$$

We have defined $p((\alpha_i,\beta_i)|\cdot)$, $p(h_i|\cdot)$ in Section 2.2, and it follows from the definition of e_i as a group allocation vector that it has the multinomial $M(1,\pi)$ distribution.

For the population proportions π, we use the symmetric Dirichlet prior $\pi\sim\mathcal{D}(1)$, which implies an uninformative uniform distribution over all possible values of π.

For the population mean and variance of the factor loadings and the nonzero components of the alpha distribution, we use the independent Normal-inverse-Wishart distribution, i.e.,

$$
\begin{aligned}
\mu^q_{\alpha,k}\left|\underline{\kappa}_\alpha,\underline{K}_\alpha\right. &\sim \mathcal{N}\left(\underline{\kappa}_\alpha,\underline{K}_\alpha\right) \qquad & \left(V^q_{\alpha,k}\right)^{-1}\left|\underline{\lambda}_\alpha,\underline{\Lambda}_\alpha\right. &\sim \mathcal{W}\left(\underline{\lambda}_\alpha,\underline{\Lambda}_\alpha^{-1}\right) \\
\mu_\beta\left|\underline{\kappa}_\beta,\underline{K}_\beta\right. &\sim \mathcal{N}\left(\underline{\kappa}_\beta,\underline{K}_\beta\right) \qquad & V_\beta^{-1}\left|\underline{\lambda}_\beta,\underline{\Lambda}_\beta\right. &\sim \mathcal{W}\left(\underline{\lambda}_\beta,\underline{\Lambda}_\beta^{-1}\right),
\end{aligned}
$$

where $\underline{\kappa}_\alpha$, $\underline{K}_\alpha$, $\underline{\lambda}_\alpha$, $\underline{\Lambda}_\alpha$ and $\underline{\kappa}_\beta$, $\underline{K}_\beta$, $\underline{\lambda}_\beta$, $\underline{\Lambda}_\beta$ are hyperparameters. We set $\underline{\lambda}_\alpha=1$ and, to gain flexibility, for $\phi:=\left(\underline{\kappa}_\alpha,\underline{K}_\alpha,\underline{\Lambda}_\alpha\right)$ we introduce an additional hierarchical layer such that ϕ is a random parameter to be estimated from the data and depends on deeper hyperparameters.[36] In particular, we use $\underline{\kappa}_\alpha\left|\underline{\kappa}_{\underline{\kappa}_\alpha},\underline{K}_{\underline{\kappa}_\alpha}\right.\sim\mathcal{N}\Big(\underline{\kappa}_{\underline{\kappa}_\alpha},\underline{K}_{\underline{\kappa}_\alpha}\Big)$, $\underline{K}_\alpha^{-1}\left|\underline{\lambda}_{\underline{K}_\alpha},\underline{\Lambda}_{\underline{K}_\alpha}\right.\sim\mathcal{W}\Big(\underline{\lambda}_{\underline{K}_\alpha},\underline{\Lambda}^{-1}_{\underline{K}_\alpha}\Big)$, and $\underline{\Lambda}_\alpha\left|\underline{\lambda}_{\underline{\Lambda}_\alpha},\underline{\Lambda}_{\underline{\Lambda}_\alpha}\right.\sim\mathcal{W}\Big(\underline{\lambda}_{\underline{\Lambda}_\alpha},\underline{\Lambda}^{-1}_{\underline{\Lambda}_\alpha}\Big)$, where we set $\underline{\kappa}_{\underline{\kappa}_\alpha}=0$, $\underline{K}_{\underline{\kappa}_\alpha}=100$, and $\underline{\lambda}_{\underline{K}_\alpha}=\underline{\Lambda}_{\underline{K}_\alpha}=\underline{\lambda}_{\underline{\Lambda}_\alpha}=\underline{\Lambda}_{\underline{\Lambda}_\alpha}=1$ to make the priors weak. For the hyperparameters for the mean and variance of the factor loadings we use fixed values because, with only one pair $\left(\mu_\beta,V_\beta\right)$, we cannot estimate them from the data. In particular, we set $\underline{\kappa}_\beta=0$, $\underline{K}_\beta=100$, $\underline{\lambda}_\beta=\text{rank}\left(V_\beta\right)$, and $\underline{\Lambda}_\beta=I$.

For K^- and K^+, a common choice (e.g., Stephens, 2000) is the Poisson $\mathcal{P}(1)$ such that $p(K)\propto\frac{1}{K!}$, which penalizes mixtures with more components. For convenience of computation and presentation, this distribution is often truncated; we truncate to $K^-_{\max}=K^+_{\max}=4$.

---

[35] Note that we impose some natural conditional independencies which simplify the density, e.g., $p\left(e_i\left|K^-,K^+,\pi,\left\{\left(\mu^q_{\alpha,k},V^q_{\alpha,k}\right)\right\},\left(\mu_\beta,V_\beta\right),\kappa_h,\lambda_h,\{\alpha_i,\beta_i,h_i\}\right.\right)=p\left(e_i\,|\,\pi\right)$.

[36] We fix $\underline{\lambda}_\alpha$ to avoid over-complicating the analysis, but it could also be made random.

For the parameters of the Gamma distribution of the error precisions h_i, we use the Miller (1980) prior

$$p\left(\kappa_h, \lambda_h \,\middle|\, \underline{p}, \underline{q}, \underline{r}, \underline{s}\right) \propto \frac{\underline{p}^{\kappa_h - 1} e^{-\frac{\underline{q}}{\lambda_h}}}{\left[\Gamma\left(\kappa_h\right)\right]^{\underline{r}} \lambda_h^{\kappa_h \underline{s} + 2}},$$

with Γ the gamma function and $\underline{p}$, $\underline{q}$, $\underline{r}$, $\underline{s}$ the hyperparameters. We set these to correspond to a 'prior sample' of 3 observations, one each at the low, middle, and high end of the distribution of the $\hat{h}_i$ estimated from fund-level OLS regressions of Equation 1; in particular, we use $\underline{p} = 2 \times 10^9$, $\underline{q} = 10^5$, $\underline{r} = \underline{s} = 3$.[37]

B Estimation algorithm

Here, we describe the MCMC algorithm used to estimate the model in Section 2. Before we proceed, we define parameters $\rho_i := \left(\alpha_i, \beta_i'\right)'$ and regressors $x_t := \left(1, F_t'\right)$, and vertically stack for each i all the r_{it}, x_t, ε_{it} into r_i, X_i, ε_i respectively. Then we vertically stack r_i, ρ_i, ε_i, and create a block diagonal matrix from the X_i so Equation 1 becomes $r = X\rho + \varepsilon$.

We are interested in making draws from the joint posterior distribution

$$p\left(K^-, K^+, \pi, \{e_i\}, \phi, \left\{\theta^q_{\rho,k}\right\}, \{\rho_i\}, \kappa_h, \lambda_h, \{h_i\} \middle| \, r, X\right),$$

which is proportional to the likelihood $p(r \,|X, \{\rho_i\}, \{h_i\})$ times the joint prior in Equation A.1 (augmented with ϕ, as explained in Section A). We use an MCMC algorithm to construct a Markov chain whose stationary distribution converges to the posterior. For a model with *known* numbers of negative and positive components K^-, K^+, we use a Gibbs sampler to make draws. For a model with *unknown* numbers of components, we use the Reversible Jump MCMC methodology to explore the model space. In the following sections, we first describe the sampler that draws from the posterior in the former case, and then we present the additional steps necessary to explore the model space in the latter case.

B.1 Without model specification uncertainty

In the first iteration of the sampler, we initialize all parameters; starting values do not matter if the sampler explores the whole posterior (see Geweke, 1999). Repeating the estimation with a variety of

[37] In the paper, we also present results from a hierarchical normal model; here we briefly discuss the priors we use for this model. The priors for parameters that are common between the two models (i.e., β_i, $\left(\mu_\beta, V_\beta\right)$, h_i, κ_h, and λ_h) are the same in both cases, as presented above. In the hierarchical normal model, $\alpha_i \sim \mathcal{N}\left(\mu_\alpha, V_\alpha\right)$ and the priors for $\left(\mu_\alpha, V_\alpha\right)$ also take the Normal-inverse-Wishart form presented above, with the difference that there is no additional hierarchical layer for ϕ, since there is only one pair $\left(\mu_\alpha, V_\alpha\right)$. Instead, we set $\underline{\kappa}_\alpha = 0$, $\underline{K}_\alpha = 100$, $\underline{\lambda}_\alpha = 1$, and $\underline{\Lambda}_\alpha = 10^{-5}$. Setting $\underline{K}_\alpha$ large and $\underline{\lambda}_\alpha$ small makes the priors very weak and renders the choice of $\underline{\kappa}_\alpha$ immaterial and the choice of $\underline{\Lambda}_\alpha$ less influential. The value of $\underline{\Lambda}_\alpha$ is chosen to be of a similar order of magnitude as our a priori belief of V_α (corresponding to a standard deviation of 1% to 5% in annual terms); a sensitivity analysis that varies the value of $\underline{\Lambda}_\alpha$ by several orders of magnitude shows that the posterior is robust to this value.

starting values drawn from the prior, our results are always very similar.

Given the values of the parameters π, $\{e_i\}$, ϕ, $\left\{\theta^q_{\rho,k}\right\}$, $\{\rho_i\}$, κ_h, λ_h, $\{h_i\}$ in iteration m, in iteration $m+1$ the Gibbs sampler sequentially draws from the following conditional posteriors: $p\left(\pi \left| \{e_i\}^{(m)}\right.\right)$, $p\left(\{\rho_i\}, \{e_i\} \left| r, X, \{h_i\}^{(m)}, \pi^{(m+1)}, \left\{\theta^q_{\rho,k}\right\}^{(m)}\right.\right)$, $p\left(\left\{\theta^q_{\rho,k}\right\} \left| \{e_i\}^{(m+1)}, \{\rho_i\}^{(m+1)}, \phi^{(m)}\right.\right)$, $p\left(\phi \left| \left\{\theta^q_{\rho,k}\right\}^{(m+1)}\right.\right)$, $p\left(\{h_i\} \left| r, X, \{\rho_i\}^{(m+1)}, \kappa_h^{(m)}, \lambda_h^{(m)}\right.\right)$, and $p\left(\kappa_h, \lambda_h \left| \{h_i\}^{(m+1)}\right.\right)$.

These conditional posteriors follow from the assumptions in Sections 2.2, 2.3. In particular, suppressing iteration superscripts for convenience, we have the following:

- Letting $\overline{\pi} = \underline{\pi} \cdot \mathbf{1} + \sum_{i=1}^{I} e_i$, the conditional posterior for π has the Dirichlet distribution:

$$\pi \left| \{e_i\} \right. \sim D\left(\overline{\pi}\right).$$

- The conditional posterior for (ρ_i, e_i) is proportional to the product of the likelihood, which follows from Equation 1 and the distribution of the error, and the joint prior density of ρ_i and e_i which follows from the assumptions in Section 2.2, i.e.,[38]

$$p\left(\rho_i, e_i \left| r_i, X_i, h_i, \pi, \left\{\theta^q_{\rho,k}\right\}\right.\right) \propto p\left(r_i \left| X_i, \rho_i, h_i\right.\right) \cdot p\left(\rho_i \left| e_i, \left\{\theta^q_{\rho,k}\right\}\right.\right) \cdot p\left(e_i \left| \pi\right.\right).$$

  This does not have a convenient form, so we use the Metropolis-Hastings algorithm to make draws. That is, we use a random-walk candidate generating density q (see Chib and Greenberg, 1995) to generate a candidate $\left(\rho_i', e_i'\right)$ (taking care that, e.g., $\alpha_i' < 0$ if fund i is allocated to the negative-alpha funds), which is then accepted or rejected probabilistically so that the Markov chain generally moves toward areas of the parameter space with high posterior. The probability of accepting the candidate draw $\left(\rho_i', e_i'\right)$ is

$$\min\left\{\frac{p\left(\rho_i', e_i' \left| r_i, X_i, h_i, \pi, \left\{\theta^q_{\rho,k}\right\}\right.\right) q\left(\rho_i, e_i\right)}{p\left(\rho_i, e_i \left| r_i, X_i, h_i, \pi, \left\{\theta^q_{\rho,k}\right\}\right.\right) q\left(\rho_i', e_i'\right)}, 1\right\}.$$

- With independent Normal-inverse-Wishart priors, the conditional posterior for $\theta^q_{\rho,k}$ is

$$\mu^q_{\rho,k} \left| \{e_i\}, \{\rho_i\}, V^q_{\rho,k}, \phi \right. \sim \mathcal{N}\left(\overline{\kappa}^q_{\rho,k}, \overline{K}^q_{\rho,k}\right) \qquad \left(V^q_{\rho,k}\right)^{-1} \left| \{e_i\}, \{\rho_i\}, \mu^q_{\rho,k}, \phi \right. \sim \mathcal{W}\left(\overline{\lambda}^q_{\rho,k}, \left(\overline{\Lambda}^q_{\rho,k}\right)^{-1}\right)$$

  where, letting $\tilde{\rho}_i := \left(\ln|\alpha_i|, \beta_i'\right)'$, the posterior hyperparameters are

$$\overline{K}^q_{\rho,k} = \left(\underline{K}_\rho^{-1} + \left(V^q_{\rho,k}\right)^{-1}\left(\sum_{i=1}^{I} e^q_{i,k}\right)\right)^{-1} \qquad \overline{\Lambda}^q_{\rho,k} = \underline{\Lambda}_\rho + \sum_{i=1}^{I} e^q_{i,k}\left(\tilde{\rho}_i - \mu^q_{\rho,k}\right)\left(\tilde{\rho}_i - \mu^q_{\rho,k}\right)'$$

$$\overline{\kappa}^q_{\rho,k} = \overline{K}^q_{\rho,k}\left(\underline{K}_\rho^{-1}\underline{\kappa}_\rho + \left(V^q_{\rho,k}\right)^{-1}\sum_{i=1}^{I}\left(e^q_{i,k}\tilde{\rho}_i\right)\right) \qquad \overline{\lambda}^q_{\rho,k} = \underline{\lambda}_\rho + \sum_{i=1}^{I} e^q_{i,k},$$

[38] We present the algorithm for the more general case in which α_i and β_i are drawn jointly as ρ_i.

with $\underline{\kappa}_\rho, \underline{K}_\rho, \underline{\lambda}_\rho, \underline{\Lambda}_\rho$ the prior parameters and $\phi := \left(\underline{\kappa}_\rho, \underline{K}_\rho, \underline{\Lambda}_\rho\right)$ random and drawn next.

- With $p(\phi) = p\left(\underline{\kappa}_\rho\right) p\left(\underline{K}_\rho\right) p\left(\underline{\Lambda}_\rho\right)$, the conditional posterior for ϕ consists of:

$$\begin{aligned}
\underline{\kappa}_\rho \left| \left\{\mu^q_{\rho,k}\right\}, \underline{K}_\rho \right. &\sim \mathcal{N}\left(\overline{\kappa}_{\underline{\kappa}_\rho}, \overline{K}_{\underline{\kappa}_\rho}\right) \\
\underline{K}^{-1}_\rho \left| \left\{\mu^q_{\rho,k}\right\}, \underline{\kappa}_\rho \right. &\sim \mathcal{W}\left(\overline{\lambda}_{\underline{K}_\rho}, \overline{\Lambda}^{-1}_{\underline{K}_\rho}\right) \\
\underline{\Lambda}_\rho \left| \left\{V^q_{\rho,k}\right\} \right. &\sim \mathcal{W}\left(\overline{\lambda}_{\underline{\Lambda}_\rho}, \overline{\Lambda}^{-1}_{\underline{\Lambda}_\rho}\right),
\end{aligned}$$

where, with $\left(\underline{\kappa}_{\underline{\kappa}_\rho}, \underline{K}_{\underline{\kappa}_\rho}\right)$, $\left(\underline{\lambda}_{\underline{K}_\rho}, \underline{\Lambda}_{\underline{K}_\rho}\right)$, $\left(\underline{\lambda}_{\underline{\Lambda}_\rho}, \underline{\Lambda}_{\underline{\Lambda}_\rho}\right)$ the prior parameters for $\underline{\kappa}_\rho$, $\underline{K}_\rho$, $\underline{\Lambda}_\rho$:

$$\begin{aligned}
\overline{K}_{\underline{\kappa}_\rho} &= \left(\underline{K}^{-1}_{\underline{\kappa}_\rho} + \underline{K}^{-1}_\rho \sum_{q,k} 1\right)^{-1} & \overline{\kappa}_{\underline{\kappa}_\rho} &= \overline{K}_{\underline{\kappa}_\rho}\left(\underline{K}^{-1}_{\underline{\kappa}_\rho}\underline{\kappa}_{\underline{\kappa}_\rho} + \underline{K}^{-1}_\rho \sum_{q,k} \mu^q_{\rho,k}\right) \\
\overline{\lambda}_{\underline{K}_\rho} &= \underline{\lambda}_{\underline{K}_\rho} + \sum_{q,k} 1 & \overline{\Lambda}_{\underline{K}_\rho} &= \underline{\Lambda}_{\underline{K}_\rho} + \sum_{q,k}\left(\mu^q_{\rho,k} - \underline{\kappa}_\rho\right)\left(\mu^q_{\rho,k} - \underline{\kappa}_\rho\right)' \\
\overline{\lambda}_{\underline{\Lambda}_\rho} &= \underline{\lambda}_{\underline{\Lambda}_\rho} + \underline{\lambda}_\rho \sum_{q,k} 1 & \overline{\Lambda}_{\underline{\Lambda}_\rho} &= \underline{\Lambda}_{\underline{\Lambda}_\rho} + \sum_{q,k}\left[\left(V^q_{\rho,k}\right)^{-1}\right].
\end{aligned}$$

- Letting $\overline{\kappa}_{h,i} = \kappa_h + \frac{T_i}{2}$ and $\frac{1}{\overline{\lambda}_{h,i}} = \frac{1}{\lambda_h} + \frac{(r_i - X_i \rho_i)'(r_i - X_i \rho_i)}{2}$, the conditional posterior for h_i is

$$h_i \left| r_i, X_i, \rho_i, \kappa_h, \lambda_h \right. \sim \mathcal{G}\left(\overline{\kappa}_{h,i}, \overline{\lambda}_{h,i}\right).$$

- The conditional posterior for (κ_h, λ_h) is

$$p\left(\kappa_h, \lambda_h \left| \{h_i\} \right.\right) \propto \frac{\overline{p}^{\kappa_h - 1} e^{-\frac{\overline{q}}{\lambda_h}}}{\left[\Gamma(\kappa_h)\right]^{\overline{r}} \lambda_h^{\kappa_h \overline{s} + 2}},$$

where $\overline{p} = \underline{p} \prod_{i=1}^I h_i$, $\overline{q} = \underline{q} + \sum_{i=1}^I h_i$, $\overline{r} = \underline{r} + I$, and $\overline{s} = \underline{s} + I$, and $\underline{p}$, $\underline{q}$, $\underline{r}$, $\underline{s}$ are the prior parameters for κ_h, λ_h. We make draws using the acceptance-rejection algorithm.

We note that, in mixture models, the posterior is invariant to permutations of the components' labels, so inference is problematic for parameters that are not invariant to component relabeling in the MCMC draws. We circumvent this issue in two ways. First, we focus on inferences that *are* invariant to relabeling, e.g., on the population proportions π^0, π^-, π^+, and the density of alpha and the factor loadings. Second, to conduct inferences that are not invariant to relabeling, e.g., the distribution parameters $\left\{\left(\mu^q_{\alpha,k}, V^q_{\alpha,k}\right)\right\}$, we achieve a unique labeling by retrospectively relabeling components in the MCMC draws so the marginal posteriors of parameters of interest are close to unimodality (for a similar approach, see Stephens, 1997). Thus, we do not need to impose artificial identifiability restrictions through priors, which do not guarantee a unique labeling and may produce biased estimates (see Celeux, 1998).

B.2 With model specification uncertainty

In this case, our estimation adapts the Reversible Jump MCMC method (Green, 1995; Richardson and Green, 1997). We construct a Markov chain that lives on the space

$$\Theta^0 \times \bigcup_{K^-=1}^{K^-_{\max}} \bigcup_{K^+=1}^{K^+_{\max}} \left(\mathcal{M}_{K^-,K^+} \times \Theta^-_{K^-} \times \Theta^+_{K^+}\right),$$

with $\mathcal{M}_{K^-,K^+}$ the model with K^- negative and K^+ positive components, Θ^0, $\Theta^-_{K^-}$, $\Theta^+_{K^+}$ the ranges of $\vartheta^0 := \pi^0$, $\vartheta^-_{K^-} := \left\{\left(\pi^-_k, \theta^-_k\right)\right\}$, $\vartheta^+_{K^+} := \left\{\left(\pi^+_k, \theta^+_k\right)\right\}$, and $\theta^q_k := \left(\mu^q_{\alpha,k}, V^q_{\alpha,k}\right)$ for $q \in \{-,+\}$. It makes moves from a model, e.g., $\left(\mathcal{M}_{i,j}, \vartheta^0, \vartheta^-_i, \vartheta^+_j\right)$, to another, e.g., $\left(\mathcal{M}_{i',j'}, \vartheta^0, \vartheta^-_{i'}, \vartheta^+_{j'}\right)$, designed to achieve *detailed balanced* so it converges to the posterior.

To estimate this model, we add the following steps in each iteration of our algorithm:

1. With probability $b^-_{K^-,K^+}$, $b^+_{K^-,K^+}$ we propose to split in two a negative or positive component, respectively; with probability $d^-_{K^-,K^+}$,$d^+_{K^-,K^+}$ we propose merging two components.
2. With probability $b^-_{K^-,K^+}$,$b^+_{K^-,K^+}$ we propose the birth of a negative or positive component; with probability $d^-_{K^-,K^+}$,$d^+_{K^-,K^+}$ we propose the death of an empty component.

First, $b^-_{K^-,K^+} + b^+_{K^-,K^+} + d^-_{K^-,K^+} + d^+_{K^-,K^+} = 1$. Also, $d^-_{1,K^+} = b^-_{K^-_{\max},K^+} = d^+_{K^-,1} = b^+_{K^-,K^+_{\max}} = 0$, because merge/death and split/birth moves are not permitted when the number of components is minimal and maximal, respectively. Finally, we propose all other permitted moves with equal probability. We denote by $m_h(\mathcal{M})$ the probability of proposing a specific move of type $h \in \{b, d, m, s\}$. For example, the probability of merging two specific negative components is the probability $d^-_{K^-+1,K^+}$ of choosing to merge negative components times the probability of picking these out of $K^- + 1$ components, so $m_m\left(\mathcal{M}_{K^-+1,K^+}\right) = \frac{d^-_{K^-+1,K^+}}{\left(K^-+1\right)K^-}$. Similarly, $m_s\left(\mathcal{M}_{K^-,K^+}\right) = \frac{b^-_{K^-,K^+}}{K^-}$, $m_b\left(\mathcal{M}_{K^-,K^+}\right) = b^-_{K^-,K^+}$, $m_d\left(\mathcal{M}_{K^-+1,K^+}\right) = \frac{d^-_{K^-+1,K^+}}{K^-_0+1}$, where $K^-_0 + 1$ is the number of empty negative components in model $\mathcal{M}_{K^-+1,K^+}$.

For convenience, we let $\theta := \left(0, \left\{\theta^-_k\right\}, \left\{\theta^+_k\right\}\right)'$ and define for each fund i the categorical allocation variable $S_i \in \left\{1, \ldots, 1 + K^- + K^+\right\}$ which corresponds to e_i, i.e., $S_i = k \Leftrightarrow e_{i,k} = 1$; we collect all S_i in S.

We present the split/birth move for negative components starting from a model $\mathcal{M}_{K^-,K^+}$ with parameters $\left(\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\right)$ and allocations S, and the merge/death move for negative components starting from a model $\mathcal{M}_{K^-+1,K^+}$ with parameters $\left(\vartheta^0, \vartheta^-_{K^-+1}, \vartheta^+_{K^+}\right)$ and allocations $\hat{S}$.[39] Moves for positive components are formulated analogously.

[39]A ˆ denotes parameters in model $\mathcal{M}_{K^-+1,K^+}$ that are different from those in model $\mathcal{M}_{K^-,K^+}$, while it is omitted for parameters that only exist in model $\mathcal{M}_{K^-+1,K^+}$; e.g., we write $\hat{S}$ but $\vartheta^-_{K^-+1}$ rather than $\hat{\vartheta}^-_{K^-+1}$.

B.2.1 Split and merge moves

The split move involves the following:

- Choose a component k^* to split to two new components, which we label k_1 and k_2.
- Draw random vector u and calculate proposal $\vartheta^-_{K^-+1} = g_{K^-,K^-+1}\left(\vartheta^-_{K^-}, u\right)$, where g_{K^-,K^-+1} must ensure this move is reversible, i.e., $\left(\vartheta^-_{K^-}, u\right) = g^{-1}_{K^-,K^-+1}\left(\vartheta^-_{K^-+1}\right)$. Like Richardson and Green (1997), we draw $u = \left(u_1, u_2'\right)'$ from $q_{K^-,K^-+1}(u) := q_1(u_1)q_2(u_2)$ for some q_1, q_2, set $\pi^-_{k_1} = u_1\pi^-_{k^*}$, $\pi^-_{k_2} = 1-\pi^-_{k_1}$ and $\theta^-_{k_1} = g_1\left(\pi^-_{k^*}, u_1, \theta^-_{k^*}, u_2\right)$, $\theta^-_{k_2} = g_2\left(\pi^-_{k^*}, u_1, \theta^-_{k^*}, u_2\right)$ such that the first two moments are preserved.
- Reallocate funds in k^* to components k_1, k_2 using density $q\left(\hat{S}\left|S, \vartheta^0, \vartheta^-_{K^-+1}, \vartheta^+_{K^+}\right.\right)$.
- Accept the move with probability $\min\{1, A\}$, where A is the product of these ratios:[40]

$$\text{likelihood ratio} = \prod_i p\left(\alpha_i \left|\theta_{\hat{S}_i}\right.\right) / p\left(\alpha_i \left|\theta_{S_i}\right.\right)$$

$$\text{prior ratio} = \frac{p\left(\hat{S}\left|\vartheta^0, \vartheta^-_{K^-+1}, \vartheta^+_{K^+}, \mathcal{M}_{K^-+1,K^+}\right.\right) p\left(\vartheta^0, \vartheta^-_{K^-+1}, \vartheta^+_{K^+}\left|\mathcal{M}_{K^-+1,K^+}\right.\right) p\left(\mathcal{M}_{K^-+1,K^+}\right)}{p\left(S\left|\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}, \mathcal{M}_{K^-,K^+}\right.\right) p\left(\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\left|\mathcal{M}_{K^-,K^+}\right.\right) p\left(\mathcal{M}_{K^-,K^+}\right)}$$

$$\text{proposal ratio} = \frac{q\left(S\left|\hat{S}, \vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\right.\right) q_{K^-+1,K^-}\left(\vartheta^-_{K^-}\left|\vartheta^-_{K^-+1}\right.\right) m_m\left(\mathcal{M}_{K^-+1,K^+}\right)}{q\left(\hat{S}\left|S, \vartheta^0, \vartheta^-_{K^-+1}, \vartheta^+_{K^+}\right.\right) q_{K^-,K^-+1}\left(\vartheta^-_{K^-+1}\left|\vartheta^-_{K^-}\right.\right) m_s\left(\mathcal{M}_{K^-,K^+}\right)}.$$

The reverse merge move involves the following:

- Choose a pair k_1, k_2 of components to be combined to form the new component k^*.
- Calculate a proposal $\vartheta^-_{K^-}$ using $\left(\vartheta^-_{K^-}, u\right) = g^{-1}_{K^-,K^-+1}\left(\vartheta^-_{K^-+1}\right)$.
- Reallocate to component k^* funds belonging to k_1 or k_2.
- Accept the proposed move with probability $\min\left\{1, \frac{1}{A}\right\}$, where A is as defined above.

To calculate A, we determine all proposals. In a split move we reallocate fund i from k^* to, e.g., k_1 with $q\left(\hat{S}_i = k_1\left|S_i = k^*, \vartheta^0, \vartheta^-_{K^-+1}, \vartheta^+_{K^+}\right.\right) := \pi^-_{k_1} p\left(\alpha_i\left|\theta^-_{k_1}\right.\right) / \left[\pi^-_{k_1} p\left(\alpha_i\left|\theta^-_{k_1}\right.\right) + \pi^-_{k_2} p\left(\alpha_i\left|\theta^-_{k_2}\right.\right)\right]$, so

$$q\left(\hat{S}\left|S, \vartheta^0, \vartheta^-_{K^-+1}, \vartheta^+_{K^+}\right.\right) = \prod_{i:S_i=k^*} \frac{\pi_{\hat{S}_i} p\left(\alpha_i\left|\theta_{\hat{S}_i}\right.\right)}{\pi^-_{k_1} p\left(\alpha_i\left|\theta^-_{k_1}\right.\right) + \pi^-_{k_2} p\left(\alpha_i\left|\theta^-_{k_2}\right.\right)}.$$

In addition, using the Jacobian transformation, we calculate that

$$q_{K^-,K^-+1}\left(\vartheta^-_{K^-+1}\left|\vartheta^-_{K^-}\right.\right) = \frac{q_{K^-,K^-+1}(u)}{|\det(J)|} \frac{1}{(K^-+1)!},$$

[40] We write, e.g., $q_{K^-,K^-+1}\left(\vartheta^-_{K^-+1}\left|\vartheta^-_{K^-}\right.\right)$ instead of $q_{K^-,K^-+1}\left(\hat{\vartheta}^0, \vartheta^-_{K^-+1}, \hat{\vartheta}^+_{K^+}\left|\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\right.\right)$, because we only propose changes to negative components. Also, instead of $p\left(\alpha_i\left|e_i, \left\{\left(\mu^q_{\alpha,k}, V^q_{\alpha,k}\right)\right\}\right.\right)$, we simply use $p\left(\alpha_i\left|\theta_k\right.\right)$ for the density of α_i for fund i allocated to component k with distribution parameters θ_k.

with $J := \frac{\partial g_{K^-,K^-+1}\left(\vartheta^-_{K^-},u\right)}{\partial\left(\vartheta^-_{K^-},u\right)} = \frac{\partial\left(\pi^-_{k_1},\pi^-_{k_2},\theta^-_{k_1},\theta^-_{k_2}\right)}{\partial\left(\pi^-_{k^*},u_1,\theta^-_{k^*},u_2\right)}$ the Jacobian of the transformation $\left(\vartheta^-_{K^-}, u\right)$ to $\vartheta^-_{K^-+1}$. In a merge move, we choose S, $\theta^-_{k^*}$, $\pi^-_{k^*}$ deterministically, so the proposals are $q\left(S\middle|\hat{S}, \vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\right) = 1$ and $q_{K^-+1,K^-}\left(\vartheta^-_{K^-}\middle|\vartheta^-_{K^-+1}\right) = \frac{1}{K^-!}$. Note that in q_{K^-,K^-+1} and q_{K^-+1,K^-} we divide by $\left(K^-+1\right)!$ and $K^-!$ — the number of ways in which we can order the negative components — due to the exchangeability assumption for mixture components.

Using the pmf of the multinomial distribution for the (independent) allocations of funds to types, and the fact that the allocations change, but the probabilities do not, we have

$$\frac{p\left(\hat{S}\,\middle|\vartheta^0, \vartheta^-_{K^-+1}, \vartheta^+_{K^+}, \mathcal{M}_{K^-+1,K^+}\right)}{p\left(S\,\middle|\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}, \mathcal{M}_{K^-,K^+}\right)} = \prod_{i:S_i=k^*} \frac{\pi_{\hat{S}_i}}{\pi^-_{k^*}}.$$

Also, using the density of the Dirichlet distribution $\pi \sim \mathcal{D}\left(\underline{\pi}\right)$, we have

$$p\left(\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\,\middle|\mathcal{M}_{K^-,K^+}\right) = \underbrace{\frac{\pi_0^{\underline{\pi}-1} \prod\limits_{1\le k\le K^-} \left(\pi^-_k\right)^{\underline{\pi}-1} \prod\limits_{1\le l\le K^+} \left(\pi^+_l\right)^{\underline{\pi}-1}}{B\left(\underline{\pi}\cdot \mathbf{1}_{1+K^-+K^+}\right)}}_{p(\pi)} \prod_{1\le k\le K^-} p\left(\theta^-_k\middle|\phi\right) \prod_{1\le l\le K^+} p\left(\theta^+_l\middle|\phi\right),$$

with $\mathbf{1}_K$ the unit vector with K elements and $B\left(\cdot\right)$ the multivariate beta function. Thus,

$$\begin{aligned}\ln A &= \sum_{i:S_i=k^*} \ln\left\{\frac{\pi^-_{k_1} p\left(\alpha_i\middle|\theta^-_{k_1}\right) + \pi^-_{k_2} p\left(\alpha_i\middle|\theta^-_{k_2}\right)}{\pi^-_{k^*} p\left(\alpha_i\middle|\theta^-_{k^*}\right)}\right\} + \ln\frac{p\left(\mathcal{M}_{K^-+1,K^+}\right)}{p\left(\mathcal{M}_{K^-,K^+}\right)} + \ln\frac{d^-_{K^-+1,K^+}}{b^-_{K^-,K^+}} + \ln\frac{p\left(\theta^-_{k_1}\middle|\phi\right) p\left(\theta^-_{k_2}\middle|\phi\right)}{p\left(\theta^-_{k^*}\middle|\phi\right)} \\ &\quad -\ln B\left(\underline{\pi}, \left(1+K^-+K^+\right)\underline{\pi}\right) + \left(\underline{\pi}-1\right)\ln\left(\frac{\pi^-_{k_1}\pi^-_{k_2}}{\pi^-_{k^*}}\right) - \ln\left(q_1(u_1)\,q_2(u_2)\right) + \ln\left|\det\left(\frac{\partial\left(\pi^-_{k_1}, \pi^-_{k_2}, \theta^-_{k_1}, \theta^-_{k_2}\right)}{\partial\left(\pi^-_{k^*}, u_1, \theta^-_{k^*}, u_2\right)}\right)\right|.\end{aligned}$$

Last, we determine proposals q_1, q_2 for u_1, u_2 and g_{K^-,K^-+1}. We choose $u_1 \sim \mathcal{B}\left(2, 2\right)$, $u_{2,1} \sim \mathcal{B}\left(1, 1\right)$, $u_{2,2} \sim \mathcal{B}\left(1, 1\right)$, and the g_{K^-,K^-+1} (and implied g_1, g_2) from Richardson and Green (1997), which preserves the mean μ and variance σ^2 before and after the split. Thus, $\left|\det\left(J\right)\right| = \pi_{k^*}\left|\mu_{k_1} - \mu_{k_2}\right|\sigma^2_{k^*}\frac{1-u^2_{2,1}}{u_1(1-u_1)u_{2,1}}$. Inverting g_1, g_2, we can calculate u, π_{k^*}, μ_{k^*}, $\sigma^2_{k^*}$ given π_{k_1}, π_{k_2}, μ_{k_1}, μ_{k_2}, $\sigma^2_{k_1}$, $\sigma^2_{k_2}$.

B.2.2 Birth and death moves

In a birth move, we add a low-probability empty component as follows:

- Draw the probability $\pi^-_{k^*}$ of the new component k^* from a proposal density $q\left(\pi^-_{k^*}\middle|\mathcal{M}_{K^-,K^+}\right)$ that draws small values, e.g., $\pi^-_{k^*} \sim \mathcal{B}\left(1, 1 + K^- + K^+\right)$.
- Rescale existing components' probabilities using, e.g., $\hat{\pi}^+_k = \pi^+_k\left(1-\pi^-_{k^*}\right)$ for $k \neq k^*$.
- Draw the parameters $\theta^-_{k^*}$ of the new component from a proposal, e.g., the prior $p\left(\theta^-_{k^*}\middle|\phi\right)$.

In a death move, we remove an empty component — i.e., with no funds — if any:

- Randomly choose an empty component, labeled k^*, to remove.
- Rescale the probabilities of all other components so they sum to 1.

Birth and death moves form a reversible pair, so their acceptance probabilities are $\min\{1, A\}$ and $\min\left\{1, \frac{1}{A}\right\}$, respectively, where A is as defined above, except that in the proposal ratio we now write $q_{K^-+1,K^-}\left(\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+} \middle| \hat{\vartheta}^0, \vartheta^-_{K^-+1}, \hat{\vartheta}^+_{K^+}\right)$ and $q_{K^-,K^-+1}\left(\hat{\vartheta}^0, \vartheta^-_{K^-+1}, \hat{\vartheta}^+_{K^+} \middle| \vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\right)$ because we change all components, and m_d, m_b replace m_m, m_s. Next, we calculate A.

No allocations S_i change, so the likelihood ratio is 1. Also, working as above we find

$$\frac{p\left(\hat{S}\middle|\hat{\vartheta}^0,\vartheta^-_{K^-+1},\hat{\vartheta}^+_{K^+}, \mathcal{M}_{K^-+1,K^+}\right)}{p\left(S\middle|\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}, \mathcal{M}_{K^-,K^+}\right)} = \prod_{1\le i\le I} \frac{\pi_{\hat{S}_i}\left(1-\pi^-_{k^*}\right)}{\pi_{S_i}}$$

$$\frac{p\left(\hat{\vartheta}^0,\vartheta^-_{K^-+1},\hat{\vartheta}^+_{K^+}\middle| \mathcal{M}_{K^-+1,K^+}\right)}{p\left(\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\middle| \mathcal{M}_{K^-,K^+}\right)} = \frac{p\left(\hat{\pi}_0, \hat{\pi}^-_1,..., \hat{\pi}^-_{K^-}, \pi^-_{k^*}, \hat{\pi}^+_1,..., \hat{\pi}^+_{K^+}\right)\left[\prod_{1\le k\le K^-} p\left(\theta^-_k\middle|\phi\right)\right] p\left(\theta^-_{k^*}\middle|\phi\right)\left[\prod_{1\le l\le K^+} p\left(\theta^+_l\middle|\phi\right)\right]}{p\left(\pi_0, \pi^-_1,..., \pi^-_{K^-}, \pi^+_1,..., \pi^+_{K^+}\right)\left[\prod_{1\le k\le K^-} p\left(\theta^-_k\middle|\phi\right)\right]\left[\prod_{1\le l\le K^+} p\left(\theta^+_l\middle|\phi\right)\right]}.$$

Using these, together with $\hat{\pi}_0 = \pi_0\left(1-\pi^-_{k^*}\right)$, $\hat{\pi}^-_k = \pi^-_k\left(1-\pi^-_{k^*}\right)$ for $1 \le k \ne k^* \le K^-$, $\hat{\pi}^+_l = \pi^+_l\left(1-\pi^-_{k^*}\right)$ for $1 \le l \le K^+$, and the pmf of the multinomial, the prior ratio is

$$\left(1-\pi^-_{k^*}\right)^I p\left(\theta^-_{k^*}\middle|\phi\right) \frac{\left(\pi^-_{k^*}\right)^{\underline{\pi}-1}\left(1-\pi^-_{k^*}\right)^{\left(1+K^-+K^+\right)\underline{\pi}-\left(1+K^-+K^+\right)}}{B\left(\underline{\pi}, \left(1+K^-+K^+\right)\underline{\pi}\right)} \frac{p\left(\mathcal{M}_{K^-+1,K^+}\right)}{p\left(\mathcal{M}_{K^-,K^+}\right)}.$$

Last, we calculate the proposal ratio. Allocations do not change, so $q\left(S\middle|\hat{S}, \vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\right) = 1$ and $q\left(\hat{S}\middle|S, \hat{\vartheta}^0, \vartheta^-_{K^-+1}, \hat{\vartheta}^+_{K^+}\right) = 1$. Also, $q_{K^-+1,K^-}\left(\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\middle|\hat{\vartheta}^0, \vartheta^-_{K^-+1}, \hat{\vartheta}^+_{K^+}\right) = \frac{1}{K^-!K^+!}$ as there are $K^-!$ and $K^+!$ ways to order the negative and positive components, while

$$q_{K^-,K^-+1}\left(\hat{\vartheta}^0, \vartheta^-_{K^-+1}, \hat{\vartheta}^+_{K^+}\middle|\vartheta^0, \vartheta^-_{K^-}, \vartheta^+_{K^+}\right) = \frac{q\left(\pi^-_{k^*}, \theta^-_{k^*}\right)}{|\det(J)|}\frac{1}{(K^-+1)!K^+!},$$

where, as in the split move, we use the Jacobian because we propose $\left(\hat{\vartheta}^0, \vartheta^-_{K^-+1}, \hat{\vartheta}^+_{K^+}\right)$ indirectly and divide by $\left(K^-+1\right)!K^+!$ due to exchangeability. So the proposal ratio is

$$\frac{1}{q\left(\pi^-_{k^*}\middle|\mathcal{M}_{K^-,K^+}\right)\cdot p\left(\theta^-_{k^*}\middle|\phi\right)} \frac{\left(K^-+1\right) d^-_{K^-+1,K^+}}{\left(K^-_0+1\right) b^-_{K^-,K^+}}\left(1-\pi^-_{k^*}\right)^{1+K^-+K^+},$$

where we use $|\det(J)| = \left(1-\pi^-_{k^*}\right)^{1+K^-+K^+}$, $q\left(\pi^-_{k^*}, \theta^-_{k^*}\right) = q\left(\pi^-_{k^*}\middle|\mathcal{M}_{K^-,K^+}\right)\cdot p\left(\theta^-_{k^*}\middle|\phi\right)$. So

$$\begin{aligned}\ln A \quad = \quad & \ln\frac{p\left(\mathcal{M}_{K^-+1,K^+}\right)}{p\left(\mathcal{M}_{K^-,K^+}\right)} + \left(\underline{\pi}-1\right)\ln\pi^-_{k^*} + \left[I + \left(1+K^-+K^+\right)\underline{\pi}\right]\ln\left(1-\pi^-_{k^*}\right) \\ & -\ln q\left(\pi^-_{k^*}\middle|\mathcal{M}_{K^-,K^+}\right) - \ln B\left(\underline{\pi}, \left(1+K^-+K^+\right)\underline{\pi}\right) + \ln\frac{\left(K^-+1\right) d^-_{K^-+1,K^+}}{\left(K^-_0+1\right) b^-_{K^-,K^+}}.\end{aligned}$$