

NBER WORKING PAPER SERIES

RETHINKING PERFORMANCE EVALUATION

Campbell R. Harvey
Yan Liu

Working Paper 22134
<http://www.nber.org/papers/w22134>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2016

A discussion with Neil Shephard provided the genesis for this paper - we are grateful. We appreciate the comments of Yong Chen, Wayne Ferson, Juhani Linnainmaa, David Ng, Lubos Pastor, and Luke Taylor. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Campbell R. Harvey and Yan Liu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Rethinking Performance Evaluation
Campbell R. Harvey and Yan Liu
NBER Working Paper No. 22134
March 2016
JEL No. G10,G11,G12,G14,G23

ABSTRACT

We show that the standard equation-by-equation OLS used in performance evaluation ignores information in the alpha population and leads to severely biased estimates for the alpha population. We propose a new framework that treats fund alphas as random effects. Our framework allows us to make inference on the alpha population while controlling for various sources of estimation risk. At the individual fund level, our method pools information from the entire alpha distribution to make density forecast for the fund's alpha, offering a new way to think about performance evaluation. In simulations, we show that our method generates parameter estimates that universally dominate the OLS estimates, both at the population and at the individual fund level. While it is generally accepted that few if any mutual funds outperform, we find that the fraction of funds that generate positive alphas is accurately estimated at over 10%. An out-of-sample forecasting exercise also shows that our method generates superior alpha forecasts.

Campbell R. Harvey
Duke University
Fuqua School of Business
Durham, NC 27708-0120
and NBER
cam.harvey@duke.edu

Yan Liu
Texas A&M University,
College Station, TX 77843
y-liu@mays.tamu.edu

1 Introduction

In a method reaching back to Jensen (1969), most studies of performance evaluation run separate regressions to obtain the estimates for alphas and standard errors. By following this approach, each fund is treated as a distinct entity and has a fund specific alpha. This is analogous to the fixed effects model in panel regressions where a non-random intercept is assumed for each subject. We depart from the extant literature by proposing a “random effects” counterpart of the performance evaluation model (referred as the *random alpha model*). In particular, we assume that fund i 's alpha α_i is drawn independently from a common distribution.

There are many reasons for us to consider the random alpha model. First, the fund data that researchers use (particularly, hedge fund data) are likely to only cover a fraction of the entire population of funds. Therefore, with the usual caveats about sample selection in mind, it makes sense to make inference on this underlying population rather than just focusing on the available fund data. This is one of the situations where a random effects setup is preferred over a fixed effects setup in panel regression models.¹

Second, our random alpha model provides a structural approach to study the distribution of fund alphas. It not only provides estimates for the quantities that are economically important (e.g., the 5th percentile of alphas, the fraction of positive alphas), but also provides standard errors for these estimates by taking into account various sources of parameter uncertainty, in particular the uncertainty in the estimation of alphas.

Currently, there are three main approaches to performance evaluation that rely on return data alone, each having its own shortcomings. In the first method, fund-level OLS are run in the first stage and hypothesis tests are performed in the second stage. The regression t -statistics are obtained for each fund and used to test statistical significance. Adjustments are sometimes used for test multiplicity. Recent papers that follow this approach include Barras et al. (2010), Fama and French (2010), Ferson and Chen (2015), and Harvey and Liu (2015a).

There are several problems with this approach when it comes to making inference on the cross-sectional distribution of fund alphas. First, it does not allow us to extrapolate beyond the range of the t -statistics of the available data. For instance, while the observed best performer might have a t -statistic of 3.0, we do not know the fraction of funds that have a t -statistic exceeding 3.0 in the population. Second, neither single tests nor multiple tests are useful when we try to make statements about the properties of the population of alphas. For instance, one question that is

¹See, for example, Maddala (2001) and Greene (2003). Searle, Casella, and McCulloch (1992) explore the distinction between a fixed effects model and a random effects model in more details.

economically important is: what is the fraction of investment funds that generate a positive alpha? Under the hypothesis testing framework, one candidate answer is the fraction of funds that are tested to generate a significant and positive alpha. However, this answer is likely to be severely biased given the existence of many funds that generate a positive yet insignificant alpha. Indeed, these funds are likely to be classified as zero-alpha funds — funds that have an alpha that strictly equals zero under hypothesis testing. In essence, equation-by-equation hypothesis testing treats fund alphas as dichotomous variables and thus does not allow us to make inference on the cross-sectional distribution of fund alphas.

Our method allows us to estimate the underlying alpha distribution and make inference on quantities that depend on the alpha population. Meanwhile, it provides a density estimate for each fund’s alpha, making it possible to make inference on individual funds and allowing us to answer the question: did the fund outperform? We are also doing hypothesis testing at some level. Similar to the usual approach to performance evaluation, time-series uncertainty in the estimation of alphas plays a key role in our inference. However, in contrast to the standard approach, which treats each fund as a separate entity and uses the individual t -statistic of alpha to make inference, our method weights the fund specific time-series uncertainty relative to the cross-sectional uncertainty for the alpha population, allowing us to efficiently draw information from the entire alpha population to make inference on a particular fund.

The second approach to performance evaluation involves first running fund-level OLS and then trying to estimate the distribution of the fitted alphas. By doing this, it is possible to make inference on the alpha population. However, this approach fails to take into account the various sources of estimation uncertainty, rendering the inference problematic. For instance, Chen et al. (2015) try to model the cross-section of fund alphas. Since the alphas are obtained from the first stage OLS, their model cannot take into account the uncertainty in the estimation of the model parameters, in particular, the uncertainty in the estimation of alphas. Such uncertainty is important given the time-varying nature of fund returns and the fact that for some investment styles standard factor models are only able to explain a small fraction of fund return variance.²

The third approach applies Bayesian methods to learn from the alpha population. For example, Jones and Shanken (2005) impose a normal prior on the alpha for an

²Chen et al. (2015) use the standard errors of the estimated alphas to control for the estimation uncertainty of the alphas. However, these standard errors are also estimated quantities based on the first stage OLS model and therefore have estimation uncertainty. Moreover, the estimation risk for betas is also important and can materially change the estimates for alphas. Our structural estimation approach allows us to jointly estimate the alpha distribution and the regression parameters for each individual fund.

average fund and uses this to adjust for the performance of an individual fund.³ Conceptually, their approach is closely related to ours in that we also try to make inference on the alpha population. However, there are important differences. We build on the frequentist approach and do not need to impose a prior distribution on the alpha population. We also allow fund alphas to be drawn from several subpopulations, which builds on important insights in the recent literature on performance evaluation and significantly enriches the structure of the alpha population.⁴ We provide a detailed discussion of Bayesian methods and contrast them with our approach.

By using portfolio holdings data, Cohen, Coval, and Pastor (2005) provide an innovative approach that infers a manager’s skill from the skill of managers that have similar portfolio holdings. Intuitively, if two managers have the same path of holdings, their alpha estimates should be very close to each other. Cohen, Coval and Pastor weight the cross-section of historical alpha estimates by the current portfolio holdings to refine the alpha estimate of a particular fund. Their idea of learning from the cross-section of managers is similar to ours. There are several differences between their paper and ours. First, while their method learns through portfolio holdings, we learn about a particular fund’s skill by grouping funds with similar alpha estimates, after adjusting for the estimation uncertainty in the alpha estimation. Second, while current holdings are informative about future fund performance, a fund’s unconditional alpha estimate should depend on the entire history of holdings. Finally, our method relies on the return data alone and is applicable to hedge fund performance evaluation where we do not have holdings data for most funds.

Our approach relies on the construction of a joint likelihood function that depends on both the alphas and the betas. By finding the maximum-likelihood estimates (MLE) of the model parameters, we make inference on the alpha distribution, controlling for various sources of estimation uncertainty. We provide a unified framework to assess performance, factor model estimation, and parameter uncertainty.

Our empirical work begins with a simulation study that takes many realistic features of the mutual fund data into account. We show that our method generates parameter estimates that achieve both a low finite-sample bias and standard error, dominating those that are generated under OLS. The superior performance of our model applies to the alpha population as well as the individual funds. We also perform an out-of-sample exercise by estimating our model in-sample and forecasting the alphas of individual funds out-of-sample. We show that our method provides a substantial improvement over OLS with respect to forecasting accuracy.

Application of the random alpha model leads to a much different answer to: What proportion of mutual funds outperform? While the existing literature suggests few if

³Other papers that apply Bayesian methods to study fund performance include Baks, Metrick, and Wachter (2001), Pástor and Stambaugh (2002a,b), Stambaugh (2003), Avramov and Wermers (2005), Busse and Irvine (2005), and Kosowski, Naik, and Teo (2007).

⁴See Barras, Scaillet, and Wermers (2010), Ferson and Chen (2015), and Chen, Cliff, and Zhao (2015).

any funds are deemed to outperform, our results suggest that over 10% of funds in the 1983-2011 time period generate positive risk-adjusted performance.

In the usual fund by fund regressions, about 29% of funds have positive alphas and 0-1% have positive significant alphas. Our alphas are different than the OLS alphas in that we have a structural model of the alpha distribution. Our method allows us to shrink positive alphas towards zero through the cross-sectional learning effect given that the median fund has a negative alpha. However, even after shrinking positive alphas towards zero, we still find that 10.6% of funds generate positive alphas. This 10.6% is accurately estimated in our framework — its 95% confidence bound is [9.5%, 12.3%]. Notice that this 10.6% is a pure statement about the probability of drawing a positive alpha from the alpha population, which, thanks to our structural framework, can be estimated by the random alpha model. It does not apply to the individual significance of an alpha from the perspective of hypothesis testing. Overall, our higher proportion of funds with positive alphas is likely due to the fact that our structural approach is more powerful in identifying small magnitude alphas.

From a methodological perspective, we propose a new procedure to efficiently estimate our structural model. It builds on and extends the standard Expectation-Maximization algorithm that allows us to sequentially learn about fund alphas (which are treated as missing observations) and estimate model parameters. Our method is important in that it allows us to capture the heterogeneity in fund characteristics in the cross-section. While we focus on performance evaluation in the current paper, a contemporaneous paper (Harvey and Liu, 2016b) builds on the insight of our model and studies the predictability of alpha. We expect our technique to be useful in other applications as well.

Our paper is organized as follows. In the second section, we present our model. In the next section, we discuss the estimation method for our model and provide a simulation study. In the fourth section, we apply our framework to mutual funds to make inference on the distribution of fund alphas. Some concluding remarks are offered in the final section.

2 Model

2.1 The Likelihood Function

For ease of exposition, suppose we have a $T \times N$ balanced panel of fund returns, T denoting the number of monthly periods and N denoting the number of funds in the cross-section. Importantly, balanced data is not required in our framework. As we shall see later, both our model and its estimation can be easily adjusted for unbalanced panel data.

Suppose we are evaluating fund returns against a set of K benchmark factors. Fund excess returns are modeled as

$$r_{i,t} = \alpha_i + \sum_{j=1}^K \beta_{ij} f_{j,t} + \varepsilon_{i,t}, \quad i = 1, \dots, N; \quad j = 1, \dots, K; \quad t = 1, \dots, T, \quad (1)$$

where $r_{i,t}$ is the excess return for fund i in period t , α_i is the alpha, β_{ij} is fund i 's risk loading on the j -th factor $f_{j,t}$, and $\varepsilon_{i,t}$ is the residual.

To simplify the exposition, let us introduce some notation. Let $R_i = [r_{i,1}, r_{i,2}, \dots, r_{i,T}]'$ be the excess return time-series for fund i . The panel of excess returns can be expressed as $\mathcal{R} = [R_1, R_2, \dots, R_N]'$. Let $\beta_i = [\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,K}]'$ be the risk loadings for fund i . We collect the cross-section of risk loadings into the vector $\mathcal{B} = [\beta'_1, \beta'_2, \dots, \beta'_N]'$. Similarly, we collect the cross-section of alphas into the vector $\mathcal{A} = [\alpha_1, \alpha_2, \dots, \alpha_N]'$. Let the standard deviation for the residual returns of the i -th fund be σ_i . We collect the cross-section of residual standard deviations into the vector $\Sigma = [\sigma_1, \sigma_2, \dots, \sigma_N]'$. Finally, let θ be the parameter vector that describes the population distribution of the elements in \mathcal{A} .

Under the model assumptions, the likelihood function of the model is

$$f(\mathcal{R}|\theta, \mathcal{B}, \Sigma) = \int f(\mathcal{R}, \mathcal{A}|\theta, \mathcal{B}, \Sigma) d\mathcal{A} \quad (2)$$

$$= \int f(\mathcal{R}|\mathcal{A}, \mathcal{B}, \Sigma) f(\mathcal{A}|\theta) d\mathcal{A}, \quad (3)$$

where $f(\mathcal{R}, \mathcal{A}|\theta, \mathcal{B}, \Sigma)$ is the complete data likelihood function (that is, the joint likelihood of both returns \mathcal{R} and alphas \mathcal{A}), $f(\mathcal{R}|\mathcal{A}, \mathcal{B}, \Sigma)$ is the conditional likelihood of returns given the cross-section of alphas and model parameters, and $f(\mathcal{A}|\theta)$ is the conditional density of the cross-section of alphas given the parameters that govern the alpha distribution.

Notice that the likelihood function of the model does not depend on the cross-section of alphas (i.e., \mathcal{A}). This is because, in our model, \mathcal{A} is treated as missing data and needs to be integrated out of the complete likelihood function $f(\mathcal{R}, \mathcal{A}|\theta, \mathcal{B}, \Sigma)$. However, once we obtain the estimates of the model parameters, the conditional distribution of \mathcal{A} can be obtained through the Bayes' law:

$$f(\mathcal{A}|\mathcal{R}, \hat{\theta}, \hat{\mathcal{B}}, \hat{\Sigma}) \propto f(\mathcal{R}|\mathcal{A}, \hat{\mathcal{B}}, \hat{\Sigma}) f(\mathcal{A}|\hat{\theta}). \quad (4)$$

This enables to us to evaluate the performance of each individual fund. Our approach to making inference on individual funds is distinctively different from current methods. The two existing approaches, as mentioned previously, draw their inference based on either the time-series likelihood (i.e., $f(\mathcal{R}|\mathcal{A}, \mathcal{B}, \Sigma)$) as in Barras et al.

(2010), Fama and French (2010), and Ferson and Chen (2015), or the cross-sectional likelihood (i.e., $f(\mathcal{A}|\theta)$) as in Chen et al. (2015). Our method, as shown in (4), combines information from both types of likelihoods, leading to a more informative inference.

Assuming that the residuals (i.e., $\varepsilon_{i,t}$'s) are independent both across funds and across time, the likelihood function can be written as:

$$f(\mathcal{R}|\theta, \mathcal{B}, \Sigma) = \int \prod_{i=1}^N f(R_i|\alpha_i, \beta_i, \sigma_i) f(\alpha_i|\theta) d\mathcal{A}, \quad (5)$$

$$= \prod_{i=1}^N \int f(R_i|\alpha_i, \beta_i, \sigma_i) f(\alpha_i|\theta) d\alpha_i. \quad (6)$$

Our goal is to find the maximum-likelihood estimate (MLE) of θ , which is the focus of the paper, along with other auxiliary parameters (i.e., \mathcal{B} and Σ) that govern the return dynamics of each individual fund. To obtain an explicit expression for the likelihood function, we assume that the residuals are normally distributed.

Residual independence is not a key assumption for our model. When there is residual dependency, the model will be misspecified. The likelihood function becomes the quasi-likelihood function. Our QMLE still makes sense as the parameters governing the dependency structure are treated as auxiliary parameters with respect to the goal of our analysis. Despite the model misspecification, in theory, the QMLE is still consistent in that it gives asymptotically unbiased estimates. It will be less efficient compared to the MLE of a correctly specified model. In our simulation study, we consider residual dependency and quantify the loss in efficiency.

2.2 The Specification of the Alpha Distribution

What is a good specification for the alpha distribution, which we denote as Ψ ? First, the density of Ψ needs to be flexible enough to capture the true underlying distribution for alpha. For instance, from both a theoretical and an empirical standpoint, two groups of fund managers could exist, one group consisting of skilled managers, and the other consisting of unskilled managers. Alternatively, we could think of five groups of managers (i.e., top, skilled, neutral, unskilled, and bottom), similar to the five star evaluation system used by Morningstar. These concerns suggest that the density of Ψ should be able to display a multi-modal pattern, the density associated with each mode capturing the alpha distribution generated by a particular group of managers.⁵

⁵Our specification of Ψ makes it possible for the density to display a multi-modal pattern. However, under certain parameterizations, a unimodal pattern is also possible. Our model estimation will help us determine what pattern is most consistent with the data.

On the other hand, having a flexible distribution does not mean that the distribution should be complicated. In fact, the very principle of regularization in statistics is to have parsimonious models to avoid overfitting.⁶ Hence, without sacrificing too much flexibility, we would like a distribution that is simple and interpretable.

Driven by these concerns, we propose to model the alpha distribution by a Gaussian Mixture Distribution (GMD) — a weighted sum of Gaussian distributions — that is widely in science and medical research to model population heterogeneity. A one-component GMD is just a standard Gaussian distribution. The two-component GMD is a mixture of two Gaussian distributions and allows for considerable heterogeneity:

$$Y = (1 - I) \cdot Y_e + I \cdot Y_h,$$

where Y is the random variable that follows the GMD, and I , Y_e and Y_h are independent random variables.⁷ I is an indicator variable that takes a value of 0 or 1, and it is parameterized by π , which is the probability for it to equal 1 (i.e., $Pr(I = 1) = \pi$). Y_e and Y_h are normally distributed variables that are parameterized by (μ_e, σ_e^2) and (μ_h, σ_h^2) . To achieve model identification, we assume $\mu_e < \mu_h$.

In our context, the model has a simple interpretation. With probability $1 - \pi$, we draw a manager from the population of unskilled managers (that is, $I = 0$), who on average generate an alpha of μ_e ('e' = low alpha). With probability π , the manager is drawn from the population of skilled managers (that is, $I = 1$), who on average generate an alpha of μ_h ('h' = high alpha). The overall population of alpha is thus modeled as the mixture of the two normal distributions.

The two-component model can be easily generalized to multi-component models. For a general L -component GMD, we order the means of its component distributions in ascending order (i.e., $\mu_1 < \mu_2 < \dots < \mu_L$) and parameterize the probabilities of drawing from each component distribution as

$$\pi = (\pi_1, \pi_2, \dots, \pi_L)', \quad \sum_{l=1}^L \pi_l = 1.$$

With enough number of components in the model, the GMD is able to approximate every density with arbitrary accuracy, the fact of which partly explains its popularity. However, the model becomes more difficult to identify when the number of components gets large.⁸ Therefore, between two models that produce similar likelihood values, we prefer the parsimonious model. We rely on our simulation framework

⁶See, for example, Bickel and Li (2006).

⁷For applications of the Gaussian Mixture Distribution in finance, see Gray (1996) and Bekaert and Harvey (1995).

⁸See, for example, Figueiredo and Jain (2002) for a discussion on the identifiability problem for a GMD and a potential solution.

to perform formal hypothesis testing on the candidate models and to select the best model.⁹ The idea of using a mixture distribution to model the cross-section of fund alphas has also been explored by the recent literature on performance evaluation, e.g., Chen et al. (2015). However, we offer a new path that takes the various sources of estimation uncertainty into account.

2.3 The Identifiability and Interpretability of Ψ

The recent literature on investment fund performance evaluation attempts to group funds into different categories. For example, Barras, Scaillet and Wermers (2010) and Ferson and Chen (2015) assume that funds are drawn from a few subpopulations, with “good” and “bad” managers coming from distinct subpopulations. Our parameterization of Ψ also bears this simple interpretation of a multi-population structure for the alpha distribution. However, different from Barras, Scaillet and Wermers (2010) and Ferson and Chen (2015), our structural estimation approach allows to take the estimation risk into account when we classify funds into distinct performance groups. Our empirical results show that our approach makes a material difference in the classification outcome.

Alternatively, we can think of Ψ as a parametric density to approximate the distribution of the population of fund alphas. The GMD is a flexible and widely used parametric family to approximate unknown densities. As in most density estimation problems, we are facing a tradeoff between accuracy and overfitting. In our application, we pay special attention to the overfitting issue. In particular, we perform a simulation-based model selection procedure to choose a parsimonious model. This allows us to use the simplest structure — provided that it adequately models the alpha distribution — to summarize the alpha population. This also makes it easier to interpret the composition of the alpha population.

To think about the identification of Ψ in our model, we first focus on an extreme case. Suppose we have an infinitely long time-series for each fund so that there is no estimation uncertainty in alpha. In this case, our model will force Ψ to approximate the cross-section of “true” alphas. Suppose the left tail of the alpha distribution is very different from the right tail. This suggests that a single component GMD is probably not enough to capture the asymmetry in the two tails. A two-component GMD may be a better candidate. Intuitively, we can first fit a normal distribution for the alpha observations that fall below a certain threshold and another normal distribution for the alpha observations that fall above a certain threshold (these two

⁹Another benefit in using the GMD is that it reduces the computational burden for the estimation of our model. In particular, when the components in \mathcal{A} follow a GMD and the returns \mathcal{R} follow a normal distribution conditional on \mathcal{A} , we show in Appendix A that the conditional distribution of the components in \mathcal{A} given \mathcal{R} is also a GMD. This makes it easy for us to simulate from the conditional distribution of \mathcal{A} given \mathcal{R} , which is the key step for the implementation of the EM algorithm that we use to estimate our model.

thresholds are not necessarily equal). We then mix these two distributions in a way that the mixed distribution approximates the middle part of the alpha distribution well, that is, the alpha distribution that covers the non-extreme alphas.

In reality, we have a finite return time-series. This introduces estimation uncertainty in both the alphas and the other OLS parameters. As a result, instead of fitting the cross-section of “true” alphas, our method tries to fit the cross-section of the distributions of the alphas, each distribution corresponding to the estimation problem of the alpha of an individual fund and capturing estimation risk. However, our previous discussion on the identification of Ψ when “true” alphas are available is still valid. In particular, the parameters in Ψ are identified by capturing the departure of the alpha distribution from a single normal distribution, only that this time the alpha distribution is no longer the distribution of “true” alphas but a mixed distribution of the estimated distributions of the alphas.

More rigorously, the parameters in Ψ can be shown to be identified through high order moments of the alpha population. For example, for a two-component GMD, its five parameters can be estimated by matching the first five sample moments of the data with the corresponding moments of the model.¹⁰ Despite its intuitive appeal, the moments-based approach cannot weight different moments efficiently. Our likelihood-based approach is able to achieve estimation efficiency. In our simulation study, where we experiment with a two-component GMD, the model parameters seem to be well identified and accurately estimated.

2.4 Discussion

The usual hypothesis testing framework with respect to making inference on the population of fund alphas presents a number of challenges. While hypothesis testing may be useful when we want to test the significance of a single fund, we need to make adjustment for test multiplicity when the same test is performed on many funds.¹¹ Hypothesis testing is less useful when we try to make inference on the entire alpha population. This is because hypothesis testing, by testing against a common null hypothesis (e.g., alpha equals zero), essentially treats fund alphas as dichotomous variables while more realistically they should be continuous. Our random alpha model assumes that the true alpha is a continuous variable and provides density estimates that can be used to evaluate each individual fund (similar to hypothesis testing) as well as the alpha population.

Hypothesis testing also places too much weight on the statistical significance of individual alphas and overlooks their economic significance from a population per-

¹⁰See Cohen (1967) and Day (1969) for the derivation of a two-component GMD based on the method of moments approach.

¹¹For recent papers on investment fund performance evaluation that emphasize multiple hypothesis testing, see Barras et al. (2010), Fama and French (2010), and Ferson and Chen (2015).

spective. For example, suppose we have two funds that both have a t -statistic of 1.5. One has an alpha of 20% (per annum) and the other has an alpha of 2% (per annum). Should we treat them the same? We think not. The 20% alpha, albeit volatile, tells us more about the plausible realizations of alphas in the cross-section than the 2% alpha.¹² Following the standard hypothesis testing framework, we not only ignore the difference in magnitude between the two alphas, but we also classify both funds as zero-alpha funds, causing an unnecessary loss of information regarding the cross-sectional distribution of alphas.

Our critique of the usual hypothesis testing approach is consistent with the recent advances in statistics, and in particular in machine learning, that emphasize *regularization*.¹³ In general, regularization refers to the process of introducing additional information or constraints to achieve model simplification that often helps prevent model overfitting. In the context of our application, we have a complex dataset given the multidimensional nature of the cross-section of investment funds. The standard hypothesis testing approach, by treating each fund as a separate entity and running equation-by-equation (that is, fund-by-fund) OLS to obtain a separate t -statistic to summarize its performance, does not reduce the complexity of the dataset. In contrast, our framework imposes a parametric distribution on the cross-section of alphas and thereby substantially reduces the model complexity. It is unlikely to produce a cross-sectional fit that is as good as the equation-by-equation OLS. However, the better fit by the equation-by-equation estimation may reflect overfitting, which means that the estimated cross-sectional distribution of alphas may be a poor estimate of the future distribution. Our method seeks to avoid overfitting with the goal of getting the best forecast of the future distribution.

At the core of our method is the idea of extracting information from the cross-section of funds. This information can be used both to make inference on the alpha population and to refine our inference on a particular fund. To motivate the idea, we use two examples throughout our paper. The first example is what we call a *one-cluster* example. Suppose all the funds in the cross-section generate an alpha of approximately 2% per annum and the standard error for the alpha estimate is about 4%. Since the t -statistics are all approximately 0.5 ($=2\%/4\%$), which is not even high enough to surpass the single test t -statistic cutoff of 2.0, let alone the multiple testing adjusted cutoffs, we would declare all the funds to be zero-alpha funds. Using our method, the estimate of the mean of the alpha population would be around 2%. In this case, we think our approach provides a better description of the alpha population than the usual hypothesis testing approach. Declaring all the funds to be zero-alpha funds ignores information in the cross-section.

While the one-cluster example illustrates the basic mechanism of our approach, it is too special. Indeed, a simple regression that groups all the funds into an index and

¹²While some investment funds can use leverage to amplify gains and losses, they also face leverage constraints. Therefore, 20% tells us more about the tails of the alpha distribution than 2%.

¹³For recent survey studies on regularization, see Fan and Lv (2010) and Vidaurre, Bielza, and Larrañaga (2013).

tests the alpha of the fund index will also generate a positive and significant estimate for the mean of the alpha population. This motivates the second example, which we call the *two-cluster* example. For the two-cluster example, suppose we have half of the funds having an alpha estimate of approximately 2% per annum and the standard error for the alpha estimate is about 4%. The other half have an alpha estimate of approximately -2% per annum and also have a standard error of about 4%. Similar to the one-cluster example, no fund is statistically significant individually. However, we throw information away if we declare all the funds to be zero-alpha funds. Different from the one-cluster example, if we group all the funds into an index and estimate the alpha for the index fund, we will have an alpha estimate close to zero. In this case, the index regression approach does not work for this example as it fails to recognize the two-cluster structure of the cross-section of fund alphas. Our approach allows us to take this cluster structure into account and make better inference on the alpha population.

The one-cluster and two-cluster examples are special cases of the alpha distributions that our framework can take into account. They correspond to essentially a point mass distribution at 2% and a discrete distribution that has a mass of 0.5 at -0.2% and 0.5 at 0.2%, respectively. Our general framework uses the GMD to model the alpha distribution and seeks to find the best fitting GMD under a penalty for model parsimony. It therefore extracts information from the entire cross-section of alphas.

After we estimate the distribution for the cross-section of alphas, we can use this distribution to refine the estimate of each individual fund's alpha. For instance, for the one-cluster example, knowing that most alphas cluster around 2.0% will pull our estimate of an individual fund's alpha towards 2.0% and away from zero. Similarly, for the two-cluster example, knowing that the alphas cluster at -2.0% and 2.0% with equal probabilities will pull our estimate of a negative alpha towards -2.0% and a positive alpha towards 2.0%, and both away from zero. In our general framework, after we identify the GMD that models the alpha cross-section, we use it to update the density estimate of each fund's alpha, thereby using cross-sectional information to refine the alpha estimate of each individual fund.

We now discuss the details of our model. To see how our method takes estimation uncertainty into account, we focus on the likelihood function in (6) (that is, $\prod_{i=1}^N \int f(R_i|\alpha_i, \beta_i, \sigma_i) f(\alpha_i|\theta) d\alpha_i$). Suppose we already have an estimate of \mathcal{B} and Σ (e.g., the OLS estimate) and seek to find the estimate for θ . Notice that $f(R_i|\alpha_i, \beta_i, \sigma_i)$, the likelihood function of the returns of fund i , can be viewed as a probability density on α_i . In particular, under normality of the residuals, we have

$$f(R_i|\alpha_i, \beta_i, \sigma_i) \equiv w(\alpha_i) \propto \exp\left\{-\frac{[\alpha_i - \frac{\sum_{t=1}^T (r_{it} - \beta_i' f_t)]^2}{2\sigma_i^2/T}}\right\}, \quad (7)$$

where $f_t = [f_{1,t}, f_{2,t}, \dots, f_{K,t}]'$ is the vector of factor returns at time t . Viewing in this way, $\int f(R_i|\alpha_i, \beta_i, \sigma_i)f(\alpha_i|\theta)d\alpha_i = \int w(\alpha_i)f(\alpha_i|\theta)d\alpha_i$ is a weighted average of $f(\alpha_i|\theta)$, with the weights (i.e., $w(\alpha_i)$) given in (7).

When σ_i/\sqrt{T} is small, that is, when there is little uncertainty in the estimation of α_i , $w(\alpha_i)$ will be concentrated around its mean, i.e., $\frac{\sum_{t=1}^T(r_{it}-\beta_i'f_t)}{T}$. In fact, when $\sigma_i \rightarrow 0, i = 1, \dots, N$ and when \mathcal{B} and $\mathbf{\Sigma}$ are set at their OLS estimates, the likelihood function in (6) converges to $\prod_{i=1}^N f(\hat{\alpha}_i^{OLS}|\theta)$ — the likelihood function when the alphas are exactly set at their OLS estimates. Therefore, ignoring the time-series uncertainty in the estimation of the alphas, the likelihood function collapses to the likelihood function constructed under the traditional approach, that is, running equation-by-equation OLS first and then estimating the distribution for the fitted alphas. This is the approach taken by Chen et al. (2015). Our approach, by using a weighting function $w(\alpha_i)$ that depends on σ_i/\sqrt{T} , allows us to take the time-series uncertainty in the estimation of the alpha into account.

Moreover, the weighting function $w(\alpha_i)$ is fund specific, that is, $w(\alpha_i)$ depends on the particular level of estimation uncertainty for α_i (i.e., σ_i/\sqrt{T}). Therefore, the likelihood function in (6) allows different weighting functions for different funds. This is important given the cross-sectional heterogeneity in estimation uncertainty across funds, in particular across investment styles.

Our approach offers more than just taking the estimation uncertainty for α_i (i.e., σ_i/\sqrt{T}) into account. As it shall become clear in later sections, our estimates of both α_i and σ_i^2 not only rely on fund i 's time series, but also use information from the cross-sectional distribution of the alphas. Hence, in our framework, the OLS t -statistic is not an appropriate metric to summarize the significance of fund alphas. Both its numerator and denominator need to adjust for the information in the alpha population.

On the other hand, our knowledge about the alpha population helps refine our estimates of the risk loadings and the residual variances. Suppose we already have an estimate of θ and seek to estimate \mathcal{B} and $\mathbf{\Sigma}$. We again focus on the likelihood function $\int f(R_i|\alpha_i, \beta_i, \sigma_i)f(\alpha_i|\theta)d\alpha_i$, but instead view $f(\alpha_i|\theta)$ as the weighting function. $f(\alpha_i|\theta)$ tells us how likely it is to observe a certain α_i from a population perspective. If α_i is unlikely to occur over a certain range, the likelihood function will downweigh this range relative to other ranges over which the occurrence of alpha is more plausible. In the extreme case when we have perfect knowledge about the alpha of a certain fund (say, $\hat{\alpha}_i^0$), the likelihood function becomes $f(R_i|\hat{\alpha}_i^0, \beta_i, \sigma_i)$, essentially the likelihood function for a linear regression model when the intercept is fixed. In general, the MLE of β_i and σ_i will be different from their unconstrained OLS estimates, reflecting our knowledge about the alpha population.

3 Estimation

3.1 A New Expectation-Maximization Framework

A direct maximization of (6) is difficult. The size of the parameter space is large and the likelihood function involves high-dimensional integrals. We offer a new implementation of the well-known Expectation-Maximization (EM) algorithm to facilitate the computation.

The idea of the EM algorithm is to treat the cross-section of alphas as missing observations and iteratively update our knowledge of the alpha distribution and the model parameters. With this approach, parameter estimates and learning about the missing observations can be done sequentially. In the context of our application, manager skill (i.e., alphas) are missing observations. In the “expectation” step of the EM algorithm, for a given set of parameter values,¹⁴ we fill in the missing observations with random draws from the conditional distribution of alphas given the parameter values. We calculate the averaged value of the likelihood function across these random draws. Essentially, at this step, we learn about manager skill to the best of our knowledge of the model parameters and update the likelihood function accordingly. In the “maximization” step of the algorithm, we maximize the updated likelihood function, which takes into account our recently updated information about manager skill. We obtain a new set of parameter estimates. These parameter estimates are subsequently fed into another “expectation” step to start a new round of estimation. The “expectation” step and the “maximization” step are performed iteratively to arrive at the MLE.

From a methodological perspective, our framework contributes to the literature on EM algorithm by allowing heterogeneous funds in the cross-section and simultaneously estimating fund specific parameters and other structural parameters.¹⁵ In particular, we allow both factor loadings and residual standard deviations to be fund

¹⁴In our model, parameter values refer to fund specific factor loadings, residual standard deviations, and parameters that govern the alpha population. The given set of parameter values could be the initial set of parameters to start the entire algorithm, for which a reasonable choice is the factor loadings and residual standard deviations from the equation-by-equation OLS estimates. It could also be the optimization outcome following the intermediate step (i.e., the “maximization” step) of the algorithm.

¹⁵See Dempster, Laird, and Rubin (1977) for the original paper that proposes the EM algorithm. See McLachlan and Krishnan (2007) for a more detailed discussion of the algorithm and its extensions. Different from these papers on the EM algorithm, our method allows for heterogeneous factor loadings and residual standard deviations in the cross-section. Chen, Cliff, and Zhao (2015) use a modified EM algorithm to group funds into different categories. They employ a two-step estimation procedure to first estimate the equation-by-equation OLS and then use the fitted alphas or the t -statistics of alphas to classify funds. We put fund specific variables on an equal footing with other structural parameters and simultaneously estimate the model parameters. As a result, we take into account the estimation uncertainty for fund specific variables, including both the factor loadings and the residual standard deviations.

specific and update the entire cross-section of fund specific variables along with other structural parameters in the maximization step of the EM algorithm. This is an important and necessary extension for the purpose of our application as we know there is estimation uncertainty as well as a large amount of heterogeneity in the risk-taking behavior of mutual funds. Failing to take either the heterogeneity or the estimation uncertainty into account may bias our estimate of the alpha population. On the other hand, allowing fund heterogeneity does not compromise the simplicity and the intuitive appeal of the standard EM algorithm. We show that our new algorithm simply embeds a constrained OLS estimate for fund specific parameters (i.e., factor loadings and residual standard deviations) into an otherwise standard EM algorithm. This greatly reduces the computational burden of our model. We provide a comprehensive simulation study to demonstrate the performance of our estimation procedure.

While we apply our framework to study fund performance in the current paper, we expect its general insight to be useful in other applications as well. Harvey and Liu (2016b) modifies the framework in this paper to study the predictability of alpha.

4 Estimation Procedure

We discuss the idea of the algorithm in the main text and describe the details in Appendix A. The following steps describe the procedure of the EM algorithm:

Step I Let $\mathcal{G} = [\theta', \mathcal{B}', \Sigma']'$ denote the collection of parameters to be estimated. We start at some parameter value $\mathcal{G}^{(0)}$. A sensible initial choice is the equation-by-equation OLS estimate for \mathcal{B} and Σ , and the MLE for θ based on the fitted OLS alphas.

Step II After the k -th iteration of the algorithm, suppose the model parameters are estimated as $\mathcal{G}^{(k)}$. We calculate the expected value of the log complete likelihood function, with respect to the conditional distribution of \mathcal{A} given the current parameter values and \mathcal{R} , i.e.,

$$L(\mathcal{G}|\mathcal{G}^{(k)}) = E_{\mathcal{A}|\mathcal{R},\mathcal{G}^{(k)}}[\log f(\mathcal{R}, \mathcal{A}|\mathcal{G})], \quad (8)$$

$$= E_{\mathcal{A}|\mathcal{R},\mathcal{G}^{(k)}}\left[\sum_{i=1}^N \log f(R_i|\alpha_i, \beta_i, \sigma_i)f(\alpha_i|\theta)\right]. \quad (9)$$

It is very likely that $L(\mathcal{G}|\mathcal{G}^{(k)})$ will not have a closed-form expression. But a variant of the EM algorithm — named the Monte Carlo EM algorithm — recommends replacing the expectation with the sample mean, where the sample is generated by simulating from the distribution of $\mathcal{A}|\mathcal{R}, \mathcal{G}^{(k)}$.¹⁶ We draw $M(=$

¹⁶See Greg, Wei and Tanner (1990), McCulloch (1997), and Booth and Hobert (1999).

100) \mathcal{A} 's from the distribution $\mathcal{A}|\mathcal{R}, \mathcal{G}^{(k)}$ and approximate the expectation in (9) by its sample counterpart:¹⁷

$$\widehat{L}(\mathcal{G}|\mathcal{G}^{(k)}) = \frac{1}{M} \sum_{m=1}^M \left[\sum_{i=1}^N \log f(R_i|\alpha_i^m, \beta_i, \sigma_i) f(\alpha_i^m|\theta) \right]. \quad (10)$$

Step III We need to find parameter values that maximize $\widehat{L}(\mathcal{G}|\mathcal{G}^{(k)})$ and update the parameter estimate as $\mathcal{G}^{(k+1)}$. This is usually not easy if the dimension of the parameter space is high. However, in our context, there is a simple solution. An inspection of (10) shows that $(\mathcal{B}', \mathbf{\Sigma}')$ and θ can be updated separately. More specifically, (10) can be written as

$$\widehat{L}(\mathcal{G}|\mathcal{G}^{(k)}) = \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \log f(R_i|\alpha_i^m, \beta_i, \sigma_i) + \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^N \log f(\alpha_i^m|\theta). \quad (11)$$

Notice that $\widehat{L}(\mathcal{G}|\mathcal{G}^{(k)})$ splits into two parts, one involving \mathcal{B} and $\mathbf{\Sigma}$, and the other involving θ . We therefore can maximize $\widehat{L}(\mathcal{G}|\mathcal{G}^{(k)})$ by separately maximizing these two parts.

Step IV With the new parameter estimate $\mathcal{G}^{(k+1)}$ obtained in *Step III*, we return back to *Step II* and start the $(k+1)$ -th iteration. We iterate between *Step II* and *Step III* until the parameter estimates converge.

The EM algorithm provides a tractable approach to find the MLE. It breaks the multi-dimensional optimization problem into smaller steps that are manageable. In theory, the EM estimator is guaranteed to converge to at least a local optimum of the likelihood function.¹⁸ It has been successfully applied to panel regression models with random effects when the the random effects do not follow a standard distribution.¹⁹ However, our model falls out of the realm of the standard application of the EM algorithm to panel regression models in that we allow heterogeneous risk loadings across funds. Therefore, the question remains as to whether the algorithm performs well in our application. We provide a detailed simulation study to evaluate the performance of the EM algorithm.

¹⁷A larger number of M gives us a closer approximation to the expectation in equation 9. However, it also increases the computational burden. We find that $M = 100$ gives us an estimate of θ (notice that the estimates of \mathcal{B} and $\mathbf{\Sigma}$ do not depend on M , as shown in Appendix A) that is very close to that under, say, $M = 1,000$. This is due to the fact that we have a large cross-section of alphas so an insufficient sampling of the alpha distribution for individual funds do not have a large impact on the optimization outcome. We therefore set $M = 100$ to save computational time.

¹⁸See Wu (1983) for the convergence properties of the EM algorithm.

¹⁹See Chen, Zhang and Davidian (2002).

We pay particular attention to the local optimum issue and construct a sequential estimation procedure to maximize the chance that our estimator converges to the global optimum. In particular, we first try a large number of randomly generated vectors of parameters to start the algorithm. Under a mild convergence threshold, we obtain many sets of initial parameter estimates. Some of these estimates may correspond to a local optimum. We then select the top performers among these estimates and apply tougher convergence thresholds to sequentially identify the global optimum. See Appendix B for the details of the implementation of our algorithm.

The steps of the EM algorithm make intuitive sense. They build on the idea that our knowledge about the cross-section of alphas and the model parameters can be sequentially updated. In *Step I*, we start with some initial parameter estimates, possibly the standard OLS estimates. In *Step II*, given our starting estimates of the model parameters, we calculate the expected value of the log likelihood function conditional on the distribution of the alphas. An intuitive way to think about this step is to replace $\mathcal{A}|\mathcal{R}, \mathcal{G}^{(k)}$ with the best estimate of \mathcal{A} given \mathcal{R} and $\mathcal{G}^{(k)}$.²⁰ By doing this, we are trying to come up with our best guess of the missing alphas given the return data and the model parameters. This is the step where we update our knowledge about the cross-section of alphas given our current estimates of the model parameters. In *Step III*, pretending that the estimated alphas in *Step II* are the true alphas, we have complete data and can easily estimate the model parameters. This is the step where we update our knowledge about the risk loadings and the residual variances (i.e., \mathcal{B} and $\mathcal{\Sigma}$). It is through the iterations between *Step II* and *Step III* that our estimates of the model parameters get refined.

More insight can be gained into the EM algorithm by specifying the parametric distribution Ψ . In *Step II*, assuming a Gaussian Mixture Distribution, Appendix A shows that the conditional distribution of \mathcal{A} given the current parameter values (denoted as $\hat{\mathcal{G}}$) and \mathcal{R} can be characterized as the distribution for N independent variables, with the i -th variable α_i following a fund specific GMD that is parameterized by $\tilde{\theta}_i = (\{\tilde{\pi}_{i,l}\}_{l=1}^L, \{\tilde{\mu}_{i,l}\}_{l=1}^L, \{\tilde{\sigma}_{i,l}^2\}_{l=1}^L)$:

$$\tilde{\mu}_{i,l} = \left(\frac{\hat{\sigma}_l^2}{\hat{\sigma}_l^2 + \hat{\sigma}_i^2/T}\right)\bar{a}_i + \left(\frac{\hat{\sigma}_i^2/T}{\hat{\sigma}_l^2 + \hat{\sigma}_i^2/T}\right)\hat{\mu}_l, \quad (12)$$

$$\tilde{\sigma}_{i,l}^2 = \frac{1}{1/\hat{\sigma}_l^2 + 1/(\hat{\sigma}_i^2/T)}, \quad (13)$$

$$\tilde{\pi}_{i,l} = \frac{\hat{\pi}_l \phi(\bar{a}_i - \hat{\mu}_l, \hat{\sigma}_l^2 + \hat{\sigma}_i^2/T)}{\sum_{l=1}^L \hat{\pi}_l \phi(\bar{a}_i - \hat{\mu}_l, \hat{\sigma}_l^2 + \hat{\sigma}_i^2/T)}, \quad l = 1, 2, \dots, L, \quad (14)$$

where

$$\bar{a}_i \equiv \sum_{t=1}^T (r_{it} - \hat{\beta}'_i f_t)/T,$$

²⁰See Neal and Hinton (1998) for a more rigorous interpretation of the EM algorithm.

and $\phi(\mu, \sigma^2)$ is the density of the normal distribution $\mathcal{N}(0, \sigma^2)$ evaluated at μ .

We can think of \bar{a}_i as the fitted alpha when β_i is fixed at $\hat{\beta}_i$. It would be the OLS estimate of alpha if $\hat{\beta}_i$ were the OLS estimate of β_i . The variance of the time-series residuals is fixed at $\hat{\sigma}_i^2$. Taken together, \bar{a}_i and $\hat{\sigma}_i^2/T$ can be interpreted as the alpha estimate and its variance based on time-series information. On the other hand, $\hat{\theta}_i = (\{\hat{\pi}_{i,l}\}_{l=1}^L, \{\hat{\mu}_{i,l}\}_{l=1}^L, \{\hat{\sigma}_{i,l}^2\}_{l=1}^L)$ is the current parameter vector governing the GMD for the cross-sectional distribution of the alphas. Therefore, (12), (13) and (14) update our estimates of the alphas by combining time-series and cross-sectional information.

We start with a L -component GMD specification for the alpha population. The updated alpha distribution for each individual fund is also a GMD with the same number of components. However, the parameters that govern the GMD will be different across funds. For each of L component distributions for the fund specific GMD, the mean (i.e., $\tilde{\mu}_{i,l}$) is a weighted average of the fitted time-series alpha and the original mean for the GMD, the variance (i.e., $\tilde{\sigma}_{i,l}^2$) is the harmonic average of the time-series variance and the original variance for the GMD, and the drawing probability (i.e., $\tilde{\pi}_{i,l}$) weights the original probability by $\phi(\bar{\mu}_i - \hat{\mu}_l, \hat{\sigma}_l^2 + \hat{\sigma}_i^2/T)$, which depends on the distance between $\bar{\mu}_i$ and $\hat{\mu}_l$ (i.e., $|\bar{\mu}_i - \hat{\mu}_l|$) and the average of the variances $\hat{\sigma}_l^2 + \hat{\sigma}_i^2/T$.

Holding everything else constant, a lower time-series variance (i.e., $\hat{\sigma}_i^2/T$) pulls both the updated mean and variance closer to their time-series estimates, thereby overweighing time-series information relative to cross-sectional information. On the other hand, a smaller distance between $\bar{\mu}_i$ and $\hat{\mu}_l$ implies a higher drawing probability (i.e., $\tilde{\pi}_{i,l}$), which means that compared to the original GMD, we are now more likely to draw from the component distribution that has a mean that is closer to $\bar{\mu}_i$. Hence, we revise our estimate of the cross-sectional distribution based on time-series information. The expressions in (12), (13) and (14) bear intuitive interpretations as to how we update the alpha estimates based on both time-series and cross-sectional information. This synthesis of information is important as it allows us to obtain the most informative estimate of the \mathcal{A} distribution, which is then used to evaluate the likelihood function as in *Step II* of the EM algorithm. It also distinguishes our method from existing approaches that only rely on one source of information, either cross-sectional or time-series.

Similar ideas that pool information across funds to make better inference on fund performance have been proposed in Jones and Shanken (2005) and Cohen, Coval, and Pastor (2005).²¹ Jones and Shanken rely on a Bayesian framework and impose a normal prior on the alpha distribution. Imposing a one-component GMD in our model, the posterior distribution of the alpha in their framework has an expression that is similar to ours. For instance, similar to equation (12), the posterior mean for the alpha of fund i in their framework is also a weighted average of the OLS alpha and the prior mean of the alpha population. However, the important difference between

²¹See Huij and Verbeek (2003) for a shrinkage approach that is similar to Jones and Shanken (2005). Harvey and Liu (2015d) apply a similar idea to the selection of risk factors.

the two approaches is that the mean of the alpha population in Jones and Shanken is fixed a priori, whereas in our framework it is a freely estimated parameter.²² As such, their model is better used to assess relative fund performance whereas ours can be used both for relative and absolute performance evaluation. The same problem exists in Jones and Shanken for the variance of each individual fund. On the other hand, different from Cohen, Coval, and Pastor, which allows one to learn through the portfolio holdings of managers, we learn about the skill of a particular manager by grouping funds with similar alpha estimates, after adjusting for the estimation uncertainty in the alpha estimation.

Another way to interpret the formulas in equations (12)-(14) is to consider the extreme case and assume that we have a single component GMD (that is, $L = 1$), and moreover, its mean is zero (that is, $\hat{\mu}_1 = 0$). In this case, we link the t -statistic of fund i 's alpha (defined as $\tilde{\mu}_i/\tilde{\sigma}_i$) with its OLS t -statistic (defined as $\bar{a}_i/\sqrt{\hat{\sigma}_i^2/T}$) through:

$$\frac{\tilde{\mu}_i}{\tilde{\sigma}_i} = \frac{\bar{a}_i}{\sqrt{\hat{\sigma}_i^2/T}} \times \sqrt{\frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 + \hat{\sigma}_i^2/T}}. \quad (15)$$

Notice that $\sqrt{\frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 + \hat{\sigma}_i^2/T}} < 1$ and the larger the time-series variance (that is, $\hat{\sigma}_i^2/T$) is relative to the cross-sectional variance (that is, $\hat{\sigma}_1^2$), the smaller this number becomes. Therefore, when the average alpha is zero in the population, we discount the OLS t -statistic with a discount factor that equals $\sqrt{\frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 + \hat{\sigma}_i^2/T}}$. More time-series uncertainty results in a harsher discount.

The idea of discounting OLS t -statistic is consistent with the idea of multiple testing adjustment, which has recently gained attention in both performance evaluation and asset pricing in general.²³ However, the mechanism in our model to deflate t -statistics is different from standard multiple testing approaches. Our model, by treating the alpha of an investment fund as random, takes into account the cross-sectional uncertainty in alpha from a population perspective. Multiple testing methods, by treating the alpha as a fund specific variable (that is, a fixed effect), adjust t -statistics by having a more stringent Type I error threshold. Despite the methodological difference, these two fundamentally different approaches arrive at the same conclusion — we need to apply a “haircut” to the individual t -statistics of fund alphas.

²²Bayesians may suggest the use of uninformative priors, both for fund alphas and risk loadings. However, Kass and Wasserman (1996) remind us that it is a dangerous practice to put faith in any default choice of prior, especially when the sample size is small (relative to the number of parameters). The sample size concern seems to be particularly relevant for the mutual fund and hedge fund data in that we have a large cross-section of risk loadings to estimate. Any distortion from the prior specifications of the risk loadings will feed into the estimation of the alpha population.

²³For recent finance applications of multiple hypothesis testing in asset pricing, see Barras et al. (2010), Fama and French (2010), and Ferson and Chen (2015), Harvey, Liu, and Zhu (2016), and Harvey and Liu (2015b,c).

In *Step III*, we update our parameter estimates based on the conditional distribution of the alphas. This is done in two steps. We first update the OLS parameters except for the regression intercepts, and then update θ — the parameter vector that governs the alpha population.

For the update of the OLS parameters (see Appendix A), we derive analytical expressions for the MLE of β_i and σ_i^2 . In particular, let $m(\alpha_i) = E_{\mathcal{A}|\mathcal{R},\mathcal{G}^{(k)}}(\alpha_i)$ and $var(\alpha_i) = Var_{\mathcal{A}|\mathcal{R},\mathcal{G}^{(k)}}(\alpha_i)$ be the conditional mean and variance of α_i . The MLE of β_i can be found as the regression coefficients obtained by projecting the return time-series (i.e., $\{r_{i,t}\}_{t=1}^T$) onto the factor time-series (i.e., $\{f_t\}_{t=1}^T$), fixing the regression intercept at $m(\alpha_i)$. Therefore, the MLE of β_i in our model differs from the usual OLS estimate in that the regression intercept is forced to equal $m(\alpha_i)$, the population mean of α_i given our current knowledge about the alpha distribution (i.e., $\mathcal{A}|\mathcal{R},\mathcal{G}^{(k)}$).

The MLE of σ_i^2 can be found by fixing β_i at its MLE (i.e., $\check{\beta}_i$). In particular, define

$$\overline{\varepsilon}_i^2 \equiv \frac{1}{T} \sum_{t=1}^T (r_{it} - \check{\beta}_i^t f_t - m(\alpha_i))^2, \quad (16)$$

as the fitted residual mean squared error. Then the MLE of σ_i^2 is given by

$$\check{\sigma}_i^2 = \overline{\varepsilon}_i^2 + var(\alpha_i). \quad (17)$$

Notice that if we use $(\sigma_i^2)^{MLE}$ to denote the MLE of the residual variance for the standard regression model that projects the time-series of returns (i.e., $\{r_{i,t}\}_{t=1}^T$) onto $\{f_t\}_{t=1}^T$, then we must have

$$\overline{\varepsilon}_i^2 \geq (\sigma_i^2)^{MLE}$$

since the standard regression model seeks to minimize the sum of squared residuals without any parameter constraints. Therefore, two effects make the MLE of the residual variance (i.e., $\check{\sigma}_i^2$) in our model larger than the standard model MLE (i.e., $(\sigma_i^2)^{MLE}$). First, $\overline{\varepsilon}_i^2$ is no less than $(\sigma_i^2)^{MLE}$ because we are considering a regression model whose intercept is fixed at $m(\alpha_i)$. Second, there is uncertainty in α_i as captured by $var(\alpha_i)$, which depends on the parameters given in (12), (13) and (14) of the updated GMD (see Appendix A). Since, as discussed previously, the updated GMD takes both time-series and cross-sectional information into account, $var(\alpha_i)$ also incorporates information about the cross-sectional dispersion of the alphas.

These two effects implied by our model make intuitive sense as they allow us to learn from both the mean and the variance of the alpha population. Additionally, the learning effect is more pronounced in small samples and will go away when we have a long enough time series of returns. This can be easily seen from the formulas of our algorithm. When T goes to infinity and based on equation (12)-(14), the alpha distribution collapses to the point mass at \bar{a}_i , which is the estimate based on time-series information only. This implies that $m(\alpha_i) = \bar{a}_i$ and $var(\alpha_i) = 0$. As a result, our MLE of β_i and σ_i converge to their OLS estimates. The fact that our method

implies differential adjustment to the alpha estimate between small and large samples makes it an attractive method for performance evaluation, where a large fraction of funds have short time series.

For the update of θ , we seek for the parameter vector θ of a GMD that best describes the alpha distribution. The optimization problem we are solving is:

$$\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \log f(\alpha_i^m | \theta), \quad (18)$$

where $\{\alpha_i^m\}_{m=1}^M$ are randomly generated samples from the conditional distribution of α_i given \mathcal{R} and $\mathcal{G}^{(k)}$. If there were just one fund in the cross-section, then $\hat{\theta}$ will approximately equal the parameters that govern the GMD for a single fund that are given in equation (12)-(14). With multiple funds in the cross-section, we have multiple GMD's, each one governing the alpha distribution of a particular fund. Our method tries to find the best θ that describes the cross-section of GMD's, which can be viewed as a mixture distribution that chooses a fund with equal probability from the cross-section of funds and, conditional on a fund being chosen, draws an alpha from the fund's GMD. Notice that this mixture distribution in our model is very different than the alpha distribution in the equation-by-equation OLS model, where it is simply the cross-section of fitted alphas. Our method allows us to capture the estimation risk of each fund's alpha and leads to a more informed estimate of the alpha distribution.

One concern about our model estimation is the large number of parameters to estimate. Indeed, since we allow heterogeneity in fund risk loadings and residual variances, the number of parameters grow almost proportionally with the number of funds in the cross-section. However, the set of parameters that grow with the number of funds are auxiliary parameters that govern the time-series dynamics of each individual fund. The key parameter set of interest — θ that parameterizes Ψ — does not change with the size of the cross-section. Intuitively, each additional fund added to the cross-section, while creating a new set of parameters to estimate for its time-series dynamics, will provide additional information for us to estimate θ . We show in the simulation study that θ is accurately estimated when we have a large cross-section.

4.1 A Simulation Study

4.1.1 Simulation Design

We provide a simulation study to examine the performance of the random alpha model and compare it with the standard equation-by-equation OLS model.

We use mutual fund data as an example. For a detailed description of the mutual fund data, see the next section where we apply our method to both mutual funds and hedge funds. For our simulation study, we require that a fund has at least eight months of return observations. This allows us to have enough time series to estimate the factor model and is consistent with the existing literature (e.g., Fama and French, 2010, Ferson and Chen, 2015). Imposing this constraint, we have 3,619 funds in the cross-section covering the 1983–2011 period. We obtain monthly returns for these funds. Except for the restriction on sample length, we do not impose any further restrictions on the data and we use all the funds in the data for our simulation study. As a result, the cross-section for our simulation study is as large as that for the real applications. This allows us to provide a more realistic evaluation of the performance of our model.

With this sample of mutual funds, we run equation-by-equation OLS based on the full sample to obtain the initial estimates for \mathcal{B} and Σ (i.e., \mathcal{B}^* and Σ^*). We also obtain the initial fitted alphas. We specify the number of component distributions for the GMD and apply it to these fitted alphas and obtain the estimate for θ (i.e., θ^*). We collect these parameter estimates into $\mathcal{G}^* = [\theta^*, \mathcal{B}^*, \Sigma^*]'$. \mathcal{G}^* will be the true underlying parameter vector that governs the data generating process. Special attention is paid to funds that do not have enough data to cover the entire sample period. In our simulations, we make sure that the simulated returns for these funds cover the same time periods as the original fund data.

We need to make a choice for the number of component distributions for the GMD in our simulation study. Notice that our goal is not to find the best fitting GMD to the OLS alphas, but to obtain a parameter set to initiate the simulation study. A one-component GMD (i.e., a single normal distribution) is obviously the simplest GMD one can specify, but it may be considered too special for a simulation study. We therefore specify a two-component GMD — the simplest multi-component GMD one can have.²⁴

Besides the number of components for the GMD, the particular value of \mathcal{G}^* is not essential for our simulation study. We could use an arbitrary set of parameters as the underlying parameter vector that governs the data generating process. However, the use of \mathcal{G}^* makes our simulation study more realistic as we are using the actual fund cross-section to extract the model parameters. It takes the cross-sectional heterogeneity in risk loadings into account as well as captures the multi-population structure for

²⁴The results of our simulation results to a large degree do not depend on the initial model we choose for the GMD. We also try a three-component GMD. The results are qualitatively similar in the sense that at the population level, the bias and variance (as measured by RMSE) for the population parameters that are implied by the random alpha model are much smaller than those that are implied by the equation-by-equation OLS, and that at the individual fund level, the random alpha model generates alpha estimates that are more precise and less volatile than the OLS model. We therefore expect the performance of the random alpha model to dominate that of the equation-by-equation OLS under alternative parameter configurations.

a plausible set of alphas (i.e., the alpha estimates based on the equation-by-equation OLS).

Table 1 reports the summary statistics of the parameter vector θ^* . The two-component GMD separates the cross-section of fitted OLS alphas into two groups. The first group has a mean that is mildly negative (-1.27% , per annum) and a relatively small standard deviation (2.61%), and the second group has a mean that is very negative (-6.44% , per annum) and a large standard deviation (16.95%). It is less frequent for an alpha to fall into the second group as its drawing probability is only 4.0% . Our model estimates are roughly consistent with the empirical evidence documented by the literature using equation-by-equation OLS. A large fraction of mutual funds exhibit alphas that are close to zero while a small fraction of funds seem to significantly underperform.

Table 1: **Parameter Vector (θ^*) for the Simulated Model**

Parameter vector (θ^*) for the simulated model. We run equation-by-equation OLS for a cross-section of 3,619 mutual funds that at least have eight months of return observations for the 1983-2011 period. We obtain the cross-section of fitted alphas. We then fit a two-component GMD on these alphas. μ_l and σ_l are the (annualized) mean and the (annualized) standard deviation for the l -th component normal distribution, and π_l is the probability for drawing from the l -th component, $l = 1, 2$.

	First component ($l = 1$)	Second component ($l = 2$)
$\mu_l(\%)$	-6.443	-1.273
$\sigma_l(\%)$	16.951	2.606
π_l	0.040	0.960

Based on \mathcal{G}^* , we simulate D ($=100$) panels of fund returns, each one having the same size as the original data panel.²⁵ In particular, for each fund i , we randomly generate its alpha based on the GMD that is parameterized by θ^* . We then generate n_i $\mathcal{N}(0, (\sigma_i^2)^*)$ random variables, where n_i is the sample size for fund i in the original data. These random variables will be the simulated return residuals. Together with the randomly generated alpha and the factor loadings β_i^* , these residuals enable us to construct the simulated return series for fund i . To examine how residual correlation affects our results, we allow the cross-section of residuals to be contemporaneously correlated with a correlation coefficient of ρ .

²⁵We currently fix D at 100 to save computational time. We will later increase D to 1,000.

4.1.2 The Alpha Population

For each of the simulated return panel, we estimate our model, thereby obtaining D sets of model estimates. Table 2 summarizes these estimates and compares with the estimates of the standard OLS model, that is, we first run equation-by-equation OLS and then fit a GMD for the cross-section of OLS alphas.²⁶

Based on the results in Table 2, the random alpha model stands out as superior to the standard OLS model. In particular, its finite sample biases are uniformly smaller (in absolute value) than those of the OLS estimates.

We first focus on Panel A and examine the estimates for the means of the component distributions. For the first component, for which the group mean is very negative and the drawing probability is small (4.0%), the bias for the OLS model is 1.25% whereas the bias for the random alpha model is 0.20%. The estimation uncertainty (RMSE) for the OLS model (2.31%) is also higher than the random alpha model (1.44%). For the second component, which happens much more frequently than the first group (drawing probability is 96.0%), the OLS model and the random alpha model have similar performance with respect to the mean. Although the OLS model is inferior than the random alpha model by making less precise and more noisy alpha forecasts for individual funds (as we shall see later), when we pool the cross-section of funds together to estimate the overall population mean, the noise at the individual fund level cancels out and the OLS model does not seem to be significantly worse than the random alpha model in terms of the estimation of the population mean. This is particularly the case in the estimation of the second group as we have more observations that fall into that group so the cancelation effect is stronger. For the first group, for which we have fewer observations, the random alpha model appears to be a better model in estimating the group mean.

Turning to the estimates of the variances, the contrast in model performance is starker. For example, we reduce the absolute values of the biases for the estimates of the standard deviations of the component normal distributions from 25% (= 4.27/16.95) and 33% (=0.86/2.61) to 0.4% (= 0.07/16.95) and 0.4% (= 0.01/2.61), respectively. Therefore, the OLS model does not seem to be able to yield consistent estimates of the standard deviations for the component distributions. Indeed, in our simulations, the OLS model frequently overestimates the standard deviations of the component normal distributions. This is not surprising since, by ignoring the time-series uncertainty in the estimation of the fund specific alphas, it attributes all the variation in the cross-section of the fitted alphas to the variation of the alpha population.²⁷ On the other hand, by taking both sources of uncertainty (i.e., time-series

²⁶Note that $\pi_1 + \pi_2 = 1$. However, we present both for completeness. Summary statistics for π_1 and π_2 in general will not sum up to one as we are averaging over the simulations.

²⁷The OLS model in our simulation study is the simplest two-stage model one can have by first running equation-by-equation OLS and then fitting the cross-section of estimated alphas. Chen, Cliff, and Zhao (2015) propose a generalization of this model by taking the estimated OLS variances as given and feeding them into the estimation of the GMD. Their paper therefore partially takes the

uncertainty for individual fund returns and cross-sectional uncertainty for the alpha population) into account, the random alpha model does not seem to be significantly biased, and is able to estimate the parameters that govern the alpha distribution with high precision.

When return residuals are correlated and based on our approach, the estimates for the means of the component normal distributions become more variable while there are no material changes for the estimates of the other parameters. The more variable estimates for the means are expected as we have less information in the cross-section when return residuals are correlated. For example, compared to the case when $\rho = 0$ in Panel A, when $\rho = 0.2$ in Panel B, the RMSE for μ_1 increases from 1.44% to 1.70% for the random alpha model. The increased estimation uncertainty is the price we have to pay for misspecifying the model likelihood function. However, the increase seems small for reasonable levels of residual correlations, especially for the random alpha model. Barras, Scaillet, and Wermers (2010) document that the average pairwise correlation between the four-factor model residuals is 0.08. We think our specification of $\rho = 0.4$ is a conservative upper bound for the average level of the residual correlation.

Overall, our results in Table 2 suggest that the OLS model, by first running equation-by-equation OLS regressions to obtain the estimated alphas and then fitting a parametric distribution to these alphas, is severely biased in estimating the parameters that govern the cross-sectional alpha distribution. The random alpha model, by explicitly modeling the underlying alpha distribution, seems to be able to provide consistent and more precise parameter estimates.

We have shown that the random alpha model produces superior parameter estimates for the alpha population in comparison with the OLS model. Based on these parameter estimates, we can calculate several important statistics that summarize the alpha population. Not surprisingly, the random alpha model produces more accurate and less volatile estimates for these statistics than the OLS model, as shown in Table 3.²⁸

Both methods generate similar results regarding the overall population mean of the alpha distribution. Under the assumption of the GMD, the overall population mean is simply the individual means of the two component distributions weighted by the corresponding drawing probabilities. Given that the two methods generate similar mean estimates for the second component of the GMD (as shown in Table 2) and that it is more likely for an alpha to come from the second component (drawing probability equals 96.0%), it is not surprising that the two methods have similar

time-series uncertainty into account. However, there are other important sources of estimation risk that cannot be addressed in their framework, e.g., the estimation of risk loadings and the estimation of residual variances themselves. Our structural approach allows us to take all of these sources of estimation risk into account.

²⁸Given the similarity in model performance across different levels of residual correlations, we set the level of residual correlation at zero for the rest of the analysis in this section. We have tried alternative correlation specifications and they do not change our results in any important way.

Table 2: **A Simulation Study: Parameter Estimates for the Alpha Population**

Model estimates in a simulation study. We fix the model parameters at \mathcal{G}^* (Table 1) and generate D sets of data sample. For each set of data sample, we estimate our model using both the proposed random alpha model (“RA”) and the standard equation-by-equation OLS (“OLS”). ρ is the assumed level of correlation among the cross-section of return residuals. For a given parameter γ , let γ_d be the model estimate based on the d -th simulation run, $d = 1, 2, \dots, D$. “True” reports the assumed true parameter value given in \mathcal{G}^* . “Bias” reports the difference between the average of the simulated parameter estimates and the true value, that is, $(\sum_{d=1}^D \gamma_d)/D - \gamma$. “RMSE” reports the square root of the mean squared estimation error, that is, $\sqrt{\sum_{d=1}^D (\gamma_d - \gamma)^2/D}$. “ $p(10)$ ” reports the 10th percentile of the parameter estimates and “ $p(90)$ ” reports the 90th percentile of the parameter estimates. μ_l and σ_l are the (annualized) mean and the (annualized) standard deviation for the l -th component normal distribution, and π_l is the probability for drawing from the l -th component, $l = 1, 2$.

		$\rho = 0$		$\rho = 0.2$		$\rho = 0.4$	
		RA	OLS	RA	OLS	RA	OLS
$\mu_1(\%)$ (True = -6.443)	Bias	0.020	1.252	-0.249	1.062	-0.103	1.141
	RMSE	1.439	2.309	1.704	2.294	2.129	2.865
	$p(10)$	-8.322	-7.880	-8.958	-7.377	-9.633	-9.073
	$p(90)$	-4.614	-2.877	-4.477	-3.326	-4.063	-2.352
$\sigma_1(\%)$ (True = 16.951)	Bias	-0.067	4.268	-0.170	2.885	-0.111	3.887
	RMSE	1.177	8.336	1.131	5.535	1.086	7.812
	$p(10)$	15.323	14.914	15.327	15.315	15.383	15.517
	$p(90)$	18.513	28.960	18.058	26.058	18.295	31.047
π_1 (True = 0.040)	Bias	0.000	0.013	-0.001	0.014	-0.001	0.013
	RMSE	0.005	0.018	0.005	0.018	0.006	0.018
	$p(10)$	0.033	0.037	0.033	0.042	0.032	0.038
	$p(90)$	0.046	0.069	0.045	0.068	0.047	0.070
$\mu_2(\%)$ (True = -1.273)	Bias	0.005	0.004	-0.045	-0.051	0.059	0.059
	RMSE	0.059	0.064	0.530	0.577	0.874	0.956
	$p(10)$	-1.340	-1.348	-1.947	-2.021	-2.345	-2.507
	$p(90)$	-1.195	-1.186	-0.687	-0.661	-0.122	-0.022
$\sigma_2(\%)$ (True = 2.606)	Bias	-0.005	0.857	-0.058	0.790	-0.100	0.765
	RMSE	0.065	0.861	0.094	0.794	0.150	0.777
	$p(10)$	2.514	3.363	2.464	3.287	2.381	3.231
	$p(90)$	2.680	3.590	2.623	3.495	2.683	3.570
π_2 (True = 0.960)	Bias	0.000	-0.013	0.001	-0.014	0.001	-0.013
	RMSE	0.005	0.018	0.005	0.018	0.006	0.018
	$p(10)$	0.954	0.931	0.955	0.933	0.953	0.930
	$p(90)$	0.967	0.963	0.967	0.958	0.968	0.962

Table 3: **A Simulation Study: Population Statistics**

Population statistics based on the model estimates in a simulation study. We fix the model parameters at \mathcal{G}^* (Table 1) and generate D sets of data sample. For each set of data sample, we estimate our model using both the proposed random alpha model (“random alpha”) and the standard equation-by-equation OLS (“OLS”). We then calculate several summary statistics for the alpha population for both models based on the estimated model parameters. “Mean” is the mean of the alpha distribution. “Stdev.” is the standard deviation of the alpha distribution. “Iqr.” is the interquartile range of the alpha distribution. “ $p10$ ” is the 10th percentile of the alpha distribution. The other percentiles are similarly defined. “True” reports the population statistics based on the true model. “Estimate” reports the averaged estimate of the population statistics across the D sets of simulations. “RMSE” reports the square root of the mean squared estimation error, that is, $\sqrt{\sum_{d=1}^D (s_d - s)^2 / D}$, where s is the true statistic and s_d is the estimated statistic based on the d -th simulated sample. Residual correlation is set at zero.

		Random alpha	OLS
Mean(%) (True = -1.477)	Estimate	-1.470	-1.468
	RMSE	0.078	0.099
Stdev.(%) (True = 4.350)	Estimate	4.330	5.864
	RMSE	0.179	1.728
Iqr.(%) (True = 3.511)	Estimate	3.646	4.932
	RMSE	0.205	1.430
$p5$ (%) (True = -6.223)	Estimate	-6.126	-7.898
	RMSE	0.206	1.688
$p10$ (%) (True = -4.946)	Estimate	-4.884	-6.178
	RMSE	0.165	1.243
$p50$ (%) (True = -1.435)	Estimate	-1.295	-1.288
	RMSE	0.179	0.188
$p90$ (%) (True = 2.077)	Estimate	2.208	3.450
	RMSE	0.182	1.382
$p95$ (%) (True = 3.353)	Estimate	3.299	4.978
	RMSE	0.147	1.634

estimates for the overall population mean. However, the random alpha model has a far better estimate of the dispersion of the alpha distribution than the OLS model. For example, the dispersion for the underlying true model is 4.35%. The average estimate for the random alpha model is 4.33%, and the RMSE is 0.18%. In contrast, the OLS model overestimates the dispersion by 35% ($= (5.86 - 4.35)/4.35$) and the RMSE

is 1.73%. This difference in model performance is also reflected in the estimation of the percentiles of the alpha population. For example, the average estimate for the 10-th percentile based on the random alpha model is -4.88% , which is very close to the true value (-4.95%). In contrast, the OLS model has an estimate that is 27% ($= |-6.18 - (-4.88)|/4.88$) lower.

4.1.3 Individual Funds

Having discussed the simulation results regarding the alpha population, we now turn to the inference of each individual fund. As mentioned previously, our method allows us to make inference on each individual fund through equation (4). More specifically, given a set of parameter estimates, the density forecast of an individual fund is given by equations (12)-(14).

In order to evaluate relative model performance, we need to choose a few statistics that summarize a model's forecasting accuracy at the individual fund level. We concentrate on two statistics. The first focuses on the point estimates. In particular, the absolute deviation (AD) calculates the absolute distance between the alpha estimate and the true alpha value. The second reflects estimation uncertainty. We calculate the length of the confidence interval that is constructed to cover the true alpha value with a certain probability. Notice that the t -statistic is not appropriate in our simulation framework since, by assumption, fund alphas are nonzero. For example, suppose the true alpha is 5% per annum for a certain fund and the point estimates based on the random alpha model and the OLS are 4% and 7%, respectively. Additionally, suppose the standard errors for the two models are the same. Clearly, the random alpha model is a better model as it provides a more accurate point estimate without raising the standard error. However, the OLS t -statistic will be higher than that based on the random alpha model, suggesting a more significant finding under the OLS. This is misleading. We therefore avoid the use of the t -statistic and separately show the improvement of our model over the OLS for the numerator and the denominator of the t -statistic, that is, the point estimate and the length of the confidence interval, both of which can be easily obtained through the density forecast of the random alpha model. Ideally, a better performing model will imply both a more accurate point estimate and a shorter confidence interval.

Table 4 reports the results. In terms of point estimates, the average distance between the model estimate and the true alpha (as measured by the mean absolute deviation) is 1.29% for the random alpha model, which is about two-thirds ($=1.29/1.85$) of that for the OLS model. In terms of estimation uncertainty, both methods generate confidence intervals that roughly achieve the pre-specified coverage rate (i.e., the probability for the confidence interval to contain the true alpha value) of 90% and 95%. However, the length of the confidence interval generated under the random alpha model is on average much shorter than that generated under the OLS model. For instance, under 95% significance, the median length is 5.72% for the random al-

pha model, which is 22%(= $(7.34 - 5.72)/7.34$) shorter than that of the OLS model. Therefore, at the individual fund level, the random alpha model is able to generate alpha estimates that are both more precise and less variable than the OLS model. Its improvement over the OLS model seems substantial from an economic perspective.

Table 4: **A Simulation Study: Individual Funds**

Summary statistics on model performance at the individual fund level. We fix the model parameters at \mathcal{G}^* (Table 1) and generate D sets of data sample. For each set of data sample, we estimate our model using both the proposed random alpha model (“random alpha”) and the standard equation-by-equation OLS (“OLS”). For the random alpha model, given the parameter estimates, we use equations (12)-(14) to first construct the density forecast for each individual fund, and then obtain the point estimate and the confidence interval. For OLS, its point estimate is the estimate for the intercept, and its confidence interval is constructed using the point estimate and the standard error for the intercept. “Mean absolute deviation” is the averaged (across simulations) mean absolute distance between the estimated alpha and the true alpha for the cross-section of funds. “Stdev. of mean absolute deviation” is the averaged (across simulations) standard deviation of the absolute distance between the estimated alpha and the true alpha for the cross-section of funds. “Length, p ” reports the averaged (across simulations) p -th percentile of the length of the 90% (or 95%) confidence intervals for the cross-section of funds. “Coverage probability” reports the averaged (across simulations) probability for the 90% (or 95%) confidence intervals to cover the true alpha values for the cross-section of funds. Other variables are similarly defined. Residual correlation is set at zero.

		Random alpha	OLS
	Mean absolute deviation(%)	1.289	1.851
	Stdev. of mean absolute deviation(%)	1.196	3.336
90% confidence interval	Length, $p10$ (%)	2.932	3.297
	Length, $p50$ (%)	4.793	6.161
	Length, $p90$ (%)	6.938	12.461
	Coverage probability	0.882	0.893
95% confidence interval	Length, $p10$ (%)	3.496	3.929
	Length, $p50$ (%)	5.719	7.341
	Length, $p90$ (%)	8.327	14.848
	Coverage probability	0.938	0.944

Overall, our results suggest that the random alpha model dominates the equation-by-equation OLS, both in terms of modeling the alpha cross-section and in terms of making inference on a particular fund’s alpha. Hence, under the assumption that fund alphas can be viewed as coming from an underlying distribution, there seems to be no reason to use the equation-by-equation OLS again for performance evaluation.

5 Results

5.1 Mutual Funds

We now apply our method to study mutual funds.²⁹ We obtain the mutual fund data used in Ferson and Chen (2015). Their fund data is from the Center for Research in Security Prices Mutual Fund database. They focus on active, domestic equity funds covering the 1984-2011 period. To mitigate omission bias (Elton, Gruber and Blake, 2001) and incubation and back-fill bias (Evans, 2010), they apply several screening procedures. They limit their tests to funds that have initial total net assets (TNA) above \$10 million and have more than 80% of their holdings in stock. They also combine multiple share classes. We require that a fund has at least eight months of return observations to enter our test. This leaves us with a sample of 3,619 mutual funds for the 1984-2011 period.³⁰ We use the four-factor model in Fama and French (1993) and Carhart (1997) as our benchmark model.

5.1.1 Parameter Estimates and Model Selection

A central issue is how we choose the number of components for the GMD that models the alpha distribution in the cross-section. A more complex model (i.e., a model with more component distributions) can potentially provide a better approximation to the underlying alpha distribution, but may overfit, leading to a model that has inferior forecasts out of sample. Standard model selection criteria (e.g., the Akaike information criterion or the Bayesian information criterion) may not work well in our context as they rely on asymptotic approximations. In our application, since the number of parameters grow with the number of funds in the cross-section, it is unclear what size of the cross-section would be regarded as large enough to warrant asymptotic approximations. To have a rigorous model selection framework that takes many aspects of our application into account (e.g., unbalanced panel, large number of model parameters), we use a simulation-based model selection approach.³¹

Consider two nested models M_0 and M_1 , with M_1 being the bigger model. For example, in our context, a GMD with a single component distribution will be nested within a two-component GMD specification as, by setting the drawing probability for one of the component distributions to zero, the latter collapses to the former. To distinguish between M_0 and M_1 , we need a metric that evaluates relative model performance. Given that our estimation relies on the MLE, a natural choice is the

²⁹In future research, we will apply our method to hedge fund returns.

³⁰We thank Yong Chen for providing us with the mutual fund data used in Ferson and Chen (2015).

³¹For a similar approach that bootstraps likelihood ratios to test the number of components in a GMD, see Feng and McCulloch (1996).

likelihood-ratio statistic, which measures the difference in likelihoods between the two candidate models. The likelihood-ratio statistic is also a key ingredient for many popular model selection criteria. In particular, let L_0 (L_1) be the value of the likelihood function evaluated at the model estimates for M_0 (M_1). The likelihood-ratio statistic (LR) is defined as:

$$\text{LR} = -2(\log L_0 - \log L_1). \quad (19)$$

When the bigger model (i.e., M_1) provides a substantial improvement over the smaller model (i.e., M_0), LR will be large and positive. Therefore, a large likelihood-ratio statistic provides evidence against the smaller model.

We simulate to find the cutoff value for LR. We first estimate M_0 and obtain its parameter estimates. Assuming M_0 is the true model, we simulate normally distributed return innovations to generate $D = 100$ return panels, similar to what we do in the simulation study. For each panel, we estimate both M_0 and M_1 , and calculate the LR statistic. The 5th percentile of these LR statistics will be used as the cutoff for the LR statistic.

We incrementally select the best performing parsimonious model. We first estimate a one-component and a two-component model. Based on the parameter estimates, the LR statistic between the two models is calculated to be $41.79 (\times 10^{-6})$. Assuming that the one-component model is true and simulating the model based on its parameter estimates, the 5th percentile of the LR statistic is found to be $6.85 (\times 10^{-6})$, which is smaller than the realized likelihood statistic. Therefore, the two-component model presents a significant improvement over the one-component model.

Next, we estimate a three-component model. The LR statistics between the two-component model and the three-component model is calculated to be $4.71 (\times 10^{-6})$. This time, assuming that the two-component model is true and simulating the model based on its parameter estimates, the 5th percentile of the LR statistic is $14.30 (\times 10^{-6})$. Hence, the realized LR statistic is less than the simulated LR cutoff, suggesting that we do not have enough evidence to discard the simpler two-component model.

Given the rejection of the three-component model, we do not need to further consider the four-component model as its incremental contribution to the three-component model is likely to be even smaller than the incremental contribution of three-component model to the two-component model. We therefore select the two-component model as the final model. It is the most parsimonious model that still provides an adequate description of the cross-sectional distribution of fund alphas.

Our finding of a two-group categorization of mutual fund managers is consistent with the recent literature on mutual fund performance evaluation. For example, Barras et al. (2010) use the false discovery approach to control for multiple testing and find that 75% of the funds are zero-alpha funds and 24% are unskilled (i.e.,

significantly negative).³² The remaining 1% appear to be skilled but are statistically indistinguishable from zero. We also find that a two-group classification is sufficient to describe the universe of fund managers. In particular, unlike for underperformers, we do not need a third component distribution to model outperformers.

5.1.2 Evaluating the Population of Fund Performance

Table 5, Panel A shows the parameter estimates for the GMD that describes the alpha population. Panel B reports the estimates for several important population statistics.

The results in Panel A show substantial differences from the results in Table 1, where we first obtain OLS alphas and then estimate the GMD that best describes the fitted alphas. For example, in Table 1, the probability for drawing an alpha from the “bad” group (4.0%) is much lower than the probability in Table 5, Panel A (28.3%). However, conditional on drawing from this group, the alpha realization can be much worse (i.e., negative) for Table 1 than for Panel A, since the “bad” group in Table 1 has both a lower mean (−6.44%, per annum) and a much higher standard deviation (16.95%, per annum) than parameters that govern the “bad” group in Panel A. These differences in parameter estimates reflect the differential treatment of estimation risk between the equation-by-equation OLS and our model.

Since the estimated GMD is composed of two component distributions, it may be difficult to see how the differences in parameters for a single component affect the overall distribution. A better way is to look at the population statistics, as shown in Panel B of Table 5. There are again substantial differences between the results in Table 1 and Panel B. First, by taking estimation risk into account, the overall population mean in Panel B is −1.14% and its 95% confidence bound is [−1.19%, −1.08%]. This estimate of the population mean is significantly higher than the estimate in Table 1 (−1.47%). Both the standard deviation and the inter-quantile range are also much lower in Panel B than in Table 1. Therefore, by taking estimation risk into account, we are able to obtain a more concentrated estimate for the alpha distribution than the equation-by-equation OLS.

³²Barras et al. (2010) study 2,076 funds covering the 1975–2006 period. So their sample is somewhat different from ours. However, given the 23 years overlap between our samples, we believe their estimates should roughly apply to our sample as well.

Table 5: **The Alpha Population: Mutual Funds**

Model estimates and population statistics for mutual funds. For a cross-section of 3,619 mutual funds covering the 1983–2011 period, we estimate our model, which is based on a two-component GMD specification for the alpha population. Assuming the estimated model is the true underlying model, we simulate to find the percentiles of both the parameter estimates and the population statistics. Panel A reports the parameter estimates for the model. μ_l and σ_l are the (annualized) mean and the (annualized) standard deviation for the l -th component normal distribution, and π_l is the probability for drawing from the l -th component, $l = 1, 2$. Panel B reports the estimated population statistics for the alpha distribution. “Mean” is the mean of the alpha distribution. “Standard deviation” is the standard deviation of the alpha distribution. “Interquartile range” is the inter-quartile range of the alpha distribution. “10th percentile” is the 10th percentile of the alpha distribution. The other percentiles are similarly defined. For both Panel A and B, “ $p(5)$ ” and “ $p(95)$ ” report the 5th and 95th percentiles of the variable of interest across simulations, respectively.

Panel A: Parameter Estimates for the Alpha Population			
	Estimate	$p(5)$	$p(95)$
$\mu_1(\%)$	-2.277	-2.301	-1.948
$\sigma_1(\%)$	1.513	1.424	1.654
π_1	0.283	0.280	0.330
$\mu_2(\%)$	-0.685	-0.748	-0.894
$\sigma_2(\%)$	0.586	0.569	0.615
π_2	0.717	0.670	0.720
Panel B: Population Statistics for the Alpha Population			
	Estimate	$p(5)$	$p(95)$
Mean(%)	-1.135	-1.189	-1.075
Standard deviation(%)	1.185	1.121	1.247
Interquartile range(%)	1.142	1.085	1.234
5th percentile(%)	-3.689	-3.803	-3.445
10th percentile(%)	-2.862	-2.966	-2.652
50th percentile(%)	-0.894	-0.935	-0.851
90th percentile(%)	0.012	-0.016	0.096
95th percentile(%)	0.287	0.222	0.390
Fraction of positive alphas	0.106	0.095	0.123

Figure 1 plots the density for the estimated alpha distribution as well as the empirical density for the OLS estimates. The density for the OLS fitted alphas is left skewed, indicating that there are more managers with large negative alphas than there are managers with large positive alphas. Our model estimation picks this up by having a separate component distribution that mostly covers negative alpha values. Allowing multiple component distributions gives our model the flexibility to capture the departure from normality in the data. Our results on model selection also show that it is both necessary (i.e., statistically significant) and sufficient to have this separate component distribution.

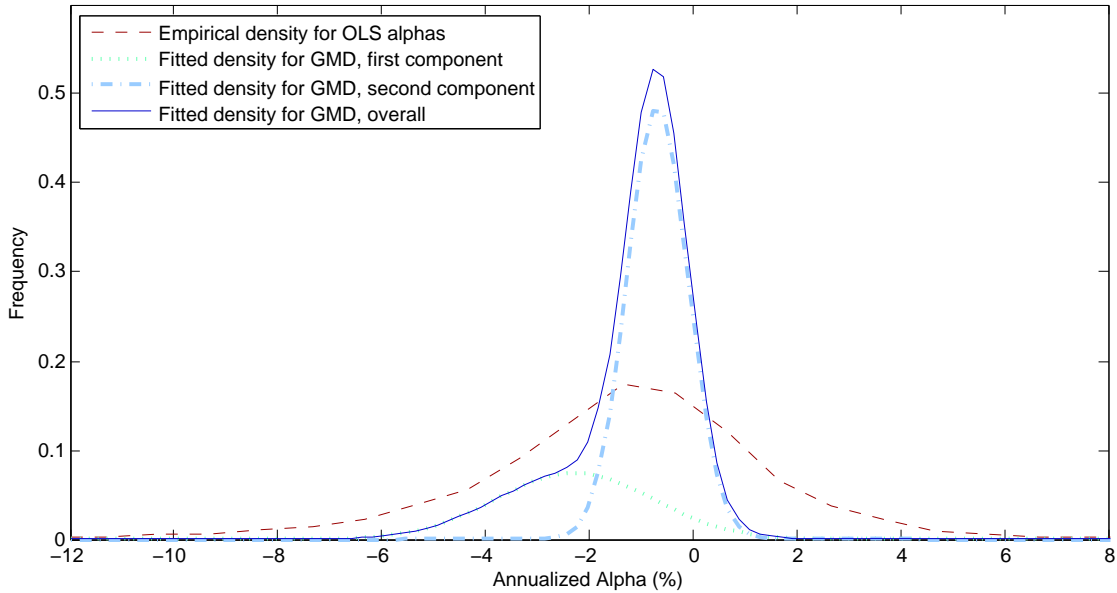
Another important observation from Figure 1 is that our method does not try to fit the OLS alphas. In fact, the overall density for the estimated GMD is more concentrated around its population mean than the empirical density for the OLS alphas. This is because our method allows us to downweigh noisy alpha estimates of individual funds when trying to make inference on the alpha population. Extreme alpha estimates based on OLS are more likely to happen for funds with a short sample, more variable risk loadings, and/or more noisy return residuals. Our structural approach allows us to take these sources of estimation risk into account.

Our method allows us to make inference on important population characteristics by deviating from the usual fund by fund hypothesis testing framework. For example, we estimate the fraction of funds generating positive alphas to be 10.6%. This is in contrast with Barras et al. (2010), who use the multiple testing approach and find that less than 1% of funds generate a positive yet statistically insignificant alpha. To interpret the difference between our results and those in Barras et al. (2010), we need to bear in mind the difference between our method and the usual hypothesis testing. Hypothesis testing, by testing against the null hypothesis that fund alphas are zero, places more prominence on alpha equalling zero than alternative values. Our method assumes that the alpha distribution is continuous and tries to back out this distribution. It is therefore more appropriate to provide inference on population characteristics.

We will likely have more power in identifying alphas with a small magnitude in our framework than hypothesis testing, provided that our parametric assumption of the alpha distribution is a good approximation of reality. For example, for the one-cluster example that we introduced previously, we assume that all the funds in the cross-section generate an alpha of approximately 2% per annum and the standard error for the alpha estimate is about 4%. Under the usual hypothesis testing approach, none of the funds is statistically significant individually. Using our approach, the estimate of the mean of the alpha population would be around 2%. For this example, we think our approach provides a better description of the alpha population. Declaring all the funds to be zero-alpha funds misses important information in the cross-section and leads to a large loss in test power.

Our results shed light on the important question of luck vs. skill for mutual fund managers. For example, we estimate that the 95th percentile of the cross-

Figure 1: **Alpha Distribution for the Mutual Fund Population**



Density plots for the alpha population. For a cross-section of 3,619 mutual funds covering the 1983–2011 period, we estimate our model, which is based on a two-component GMD specification for the alpha population. The solid line shows the density for the estimated GMD. The dotted line shows the density for the first component of the GMD that has a negative mean. The dash-dotted line shows the density for the second component of the GMD that has a positive mean. We also estimate the equation-by-equation OLS. The dashed line shows the empirical density for the fitted OLS alphas.

section of alphas is 0.29% per annum. This number is accurately estimated as the 95% confidence interval is from 0.22% to 0.39%.³³ Therefore, at least 5% of funds are generating a positive alpha. From the hypothesis testing perspective, 0.29% is not a big alpha and would likely to be overwhelmed by the standard error for a typical fund. This would lead to an insignificant t -statistic and the conclusion that almost no fund has skill, either from a single testing or a multiple testing perspective. Our model offers a different way to interpret this 0.29%. Since we have a large number of funds clustered around the 95th percentile in terms of fund performance, pooling information across these funds should give us a good estimate of the average performance among these funds. It is true that viewed in isolation, none of these funds seems to display a significant alpha. But it would be misleading to conclude

³³In our framework, the 95th percentile of 0.29% is *significant* given that the lower bound of the 95% confidence interval is above zero. However, our interpretation of significance should not be confounded with the significance of individual funds that belong to the top 5% of alphas from the perspective of the usual fund by fund hypothesis testing.

that all of these funds are zero-alpha funds. A superior approach is to recognize this population structure and explicitly model and estimate the underlying alpha distribution that individual alphas are drawn from.

Notice that our estimate (0.29%) of the 95th percentile of the alpha distribution is substantially lower than the estimate (3.07%) based on equation-by-equation OLS. This stems from the shrinkage effect that we mentioned previously. Two features contribute to the shrinkage effect. First, since the median fund generates a negative alpha, cross-sectional learning forces us to pool alphas that are different from the population mean towards the population mean. Second, large positive alphas are usually generated with a higher level of residual standard deviation than large negative alphas with the same magnitude. For example, the mean residual standard deviation for funds with alphas above the 95th percentile (i.e., 3.07%) is 7.4% (per annum) whereas the mean residual standard deviation for funds with alphas below -3.07% is 5.9%. Intuitively, in a competitive market, it is more difficult to generate a positive alpha than a negative alpha with the same magnitude. As a result, our method downweights the time-series information of funds with positive alphas more aggressively than funds with negative alphas with the same magnitude. These two features reinforce each other and generate the large discounts for positive alphas within our structural framework.

Linking to the existing literature, three approaches are proposed to evaluate mutual fund performance. The first method uses the extreme test statistics and tries to evaluate the significance of the best/worst funds, while controlling for test multiplicity (see, for example, Kosowski et al. 2006, Fama and French, 2010, Harvey and Liu, 2015a). It is based on hypothesis testing and its null hypothesis is that each fund has a zero alpha. It is designed to answer the question of whether there exists any fund that significantly outperforms/underperforms and cannot further classify funds into different performance groups. Using this approach, Kosowski et al. (2006) find that there exist managers that significantly outperform. Refining the method in Kosowski et al. (2006) to control for cross-sectional dependency, Fama and French (2010) find no outperforming funds.

The second approach tries to classify funds into broad categories. Papers that follow this approach include Barras et al. (2010) and Ferson and Chen (2015). The assumption of this approach is less stringent than the assumption under the previous approach in that not all funds need to have a zero alpha. Certain funds can have nonzero alphas and this approach tries to control the false discovery rate at 5%. Using this approach, Barras et al. (2010) find that about 75% of funds are zero-alpha funds. Ferson and Yong (2015) refine this method by allowing a non-zero probability for true alphas to disguise themselves as zero, and find that 50% or fewer have zero alphas. Neither paper finds evidence of funds that significantly outperform.

From an methodological perspective, there are several important differences between our approach and the false classification (FC) method in Barras et al. (2010) and Ferson and Chen (2015). The FC approach, being essentially a variant of the

usual hypothesis testing framework, postulates that fund alphas can only take a small number of values. While this offers a simplification of the inference problem, there is no particular reason to think that fund alphas can only take a few specific values. As a result, if a fund has a true alpha that is very different from these assumed values, the estimation error by assigning this fund to any particular alpha group might be large. Our approach allows us to realistically model the alpha population as following a continuous distribution, thereby reducing the estimation error in the FC approach where fund alphas are forced to take a small number of values.

Second, the loss functions in our approach and the FC method are different. FC relies on the multiple hypothesis testing approach and aims to strike a balance between Type I (i.e., false discovery rate) and Type II error rates. Our maximum likelihood-based approach tries to find the best parametric model that fits the data through optimally weighting the likelihood from fitting the panel of return time-series and the likelihood from fitting the cross-section of alphas. Hence, a material advantage of our framework is that it allows us to take into account the parameter uncertainty in estimating both fund alphas and other OLS parameters (i.e., factor loadings and residual standard deviations) when we try to fit the cross-section of estimated alphas. On the other hand, our structural approach also allows us to address the Type I error concern that is the focus of the FC method. In particular, assuming all funds have a zero alpha, if we estimate the alphas of a thousand funds, on average 25 funds will appear to have a significant positive alpha from a single test perspective. In our framework, these 25 funds will likely not have a significant positive alpha as the posterior distribution of alpha weights the information from the time-series (which is what the single test p -values are based on) by using information from the alpha cross-section. Since our estimate of the mean of the alpha population will likely be zero, learning across funds allows us to downwardly adjust the significance of each individual fund, leading us to correctly declare the 25 funds as insignificant. Equation (15) shows the precise formula for how our model adjusts the statistical significance of individual funds when the alpha population has a zero mean.

The third approach, as taken by our paper, is to treat alphas as continuous and try to estimate the underlying distribution for alphas. We deviate from the usual hypothesis testing approach in that we do not think an alpha of zero is any different than an alpha of other value. Another salient feature of our model is that we take various sources of estimation risk into account.

One can think of the three approaches as following an order that tries to obtain a finer and finer understanding of the alpha distribution. The first approach tries to answer the very basic question of whether there exists any fund that has a non-zero alpha. If the answer is yes, we proceed to the second approach to classify funds into broad categories. Finally, viewing alphas as coming from an underlying distribution, we use the third approach to provide a more precise description of this distribution.

Fundamentally, our approach is different from the first two approaches that rely on fund by fund hypothesis testing. In the context of performance evaluation, we have

multiple funds in the cross-section so we have to perform multiple hypothesis tests. However, compared to a single hypothesis test, there are many pitfalls associated with performing multiple hypothesis tests, some of which are not well understood by the literature. For example, the definition of test power is ambiguous given the multi-dimensional nature of the hypothesis testing problem.

Viewing fund alphas as coming from an underlying distribution, our model estimates suggest that mutual fund managers are doing better than what people have previously thought. We estimate that more than 10% of funds are generating a positive alpha. Our estimate is higher than those reported in the literature and likely due to the fact our structural approach has more power in identifying small but non-negligible alphas. If decreasing return to scale were the underlying economic mechanism that drives alpha dynamics (Berk and Green, 2004), then small but positive alphas are usually associated with large funds. Given that larger funds have a greater impact on the mutual fund industry than smaller funds, it would be a mistake to label these funds as zero alpha funds from an economic perspective.

5.1.3 Individual Fund Evaluation: In-sample

We use our estimated model to make inference on the alphas of individual funds. Given a set of parameter estimates, which use the information from the cross-section of funds, we are able to refine the alpha estimate of an individual fund that is based on time-series information alone, providing a more informative alpha estimate for an individual fund. The intuition is given in the one-cluster and two-cluster examples that we introduced previously. For example, for the two-cluster example, we assume that half of the funds have an alpha estimate of approximately 2% per annum and the standard error for the alpha estimate is about 4%. The other half have an alpha estimate of approximately -2% per annum and also have a standard error of about 4%. Our model is able to recognize the two-cluster structure of the alpha population. Knowing that the alphas cluster at -2% and 2% with equal probabilities, we will pull the estimate of a negative alpha towards -2% and a positive alpha towards 2% , and both away from zero.

The formulas that provide density forecasts for individual funds are given in (12)-(14). We compare our model with the equation-by-equation OLS both from an in-sample fit and an out-of-sample forecasting perspective.

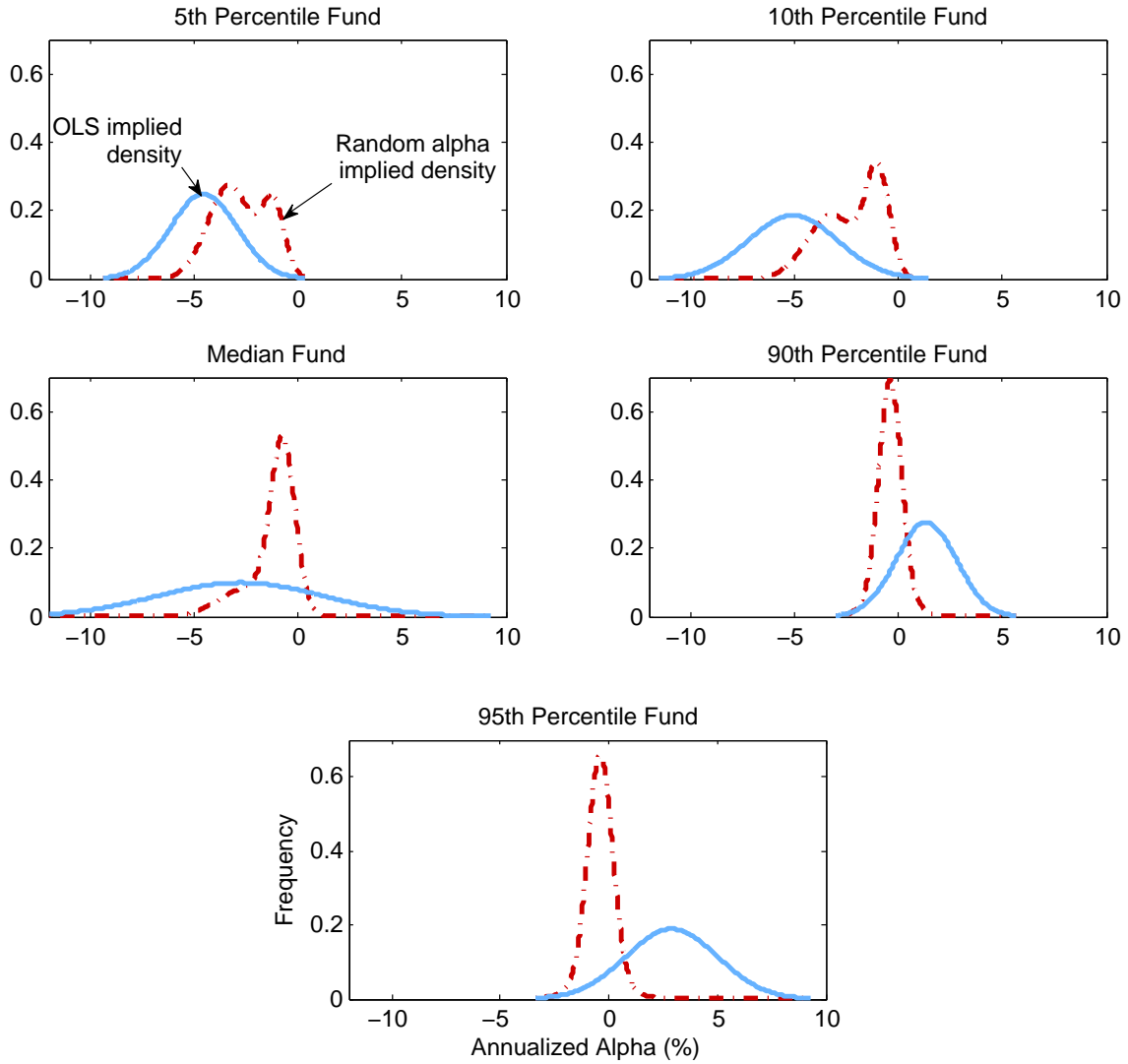
Focusing on in-sample fitting, Figure 2 shows the density forecasts based on our model for several exemplar funds. In particular, we rank funds by the t -statistics of their OLS alpha estimates and choose several funds that represent different percentiles of the cross-section of t -statistics.

We see several noticeable differences between our density forecasts and the forecasts based on OLS. First, there is a shrinkage effect where the means of our forecasts pull the OLS means towards the overall population mean. This is the cross-sectional

learning effect that we mentioned previously. Knowing the alpha distribution of other funds helps us make better inference on the alpha of a particular fund. Its OLS alpha estimate based on time-series information alone needs to be adjusted for the information in the cross-section. The shrinkage effect seems particularly strong for funds with large positive OLS alphas. This is because we are more likely to observe a negative alpha than a positive alpha for the alpha population. In addition, as we mentioned previously, large positive alphas are usually associated with a larger residual standard deviation than negative alphas with the same magnitude. The cross-sectional learning effect therefore shrinks a positive alpha towards the population mean by more than what it shrinks a negative alpha with the same magnitude towards the population mean.

Second, the dispersion for the density forecast of our model is uniformly lower than that based on the OLS density forecast. This is consistent with our simulation study where we show that the average length of the confidence interval based on our method is substantially lower than that based on the OLS. Intuitively, our density forecast combines information from both the cross-section and the time-series so it is less dispersed than the OLS density forecast, which only uses the time-series information. (13) makes this intuition more precise. Suppose we have a single component distribution for the GMD, then the variance of a fund's alpha estimate following our approach is always smaller than its variance based on time-series information alone.

Figure 2: Alpha Distributions for Individual Mutual Funds



Density plots for individual funds. For a cross-section of 3,619 mutual funds covering the 1983–2011 period, we estimate our model, which is based on a two-component GMD specification for the alpha population. We also estimate the equation-by-equation OLS. We rank the cross-section of funds based on the t -statistics of their OLS alpha estimates and choose five funds whose t -statistics are the closest to the 5th, 10th, 50th, 90th, and 95th percentiles of the cross-section of t -statistics. Based on our model estimate, we plot the density estimates for these funds using (12)–(14). We also plot the density estimates for the OLS alphas.

Table 6: **Differences in Density Forecasts between OLS and the Random Alpha Model**

Differences in density forecasts between the OLS and the random alpha model. For a cross-section of 3,619 mutual funds covering the 1983–2011 period, we estimate our model, which is based on a two-component GMD specification for the alpha population. We also estimate the equation-by-equation OLS. We group funds into several groups based on the t -statistics of their OLS alpha estimates (denoted as t_{α}^{OLS}). We calculate the average difference in point estimates and confidence intervals between the random alpha model and the OLS model. “Diff. in mean” reports the average difference in the mean forecast between our model and the OLS. “% diff. in CI(90)” and “% diff. in CI(95)” report the percentage differences in the length of the 90% and 95% confidence intervals between our model and OLS, respectively. “# of funds” reports the number of funds for each t -statistic category.

t_{α}^{OLS}	Diff. in mean (%)	% diff. in CI(90)	% diff. in CI(95)	# of funds
$(-\infty, -2.0)$	3.391	−30.8%	−32.7%	523
$[-2.0, -1.5)$	2.352	−42.1%	−42.5%	391
$[-1.5, 0)$	0.688	−54.9%	−53.3%	1,640
$[0, 1.5)$	−2.052	−63.8%	−61.7%	906
$[1.5, 2.0)$	−3.774	−63.2%	−61.0%	98
$[2.0, \infty)$	−5.722	−64.5%	−61.8%	61

Finally, our density forecasts display non-normality, especially for funds with a negative mean estimate for alpha. For funds with a positive mean estimate, although the density looks unimodal, it is still a mixture distribution of two normal densities. This shows the flexibility of the GMD specification to capture different shapes of a probability density function. It also makes sense to have a non-normal density forecast for individual funds if the underlying distribution for the alpha population is non-normally distributed. If this underlying distribution is more heavy-tailed and skewed than the normal distribution, then the density forecasts for individual funds should be able to reflect these non-normal features for the alpha population.

Table 6 summarizes the differences in both point estimates and confidence intervals between our model and the OLS. We group funds into different categories based on their OLS t -statistics and calculate the average difference between our model estimates and the OLS model estimates.

Focusing on the mean estimates, we see the differential impact of the shrinkage effect across different t -statistic groups. For example, for funds with an OLS t -statistic below -2.0 , on average our model pulls the OLS alpha estimate closer to zero by 3.4% per annum. At the other extreme, for funds with significantly positive

OLS alpha estimates (i.e., OLS t -statistic > 2.0), we on average pull their alpha estimates closer to zero by 5.7% per annum. The shrinkage effect seems to be more pronounced for funds with large positive alpha estimates. This is attributable to, as we mentioned previously, the differential treatment of positive and negative alphas by the cross-sectional learning effect since we are more likely to observe a negative alpha than a positive alpha for the alpha population and a large positive alpha is usually generated with more uncertainty than a negative alpha with the same magnitude.

For confidence intervals, our model is able to shrink the 90% and 95% confidence intervals by at least 30% of the corresponding OLS confidence intervals. The reductions in estimation uncertainty seem substantial and are consistent with our results in the simulation study (see Table 4), in which we show that the reduction in the length of the confidence interval is not accompanied by a loss in the coverage rate. In fact, we are able to achieve a pre-specified coverage rate (i.e., 90% or 95%) with a much shorted confidence interval.

The difference between Table 6 and Table 4 is that, unlike in the simulation study, we no longer observe the true alpha for each individual fund. To better assess the power of our approach, we perform an out-of-sample forecasting exercise in the next section.

5.1.4 Individual Funds Evaluation: Out-of-sample

We perform an out-of-sample analysis of our method by splitting our data into an in-sample estimation period and an out-of-sample holdout period. Notice that this is not a true out-of-sample test as we have experienced the data. One way to interpret our results is to assume that someone tries to assess the predictive power of our model by following a simple strategy. She estimates our model at the end of the in-sample period and uses the model estimates to forecast returns for the out-of-sample period. We try to evaluate such a strategy from a historical perspective.

Our sample runs from 1984 to 2011. We partition our sample into two parts, with the first two-thirds as the estimation period and the last one-third as the out-of-sample testing period. This way of partitioning the sample makes sure that we have a long enough in-sample period to have a reasonable model estimate.

For the in-sample period (i.e., 1984-2001), we estimate both our model and the equation-by-equation OLS. Based on our model estimates, we construct a density forecast for each fund's alpha and use the mean of this density forecast to predict fund alpha in the future. For OLS, we use its in-sample alpha estimate to forecast its alpha in the future. The future alpha for each fund is obtained by running equation-by-equation OLS for the out-of-sample period (i.e., 2002-2011). Notice that the out-of-sample alpha is an estimated alpha and may not represent the true alpha.

For the in-sample period (i.e., 1984-2001), similar to our requirement for the full-sample estimate, a fund needs to have at least eight monthly observations to be considered in our estimation. This leaves us with 1,765 funds. Additionally, in order to have a valid alpha proxy for the out-of-sample period, we again require a fund to have at least eight monthly observations for the out-of-sample period. This further requirement leaves us with 1,488 funds for the out-of-sample period. To sum up, our in-sample estimation is based on 1,765 funds. Among these funds, 1,448 will be used in out-of-sample testing.

Table 7, Panel A shows the in-sample model estimates, and Panel B shows the out-of-sample forecasting performance. Focusing on Panel A, there are noticeable differences between the parameter estimates for the 1984-2001 period and for full sample period (see Table 5). Compared with the estimates in Table 5, it is less likely (drawing probability = 1.2%) to draw the alpha from the group with a very negative mean. However, conditional on drawing from this group, the alpha dispersion (15.15%) is much higher than the corresponding dispersion in Table 5 (1.51%). For the group with a mildly negative mean, its mean (-0.35%) is higher than the corresponding mean in Table 5 (-0.69%). At least two factors contribute to these differences in model estimates. First, the average fund return (and OLS alpha) is significantly higher for the in-sample period than for the full sample period. Second, compared to the full sample estimation, we have fewer funds for the in-sample estimation. This implies a lesser degree of learning across funds and may cause a larger estimate for the dispersion of the alpha distribution. Despite these differences between the subsample and the full sample estimation, it remains interesting to see how our model performs out-of-sample.

Panel B shows the out-of-sample forecasting results. We again group funds based on their in-sample OLS t -statistics and present the average forecast error for each group. Our model seems to provide a better alpha forecast for all except one group of funds. The improvement of our model over the OLS is substantial. For example, for the 610 funds that have an in-sample t -statistic between zero and 1.5, our model is able to reduce the average forecast error from 5.54% to 2.61% (per annum). The reduction in forecast error is more pronounced for funds with large (absolute) OLS t -statistics. This is consistent with our finding based on the full sample estimation that the shrinkage effect is stronger for funds with large (absolute) OLS t -statistics. Across all groups of funds, the average percentage reduction in forecast error is 48% ($= (5.17\% - 2.71\%)/5.17\%$). Therefore, our model is able to provide substantially better out-of-sample alpha forecasts compared to the OLS model.

Table 7: **Out-of-sample Forecasts for Mutual Funds**

In-sample model estimates (1984-2001) and out-of-sample forecasts (2002-2011) based on OLS and the random alpha model. We partition the mutual fund data into two parts and use the first part (1984-2001) for in-sample model estimation and the second part (2002-2011) for out-of-sample testing. For the in-sample period, we require a fund to have at least eight monthly observations. This leaves us with 1,765 funds. We estimate both our model and the equation-by-equation OLS based on these 1,765 funds. Panel A shows the parameter estimates for the random alpha model. μ_l and σ_l are the (annualized) mean and the (annualized) standard deviation for the l -th component normal distribution, and π_l is the probability for drawing from the l -th component, $l = 1, 2$. For out-of-sample testing, we additionally require a fund to have at least eight monthly observations for the out-of-sample period. 1,448 out of the 1,765 funds satisfy this additional requirement. We evaluate the out-of-sample forecasting performances of models based on these 1,448 funds. In particular, based on the in-sample estimates for our model, we construct a density forecast for each fund's alpha and use the mean of this density forecast to predict fund alpha in the future. For OLS, we use its in-sample alpha estimate to forecast its alpha in the future. The future alpha for each fund is obtained by running equation-by-equation OLS for the out-of-sample period. Panel B shows the forecasting results for both OLS and the random alpha model. " t_{α}^{OLS} " denotes the in-sample t -statistic for the alpha estimate of the OLS model. "OLS forecast error (%)" calculates the average absolute forecast error (i.e., the alpha forecast based on the in-sample model minus the out-of-sample OLS alpha estimate) for OLS within a group of funds. "RA forecast error (%)" calculates the average absolute forecast error for the random alpha model within a group of funds.

Panel A: In-sample Model Estimates, 1984–2001

Parameters	Estimate
μ_1 (%)	-2.935
σ_1 (%)	15.146
π_1	0.012
μ_2 (%)	-0.354
σ_2 (%)	1.065
π_2	0.988

Panel B: Out-of-sample Forecasts, 2002–2011

In-sample, t_{α}^{OLS}	OLS forecast error (%)	RA forecast error (%)	# of funds
$(-\infty, -2.0)$	6.613	3.286	64
$[-2.0, -1.5)$	3.699	3.089	75
$[-1.5, 0)$	2.916	2.748	565
$[0, 1.5)$	5.542	2.606	610
$[1.5, 2.0)$	10.469	2.381	87
$[2.0, \infty)$	12.022	2.766	87
Overall	5.165	2.710	1,488

6 Other Issues

6.1 The Bayesian Approach

Bayesian methods have been applied to study fund performance. For example, Pástor and Stambaugh (2002) and Kosowski, Naik and Teo (2007) use information in seemingly unrelated assets to improve on the precision of performance estimates. Baks, Metrick, and Wachter (2001) make inference on mutual funds' alphas using informative priors about individual fund alphas. However, these studies focus on the inference of each individual fund and cannot make inference on the overall alpha population. Moreover, they do not allow us to learn from the entire alpha population to refine the estimate of each individual fund's alpha. As a result, there is a loss of information in making efficient inference on fund alphas.

Another concern for this strand of literature, as pointed out by Jones and Shanken (2005) and Busse and Irvine (2006), is that the prior specification greatly affects the predictive accuracy of Bayesian alphas. Many mutual funds in our sample have a short time-series. The estimation uncertainty for the alpha seems high relative to its point estimate, making the absolute value of the t -statistic small. In this situation, a prior specification for alphas, no matter how uninformative it is, will likely weigh heavily on the estimation of fund alphas. However, among all the prior specifications one can choose, which one is the best? It is data mining (or model mining) in nature if we chose the best prior that seems to fit the data, either in-sample or out-of-sample. Therefore, although the Bayesian approach implies a shrinkage effect that is similar to ours, the inherent subjectivity of the choice of the prior and the potentially large impact of this choice on inference makes us hesitant to apply Bayesian methods to performance evaluation. Our model offers a frequentist framework and does not rely on the choice of a prior distribution.

Among the papers that apply Bayesian methods, Jones and Shanken (2005) is the closest to ours. They specify a normal prior for the alpha population in the cross-section and allow diffuse and heterogeneous priors on the other OLS parameters (that is, risk loadings and residual variances). As for most Bayesian models, the choice of the diffuse prior, or any kind of uninformative prior, is not without consequences. In particular, Kass and Wasserman (1996) show that it is a dangerous practice to put faith in any default choice of prior, especially when the sample size is small (relative to the number of parameters). The issue seems particularly relevant for the estimation of risk loadings since we usually have a short time-series for fund returns (e.g., 24.5% of mutual funds in our sample have no greater than 36 return observations). Any distortion resulting from the prior specifications of the cross-section of risk loadings will feed into the estimation of the alpha population. In contrast, our model follows a frequentist framework and does not require any prior knowledge about parameters of interest.

The second advantage of our model is that it allows the use of the GMD to flexibly model the alpha population. This is not a trivial extension of the single normal prior assumption in Jones and Shanken (2005). As documented by Barras, Scaillet, and Wermers (2010), Ferson and Chen (2015), and Chen, Cliff, and Zhao (2015), investment fund managers are better classified as coming from a few subpopulations. It is therefore important to have a parametric specification for the alpha distribution that is able to accommodate this subpopulation structure. The Bayesian framework per se does not preclude a multi-population modeling of the alpha population. However, it is not clear how to impose an uninformative prior while at the same time generating a posterior distribution that features a multi-population structure. In addition, a multi-population specification will likely force us to use non-conjugate priors, which will significantly increase the computational burden of the Bayesian methods.

6.2 Sample Selection Bias

As with all approaches to performance evaluation, sample selection may bias our results. On the one hand, studies that condition on fund survival overestimate fund performance, see Brown, Ibboson, Ross (1992), Elton, Gruber, and Blake (1996), and Carhart, Carpenter, Lynch, and Musto (2002). On the other hand, reverse-survivorship may understate fund performance, see Linnainmaa (2013).

We believe the bias will likely be smaller in our framework compared to the standard equation-by-equation OLS. For example, when there is reverse-survivorship bias, a skilled fund may drop out of sample after having a bad (unlucky) shock. This makes its in-sample alpha an understatement of its true population value. Hence, using the equation-by-equation OLS, if we take the average of the cross-section of fitted alphas, this average will underestimate the overall population mean if there is reverse-survivorship bias. Funds that have a shorter history and a higher level of idiosyncratic volatility are more likely to drop out after experiencing a bad shock. In our framework, the importance of these funds is downwardly weighted. We know their alpha estimates are more noisy so we put less weight on them in terms of learning about the alpha population.

6.3 Random Alpha Model vs. Multiple Hypothesis Testing

By treating the alpha of an investment fund as random, our model takes into account the cross-sectional uncertainty in alpha from a population perspective and helps deflate the fund alpha and its t -statistic, thereby imposing a more conservative inference on the fund alpha. This is consistent with the idea of multiple testing that has been applied to performance evaluation (see, Barras et al., 2010, Fama and French, 2010, and Ferson and Chen, 2015) and to asset pricing in general (see, Harvey, Liu, and

Zhu, 2016, and Harvey and Liu, 2015b,c). What is the connection between the two methods?

Suppose a researcher wants to test the effectiveness of a drug for all patients. The researcher divides the sample into a female group and a male group and separately tests the effectiveness of the drug. Since two tests have been tried, the chance of finding a significant result is higher than the case with a one shot test. The researcher can apply a multiple testing adjustment to these two tests so that the overall error rate, however defined, is controlled at a pre-specified level. However, it does not make sense to use the model in our paper since there are a limited number of gender types in the population (i.e., we do not have hundreds of gender types). It is not appropriate to view the means of the two groups — male and female — as coming from an underlying distribution as there are only two samples from this distribution.

The random alpha model applies when it is plausible to view the objects in the cross-section as coming from a certain underlying population. For fund alphas, it makes sense to think that the alphas for different funds are not independent of each other since there are limited investment opportunities in the financial market and funds compete with each other to generate alphas.³⁴

Despite their similarities in discounting fund alphas and their t -statistics, the two models are fundamentally different. The multiple testing approach, and hypothesis testing in general, treats the fund alpha as a dichotomous variable (that is, zero vs. nonzero). Its objective function is also about controlling the probability or the fraction of false discoveries, that is, a zero alpha fund being incorrectly classified as a nonzero fund. On the other hand, the random alpha model preserves the continuity of the alpha distribution. Its objective function is the goodness-of-fit of a parametric model to the data. While the hypothesis testing framework is useful to roughly classify investment managers into different groups, the random alpha model is designed to provide inference on the alpha population as well as refining inference about a particular fund.

6.4 Misspecification of the Factor Model

Inference on fund alphas both at the population and at the individual fund level is contingent upon the benchmark model being used. For instance, for mutual funds performance evaluation, suppose the true benchmark model a five-factor model that includes the Fama and French (1993) and Carhart (1997) four factors. Then misspecifying the benchmark model as the four-factor model will likely lead to biased alpha estimates, both for the alpha population and for the individual funds.³⁵

³⁴See French (2008) for a similar argument on the competitiveness of the investment funds industry.

³⁵Harvey and Liu (2016a) examine the distortions in asset pricing tests when factor models are misspecified.

The concern about model risk is to some extent alleviated by considering the random alpha model. Using the aforementioned five-factor model example, suppose the fifth factor — the factor that is missing from the four-factor model — only applies to a small fraction of funds. By using a misspecified four-factor model, the equation-by-equation OLS will imply biased alpha estimates for this small fraction of funds. Under the random alpha model, we are able to learn from the entire cross-section of funds, including those that are not exposed to the fifth factor. As a result, the bias in the alpha estimates for the small fraction of funds that are exposed to the fifth factor is likely to be lower under the random alpha model than under the OLS model.

When the benchmark model is missing a factor that applies to the majority of funds, it is unlikely that any performance evaluation model performs well. One therefore needs to be cautious when trying to interpret the results of our paper. Our inference relies on a pre-specified benchmark model for performance evaluation and is sensitive to this choice. Harvey and Liu (2016b) explore this issue in greater detail and provide alpha forecasts that take into account the choice of the benchmark model.

Another possible misspecification of the factor model assumes a constant beta while the true beta is time-varying. If fund-level characteristics and macroeconomic variables can be used as instruments to model time-varying betas, then the static factor model considered in our current paper would be missing factors that interact these instruments with the benchmark factors. Harvey and Liu (2016c) study the impact of beta variability for performance evaluation adapting the framework in this paper to model dynamic risk exposures.

6.5 Time-varying Alphas

While our paper focuses on unconditional alphas, we can use fund-level characteristics as instruments to study conditional alphas. Jones and Mo (2016) show that a number of firm characteristics help forecast the cross-section of fund alphas. They also find that the performance of many of these characteristics in explaining fund alphas deteriorates through time. Our model can be easily extended to take into account the predictability and the variation in predictability of fund returns by using fund characteristics. Our framework allows one to make inference by drawing information from the entire cross-section, which can potentially improve the out-of-sample predictability of fund alphas. This is further explored in Harvey and Liu (2016b).

7 Conclusions

How do we evaluate investment fund managers? This is a question that bears important economic consequences for wealth management and capital reallocation. Our

paper proposes a structural estimation approach to answer this question. Viewing fund alphas as coming from an underlying population, our model first backs out the distribution of the alpha population and then uses this distribution to refine the alpha estimate for each individual fund. By drawing on information from the cross-section of alphas, we show that our model is able to generate more accurate alpha estimates, both in-sample and out-of-sample.

The idea of our model is likely to be useful for other applications. Essentially, when there is cross-sectional heterogeneity and when it is appropriate to view the effects as coming from a certain population, we can apply our model to make inference on both the population and the individual effects. Our use of the GMD is also flexible enough to approximate a variety of parametric distributions for the population.

Our framework can be extended along several important directions. First, while we treat the alpha of a particular fund as fixed across time, we can relax this assumption by allowing fund alphas to be time-varying. This allows us to study performance persistence from a population perspective. Second, to capture the time variation in risk loadings, we can allow betas to be time-varying as well, possibly through the dependence of fund risk loadings on macroeconomic and financial variables. We expect the random alpha model framework to be a fruitful area of future research for performance evaluation and asset pricing in general.

References

- Avramov, D., and R. Wermers. 2005. Investing in mutual funds when returns are predictable. *Journal of Financial Economics* 81, 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, 45–85.
- Barras, L., O. Scaillet, and R. Wermers. 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance* 65, 179–216.
- Bekaert, G., and C. R. Harvey. 1995. Time-varying world market integration. *Journal of Finance* 50, 403–444.
- Berk, J. B., and R. C. Green. 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112, 1269–1295.
- Bickel, P. J., and B. Li. 2006. Regularization in statistics. *Test* 15, 271–344.
- Booth, J. G. and J. P. Hobert. 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society. Series B.* 61, 265–285.
- Brown, S. J., W. Goetzmann, R. G. Ibbotson, and S. A. Ross. 1992. Survivorship bias in performance studies. *Review of Financial Studies* 5, 553–580.
- Busse, J. A., and P. J. Irvine. 2006. Bayesian alphas and mutual fund persistence. *Journal of Finance* 61, 2251–2288.
- Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82.
- Carhart, M. M., J. N. Carpenter, A. W. Lynch, and D. K. Musto. 2002. Mutual fund survivorship. *Review of Financial Studies* 15, 1439–1463.
- Chen, J., D. Zhang, and M. Davidian. 2002. A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics* 3, 347–360.
- Chen, Y., M. Cliff, and H. Zhao. 2015. Hedge funds: The good, the bad, and the lucky. *Journal of Financial and Quantitative Analysis*, Forthcoming.
- Cohen, A. C. 1967. Estimation in mixtures of two normal distributions. *Technometrics* 9, 15–28.
- Cohen, R. B., J. D. Coval, and L. Pástor. 2005. Judging fund managers by the company they keep. *Journal of Finance* 60, 1057–1096.

- Day, N. E. 1969. Estimating the components of a mixture of normal distributions. *Biometrika* 56, 463–474.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39, 1–38.
- Elton, E. J., M. J. Gruber, and C. R. Blake. 2001. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases, *Journal of Finance* 56, 2415–2430.
- R. B., Evans. 2010. Mutual fund incubation. *Journal of Finance* 65, 1581–1611.
- Fama, E. F., and K. R. French. 2010. Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance* 65, 1915–1947.
- Fan, J., and J. Lv. 2010. A selective overview of variable selection in high dimensional feature space. *Statistical Sinica* 20, 101–148.
- Feng, Z. D., and C. E. McCulloch. 1996. Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society. Series B*, 609–617.
- Ferson, W., and Y. Chen. 2015. How many good and bad fund managers are there, really? *Working Paper*.
- Figueiredo, M. A., and A. K. Jain. 2002. Unsupervised learning of finite mixture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24, 381–396.
- French, K. 2008. Presidential address: The cost of active investing. *Journal of Finance* 63, 1537–1573.
- Gray, S. F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27–62.
- Greene, W. H. 2003. *Econometric analysis*. Pearson Education India.
- Greg, C., G. Wei, and M. A. Tanner. 1990. A monte carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association* 85, 699–704.
- Harvey, C. R., Y. Liu, and H. Zhu. 2016. ... and the cross-section of expected returns. *Review of Financial Studies* 29, 5–72.
- Harvey, C. R., and Y. Liu. 2015a. Luck vs. skill and factor selection. in *The Fama Portfolio*, John Cochrane and Tobias J. Moskowitz, ed., Chicago: University of Chicago Press.
- Harvey, C. R., and Y. Liu. 2015b. Lucky factors. *Working Paper*. Available at <http://ssrn.com/abstract=2528780>.

- Harvey, C. R., and Y. Liu. 2015c. Backtesting. *Journal of Portfolio Management* 42, 13–28.
- Harvey, C. R., and Y. Liu. 2015d. A structural approach to factor selection. *Work in Progress*.
- Harvey, C. R., and Y. Liu. 2016a. Factor model uncertainty and asset pricing tests. *Work in Progress*.
- Harvey, C. R., and Y. Liu. 2016b. Predicting alpha. *Work in Progress*.
- Harvey, C. R., and Y. Liu. 2016c. Real-time performance benchmarking. *Work in Progress*.
- Huij, J., and M. Verbeek. 2007. Cross-sectional learning and short-run persistence in mutual fund performance. *Journal of Banking & Finance* 31, 973–997.
- Jensen, M. C. 1968. The performance of mutual funds in the period 1945-1964. *Journal of Finance* 23, 389–416.
- Jensen, M. C. 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* 42, 167–247.
- Jones, C., and J. Shanken. 2005. Mutual fund performance with learning across funds. *Journal of Financial Economics* 78, 507–552.
- Jones, C., and H. Mo. 2016. Out-of-sample performance of mutual fund predictors. *Working Paper*.
- Kass, R. E., and L. Wasserman. 1996. The selection of prior distributions by formal rules. *Journal of the American Statistical Association* 91, 1343–1370.
- Linnainmaa, J. T. 2013. Reverse survivorship bias. *Journal of Finance* 68, 789–813.
- Maddala, G. S. 2001. Introduction to econometrics, John Wiley and Sons Ltd. *West Sussex, England*.
- McCulloch, C. E. 1997. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* 92, 162–170.
- McLachlan, G. and T. Krishnan. 2007. *The EM algorithm and extensions*. Vol. 382. John Wiley & Sons, 2007.
- Neal, R., and G. Hinton. 1998. A view of the EM algorithm that justifies incremental, sparse, and other variants. In Jordan, M., editor, *Learning in Graphical Models*. Kluwer Academic Press.
- Pástor, L., and R. Stambaugh. 2002a. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics* 63, 315–349.

Pástor, L., and R. Stambaugh. 2002b. Investing in equity mutual funds. *Journal of Financial Economics* 63, 351-380.

Stambaugh, R. 2003. Inference about survivors. Unpublished working paper. Wharton School, University of Pennsylvania.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. John Wiley & Sons, New York.

Vidaurre, D., C. Bielza, and P. Larrañaga. 2013. A survey of L_1 regression. *International Statistical Review* 81, 361-387.

Wu, C. F. J. 1983. On the convergence properties of the EM algorithm. *Annals of Statistics* 11, 95-103.

A Implementing the EM Algorithm

A.1 Step II: Characterizing $f(\mathcal{A}|\mathcal{R}, \mathcal{G}^{(k)})$

Using Bayes' law, we have:

$$f(\mathcal{A}|\mathcal{R}, \mathcal{G}^{(k)}) \propto f(\mathcal{R}|\mathcal{A}, \mathcal{G}^{(k)})f(\mathcal{A}|\mathcal{G}^{(k)}). \quad (\text{A.1})$$

Given the independence of the residuals and the α_i 's, the right-hand side of (A.1) is the product of the likelihoods of all funds, i.e.:

$$f(\mathcal{R}|\mathcal{A}, \mathcal{G}^{(k)})f(\mathcal{A}|\mathcal{G}^{(k)}) = \prod_{i=1}^N f(R_i|\alpha_i, \mathcal{G}^{(k)})f(\alpha_i|\mathcal{G}^{(k)}).$$

Therefore, to characterize $f(\mathcal{A}|\mathcal{R}, \mathcal{G}^{(k)})$, it is sufficient for us to determine $f(R_i|\alpha_i, \mathcal{G}^{(k)})f(\alpha_i|\mathcal{G}^{(k)})$ for each fund i . For ease of exposition, we use \mathcal{G} and $\mathcal{G}^{(k)}$ interchangeably to denote the known parameters at the k -th iteration.

Under normality, we have

$$\begin{aligned} f(R_i|\alpha_i, \mathcal{G}^{(k)}) &\propto \exp\left\{-\frac{\sum_{t=1}^T (r_{it} - \alpha_i - \beta'_i f_t)^2}{2\sigma_i^2}\right\}, \\ &\propto \exp\left\{-\frac{[\alpha_i - \frac{\sum_{t=1}^T (r_{it} - \beta'_i f_t)}{T}]^2}{2\sigma_i^2/T}\right\}, \end{aligned}$$

which can be viewed as the probability density for α_i . Moreover, it can be recognized as a normal density with mean $\bar{a}_i \equiv \sum_{t=1}^T (r_{it} - \beta'_i f_t)/T$ and variance σ_i^2/T , i.e., $\mathcal{N}(\bar{a}_i, \sigma_i^2/T)$.

By assumption, $f(\alpha_i|\mathcal{G}^{(k)})$ is the density for a GMD that is parameterized by $\theta = (\{\pi_l\}_{l=1}^N, \{\mu_l\}_{l=1}^N, \{\sigma_l^2\}_{l=1}^N)$. It can be shown that $f(R_i|\alpha_i, \mathcal{G}^{(k)})f(\alpha_i|\mathcal{G}^{(k)})$ — the product of a normal density (i.e., $\mathcal{N}(\bar{a}_i, \sigma_i^2/T)$) and the density for a GMD — is also a density for a GMD, whose parameters are given by

$$\begin{aligned} \tilde{\mu}_{i,l} &= \left(\frac{\sigma_l^2}{\sigma_l^2 + \sigma_i^2/T}\right)\bar{a}_i + \left(\frac{\sigma_i^2/T}{\sigma_l^2 + \sigma_i^2/T}\right)\mu_l, \\ \tilde{\sigma}_{i,l}^2 &= \frac{1}{1/\sigma_l^2 + 1/(\sigma_i^2/T)}, \\ \tilde{\pi}_{i,l} &= \frac{\pi_l \phi(\bar{a}_i - \mu_l, \sigma_l^2 + \sigma_i^2/T)}{\sum_{l=1}^L \pi_l \phi(\bar{a}_i - \mu_l, \sigma_l^2 + \sigma_i^2/T)}, \quad l = 1, 2, \dots, L, \end{aligned}$$

where $\phi(\mu, \sigma^2)$ is the density of the normal distribution $\mathcal{N}(0, \sigma^2)$ evaluated at μ .

Therefore, $f(\mathcal{A}|\mathcal{R}, \mathcal{G}^{(k)})$ can be characterized as the density for N independent variables. The i -th variable follows a GMD that is parameterized by

$$\tilde{\theta}_i = (\{\tilde{\pi}_{i,l}\}_{l=1}^L, \{\tilde{\mu}_{i,l}\}_{l=1}^L, \{\tilde{\sigma}_{i,l}^2\}_{l=1}^L).$$

A.2 Step III: Maximizing $\sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \log f(R_i|\alpha_i^m, \beta_i, \sigma_i)$

Given the independence of the residuals, we can find the MLE of \mathcal{B} and $\mathcal{\Sigma}$ fund by fund. In particular, the log-likelihood for fund i is given by

$$\frac{1}{M} \sum_{m=1}^M \log f(R_i|\alpha_i^m, \beta_i, \sigma_i) = \frac{1}{M} \sum_{m=1}^M \sum_{t=1}^T \log f(r_{it}|\alpha_i^m, \beta_i, \sigma_i), \quad (\text{A.2})$$

through which we can find the MLE of β_i and σ_i . Under the normality assumption, it can be shown that the right hand side of (A.2) can be written as

$$\frac{1}{M} \sum_{m=1}^M \sum_{t=1}^T \log f(r_{it}|\alpha_i^m, \beta_i, \sigma_i) = -\frac{T}{2} \log(2\pi\sigma_i^2) - \frac{1}{2\sigma_i^2} \left[\sum_{t=1}^T (r_{it} - \beta_i' f_t - \bar{\alpha}_i)^2 + T(\bar{\alpha}_i^2 - \bar{\alpha}_i^2) \right], \quad (\text{A.3})$$

where $\bar{\alpha}_i$ and $\bar{\alpha}_i^2$ are defined as:

$$\bar{\alpha}_i = \frac{1}{M} \sum_{m=1}^M \alpha_i^m, \quad \bar{\alpha}_i^2 = \frac{1}{M} \sum_{m=1}^M (\alpha_i^m)^2.$$

An inspection of (A.3) shows that the MLE of β_i and σ_i can be found sequentially. We find the MLE for β_i first. Notice that the MLE $\hat{\beta}_i$ is essentially the estimates of the slope coefficients for the OLS that regresses the time-series of $\{r_{it} - \bar{\alpha}_i\}_{t=1}^T$ on $\{f_t\}_{t=1}^T$. As a result, we have

$$\hat{\beta}_i = (F'F)^{-1}F'Y_i,$$

where

$$F_{(T \times K)} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_T \end{bmatrix}, \quad Y_{i(T \times 1)} = \begin{bmatrix} r_{i,1} - \bar{\alpha}_i \\ r_{i,2} - \bar{\alpha}_i \\ \vdots \\ r_{i,T} - \bar{\alpha}_i \end{bmatrix}.$$

Fixing β_i at its MLE, we take the first-order derivative of (A.3) with respect to σ_i^2 to obtain the MLE for σ_i^2 , i.e.,

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1} (r_{it} - \hat{\beta}_i' f_t - \bar{\alpha}_i)^2 + (\bar{\alpha}_i^2 - \bar{\alpha}_i^2).$$

Define $\bar{\varepsilon}_i^2 \equiv \frac{1}{T} \sum_{t=1} (r_{it} - \hat{\beta}_i' f_t - \bar{\alpha}_i)^2$ and $\widehat{Var}(\alpha_i) = (\bar{\alpha}_i^2 - \bar{\alpha}_i^2)$. The MLE of σ_i^2 can be expressed as

$$\hat{\sigma}_i^2 = \bar{\varepsilon}_i^2 + \widehat{Var}(\alpha_i). \quad (\text{A.4})$$

Note that $\{\alpha_i^m\}_{m=1}^M$ are simulated data. When the size of the simulated data is large, the sample moments in (A.4) will be close to the population moments. We therefore replace the sample moments with their population moments. This helps us obtain the exact analytical solutions for β_i and σ_i when the conditional distribution of \mathcal{A} is given in Appendix A.1. In particular, the exact MLE for β_i is:

$$\check{\beta}_i = (F'F)^{-1} F' \check{Y}_i,$$

where $\check{Y}_i = [r_{i,1} - m(\alpha_i), r_{i,2} - m(\alpha_i), \dots, r_{i,T} - m(\alpha_i)]'$ and $m(\alpha_i) = E_{\mathcal{A}|\mathcal{R},\mathcal{G}^{(k)}}(\alpha_i) = \sum_{l=1}^L \tilde{\pi}_{i,l} \tilde{\mu}_{i,l}$. The exact MLE for σ_i^2 is:

$$\check{\sigma}_i^2 = \frac{1}{T} \sum_{t=1} (r_{it} - \check{\beta}_i' f_t - m(\alpha_i))^2 + var(\alpha_i),$$

where

$$\begin{aligned} var(\alpha_i) &\equiv Var_{\mathcal{A}|\mathcal{R},\mathcal{G}^{(k)}}(\alpha_i), \\ &= \sum_{l=1}^L \tilde{\pi}_{i,l} [(\tilde{\mu}_{i,l} - m(\alpha_i))^2 + \check{\sigma}_{i,l}^2]. \end{aligned}$$

The parameter values in $\check{\theta}_i = (\{\tilde{\pi}_{i,l}\}_{l=1}^L, \{\tilde{\mu}_{i,l}\}_{l=1}^L, \{\check{\sigma}_{i,l}^2\}_{l=1}^L)'$ can be found in Appendix A.1.

A.3 Step III: Maximizing $\sum_{m=1}^M \sum_{i=1}^N \log f(\alpha_i^m | \theta)$

The optimization of $\sum_{m=1}^M \sum_{i=1}^N \log f(\alpha_i^m | \theta)$ in itself needs to invoke the EM algorithm. Our goal is to find the MLE of θ when MN observations are assumed to be drawn from the GMD that is parameterized by θ . For ease of exposition, we replace

the subscript in α_i^m with j so that $\{\alpha_i^m\}_{(i=1,\dots,N; m=1,\dots,M)} = \{\alpha_{ij}\}_{(i=1,\dots,N; j=1,\dots,M)}$. The starting value of θ is obtained from $\mathcal{G}^{(k)}$.

- Suppose the initial parameter vector is $\hat{\theta} = (\{\hat{\pi}_l\}_{l=1}^L, \{\hat{\mu}_l\}_{l=1}^L, \{\hat{\sigma}_l^2\}_{l=1}^L)$.
- Expectation Step: Compute the expected value of the indicator variable that indicates which population (e.g., the population of skilled or unskilled managers) α_{ij} is drawn from:

$$\begin{aligned}\hat{p}_{ijl} &= \widehat{Pr}(\alpha_{ij} \text{ comes from Group } l) \\ &= \frac{\hat{\pi}_l \phi(\alpha_{ij}; \hat{\mu}_l, \hat{\sigma}_l^2)}{\sum_{l=1}^L \hat{\pi}_l \phi(\alpha_{ij}; \hat{\mu}_l, \hat{\sigma}_l^2)}, \quad i = 1, \dots, N; \quad j = 1, \dots, M; \quad l = 1, \dots, L,\end{aligned}$$

where $\phi(\cdot; \mu, \sigma^2)$ is the density of the normal distribution $\mathcal{N}(\mu, \sigma^2)$.

- Maximization Step: Compute the weighted means and variances, with weights obtained from the Expectation Step:

$$\begin{aligned}\tilde{\mu}_l &= \frac{\sum_{ij} \hat{p}_{ijl} \alpha_{ij}}{\sum_{ij} \hat{p}_{ijl}}, \quad \tilde{\sigma}_l^2 = \frac{\sum_{ij} \hat{p}_{ijl} (\alpha_{ij} - \tilde{\mu}_l)^2}{\sum_{ij} \hat{p}_{ijl}}, \\ \tilde{\pi}_l &= \frac{\sum_{ij} \hat{p}_{ijl}}{MN}, \quad l = 1, \dots, L.\end{aligned}$$

- Iterate between the Expectation Step and the Maximization Step until convergence.

A.4 The Value of the Likelihood Function

We derive the value of the likelihood function given in equation (3). This is used to evaluate relative model performance.

Under the model assumptions, the overall likelihood function can be decomposed as

$$L(\mathcal{G}|\mathcal{R}) \equiv f(\mathcal{R}|\theta, \mathcal{B}, \Sigma), \quad (\text{A.5})$$

$$= \prod_{i=1}^N \int f(R_i|a_i, \mathcal{G}) f(a_i|\mathcal{G}) da_i, \quad (\text{A.6})$$

where \mathcal{G} is the model MLE. Therefore, to obtain the overall likelihood, we only need to calculate the component likelihood, that is, $\int f(R_i|a_i, \mathcal{G})f(a_i|\mathcal{G})$. Under the model assumptions, the integrand of the component likelihood can be written as:

$$\begin{aligned}
& f(R_i|a_i, \mathcal{G})f(a_i|\mathcal{G}) \\
&= \prod_{t=1}^T (2\pi\sigma_i^2)^{-1/2} \exp\left[-\frac{(r_{it} - \alpha_i - \beta'_i f_t)^2}{2\sigma_i^2}\right] \times \sum_{l=1}^L \pi_l (2\pi\sigma_l^2)^{-1/2} \exp\left[-\frac{(a_i - \mu_l)^2}{2\sigma_l^2}\right], \\
&= (2\pi\sigma_i^2)^{-T/2} \sum_{l=1}^L \pi_l (2\pi\sigma_l^2)^{-1/2} \exp\left[-\frac{\sum_{i=1}^T (r_{it} - \alpha_i - \beta'_i f_t)^2}{2\sigma_i^2} - \frac{(\alpha_i - \mu_l)^2}{2\sigma_l^2}\right], \\
&= (2\pi\sigma_i^2)^{-T/2} \sum_{l=1}^L \pi_l (2\pi\sigma_l^2)^{-1/2} \\
&\quad \times \exp\left\{-\frac{(\sigma_l^2 + \sigma_i^2/T)}{2\sigma_l^2(\sigma_i^2/T)} \left[\alpha_i - \frac{\sum_{t=1}^T (r_{it} - \beta'_i f_t) \sigma_l^2 + \mu_l \frac{\sigma_i^2}{T}}{\sigma_l^2 + \sigma_i^2/T}\right]^2\right. \\
&\quad \left. + \frac{(\sum_{t=1}^T (r_{it} - \beta'_i f_t) \sigma_l^2 + \mu_l \frac{\sigma_i^2}{T})^2}{2(\sigma_l^2 + \sigma_i^2/T)\sigma_l^2(\sigma_i^2/T)} - \frac{(\sum_{t=1}^T (r_{it} - \beta'_i f_t)^2 \sigma_l^2 + \mu_l^2 \frac{\sigma_i^2}{T})}{2\sigma_l^2(\sigma_i^2/T)}\right\}, \\
&= (2\pi\sigma_i^2)^{-T/2} \sum_{l=1}^L \pi_l (2\pi\sigma_l^2)^{-1/2} \times \sqrt{2\pi(\sigma_i^2/T)\sigma_l^2/(\sigma_l^2 + \sigma_i^2/T)} \\
&\quad \times \underbrace{\frac{1}{\sqrt{2\pi(\sigma_i^2/T)\sigma_l^2/(\sigma_l^2 + \sigma_i^2/T)}} \exp\left\{-\frac{(\sigma_l^2 + \sigma_i^2/T)}{2\sigma_l^2(\sigma_i^2/T)} \left[\alpha_i - \frac{\sum_{t=1}^T (r_{it} - \beta'_i f_t) \sigma_l^2 + \mu_l \frac{\sigma_i^2}{T}}{\sigma_l^2 + \sigma_i^2/T}\right]^2\right\}}_{\phi(\alpha_i; \mu_{0i}, \sigma_{0i}^2)} \\
&\quad \times \exp\left\{\frac{(\sum_{t=1}^T (r_{it} - \beta'_i f_t) \sigma_l^2 + \mu_l \frac{\sigma_i^2}{T})^2}{2(\sigma_l^2 + \sigma_i^2/T)\sigma_l^2(\sigma_i^2/T)} - \frac{(\sum_{t=1}^T (r_{it} - \beta'_i f_t)^2 \sigma_l^2 + \mu_l^2 \frac{\sigma_i^2}{T})}{2\sigma_l^2(\sigma_i^2/T)}\right\},
\end{aligned}$$

where $\phi(\alpha_i; \mu_{0i}, \sigma_{0i}^2)$ is the density function for a normal distribution parameterized by:

$$\begin{aligned}
\mu_{0i} &= \frac{\sum_{t=1}^T (r_{it} - \beta'_i f_t) \sigma_l^2 + \mu_l \frac{\sigma_i^2}{T}}{\sigma_l^2 + \sigma_i^2/T}, \\
\sigma_{0i}^2 &= (\sigma_i^2/T)\sigma_l^2/(\sigma_l^2 + \sigma_i^2/T).
\end{aligned}$$

Therefore, by integrating over a_i , the part involving the normal density becomes one, and we have:

$$\int f(R_i|a_i, \mathcal{G})f(a_i|\mathcal{G})da_i = (2\pi\sigma_i^2)^{-T/2} \sum_{l=1}^L \pi_l \sqrt{(\sigma_i^2/T)/(\sigma_l^2 + \sigma_i^2/T)}$$

$$\times \exp\left\{\frac{(\frac{\sum_{t=1}^T (r_{it} - \beta'_i f_t)}{T} \sigma_l^2 + \mu_l \frac{\sigma_i^2}{T})^2}{2(\sigma_l^2 + \sigma_i^2/T)\sigma_l^2(\sigma_i^2/T)} - \frac{(\frac{\sum_{t=1}^T (r_{it} - \beta'_i f_t)^2}{T} \sigma_l^2 + \mu_l^2 \frac{\sigma_i^2}{T})}{2\sigma_l^2(\sigma_i^2/T)}\right\}.$$

Define

$$\hat{\alpha}_i = \frac{\sum_{t=1}^T (r_{it} - \beta'_i f_t)}{T},$$

$$\widehat{\alpha}_i^2 = \frac{\sum_{t=1}^T (r_{it} - \beta'_i f_t)^2}{T},$$

$$w_{l,i}^c = \frac{\sigma_l^2}{\sigma_l^2 + \sigma_i^2/T},$$

$$w_{l,i}^t = 1 - w_{l,i}^c,$$

then the component likelihood can be written as

$$\int f(R_i|a_i, \mathcal{G})f(a_i|\mathcal{G})da_i = (2\pi\sigma_i^2)^{-T/2} \sum_{l=1}^L \pi_l \sqrt{w_{l,i}^t}$$

$$\times \exp\left\{\frac{(\hat{\alpha}_i w_{l,i}^c + \mu_l w_{l,i}^t)^2 - (\widehat{\alpha}_i^2 w_{l,i}^c + \mu_l^2 w_{l,i}^t)}{2[1/(1/\sigma_l^2 + 1/(\sigma_i^2/T))]} \right\}.$$

The overall likelihood can be calculated as the product of the component likelihoods of the cross-section of funds, as given in equation (A.6).

B Estimation Details

In this appendix, we detail the implementation of the estimation method that is described in Section 3.

In *Step I*, we choose a set of starting values to initialize our estimation. We have a large number of parameters that are given by $\mathcal{G} = [\theta', \mathcal{B}', \Sigma']'$. However, the time-series information of each fund helps us estimate each fund's risk loadings (i.e., β) and residual variance, providing a reasonable set of starting values. Therefore, for \mathcal{B} and Σ , we start with their equation-by-equation OLS estimates, that is:

$$\mathcal{B}^0 = \mathcal{B}^{OLS}, \quad \Sigma^0 = \Sigma^{OLS}, \quad (\text{B.1})$$

where a superscript of zero denotes the starting values.

For parameters that govern the GMD (i.e., θ), we randomly generate multiple sets of starting values to avoid local optimums. In particular, for a L -component GMD and for the L parameters that govern the means of the component distributions, we randomly choose L numbers that are uniformly distributed over the interval of $[-20\%, 20\%]$ (per annum). The boundary of 20% reflects our knowledge of the mutual fund data. Our prior is that it is unlikely to have a population of funds that are concentrated around a mean that resides outside of the $[-20\%, 20\%]$ interval. Our estimation results confirm this prior. We never obtain optimal mean estimates for the component distributions that are close to the boundaries. After randomly generating the L mean parameters, we rank them in an ascending order for model identification.

We follow a similar procedure to choose the starting values for the standard deviations of the component distributions. In particular, we randomly choose L numbers that are uniformly distributed over the interval of $[0.1\%, 20\%]$ (per annum). Again, the choices of the boundaries reflect our priors about the standard deviations of the component distributions. Our estimation results confirm that these boundaries are never violated for the optimized estimates of the standard deviations of the component distributions.

For the drawing probabilities, the selection of the starting values is more complicated than the selection of the previous two sets of parameters as we now have the parameter constraint that the sum of the L drawing probabilities should be one. We therefore follow a sequential procedure to choose the starting values. We first draw a number (i.e., p_1) that is randomly distributed over the unit interval. After drawing the first number, we draw a second number that is uniformly distributed over $[0, 1 - p_1]$. We continue in this way to draw the rest of the probabilities. In particular, after choosing the first l probabilities (i.e., $\{p_i\}_{i=1}^l$), we choose the $(l + 1)$ -th probability by drawing a number that is uniformly distributed over $[0, 1 - \sum_{i=1}^l p_i]$. Lastly, after choosing the $(L - 1)$ -th probability, the last probability is simply set as $1 - \sum_{i=1}^{L-1} p_i$.

After following the above steps, we now have a randomly generated set of initial parameter values $\mathcal{G}^0 = [(\theta^0)', (\mathcal{B}^0)', (\Sigma^0)']'$, where θ^0 contains the parameters that govern the GMD. Taking this set of parameter values as input for our algorithm, our estimation becomes automatic. In particular, starting from \mathcal{G}^0 and following *Step III-IV*, we arrive at a new set of parameters \mathcal{G}^1 . Next, starting at \mathcal{G}^1 , we follow our algorithm and arrive at \mathcal{G}^2 . We continue in this way and obtain a sequence of parameter estimates $\{\mathcal{G}^k\}_{k=0}^K$. This sequence of parameter estimates are converging as K gets larger. The speed of convergence for our algorithm seems high in that the variations in parameter values become very small after ten to fifteen iterations. To terminate the program, we set a tough threshold for the distance of the parameter estimates between adjacent iterations. In particular, we stop the program at the K -th iteration if the L_1 distance between θ^{K-1} and θ^K is within d^{lim} . To prevent the program from running too many iterations, another criterion we impose is that if the program does not stop until the K^{lim} -th iteration, we stop it at K^{lim} . The choices of d^{lim} and K^{lim} depend on whether the estimation is the intermediate step or the final step, as we shall explain next.

We have explained how our estimation works for one set of starting values. We need to try multiple sets of starting values to avoid local optimums. In particular, following the aforementioned generating procedure for starting values, we randomly generate 100 sets of starting values. For each set, we run our algorithm by setting $d^{lim} = 10^{-1}$ and $K^{lim} = 30$ and obtain 100 sets of parameter estimates. This an intermediate optimization step in which we try to save the computational time by setting d^{lim} and K^{lim} at lenient thresholds and obtain 100 sets of rough estimates. Next, we rank the 100 sets of parameter estimates by the corresponding values of the optimized likelihood function. We choose the top 20 sets and rerun our program by starting at the estimated parameter values. This time, we set $d^{lim} = 10^{-2}$ and $K^{lim} = 50$. We again rank the resulting 20 sets of parameter estimates by the corresponding values of the likelihood function. We choose the top five sets and rerun our program by starting at the estimated parameter values obtained from the previous step. This is the final step estimate and we set $d^{lim} = 10^{-3}$ and $K^{lim} = 100$. We choose the best one (in terms of the value of the likelihood function) among the five sets of estimates as our final estimate. We often see that five sets of parameter estimates in the final step are very close to each other. This assures us that the local optima have been thrown out during the intermediate steps.

C FAQ

C.1 General Questions

- *In short, what is the biggest reason to consider the random alpha model for performance evaluation?*

Quoting Searle, Casella, and McCulloch (1992), “Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population.” For performance evaluation, we are interested in both the effects themselves (that is, to evaluate which manager outperforms) and the population (that is, the underlying distribution for alphas). The random alpha model provides a suitable framework to think about both.

- *Why not do a random alpha model with a multiple testing adjustment?*

The mechanisms for the random alpha model and the multiple testing framework to discount the significance of fund alphas are different. The random alpha model forces the cross-section of alphas to fit a parametric density. Observations that are too extreme according to the fitted density are adjusted. Multiple testing adjustment invokes the hypothesis testing framework and uses the p -value to measure the distance between the estimated alpha and zero. A smaller p -value indicates a larger distance from zero and we are trying to identify alphas that are sufficiently distant from zero. It is possible to make mistakes by falsely declaring a zero alpha as nonzero. To control for the false discovery rate, we need to adjust the p -values upward. Both the quantities of interest (i.e., raw alpha vs. p -value of alpha) and the objectives (i.e., goodness-of-fit to a density vs. false discovery rate) are different between the two methods. Applying both will likely overkill the significance of fund alphas.

- *Why is MLE better than the moments-based approach for the estimation of a GMD?*

In general, we need an infinite number of moments — properly weighted — to achieve the estimation efficiency that MLE provides. For example, for a two-component GMD, although it is identified and its five parameters can be estimated using the first five sample moments alone, the sixth moment as well as other higher moments provide additional information for the estimation of the model and should be incorporated into the estimation to improve estimation efficiency.

- *How does model misspecification affect the results of the paper?*

There are different kinds of model misspecifications. A misspecification of the return residuals changes our MLE into a QMLE, which will not bias our estimates. The loss in estimation efficiency is also small, as we show in the simulation study. On the other hand, a misspecification of the factor model (e.g., omitting a true factor) in general will introduce bias for the alpha estimates. Compared to existing models, our model can to some extent alleviate the model misspecification issue, thanks to its ability to use information in the entire cross-section to provide inference. We have a discussion on this towards the end of the paper. However, both our model as well as existing models are sensitive to the issue of model misspecification. See Harvey and Liu (2016a) for an examination on how factor model misspecifications affect asset pricing tests.

- *Does it make sense to treat all funds equally? It seems that there is more information for a fund with a \$1 trillion AUM than a fund with a \$10 million AUM.*

From the perspective of making inference on the alpha population, we think that the alpha for the \$10 million fund is just as important as the alpha for the \$1 trillion AUM fund. If an investor invests \$1 million in either fund, the alpha she gets is simply the alpha for either fund. The alpha for the smaller fund will not be discounted because the fund is smaller. It is likely that the returns for smaller funds are more noisy than returns for larger funds. Our method takes the estimation uncertainty into account.

- *Why not use a three-component GMD for mutual funds in the simulation study?*

For mutual funds, a two-group separation is more in line with what the literature has found, that is, bad funds and average funds. We also did a three-component GMD specification in the simulation study. It reduces the proportion of average funds to about 80% and split the rest 20% into bad funds and very bad funds. However, as we show in the actual application for mutual funds, a two-component GMD fits the data better than a three-component GMD. Therefore, we use the simpler two-component GMD specification for the simulation study.

- *How many funds have a t -statistic over 2.0 under OLS?*

For our sample, under equation-by-equation OLS, 1.7% of funds have a single test t -statistic above 2.0. However, since we have run thousands of tests, we need to adjust for multiple testing. Applying multiple testing adjustments, few funds are found to be significantly outperforming.

Our method departs from the hypothesis testing framework and assumes a continuous distribution for fund alphas. In our framework, alphas are almost

surely not zero by construction. To see the difference between our framework and hypothesis testing, suppose we have 100 funds, each one having an OLS intercept of 1% (per annum) and a standard error of 2% (per annum). Under hypothesis testing, there is no outperformer, as none of the t -statistics is able to pass the single test t -statistic threshold, let alone the multiple testing t -statistic threshold. Under our model, we estimate the alpha distribution to be, say, normal around a mean of 1%. If we test the significance of each alpha under our model, it might as well be the case that none of the t -statistics is above 2.0, especially if the 2% standard error is high enough at the individual fund level. However, this is not evidence against our model since it is not based on hypothesis testing. In our framework, it is possible that all individual funds have a t -statistic below 2.0 while at the same time the population mean is positive and statistically different from zero.

- *What is the intuition behind the EM algorithm to refine the OLS estimates of alphas?*

Imagine that the parameters that govern the alpha population (i.e., the normal mixture distribution) are given. In the “expectation” step, we calculate the conditional distribution of alphas by mixing information from the time-series and the cross-section. Essentially, OLS estimates are adjusted for the information in the mixture distribution. More noisy OLS alpha estimates (which are likely due to higher levels of residual standard deviations) are adjusted more aggressively than less noisy OLS alpha estimates. Hence, the new alpha estimates after the “expectation” step are less noisy than the OLS estimates that are based on time-series information alone. However, these new alpha estimates should change our initial guess of the alpha population (i.e., parameters in the normal mixture distribution). As a result, in the “maximization” step, we try to find a new set of parameters that best explain these new alpha estimates. We iterate between the “expectation” step and the “maximization” step to refine our estimates of both the individual alphas and the parameters that govern the alpha population.