DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Mutual Fund Selection for Realistically Short Samples

Charlotte Christiansen, Niels S. Grønborg and Ole L. Nielsen

CREATES Research Paper 2018-36

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Mutual Fund Selection for Realistically Short Samples*

Charlotte Christiansen† **Niels S. Grønborg**‡ **Ole L. Nielsen**§

This version: December 13, 2018

*The authors acknowledge support from the Danish Council of Independent Research and CREATES. The authors are grateful for helpful comments and suggestions from Tom Engsted, Thomas Quistgaard Pedersen, and Ran Xing as well as from seminar participants at CREATES.

†DFI, CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. Phone: +45 8716 5576. Email: cchristiansen@econ.au.dk. Lund University.

‡DFI, CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. Phone: +45 8716 5312. Email: ngroenborg@econ.au.dk.

§CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. Phone: +45 8716 4825. Email: olni@econ.au.dk. Corresponding author.

Mutual Fund Selection for Realistically Short Samples

Abstract

Performance of mutual fund selection methods is typically assessed using long samples (long time series). It is, however, very often of interest how well the methods perform in shorter samples. We carry out an extensive simulation study based on an empirically motivated skill distribution. For both short and long samples, we present evidence of large differences in performance between popular fund selection methods. In an empirical analysis, we show that the differences documented by the simulations are empirically relevant.

Keywords: Mutual funds; Fund selection; Simulation; Small sample properties.

JEL Classification: G11; G12; G17; G20.

This version: December 13, 2018.

I. Introduction

How to select the best performing mutual funds *ex-ante* is a long standing problem for practitioners and academics alike. The problem is of immense interest not only because of the size of the mutual fund industry, but also because the selection of the best mutual funds has real welfare implications for individual investors. The academic literature on mutual fund selection generally documents impressive properties for long sample periods. Investors may, however, not be interested in historic fund performance many years in the past, but only care about performance in more recent periods. Similarly, researchers routinely apply the fund selection methods on shorter rolling windows in order to analyze the dynamics of fund performance. For these reasons, it is very important to understand how fund selection methods perform in small samples and to compare the performance of the various methods when used in short sample periods.

The contribution of this paper is to provide a first comparison of how various popular fund selection methods work for realistically small samples understood in the sense of short sample periods. We ask whether there are material differences in performance between various fund selection methods, when applied to small samples. The analysis is relevant because there is no guarantee that the methods, which perform best on large samples are also the ones that perform best on small samples. If a selection method only works well for large sample sizes, say 20 years of monthly returns data, we are concerned that the ability to select well-performing funds is driven, to some extent, by survivor bias.

The literature on fund selection is extensive, see e.g. Bollen and Busse (2004), Avramov and Wermers (2006), Mamaysky et al. (2008), Guercio and Reuter (2014), Groenborg et al. (2017), and the survey of Jones and Wermers (2010). In addition to the academic literature, a large industry of investment advisers offer guidance on fund selection and there are thousands of registered investment advisers in the US.

Previous literature compare different mutual fund selection methods with fairly large

samples compared to what we do in this paper. Barras et al. (2010) use a sample of 32 years, Ferson and Chen (2015) use a sample of 17 years, while Harvey and Liu (2018) use a sample of 27 years in their simulation studies. Andrikogiannopoulou and Papakonstantinou (2018) show that the analysis in Barras et al. (2010) is overly conservative using simulation studies based on samples of at least 180 monthly observations.

The methods we compare are in the spirit of Carhart (1997). The paper is concerned with fund selection methods that rely on traditional alphas, which has long been a popular performance measure. The method of Carhart (1997) compares fund returns, net of transaction costs and management fees, to a set of benchmark portfolios that mimic systematic risk exposures. In this framework, the value a fund manager adds per period to the investor is net alpha multiplied by the size of the investment.[1] The mutual fund literature has shown that prediction of alpha is related to various fund characteristics. These include: fund size (Chen et al. (2004)), industry concentration (Kacperczyk et al. (2005)), return gap (Kacperczyk et al. (2008)), active share (Cremers and Petajisto (2009)), coefficient of determination (Amihud and Goyenko (2013)), and a time-varying skill index (Kacperczyk et al. (2014)). Furthermore, a recent paper by Uppal and Zaffaroni (2017) shows that the alpha component of mispricing is a lot more important than the beta component of asset mispricing.

The mutual fund literature has identified problems with traditional alpha measures. Ferson and Lin (2014) argue that shortcomings with traditional alphas can be caused by a number of reasons such as asymmetric information between fund managers and investors, nonlinear payoffs, time-varying model parameters, and market incompleteness. These problems may imply that some covariance terms between fund returns and the pricing kernel are excluded from a traditional benchmark model (see e.g. Glode (2011)). Although the literature documents many problems with traditional alpha measures, our paper only considers fund selection methods that measure performance with traditional alphas.

[1] A different approach has recently been suggested by Berk and van Binsbergen (2015). Their measure of fund manager skill is defined as a funds' assets under management multiplied by gross alpha. This measure differs from the net-alpha approach, as it measures the value a fund manager extracts from capital markets.

The first class of models that we consider is the relatively simple bootstrap procedures of Kosowski et al. (2006) and Fama and French (2010). The second class is the false discovery rate methods of Barras et al. (2010) and Ferson and Chen (2015). Finally, we consider the novel expectation maximization approach of Harvey and Liu (2018) as the third class of selection methods.

To asses the small sample performance of fund selection methods, we set up an extensive simulation study, where funds are drawn from a Gaussian Mixture Distribution (GMD), allowing us to construct three groups of funds with bad performance, neutral performance, and good performance, respectively. The GMD poses realistic challenges in identifying which distribution a fund is drawn from. For example, the distribution of funds with neutral performance overlaps with the distribution of funds with good performance. The parameters of the GMD skill distribution are chosen based on a comprehensive empirical analysis of US equity mutual funds.

We uncover several interesting findings. Firstly, we document large differences in performance between the different fund selection methodologies when applied to small samples of 60 monthly observations. Secondly, we find that the simple bootstrap methods and the false discovery rate methods outperform the advanced expectation maximization method for small samples.

After presenting results for the small sample simulation study, we compare these to a similar large sample simulation study. We find large variations in performance across different methods and also large differences compared to the small sample setting. Therefore, inference on small sample performance based on large sample simulations can be misleading.

Lastly, we conduct an empirical analysis, where we apply each of the seven fund selection methodologies to a series of subsamples. In line with our simulation study, we show that the different approaches lead to very different results.

The remainder of this paper is structured as follows. Section II presents the different fund selection methods in detail. Section III describes our data set. In Section IV we set up our

simulation framework and present the main results for the small sample analysis. Section V represents the large sample simulation results. Section VI is devoted to an empirical comparison of the analyzed methods, while some concluding remarks are presented in Section VII. Various details are delegated to the Appendix.

II. Mutual Fund Selection Methods

This section presents the mutual fund selection methods. First in subsection II.A we introduce mutual fund selection methods, then we present our benchmark method in subsection II.B. Subsection II.C presents bootstrap methods, subsection II.D presents methods based on the false discovery rate, and subsection II.E presents the structural modeling approach based on the expectation maximization algorithm.

II.A. Introduction

When developing a fund selection methodology, it is essential to specify an underlying skill distribution. A common feature for existing mutual fund selection methods is the assumption that the true skill of a mutual fund manager is drawn from different skill distributions or groups. In line with most of the literature, we define different skill groups based on the ability of fund managers to generate net alpha, α. Net alpha is defined as the excess return (above the risk-free rate) generated by a fund after controlling for common risk factors and fees. The net alphas are estimated from an asset pricing model such as the Carhart (1997) four-factor model. More details about how we estimate the alphas in the benchmark model follow in subsection II.B. We consider skilled fund managers, unskilled fund managers, and neutral (zero-alpha) fund managers, and rely on the following definition of the three skill groups:

- Skilled fund managers have sufficient stock-picking skills to more than recover trading costs and expenses ($\alpha > 0$).

- Unskilled fund managers have insufficient stock-picking skills to recover trading costs

and expenses ($\alpha < 0$).

- Zero-alpha fund managers have stock-picking skills just sufficient to recover trading costs and expenses ($\alpha = 0$).

We assume that each group has its own skill distribution such that the skill for all funds is a composite distribution. We label the fractions of skilled, unskilled, and zero-alpha funds by π_+, π_-, and π_0, respectively. The composite skill distribution for all funds is illustrated in Figure I. The figure shows a Gaussian Mixture Distribution (GMD) for three unobserved skill distributions with $\pi_+ = 11\%$, $\pi_- = 59\%$, $\pi_0 = 30\%$. The parameters of the GMD are determined by the empirical properties of our data set of US equity mutual funds, cf. Section III. Furthermore, the figure depicts fund groups with different annualized means, α_s, and standard deviations, σ_s, for $s = +, -, 0$.[2]

[Figure I about here.]

Due to the unobserved nature of the true underlying fund manager skill, the composite skill distribution poses a major challenge with respect to inference on the cross-section of fund alphas. That is, as the three underlying skill distributions overlap for some values of alpha, a fund selection method can perceive a true zero-alpha fund as skilled due to luck. In the more extreme case, a mutual fund selection method can even confuse a truly skilled fund as unskilled as the outcome of extreme bad luck.[3] Consequently, the accuracy associated with the methods' ability to disentangle true skill from luck is essential for inference about mutual fund performance and selection.

[2] The GMD is characterized by the annualized mean parameters $\alpha_+ = 1.32\%$, $\alpha_- = -1.77\%$, and $\alpha_0 = 0\%$, and the annualized standard deviations $\sigma_+ = 0.38\%$, $\sigma_- = 0.61\%$, and $\sigma_0 = 0.50\%$.

[3] Barras et al. (2010) assume that the two extreme distributions do not overlap, whereas Ferson and Chen (2015) solve the problem by introducing power and confusion parameters.

II.B. Benchmark Model

We define risk-adjusted performance (alpha) in terms of the extension of Jensen (1968) and Fama and French (1993) introduced by Carhart (1997). The four-factor model is expressed as

$$\begin{aligned} R_{i,t} &= \alpha_i + \mathrm{b}_i R_{mt} + \mathrm{s}_i \mathrm{SMB}_t + \mathrm{h}_i \mathrm{HML}_t + \mathrm{m}_i \mathrm{MOM}_t + \varepsilon_{it} \\ &= \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{X}_t + \varepsilon_{it}, \end{aligned} \tag{1}$$

where $R_{i,t}$ is fund manager i's excess return in period t over the one-month Treasury bill rate, R_{mt} is the excess return of the market index and SMB_t, HML_t and MOM_t are the returns of factor portfolios related to size, book-to-market, and momentum, respectively. $\boldsymbol{\beta}_i$ is a vector of b_i, s_i, h_i, and m_i and $\boldsymbol{X}_t$ is a vector of the four common risk factors at time t. $\varepsilon_{i,t}$ is fund i's idiosyncratic return residual at time t, which is assumed to have a zero mean with an independent and identical normal distribution.[4] In this framework, the true stock-picking skill of mutual fund manager i is measured by its alpha, α_i.

Inference on the stock-picking ability of a mutual fund is based on a traditional two-stage procedure. First, we estimate alphas via fund-by-fund OLS regressions, and second we perform hypothesis tests under a null of zero abnormal performance ($\alpha_i = 0$). The estimate of the fraction of skilled (unskilled) funds is the fraction of funds that have a significant and positive (negative) alpha. We compute fund t-statistics with Newey and West (1987) heteroskedasticity and autocorrelation consistent standard deviation estimates. We use a 5% level of significance when not stated otherwise. This approach forms a natural benchmark for other fund selection methodologies.

The drawback of the benchmark approach is that we undertake a multiple testing problem. That is, when we simply count the number of skilled and unskilled funds based on the

[4] Ferson and Schadt (1996) propose a conditional version of the benchmark model to account for time variation in factor loadings. Kosowski et al. (2006) and Barras et al. (2010) find no material impact on their results when accounting for time-varying risk taking. In the remainder of this paper, we therefore use the unconditional benchmark model.

estimated t-statistics at a 5% significance level, the probability that we wrongfully reject at least one null hypothesis will be greater than 5%. That probability is increasing with the number of comparisons and therefore with the number of funds. Wrongfully rejecting a null hypothesis of zero alpha means labeling a fund with zero alpha as being skilled or unskilled. Controlling for the multiple comparisons is therefore essential, when attempting to select the best (or worst) mutual funds. As an additional benchmark model we extend the benchmark model above with the Bonferroni correction that accounts for the multiple comparisons by decreasing the level of significance.

Several papers address the problem of selecting the best funds. We concentrate on popular papers that measure performance by the funds' net alphas. We consider three different classes of models and compare them to the two benchmarks presented above. The first class of models are simple bootstrap methods. Kosowski et al. (2006) and Fama and French (2010) both suggest to use the bootstrap to perform more reliable inference on the performance of individual funds. The second class of models include Barras et al. (2010) and Ferson and Chen (2015) and tackles the multiple comparison problem by controlling for the false discovery rate. Finally, the framework of Harvey and Liu (2018) constitutes our third class of models and relies on a structural modeling approach.

II.C. The Bootstrap Method

Kosowski et al. (2006) and Fama and French (2010) use bootstrap selection methods in order to separate manager skill from luck.[5] The studies argue that the necessity of the bootstrap method arises due to complex distributional properties of the cross-section of fund alphas. Ignoring this, and using the benchmark model, results in rejection rates that differ from their nominal level (of 5%).

We use the two bootstrap methods to simulate null distributions of neutral performance by means of the resampling schemes introduced in the papers. We estimate alphas with the

[5]We outline both bootstrap methods step-by-step in Appendix A.

benchmark model in Equation (1). The bootstrap resampling implies that the simulated null distributions of fund returns only depend on sampling variation, as they are generated under the null of zero abnormal performance. We use the bootstrap methods to infer whether individual funds are skilled (unskilled) or lucky (unlucky). We do this by comparing the actual distribution of estimated fund $\hat{t}(\alpha)$-statistics from Equation (1) with bootstrapped distributions of fund t-statistics. We prefer to draw inferences based on t-statistics rather than on the alphas themselves as these are pivotal statistics independent of unknown population parameters.

Although the two bootstrap methods impose the same null distribution for simulated statistics, they use different resampling schemes. Kosowski et al. (2006) use a residual resampling approach for each individual fund, whereas Fama and French (2010) jointly resample both excess fund and factor returns.

The different bootstrap resampling schemes of Kosowski et al. (2006) and Fama and French (2010) result in one significant difference. Consider the benchmark model in Equation (1), then for a given bootstrap simulation run and conditional on the realization of the common risk factors, the Fama and French (2010) method takes the cross-sectional distribution of the residuals into consideration, whereas the method of Kosowski et al. (2006) assumes independence between the residuals across funds. This difference leads to wide confidence intervals for the Fama and French (2010) method relative to the Kosowski et al. (2006) method. The wider confidence intervals of the Fama and French (2010) method is a consequence of using the same time index for all funds within a given bootstrap run to capture the cross-correlation of fund returns, while the narrower confidence intervals of the Kosowski et al. (2006) method arise due to pooling over time by randomly drawing some funds' residuals in down market periods and other funds' residuals in up market periods (cf. Blake et al. (2017)). Further, Blake et al. (2014) argue that the bootstrap methods of Kosowski et al. (2006) and Fama and French (2010) can result in biased estimates. They argue that by assigning equal resampling weights to all observations, the bootstrap method puts too large weights on

observations in the tails of the true but unknown sampling distributions.

A common problem for both resampling schemes is that any autocorrelation structure in fund excess returns is lost. This problem can be mitigated by resampling blocks of timeseries instead of single time indices, cf. Politis and Romano (1994).

Finally, although both bootstrap methods account for complex distributional properties of the cross-section of alpha, we still end up with a multiple hypothesis problem when we infer the fractions of skilled, unskilled, and zero-alpha funds. The reason is that we simply count the number of rejected null hypotheses. That is, for funds with a positive alpha estimate from Equation (1), we test the null hypothesis H_0: $\alpha_i \leq 0$ against the alternative H_1: $\alpha_i > 0$. In contrast, for funds with a negative alpha estimate, we test the null hypothesis H_0: $\alpha_i \geq 0$ against the alternative H_1: $\alpha_i < 0$. Hence, when we count the number of significant funds, each at a 5% significance level, the probability that we wrongfully reject at least one null hypothesis of zero alpha will be greater than 5%. The probability is also increasing in the number of hypothesis tests that we perform.

II.D. The False Discovery Rate Method

Barras et al. (2010) and Ferson and Chen (2015) introduce variants of the false discovery rate method (FDR) for mutual fund selection.[6] We use both methods to obtain estimates for the proportions of unskilled, zero-alpha and skilled funds.

The FDR method accounts for multiplicity by utilizing a high significance level thus balancing the trade-off between Type I and Type II errors. By the use of a high significance level, the FDR method attains high statistical power which decreases the probability of committing type II errors. Although a high significance level increases the power of a test, it also increases the potential for confusion between good and bad funds. Confusion is the extreme case where a fund is perceived as skilled (unskilled), when it is in fact truly unskilled (skilled).

[6] We outline both FDR methods step-by-step in subsections in Appendix A.

Barras et al. (2010) assume that a truly skilled fund cannot be perceived as an unskilled fund and vice-versa. In comparison, Ferson and Chen (2015) explicitly model the confusion parameter. Notably, Ferson and Chen (2015) show that their model nests the model of Barras et al. (2010) when assuming power of one and zero confusion.

The FDR method of Ferson and Chen (2015) also extends the bootstrap methods by Kosowski et al. (2006) and Fama and French (2010). The two bootstrap methods simulate the cross-section of fund alphas under the null of zero abnormal performance, which accounts for multiplicity under the null that all funds have zero abnormal performance. In comparison, the method of Ferson and Chen (2015) accounts for multiplicity by allowing funds to have alpha different from zero.

II.E. The Structural Model Method

The method of Harvey and Liu (2018) takes a structural approach to model the fraction of skilled, unskilled and zero-alpha funds.[7,8] That is, Harvey and Liu (2018) model the cross-sectional alpha distribution by imposing structure with a Gaussian Mixture Distribution (GMD), which they estimate with an expectation-maximization (EM) algorithm.

Harvey and Liu (2018) argue that the benchmark model approach for fund selection is comparable to a panel regression model with fixed effects. They argue that this is the case, as a non-random alpha is assumed for each fund. The model by Harvey and Liu (2018) advocates the use of a random effects counterpart, which assumes that fund alphas are drawn independently from the same cross-sectional distribution. Their approach allows them to pool information from the cross-sectional alpha distribution to reduce noise of individual alpha estimates.

[7] Harvey and Liu (2018) only consider two skill groups, but their methodology is readily extendable to three groups.

[8] We present estimation details of the model by Harvey and Liu (2018) Appendix A.

III. Mutual Fund Data

III.A. Fund Selection

We examine monthly observations between June 1992 and December 2017. We focus on actively managed no-load open-end US domestic equity funds. The returns are measured after expense ratios and trading costs. The data is available from the Center for Research in Security Prices (CRSP) database.[9]

Our sample is originally free of survivorship bias, but we further select funds with a minimum of 60 monthly return observations and we impose a number of screens in order to mitigate omission, incubation, and back-fill biases; see Elton et al. (2001) and Evans (2010). For individual fund shareclasses, we exclude shareclasses that do not report a year of organization and we exclude shareclass return observations prior to the reported year of organization. We also exclude shareclasses that invest less than 80% of their portfolio holdings in common stock or have initial total net assets (TNA) below $5 million in their first month to enter our sample. We avoid a look-ahead bias by allowing shareclasses to remain in our sample if their TNA subsequently decrease below $5 million. Finally, we include only no-load shareclasses.

Until July 2003, the CRSP database lists different shareclasses from the same fund as unique funds. In order to combine multiple shareclasses into specific funds, we use shareclass names and share class returns to identify shareclasses from the same fund. Conditional on our selected screens, we combine multiple shareclass returns to individual fund portfolio returns. Specifically, we calculate a monthly fund net return by weighting shareclass net returns with the proportion of fund TNA represented by each shareclass at the beginning of each month. We calculate a monthly fund TNA by summing across shareclasses.[10] Subsequently, the

[9] Further details about the data set are available in Appendix B.

[10] Return observations need not be contiguous. However, after a missing return we delete the following-month return. We delete following-month returns due to CRSP database convention of filling following-month returns with the funds' cumulated return since last nonmissing return.

CRSP database matches shareclasses into fund portfolios.

Our final sample is an unbalanced panel of 1,481 funds that exist for at least 60 months during our sample period. We investigate the performance of all investment objectives, as well as performance for investment styles: aggressive growth (96), growth (849), growth and income (430) and income (106). Appendix B describes how we classify funds into the different investment styles. We start our sample in June 1992 as the category aggressive growth, conditional on our fund screens, contains its first fund with a 60 month return history from this point forward.

Table I shows the number of funds in existence, average TNA, average excess returns, average standard deviations, and average higher order moments for the full sample period as well as for six five-year subperiods. We consider five-year subperiods because this is what is typically considered by the previous literature in rolling window analysis. For the same reasons, we will later use the 60-month period to define a short sample for our simulation study. Panel A shows statistics for all investment objectives, while panels B-E show statistics for individual investment styles. We show the descriptive statistics both for all funds (at least one observation) and for funds with at least 60 observations (available in the entire subsample).

We see from Panel A that there are initially 145 funds with 60 monthly return observations in the first five years of our sample and that this number increases to 1,481 during the entire 1992-2017 period. In Panel C we see that this increase is mainly driven by a large expansion in growth funds. From Table I, we also see that aggressive growth funds (Panel B) deliver the largest average excess return over the entire evaluation period. However, we also see that this larger average excess return has a larger average standard deviation. The excess returns have small negative skewness and small positive excess kurtosis.

[Table I about here.]

III.B. Benchmark Performance Results

Table II shows annualized alphas, factor loadings, and coefficients of determination for equal-weighted portfolios from the four-factor benchmark model in Equation (1).[11] The annualized alpha for a portfolio of funds reveals whether funds on average produce risk-adjusted returns different from that implied by the four common risk factors. Similar to previous studies, the benchmark model results in a significantly negative annualized alpha equal to -1.00% for the full sample period. From Panel A, we also find that aggressive growth funds tilt towards small capitalization and low book-to-market stocks, while income funds tilt towards large capitalization- and high book-to-market stocks. In Panel B, we present annualized alpha estimates for each of the five-year periods. The alphas are decreasing (becoming more negative) over time until the last subperiod where they increase again.

[Table II about here.]

We investigate if the residuals from the benchmark model in Equaton (1) are well behaved as assumed; normality, homoskedasticity, and no first order serial correlation. We reject normality (see Jarque and Bera (1987)) for 49.43% of the funds, we reject homoskedasticity (see White (1980)) for 58.88% of the funds, and we reject first order serial independence (see Ljung and Box (1978)) for 11.21% of the funds. The fact that we find such irregularities in the residuals from the benchmark model motivates the use of bootstrap based methods.

We investigate further the 11.21% of fund residuals that suffer from first order serial dependence. For the relevant funds, we estimate the following AR(1) process:

$$\varepsilon_{i,t} = \gamma \varepsilon_{i,t-1} + u_{i,t}, \tag{2}$$

where we find an average $\hat{\gamma}$-estimate of 0.08. We use this average value as input to generate first order serial dependence in our simulation study. The first order serial correlation is

[11] The results based on using value-weighted returns are available upon request.

significant, but small, and applies to only a limited fraction of the funds. It is therefore an empirical question whether it is important to take serial correlation into account.

III.C. Parameter Inputs

We select a three-component Gaussian Mixture Distribution as the specification for our skill distribution and find the inputs for the GMD by picking empirical values using the Ferson and Chen (2015) method. We apply the method via a 60-month rolling window estimation procedure from June 1992 to December 2017, where we move the window 12 months forward for each estimation run.[12]

From the estimation procedure, we obtain 22 sets of skilled, unskilled, and zero alpha fund estimates, as well as their alpha estimates. We then take the mean across the estimates and obtain estimates for the drawing probabilities (π) and location parameters (α) of the GMD. We further obtain scale parameters (σ) for skilled and unskilled funds, by taking the standard deviation across α-estimates. This procedure implies that we only have to motivate the scale parameter of the zero-alpha fund distribution. We set this equal to $\sigma_0 = 0.50\%$ annually, as this provides overlaps between the densities of the GMD. Figure I shows the true density of the GMD based on our chosen parameter values.

We recognize that the parameter values that we use as inputs for the simulation study can affect the simulation results. For example, we find relatively high fractions of skilled funds only in the early part of of our sample, cf. Section VI Table VII. The time variation in the fractions of fund skill can therefore be used to motivate alternative drawing probabilities. Another variation is related to the numerical size of the location parameters for skilled and unskilled funds. Higher numerical values of the location parameters increase the signal-to-noise ratio in the simulated data and increases the precision of the selection methods. Finally, the scale parameters control the variation of the location parameters. Increasing the scale parameters leads to more noise in the simulated data and therefore less precision of the selection methods.

[12]Note that for the last estimation, we only move the rolling window seven months as this maintains the five-year period of the window such that the estimation period becomes January 2013 to December 2017.

IV. Small Sample Simulation Study

IV.A. Simulation Design

We evaluate the small sample properties of the mutual fund selection methods via a comprehensive Monte Carlo simulation study.[13] We generate a balanced return panel for each Monte Carlo simulation, where excess returns are simulated from the four-factor benchmark model:

$$
\begin{aligned}
R_{i,t} &= \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{X}_t + \varepsilon_{i,t}, \qquad (3)\\
\varepsilon_{i,t} &= \gamma \varepsilon_{i,t-1} + u_{i,t},\\
\boldsymbol{X}_t &\sim \mathcal{N}(0, \Sigma_X),\\
u_{i,t} &\sim \mathcal{N}(0, \Omega_u).
\end{aligned}
$$

As illustrated in Figure I, we simulate fund skill, α_i, from a three-component GMD. We use the parameter inputs from Section III.C in the simulations. That is, we generate our composite skill distribution in every Monte Carlo simulation by drawing $\pi_+ = 11\%$ skilled funds with annual $\alpha_+ \sim \mathcal{N}(1.32\%, 0.38\%^2)$, $\pi_- = 59\%$ unskilled funds with annual $\alpha_- \sim \mathcal{N}(-1.77\%, 0.61\%^2)$, and $\pi_0 = 30\%$ zero-alpha funds with annual $\alpha_0 \sim \mathcal{N}(0\%, 0.50\%^2)$.

Our skill distribution frequently draws a skilled or unskilled fund, which should provide a challenge

factor loadings from the benchmark model in Equation (1), Σ_X is an empirical covariance structure of the factor portfolio returns and Ω_u is a residual covariance matrix.

We simulate residuals in two different settings. In the first setting we assume that residuals are serially uncorrelated, $\gamma = 0$, and in the second setting we assume that residuals are generated by an AR(1) process with $\gamma = 0.08$.

For both settings we further investigate how contemporaneous residual cross-correlations impact our results by considering three correlation schemes. The reason we simulate funds

[13] We provide an in depth description of our data generating processes in Appendix C.

with contemporaneous dependence is to classify how well the mutual fund selection methods capture herding behavior, which is the tendency of funds to follow the crowd and load on similar non-priced residual risks over time (see e.g. Wermers (1999)). In the first scheme we assume zero correlation across fund residuals. In the second scheme we select a common cross-correlation coefficient ρ for all off-diagonal elements in the covariance matrix Ω_u. In the third scheme we select empirical correlation coefficients to apply the actual cross-sectional data heterogeneity. The average correlation coefficient for this scheme is 0.07, while the average of the absolute value of the correlation coefficients is 0.21.[14] Common for all schemes is that the diagonal elements of Ω_u contain variance estimates from fitted residuals of Equation (1).

IV.B. Simulation Results for Small Samples

In this set of simulation results we consider $N = 500$ funds over $T = 60$ months. So, the small sample is with respect to the time series dimension, not the cross-sectional dimension, of the data set. We present simulation results for no serial dependence in Table III and for serial dependence in Table IV.

[Table III about here.]

[Table IV about here.]

Comparing Tables III and IV, we see that their results are very similar. Therefore, allowing for serial dependence hardly changes the results. So, the small (but significant) serial correlation documented in the data set, cf. Section III.B, appears to be of no materiel importance. For this reason we concentrate on describing the results in Table III in more detail.

[14] The correlation coefficients are estimated for 929 funds that have 60 monthly return observations over the last five years of our sample.

Table III shows estimates of the simulated cross-sectional skill distributions generated by Equation (3). Skill estimates are presented for the seven mutual fund selection methods described above: benchmark model (1), benchmark model with a Bonferroni correction (2), Kosowski et al. (2006) (3), Fama and French (2010) (4), Barras et al. (2010) (5), Ferson and Chen (2015) (6), and Harvey and Liu (2018) (7). In Panel A we consider zero correlation between all funds ($\rho = 0$). In Panel B we use fixed pairwise correlation coefficients of $\rho = 0.3$, while Panel C presents results based on the empirical correlation coefficients that match actual cross-sectional data heterogeneity. We present the bias and RMSE for each simulation scenario.

The bias and RMSE are defined as follows:

$$\text{Bias}_s = \frac{1}{M}\sum_{m=1}^{M}\hat{\pi}_{m,s} - \pi_s, \tag{4}$$

and

$$\text{RMSE}_s = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\pi}_{m,s} - \pi_s)^2}, \tag{5}$$

for $s = +, -, 0$. We also summarize the average absolute bias and the average RMSE across all skill distribution estimates where Avg. Abs. Bias $= \frac{1}{S}\sum_{s=1}^{S}|\text{Bias}_s|$ and Avg. RMSE $= \frac{1}{S}\sum_{s=1}^{S}\text{RMSE}_s$.

Let us first consider the case with zero pairwise correlation, $\rho = 0$ (Panel A). The small sample simulations show that the smallest bias and RMSEs are obtained for the Fama and French (2010) method. The methods of Kosowski et al. (2006) and Ferson and Chen (2015) also follow with fairly good performance. The performance of the other methods is a lot worse as measured by their bias and RMSE. The new Harvey and Liu (2018) method has about the same bias as the benchmark model and lower bias than the adjusted benchmark model, while its RMSE is larger than both the benchmark and the adjusted benchmark. For the Harvey and Liu (2018) method, the bias is small for neutral funds and large for skilled

funds compared to the benchmark.

For a fixed pairwise correlation of $\rho = 0.3$ (Panel B), the small sample simulation results show that the preferred methods are the Kosowski et al. (2006) and Ferson and Chen (2015) methods. For $\rho = 0.3$, we also find that the bootstrap method of Kosowski et al. (2006) has a lower bias than the bootstrap method of Fama and French (2010) and that the latter method identifies too many zero-alpha funds. This result seems puzzling, as the Fama and French (2010) method should have better performance when the residual cross-correlation coefficient is high. The finding is the result of the relatively large confidence intervals of the Fama and French (2010) method. Across the small sample simulations for $\rho = 0.3$, we find for all hypothesis tests that the 90% confidence intervals of the Fama and French (2010) method are 8.3 times wider, on average, relative to those from the method of Kosowski et al. (2006). This helps explain why the bootstrap method of Fama and French (2010) identifies too many zero-alpha funds in small sample simulations when the cross-correlation between fund residuals is high.

The results based on the empirical correlation coefficients (Panel C) are similar to the results with zero correlation, namely where the lowest bias and RMSE are obtained using the Fama and French (2010) method, followed by the methods of Kosowski et al. (2006) and Ferson and Chen (2015). Again, the new Harvey and Liu (2018) method has similar performance to the two benchmark models.

Overall, the small sample simulation properties vary a lot across the seven models that we consider. We see that the Kosowski et al. (2006) and Fama and French (2010) bootstrap methods and the false discovery rate method of Ferson and Chen (2015) have the most attractive properties for realistically short sample periods.

V. Large Sample Simulation Study

We use the same simulation design as for the small sample simulation study except that we consider $T = 307$ months instead of $T = 60$. So, now we consider a sample that corresponds

to more than 25 years of time series observations which is similar to our empirical sample. The sample is large with respect to the time series dimension, not the cross-section. We focus on the case of no serial dependence. The structure and contents of Table V is similar to that of Table III.

[Table V about here.]

Table V shows that the smallest bias and RMSE are obtained for the false discovery rate methods of Barras et al. (2010) and Ferson and Chen (2015) and the bootstrap method of Fama and French (2010). The Harvey and Liu (2018) has larger average bias and average RMSE than the benchmark model. Its bias is low for neutral performance and high for skilled funds compared to the benchmark.

For $\rho = 0.3$ (Panel B), we now also find that the Fama and French (2010) bootstrap method has the lowest bias across all selection methods. In comparison to the small sample simulations in Table III, the length of the average 90% confidence intervals for the hypothesis tests of the Fama and French (2010) bootstrap method is reduced by 8.1%. The reduction in the average length of the 90% confidence intervals leads to a smaller bias, as the method now rejects more null hypotheses of zero fund skill.

We compare the large sample simulations with the small sample simulations in Table III. The Kosowski et al. (2006), Fama and French (2010), and Ferson and Chen (2015) methods have the most attractive properties for realistically short sample periods. For long samples, the Barras et al. (2010) method enters the set of preferred models and the Kosowski et al. (2006) method exits it.

Thus, the preferred methods for selecting mutual funds are different for realistically small samples and large samples. For this reason, any inference on small sample performance based on large sample simulations should be met with some skepticism.

VI. Empirical Comparison of Mutual Fund Selection Methods

We apply the selection methods more directly to the mutual fund data set. We estimate the fraction of unskilled, zero-alpha, and skilled funds, respectively, for the full sample period from 1992-2017 and for the six five-year subsamples from Table I. We compare the results across mutual fund selection methods.

[Table VI about here.]

Table VI shows the empirical results for the entire sample period. Each panel shows the estimated fractions of unskilled, neutral, and skilled funds for the different investment styles: Panel A for all investment styles, panel B for aggressive growth funds, panel C for growth funds, panel D for growth and income funds, and panel E for income funds. The fraction of unskilled, neutral, and skilled funds are about the same across investment styles for a given fund selection method. So, investment style is only of minor importance when comparing fund selection methods.

All methods except Harvey and Liu (2018) (informally) identify two skill distributions in the data, namely unskilled and neutral funds. We see this because they find that roughly zero percent of the funds are skilled, $\pi_+ \approx 0$. Only the benchmark model finds skilled funds (0.7%), however not more than we would expect to be Type I errors. Additionally, our simulation results suggest that the benchmark model overestimates the fraction of skilled funds. The finding of no skilled funds is consistent with the mutual fund literature in general, cf. Carhart (1997).

The benchmark model and the adjusted benchmark model both find that most of the funds have neutral performance (76% and 99%, respectively). The same results are found by the false discovery rate method of Barras et al. (2010) (64%). In contrast, the bootstrap methods of Kosowski et al. (2006) and Fama and French (2010) as well as the false discovery rate method of Ferson and Chen (2015) find that most of the funds are unskilled (79%, 79%, and 86%). The large fraction of unskilled funds identified by the methodology of Ferson and

Chen (2015) is noteworthy in light of our simulation results. Our results indicate that this approach underestimates the fraction of unskilled funds.

We find that there is a lot of variation in the empirical results across selection methods. This is similar to our simulation results, where the benchmark model and Barras et al. (2010) find the most zero alpha funds. This could indicate that they are overly conservative. In the simulation study we find that the bootstrap methods and the Ferson and Chen (2015) method are the most precise for small samples. These methods produce very similar results empirically by finding mostly unskilled funds and classifying the remaining funds as zero-alpha funds.

The empirical findings based on the methodology of Harvey and Liu (2018) are very different from the other six methods. This method only finds unskilled funds.[15] The simulation results for this method are also very different from the other methods, both for small and large samples.

[Table VII about here.]

Table VII shows the estimated fractions of unskilled, neutral, and skilled funds for each of the subperiods for all investment styles.

The Harvey and Liu (2018) model still only detects one skill group for all five-year periods; unskilled funds. So, here the results are unchanged compared to the full sample.

Otherwise, the fraction of unskilled, neutral, and skilled funds vary considerably across the subsamples and across selection methods.

We find some skilled funds when we consider five-year subsamples which is in contrast to the full sample period where the fraction of skilled funds was approximately zero for all selection methods and investment styles. Still with the notable exception of Ferson and Chen (2015), all models indentify very low fractions of skilled funds. The Ferson and Chen (2015) method identifies very high fractions of skilled funds in the first two sub-periods (20% and

[15]When we use a two-component GMD specification, we find that the two GMD components have identical location and scale parameter estimates. Therefore, we do not test whether the model with two GMD components provide an improvement over a one component specification.

32%), but no skilled funds in the last part of our sample. For the 1992-1997 period, the different results could partly be due to the popularization of the Fama and French (1993) SMB and HML factors around 1993. Cremers and Petajisto (2009) find that the active share of actively managed funds decreases over time. They argue that recent data is influenced by "closet indexing" among active funds. This could also indicate less skill in more recent data.

The two benchmark models and the Barras et al. (2010) method find that the majority of funds are neutral during all subsamples. Likewise, Ferson and Chen (2015) find that the majority of funds are unskilled in all sample periods. For those four methods the subsample results are similar to the entire sample results.

For the boothstrap methods of Kosowski et al. (2006) and Fama and French (2010) the results change over time. In the first two subperiods (1992-1997 and 1997-2002) the majority of funds have neutral skills, while thereafter the majority are unskilled, similar to the entire sample period. The fraction of unskilled and neutral funds for the two bootstrap methods is quite different. This is due to the larger confidence intervals of the Fama and French (2010) method compared to the Kosowski et al. (2006) method. The finding of less unskilled funds in the earlier part of our sample period is consistent with Kosowski et al. (2006), Barras et al. (2010), and Ferson and Chen (2015) who find that fund performance becomes worse in more recent data.

VII. Conclusion

We set up a simulation study based on an empirically motivated composite skill distribution and investigate the performance of seven popular fund selection methodologies. We document significant differences in performance between the methodologies for both short and for long sample periods. Furthermore, we show that the relative performance of the models differs for short and long samples. This indicates that inference on short sample performance based on simulations with long samples can be misleading.

We carry out an empirical analysis of the selection methodologies for US equity mutual

funds. We show that the differences in performance of the selection methods are empirically relevant for short subsamples as well as for a larger sample.

References

Amihud, Y. and R. Goyenko (2013). Mutual fund's r2 as predictor of performance. *The Review of Financial Studies 26*(3), 667–694.

Andrikogiannopoulou, A. and F. Papakonstantinou (2018). Reassesing false discoveries in mutual fund performance: Skill, luck, or lack of power? *Journal of Finance, forthcoming*.

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

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

Berk, J. B. and J. H. van Binsbergen (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics 118*(1), 1 – 20.

Blake, D., T. Caulfield, C. Ioannidis, and I. Tonks (2014). Improved inference in the evaluation of mutual fund performance using panel bootstrap methods. *Journal of Econometrics 183*(2), 202 – 210.

Blake, D., T. Caulfield, C. Ioannidis, and I. Tonks (2017). New evidence on mutual fund performance: A comparison of alternative bootstrap methods. *Journal of Financial and Quantitative Analysis 52*(3), 1279–1299.

Bollen, N. P. and J. A. Busse (2004). Short-term persistence in mutual fund performance. *The Review of Financial Studies 18*(2), 569–597.

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

Chen, J., H. Hong, M. Huang, and J. D. Kubik (2004). Does fund size erode mutual fund

performance? the role of liquidity and organization. *The American Economic Review 94*(5), 1276–1302.

Cremers, K. M. and A. Petajisto (2009). How active is your fund manager? a new measure that predicts performance. *The Review of Financial Studies 22*(9), 3329–3365.

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford University Press, New-York.

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. *The Journal of Finance 56*(6), 2415–2430.

Evans, R. B. (2010). Mutual fund incubation. *The Journal of Finance 65*(4), 1581–1611.

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

Fama, E. F. and K. R. French (2010, October). Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance 65*(05), 1915–1947.

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

Ferson, W. and J. Lin (2014). Alpha and performance measurement: The effects of investor disagreement and heterogeneity. *The Journal of Finance 69*(4), 1565–1596.

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

Glode, V. (2011). Why mutual funds "underperform". *Journal of Financial Economics 99*(3), 546 – 559.

Groenborg, N., A. Lunde, A. Timmermann, and R. Wermers (2017). Picking funds with confidence. *CREATES Research Paper Series 13*.

Guercio, D. D. and J. Reuter (2014). Mutual fund performance and the incentive to generate alpha. *The Journal of Finance 69*(4), 1673–1704.

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

Jarque, C. M. and A. K. Bera (1987). A test for normality of observations and regression residuals. *International Statistical Review / Revue Internationale de Statistique 55*(2), 163–172.

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

Jones, R. C. and R. Wermers (2010). Active management in mostly efficient markets. *Financial Analysts Journal 67*(6), 29–45.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp (2014). Time-varying fund manager skill. *The Journal of Finance 69*(4), 1455–1484.

Kacperczyk, M., C. Sialm, and L. Zheng (2005). On the industry concentration of actively managed equity mutual funds. *Journal of Finance 60*, 1983–2012.

Kacperczyk, M., C. Sialm, and L. Zheng (2008). Unobserved actions of mutual funds. *Review of Financial Studies 21*, 2379–2416.

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

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

Mamaysky, H., M. Spiegel, and H. Zhang (2008). Estimating the dynamics of mutual fund alphas and betas. *The Review of Financial Studies 21*(1), 233–264.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Politis, D. and J. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*, 1303–1313.

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

Uppal, R. and P. Zaffaroni (2017). Portfolio choice with model misspecification: a foundation for alpha and beta portfolios. *Working paper*.

Wermers, R. (1999, April). Mutual fund herding and the impact on stock prices. *The Journal of Finance 54*(2), 581–622.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

A. Mutual Fund Selection Methods

This appendix describes how we implement the non-benchmark mutual fund selection methods that we include in our study. Subsection A.1 describes the bootstrap method of Kosowski et al. (2006), subsection A.2 describes the bootstrap method of Fama and French (2010), subsection A.3 describes the false discovery rate method of Barras et al. (2010), subsection A.4 describes the false discovery rate method of Ferson and Chen (2015), and subsection A.5 describes the structural modeling approach by Harvey and Liu (2018).

A.1. The Kosowski et al. (2006) Bootstrap Method

To employ the Kosowski et al. (2006) bootstrap, we estimate the four-factor benchmark model in Equation (1) for fund i and save the parameter estimates, the $\hat{t}(\hat{\alpha}_i)$-statistic and the time series of fitted residuals $\{\hat{\epsilon}_{it}, t = T_{i0}, ..., T_{i1}\}$. Time indices T_{i0} and T_{i1} denote the first and last monthly return for fund i. Next, we draw with replacement a pseudo time series of residuals, $\{\hat{\epsilon}^b_{i,t_\varepsilon},\ t_\varepsilon = s^b_{T_{i0}}, ..., s^b_{T_{i1}}\}$, where b is an index for the bootstrap simulation run and time indices $s^b_{T_{i0}}, ..., s^b_{T_{i1}}$ are drawn uniformly from fund i's original time indices $T_{i0}, ..., T_{i1}$. Under the null hypothesis of zero abnormal performance, we generate pseudo excess returns:

$$R^b_{i,t} = \hat{\boldsymbol{\beta}}_i \boldsymbol{X}_t + \hat{\epsilon}^b_{i,t_\varepsilon}, \tag{6}$$

where the excess return series for fund i is generated with the original parameter estimates, factor returns in their original historical ordering, and the resampled vector of residuals.

We then regress the pseudo excess returns on the original set of factor returns,

$$R^b_{i,t} = \alpha^b_i + \boldsymbol{\beta}^b_i \boldsymbol{X}_t + \tilde{\epsilon}_{i,t}, \tag{7}$$

and save the estimate of $\hat{t}(\hat{\alpha}^b_i)$. We estimate $\hat{t}(\hat{\alpha}^b_i)$ with a Newey and West (1987) heteroskedasticity and autocorrelation adjusted standard error. Note that the true values of the saved

bootstrapped statistics are zero, as implied by the data generating process in Equation (6). To form cross-sectional bootstrap distributions of $\hat{t}(\hat{\alpha}_i^b)$ we repeat the outlined procedure for all funds $i = 1, ..., N$ across $b = 1, ..., 1,000$ bootstrap iterations.

In order to perform inference, we sort the cross-sectional bootstrapped statistics in descending order for each bootstrap simulation. This produces N probability density functions (PDFs), which solely depend on sampling variation. Inference is drawn for each fund by comparing actual $\hat{t}(\hat{\alpha}_i)$-estimates sorted in descending order with sorted bootstrap PDFs. Conditional on the sign of the sorted $\hat{t}(\hat{\alpha}_i)$-estimate, we test the null hypothesis of zero abnormal performance against either positive abnormal performance ($\alpha_i > 0$) or negative abnormal performance ($\alpha_i < 0$).

As an example, suppose that the best ex post performing fund in terms of its t-statistic has a $\hat{t}(\hat{\alpha}_i)$-estimate that is larger than the 5% upper tail cutoff value of the corresponding bootstrapped PDF. We then draw the conclusion that the top-performing fund has a higher t-statistic than what can be attributed to luck alone. We compute estimates of skilled, unskilled, and zero-alpha funds by the counts of significant and insignificant fund t-statistics. We choose a 5% upper or lower tail cutoff values conditional on the sign of $\hat{t}(\hat{\alpha}_i)$ in each hypothesis test corresponding to our significance level.

A.2. The Fama and French (2010) Bootstrap Method

To employ the method of Fama and French (2010), we estimate $\hat{\alpha}_i$ and the $\hat{t}(\alpha_i)$-statistic from the four-factor benchmark model in Equation (1) for each fund i. We uniformly draw, with replacement, a common time index for all funds, $t_{FF} = s_{T_0}^b, ..., s_{T_1}^b$, where $T_0, ..., T_1$ denote the first and the last month of the fund sample that we investigate. This is done by resampling excess returns according to the common time-index, t_{FF}, and deducting fund i's estimated alpha at all resampled time-periods where fund i has a monthly excess return observation. We drop a fund in a bootstrap simulation run if that fund has less than 12 monthly resampled returns. For fund i, we generate a pseudo excess return series under the null hypothesis of

zero abnormal performance. Further, we resample the common risk factors according to the time index t_{FF}, which allows us to regress the zero-alpha excess returns on the jointly resampled set of factor returns:

$$R^b_{i,t_{FF}} - \hat{\alpha}_i = \alpha^b_i + \boldsymbol{\beta}^b_i \boldsymbol{X}^b_{t_{FF}} + \tilde{\epsilon}_{i,t}. \tag{8}$$

From Equation (8), we save $\hat{t}(\hat{\alpha}^b_i)$ and repeat the outlined procedure for all funds $i = 1, ..., N$ across $b = 1, ..., 1,000$ bootstrap iterations. We estimate $\hat{t}(\hat{\alpha}^b_i)$ with a Newey and West (1987) heteroskedasticity and autocorrelation adjusted standard error. Inference is performed similar to the Kosowski et al. (2006) bootstrap method.

A.3. The Barras et al. (2010) False Discovery Rate Method

We employ the Barras et al. (2010) FDR method to estimate fractions of skilled, unskilled and zero-alpha funds. Barras et al. (2010) estimate the fraction of zero-alpha funds as:

$$\hat{\pi}_0(\lambda) = \frac{\hat{W}(\lambda)}{N} \frac{1}{1-\lambda}, \tag{9}$$

where λ is a significance threshold above which p-values of fund alphas are assumed to be generated by zero-alpha funds, $\hat{W}(\lambda)$ is the number of funds with an alpha p-value higher than the significance threshold λ and N is the number of funds in our sample. We select an optimal significance threshold, $\lambda^* \in [0.30, 0.70]$, following the bootstrap method by Storey (2002), which minimizes the mean squared error of the point estimator $\hat{\pi}_0(\lambda)$. The fund p-values required to compute $\hat{W}(\lambda)$ are obtained with the Kosowski et al. (2006) bootstrap procedure, where we account for potential asymmetry in funds' estimated t-statistics by calculating bootstrapped p-values as in Davidson and MacKinnon (2004). Note that to calculate p-values we do not sort the cross-sectional bootstrap statistics. Instead we calculate

double sided p-values as:

$$\hat{p}_i = 2 \cdot \min\left(\frac{1}{B}\sum_{b=1}^{B}\mathcal{I}\{\hat{t}_i^b > \hat{t}_i\}, \frac{1}{B}\sum_{b=1}^{B}\mathcal{I}\{\hat{t}_i^b < \hat{t}_i\}\right), \tag{10}$$

where $B = 1,000$ is the number of bootstraps, $\hat{t}_i$ is the empirical OLS estimate from Equation (1) for fund i, $\hat{t}_i^b$ is a bootstrapped t-statistic for fund i from the Kosowski et al. (2006) method, and $\mathcal{I}\{\cdot\}$ is an indicator function that takes on the value 1 if the statement in curly brackets is true.

To correct for false discoveries, we estimate the fraction of funds that are expected to exhibit significant performance purely by chance:

$$\hat{F}_+(\gamma) = \hat{F}_-(\gamma) = \hat{\pi}_0 \cdot \gamma/2, \tag{11}$$

where $\hat{F}_+(\gamma)$ and $\hat{F}_-(\gamma)$ are the estimated fractions of lucky and unlucky funds at a given significance level γ, respectively. We estimate the final fractions of skilled and unskilled funds by subtracting the expected fraction of funds that have significant performance purely by chance from the fractions of significantly outperforming, $\hat{S}_+(\gamma)$, and underperforming funds, $\hat{S}_-(\gamma)$:

$$\begin{aligned}\hat{\pi}_+(\gamma) &= \hat{S}_+(\gamma) - \hat{\pi}_0 \cdot \gamma/2, \\ \hat{\pi}_-(\gamma) &= \hat{S}_-(\gamma) - \hat{\pi}_0 \cdot \gamma/2.\end{aligned} \tag{12}$$

The fractions $\hat{S}_+(\gamma)$ and $\hat{S}_-(\gamma)$ are determined from the p-values calculated with Equation (10) at a given significance level γ. Similar to selecting λ^*, we select a sufficiently high significance level, $\gamma^* \in [0.30, 0.50]$, with a bootstrap method that minimizes the MSE of the point estimators $\hat{\pi}_+(\gamma)$ and $\hat{\pi}_-(\gamma)$. Further, we truncate estimated fractions of skilled and unskilled funds at zero if negative.[16]

[16] Details on the bootstrap methods for selecting λ^* and γ^* are available in an appendix available at Laurent Barras's personal web site.

A.4. The Ferson and Chen (2015) False Discovery Rate Method

Ferson and Chen (2015) propose a generalization of the FDR by Barras et al. (2010). We employ their generalization and follow their three-stage estimation procedure with a fixed significance level of $\gamma = 20\%$.

In the first estimation stage, we perform three bootstrap simulations with the Fama and French (2010) method. First, we simulate the cross-section of mutual fund alphas by imposing zero abnormal performance under the null hypothesis. This produces two critical values t_+ and t_-, where t_+ is a value above which 10% of bootstrapped t-statistics are placed when the null of zero abnormal performance is true. Similarly, t_- is a value below which 10% of bootstrapped t-statistics are placed when the null of zero abnormal performance is true.

Second, we simulate the cross-section of alphas under the alternative that funds are truly good and centered around $\alpha_+ > 0$. This simulation produces a parameter estimate, $\hat{\phi}_+$, reflecting the power of the test for good funds and a parameter estimate, $\hat{\delta}_-$, which reflects the confusion of the test for good funds. The power parameter is the fraction of bootstrapped t-statistics above t_+, while the confusion parameter is the fraction of bootstrapped t-statistics below t_-.

Third, we simulate the cross-section of alphas under the alternative that funds are truly bad and centered around $\alpha_- < 0$. Again, we obtain a power parameter estimate, $\hat{\phi}_-$, and a confusion parameter estimate $\hat{\delta}_+$. The power parameter is the fraction of bootstrapped t-statistics below t_- and the confusion parameter is the fraction of bootstrapped t-statistics above t_+.

In the second estimation stage, we apply the information contained in the power and confusion parameter estimates from the actual cross-section of funds. We denote $\hat{S}_-$ and $\hat{S}_+$ as the fraction of funds with significant under- and over performance, respectively. We compute $\hat{S}_-$ as the fraction of empirical t-statistics below t_- and we calculate $\hat{S}_+$ as the

fraction of empirical t-statistics above t_+. Then, we minimize the MSE of the equations:

$$\hat{S}_- = \hat{\pi}_0 \cdot \gamma/2 + \hat{\phi}_- \hat{\pi}_- + \hat{\delta}_- \hat{\pi}_+, \tag{13}$$
$$\hat{S}_+ = \hat{\pi}_0 \cdot \gamma/2 + \hat{\phi}_+ \hat{\pi}_+ + \hat{\delta}_+ \hat{\pi}_-.$$

We numerically minimize the MSE of Equations (13) with respect to the two unknowns $\hat{\pi}_+$ and $\hat{\pi}_-$ subject to the Kuhn-Tucker conditions $\hat{\pi}_+ \geq 0$, $\hat{\pi}_- \geq 0$ and $\hat{\pi}_+ + \hat{\pi}_- \leq 1$. This gives us estimates for fractions of skilled, unskilled, and zero-alpha funds due to the relation $\pi_0 = 1 - \pi_+ - \pi_-$.

In the third estimation stage, we search for the best values of α_- and α_+ in terms of model fit. That is, we perform a grid search across a range of alpha values, repeating the first two stages of the estimation procedure thus obtaining candidate estimates for skilled, unskilled, and zero-alpha funds. In our empirical analysis, we search across the annualized alpha values $\alpha \in \pm[0.48\%, 4.80\%]$ with increments of $\pm 0.48\%$.[17] For each set of alpha values, we obtain fund alpha t-statistics based on the implied simulated mixture distribution, which we compare to fund alpha t-statistics estimated from actual data.

A formal comparison between the simulated mixture distribution and the actual data in terms of goodness of fit is determined by the Pearson χ^2 statistic:

$$\text{Pearson } \chi^2 = \sum_{k=1}^{K} \frac{(O_k - M_k)^2}{O_k}, \tag{14}$$

where K is the number of cells, O_k is the number of fund alpha t-statistics coming from cell k in the data, and M_k is the number of fund alpha t-statistics coming from cell k using the model. The null hypothesis of the Pearson χ^2 test is that the observed frequency distribution is the same as the model frequency distribution. This implies that the lowest Pearson χ^2-statistic across our range of alpha values produces our final estimates of the fractions for

[17] In our simulation study, we search across the annualized alpha values $\alpha \in \pm[2.4\%, 7.2\%]$ with increments of $\pm 2.4\%$. We use higher values of alpha in the simulation study to obtain larger values of the power parameters ϕ_- and ϕ_+.

skilled, unskilled, and zero-alpha funds.

A.5. Harvey and Liu (2018) Noise Reduced Alpha Model

In order to estimate fractions of skilled, unskilled, and zero-alpha funds in our simulation study, we employ the Noise Reduced Alpha (NRA) model of Harvey and Liu (2018) up to a three-component GMD and apply their four-stage estimation procedure. The GMD is characterized by the following parameters:

$$\theta = (\{\pi_l\}_{l=1}^{G}, \{\mu_l\}_{l=1}^{G}, \{\sigma_l^2\}_{l=1}^{G}), \tag{15}$$

where π_l are the weights, μ_l are the means and σ_l^2 are the variances for the $G = 3$ mixture components.

In the first estimation stage, we denote $\mathcal{K} = [\theta', \mathcal{B}', \Sigma']$ as the parameters to be estimated. $\mathcal{B}$ and Σ contain the individual fund-by-fund slope estimates and residual standard error estimates from the four-factor benchmark model in Equation (1). For the first iteration, $\mathcal{K}^0$, we initialize the EM algorithm with equation-by-equation OLS estimates for $\mathcal{B}$ and Σ. For the initialization of GMD parameters, we randomly draw economically meaningful values for the mean and variance parameters, and random values for the weight components such that the relation $\sum_{l=1}^{G} \pi_l = 1$ is kept. More specific, in our empirical study we guess prior to the estimation that skilled funds have an annual mean parameter of $\mu_+ \in [0.1\%, 20\%]$, unskilled funds have an annual mean parameter of $\mu_- \in [-0.1\%, -20\%]$ and zero-alpha funds are centered around an annual mean parameter of zero. For all skill groups, we guess that the annual standard deviation of the mean parameter is $\sigma_l \in [0.1\%, 10\%]$.[18] The initial values of the component weights are selected from a sequential procedure where the first weight π_1 is set to a random value $w_1 \in [0, 1]$. The guess for the second component weight, π_2, is set to a random value over the interval $w_2 \in [0, 1 - w_1]$ such that the final guess for π_3 becomes

[18] In the simulation study, we initialize the annualized mean parameters as $\mu_+ \in [0.1\%, 5\%]$ and $\mu_- \in [0.1\%, 5\%]$. Further, we initialize the annualized standard deviations as $\sigma_l \in [0.1\%, 5\%]$.

$w_3 = 1 - w_1 - w_2$. To avoid a local optimum for the EM algorithm, we generate 100 different sets of starting values for the parameter θ in each estimation run.

The second stage of the estimation uses the parameters from the s-th iteration of the EM algorithm, $\mathcal{K}^{(s)}$, to calculate the expected value of the log complete likelihood function. The log complete likelihood function is stated with respect to the conditional fund alpha distribution $\mathcal{A}$, current parameter values, and excess fund returns denoted $\mathcal{R}$:

$$\begin{aligned} \mathrm{L}(\mathcal{K}|\mathcal{K}^{(s)}) &= \mathrm{E}_{\mathcal{A}|\mathcal{R},\mathcal{K}^{(s)}}[\log f(\mathcal{R},\mathcal{A}|\mathcal{K})], \\ &= \mathrm{E}_{\mathcal{A}|\mathcal{R},\mathcal{K}^{(s)}}[\sum_{i=1}^{N} \log f(R_{i,t}|\alpha_i,\boldsymbol{\beta}_i,\sigma_i) f(\alpha_i|\theta)]. \end{aligned} \tag{16}$$

The closed-form solution of $\mathrm{L}(\mathcal{K}|\mathcal{K}^{(s)})$ does most likely not exist. Therefore, we implement a Monte Carlo EM algorithm, which replaces the expectation in Equation (16) with the sample mean. The sample mean is calculated from simulating $M = 100$ alpha cross-section samples from the conditional fund alpha distribution $\mathcal{A}|\mathcal{R}$:

$$\hat{\mathrm{L}}(\mathcal{K}|\mathcal{K}^{(s)}) = \frac{1}{M}\sum_{m=1}^{M}[\sum_{i=1}^{N} \log f(R_{i,t}|\alpha_i^m,\boldsymbol{\beta}_i,\sigma_i) f(\alpha_i^m|\theta)]. \tag{17}$$

In the third stage of the estimation procedure, we maximize $\hat{\mathrm{L}}(\mathcal{K}|\mathcal{K}^{(s)})$ by updating parameter estimates from $\mathcal{K}^{(s)}$ to $\mathcal{K}^{(s+1)}$. This is a simple task, as we update fund parameters $\mathcal{B}$ and Σ, and parameters governing the GMD, θ, separately. That is, we rewrite Equation (17) to:

$$\hat{\mathrm{L}}(\mathcal{K}|\mathcal{K}^{(s)}) = \sum_{i=1}^{N}\frac{1}{M}\sum_{m=1}^{M} \log f(R_{i,t}|\alpha_i^m,\boldsymbol{\beta}_i,\sigma_i) + \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{N} \log f(\alpha_i^m|\theta), \tag{18}$$

where the first part depends on $\mathcal{B}$ and Σ, and the latter part depends on θ. We update $\mathcal{B}$ and Σ according to their exact Maximum Likelihood Estimate (MLE), while we update GMD parameters using responsibilities calculated from the E-step of the algorithm.[19]

[19] The formulas for parameter updates are available in Appendix A of Harvey and Liu (2018).

The fourth stage of the EM algorithm works by iterating between stages two and three. We achieve convergence at the s-th iteration of the EM-algorithm if the difference between all parameters in θ^{s-1} and θ^s are below some threshold δ^{lim} or the iteration counter s hits an upper predefined limit S^{lim}.

We start an estimation run with 100 initial guesses of θ and save computational time via three intermediate optimization steps. The first intermediate step sets soft thresholds at $\delta^{lim} = 10^{-1}$ and $S^{lim} = 30$, which produces 100 rough sets of parameter estimates as candidates for an optimum. We rank the candidates according to their optimized component log likelihood function and select the 20 best parameter sets. In the second intermediate step we set tougher thresholds at $\delta^{lim} = 10^{-2}$ and $S^{lim} = 50$, and rerun the EM-algorithm with the best parameter sets from the previous optimization step. This produces 20 candidates for an optimum, which we rank according to their optimized component log likelihood function. Finally, we select the 10 best sets of parameter estimates and rerun the EM-algorithm at the computationally demanding thresholds $\delta^{lim} = 10^{-3}$ and $S^{lim} = 100$. After this final optimization step, we obtain the optimal parameter set, $\mathcal{K}^* = [\theta^*, \mathcal{B}^*, \Sigma^*]$, according to the optimized component log likelihood function.

B. Mutual Fund Sample Description

In this appendix we provide a detailed description of our mutual fund sample. We document how we exclude passive- and load share classes, how we mitigate omission-, incubation-, and back-fill bias, how we combine multiple share classes and how we classify funds' investment objective using information obtained from the CRSP Survivor Bias Free Mutual Fund Database.

B.1. Exclusion of Passive Fund Share Classes

We exclude passive mutual fund share classes from the original data if the CRSP *index_fund_flag* equals "B", "D", or "E", corresponding to index-based funds, pure index funds, or

index fund enhanced, respectively. The variable *index_fund_flag* is available from 2003 and onwards. Therefore, we also search fund names for the keyword "index" and remove fund share classes where this is the case. Further, we also exclude exchange traded fund (ETF) share classes and exchange traded note (ETN) share classes if the CRSP *et_flag* equals either "F" or "N".

B.2. Exclusion of Load-Fund Share Classes

We obtain information on front- and rear-loads from the CRSP variables *front_load* and *rear_load*. We exclude all load share classes, i.e. share classes that have ever had a load larger than 0%.

B.3. Omission-, Incubation-, and Back-Fill Bias Screens

In order to avoid omission bias (e.g. Elton et al. (2001)) and incubation and back-fill bias (e.g. Evans (2010)) we use information from a number of CRSP database variables. First, we exclude share class return observations prior to the reported year of share class organization. This is done with the variable *first_offer_dt*, which reports the first date a fund was offered. Second, if the date where a share class was first offered is missing, then we exclude that share class from our sample. Third, we exclude share classes that initially have less than $5 million TNA. Share classes that subsequently fall below $5 million in assets under management are allowed to remain in the sample. We use the CRSP variable *mtna* to find the date where a fund share class is eligible to enter our sample.

B.4. Combining Multiple Share Classes

The CRSP database lists different share classes with the variable *crsp_fundno* and assigns share classes into fund portfolios by the variable *crsp_portno* for funds that are active after July 2003. Therefore, for fund share classes that are inactive after July 2003, we combine multiple share classes into specific fund portfolios by investigating the name and

monthly returns of individual share classes by using the CRSP variables *fund_name* and *mret*, respectively.

B.5. Fund Classification by Investment Objective

We focus on active no-load open-end US domestic equity fund share classes. We define equity share classes as classes that at some point in time have 80% or more common stocks in their portfolio and use the CRSP variable *per_com* to find the first date for funds to be eligible in our sample. Conditional on our chosen fund screens, our resulting sample of 1,481 domestic equity funds are each assigned one of four broad investment objectives. We classify funds as aggressive growth, growth, growth and income, or income funds using objective codes from Wiesenberger (1962-1993), Strategic Insight (1993-1998) and Lipper (1998-present). We select the first available investment objective to classify a fund throughout its life. That is, after combining multiple share classes, we let the first available share class investment objective determine the investment objective of the entire fund. Our classification of investment objectives is listed as follows:

- Aggressive Growth: Wiesenberger: AGG, AG; Strategic Insight: AGG; Lipper: CA.
- Growth: Wiesenberger: G, G-S, S-G, GRO, LTG; Strategic Insight: GRO; Lipper: G.
- Growth and Income: Wiesenberger: GCI, G-I, G-I-S, G-S-I, I-G, I-G-S, I-S-G, S-G-I, S-I-G, GRI; Strategic Insight: GRI; Lipper: GI.
- Income: Wiesenberger: I, I-S, IEQ, ING; Strategic Insight: ING; Lipper: EI, EIEI.

C. Simulation Study Details

This appendix provides details regarding the simulation study in Section IV.

C.1. Data Generating Process: Factor Returns and Residuals

In order to simulate factor returns in Equation (3), we draw from the normal distribution $\boldsymbol{X}_t \sim \mathcal{N}(0, \Sigma_X)$. Σ_X is the empirical covariance structure of the factor portfolio returns which we model via a Cholesky factorization. The covariance matrix Σ_X is given as:

$$\Sigma_X = \begin{bmatrix} \sigma^2_{MKT} & \sigma_{MKT,SMB} & \sigma_{MKT,HML} & \sigma_{MKT,MOM} \\ \sigma_{MKT,SMB} & \sigma^2_{SMB} & \sigma_{SMB,HML} & \sigma_{SMB,MOM} \\ \sigma_{MKT,HML} & \sigma_{SMB,HML} & \sigma^2_{HML} & \sigma_{HML,MOM} \\ \sigma_{MKT,MOM} & \sigma_{SMB,MOM} & \sigma_{HML,MOM} & \sigma^2_{MOM} \end{bmatrix}. \tag{19}$$

For this 4×4 symmetric and positive definite matrix, we find $\mathbf{C} = (c_{ij})_{i,j=1,...,4}$ s.t. $\Sigma_X = \mathbf{C}\mathbf{C}^T$ where $\mathbf{C}$ is a square lower triangular matrix. We now calculate c_{ij} recursively as:

$$\begin{aligned} c_{jj} &= \sqrt{\Sigma_{jj,X} - \sum_{s=1}^{j-1} c^2_{js}}, \\ c_{ij} &= \frac{1}{c_{jj}}(\Sigma_{ij,X} - \sum_{s=1}^{j-1} \Sigma_{is,X}\Sigma_{js,X}), \quad \text{for} \quad i > j. \end{aligned} \tag{20}$$

The implication of the equations above is that we can simulate correlated factor portfolio returns by simulating standard normal variates. That is, we can generate the four factor returns component by component at each time t as:

$$\begin{aligned} X_k &= \sum_{s=1}^{k} c_{ks} Y_s, \quad k = 1, ..., 4, \\ Y_s &\overset{i.i.d.}{\sim} \mathcal{N}(0, 1). \end{aligned} \tag{21}$$

Similar to generating correlated factor portfolio returns, we follow the same Cholesky factorization when we impose different correlation structures in fund residual series. Consider

the covariance matrix of the fund residual series:

$$\Omega_u = \begin{bmatrix} \omega_{11}^2 & \omega_{12} & \dots & \omega_{1N} \\ \omega_{21} & \omega_{22}^2 & \dots & \omega_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{N1} & \omega_{N2} & \dots & \omega_{NN}^2 \end{bmatrix}. \tag{22}$$

As an example, if we want a correlation coefficient of $\rho = 0.3$ for all non diagonal elements, then we model the covariance between fund i and j as:

$$\omega_{ij} = \rho_{ij}\sqrt{\omega_{ii}^2\omega_{jj}^2}, \tag{23}$$

where ω_{ii} and ω_{jj} are standard deviation estimates of fitted residual series for fund i and j. When we simulate with empirically motivated correlation coefficients, we estimate these for all funds with 60 return observations in the last five years of our sample (January 2013 - December 2017). We choose this period since this is when most funds exist simultaneously (929 funds). In each Monte Carlo simulation, we generate excess returns for 500 funds. Therefore, we draw 500 random funds and use their pairwise correlation coefficients in every simulated panel of excess returns.

Note that due to numerical instability, the covariance matrix in Equation (22) can be rank deficient. When this is the case, we use an eigenvalue decomposition performed by the MATLAB function *cholcov()*.

Figure I: Composite Distribution of Mutual Fund Manager Skill

This figure shows the unobserved alpha-distributions for skilled, unskilled, and zero-alpha funds (solid red, blue, and green lines respectively), and the observed composite distribution of all mutual fund alphas, (dashed black line). The figure shows the true underlying fund alpha distributions and resulting composite distribution used in the simulation study. The distributions are characterized by the following parameter values, which are obtained from empirical data: $\pi_+ = 0.11$, $\pi_- = 0.59$, $\pi_0 = 0.30$; $\alpha_+ = 1.32\%$, $\alpha_- = -1.77\%$, $\alpha_0 = 0\%$; $\sigma_+ = 0.38\%$, $\sigma_- = 0.61\%$ and $\sigma_0 = 0.50\%$. The α's and σ's are stated with their annualized values.

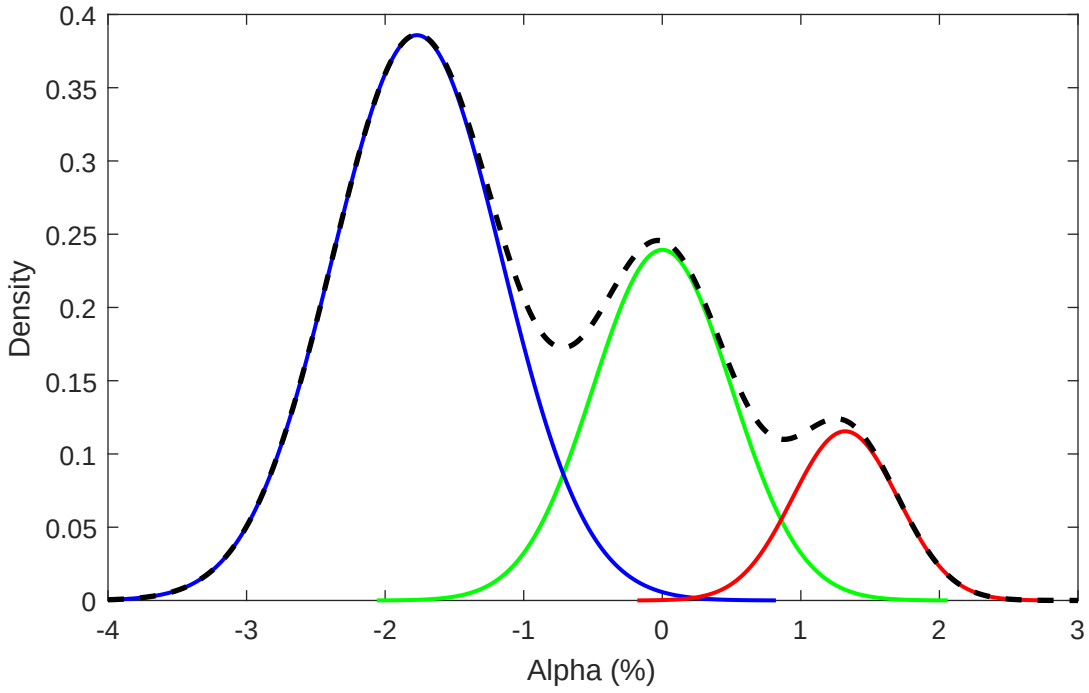

Table I: Descriptive Statistics

This table shows the number, TNA, returns in excess of the risk-free rate, standard deviations-, skewness coefficients-, and kurtosis coefficients of excess returns of U.S. no-load open-end equity funds existing between June 1992 and December 2017. The statistics are presented for all investment objectives and for investment styles: aggressive growth, growth, growth and income, and income in panels A-E, respectively. In each panel for a given subperiod, the first and second columns show the number of funds in existence, the third and fourth columns show the average fund TNA in $ million, the fifth and sixth columns show the annualized average fund excess returns in percent($12 * \text{Avg}(R_{i,t})$), the seventh and eighth columns show the annualized average standard deviations of fund excess returns in percent($\sqrt{12} * \text{Std}(R_{i,t})$), the ninth and tenth columns show the average skewness coefficients of fund excess returns, and the eleventh and twelfth columns show the average kurtosis coefficients of fund excess returns. Odd numbered columns contain statistics for all funds that have at least one return observation in a given subperiod, while even numbered columns contain statistics for all funds that have 60 return observations in a given subperiod. Each subperiod consists of 60 months and the entire evaluation period consists of 307 months.

	Funds		TNA		Exc. Ret.(%)		Std. Dev.(%)		Skewness		Kurtosis	
Period	≥1	≥60	≥1	≥60	≥1	≥60	≥1	≥60	≥1	≥60	≥1	≥60
Panel A: All Investment Styles												
92-97	466	145	284.28	325.47	10.50	10.45	11.82	10.66	-0.29	-0.23	2.73	2.92
97-02	767	284	592.93	641.60	2.41	3.24	19.97	20.10	-0.26	-0.40	3.18	3.52
02-07	958	420	672.37	721.83	6.71	7.18	11.83	12.97	-0.48	-0.61	3.45	4.17
07-12	1473	530	773.50	837.20	-0.49	-0.22	20.38	20.53	-0.38	-0.44	3.16	3.42
12-17	1503	918	1099.25	1240.23	13.61	14.03	10.00	10.53	-0.16	-0.21	3.04	3.17
13-17	1473	929	1149.42	1320.27	12.89	13.14	9.94	10.40	-0.19	-0.21	3.20	3.24
92-17	2279	1481	698.32	772.31	6.64	6.90	15.08	15.66	-0.47	-0.57	4.22	4.70
Panel B: Aggressive Growth Funds												
92-97	32	1	75.82	84.31	10.45	11.06	16.93	13.14	-0.02	-0.08	2.85	3.01
97-02	60	20	308.37	340.66	5.09	6.47	25.52	27.67	-0.18	-0.25	3.27	3.69
02-07	67	33	285.03	295.81	8.18	9.09	14.74	15.11	-0.36	-0.50	3.30	3.61
07-12	87	40	425.72	441.41	0.76	0.28	22.50	22.08	-0.47	-0.43	3.38	3.44
12-17	75	59	763.61	797.12	13.61	14.03	11.63	11.88	-0.27	-0.31	3.15	3.23
13-17	75	60	803.96	843.84	13.94	14.21	11.61	11.71	-0.29	-0.32	3.22	3.35
92-17	123	96	383.55	404.49	7.87	8.08	18.82	18.69	-0.51	-0.52	4.50	4.74
Panel C: Growth Funds												
92-97	264	87	242.55	273.74	10.50	10.52	12.22	11.26	-0.27	-0.22	2.74	2.96
97-02	434	167	535.99	577.05	2.12	3.12	20.92	20.58	-0.24	-0.41	3.15	3.50
02-07	562	246	598.93	646.95	6.72	7.21	12.08	13.18	-0.45	-0.59	3.34	4.00
07-12	868	307	750.80	821.79	-0.11	0.12	20.75	20.89	-0.36	-0.43	3.15	3.41
12-17	867	532	1057.93	1187.81	13.08	13.29	10.39	10.83	-0.15	-0.20	2.98	3.11
13-17	846	542	1104.09	1261.12	13.29	13.50	10.37	10.69	-0.17	-0.20	3.13	3.20
92-17	1315	849	651.07	720.21	6.72	7.01	15.64	16.03	-0.44	-0.54	4.16	4.64
Panel D: Growth and Income Funds												
92-97	151	53	410.64	471.95	10.51	10.15	10.22	9.72	-0.37	-0.24	2.70	2.88
97-02	234	91	773.30	829.66	2.25	2.71	17.45	17.75	-0.33	-0.42	3.24	3.55
02-07	269	126	860.01	904.65	6.68	6.65	10.98	12.12	-0.56	-0.68	3.72	4.56
07-12	409	155	915.63	976.47	-1.33	-0.89	19.45	19.59	-0.39	-0.45	3.16	3.40
12-17	439	267	1317.29	1494.75	11.59	11.51	9.18	9.85	-0.14	-0.21	3.08	3.19
13-17	429	265	1392.60	1612.02	12.28	12.55	9.08	9.74	-0.20	-0.21	3.28	3.30
92-17	658	430	872.66	960.09	6.35	6.51	13.70	14.67	-0.53	-0.63	4.32	4.78
Panel E: Income Funds												
92-97	19	4	114.32	119.43	12.22	12.68	10.18	9.31	-0.45	-0.43	2.66	2.63
97-02	39	6	474.16	532.12	3.45	3.81	15.98	16.88	-0.24	-0.34	3.07	3.08
02-07	60	15	1099.22	1309.22	4.52	6.85	10.06	12.10	-0.49	-0.70	3.38	4.85
07-12	109	28	721.80	781.44	-1.61	-0.86	19.26	19.73	-0.37	-0.41	3.11	3.66
12-17	122	60	854.97	1032.28	11.59	11.51	9.14	9.49	-0.18	-0.27	3.23	3.56
13-17	123	62	838.66	1042.09	11.42	11.45	9.04	9.32	-0.21	-0.23	3.32	3.27
92-17	183	106	653.58	759.10	6.27	6.79	13.49	13.98	-0.50	-0.61	4.18	4.85

Table II: Benchmark Mutual Fund Performance

This table shows annualized alphas for the unconditional four-factor model in Equation (1) for all funds as well as for investment styles: aggressive growth, growth, growth and income, and income. In panel A regressions for equal-weighted portfolios are based on monthly observations from June 1992 to December 2017. Panel A shows annualized estimates of intercepts ($12 * \hat{\alpha}$), regression slopes ($\hat{b}$, $\hat{s}$, $\hat{h}$ and $\hat{m}$) and adjusted-R^2. Figures in parentheses denote Newey and West (1987) heteroskedasticity and autocorrelation consistent estimates of double-sided p-values under the null hypothesis that a parameter equals zero. Panel B shows annualized intercept estimates in percent for equal-weighted portfolios of funds that have existed for at least one month or the entire five-year subperiod. Equal-weighted portfolios are rebalanced monthly in order to include all existing funds at the beginning of a given month.

Panel A: Full Sample Factor Model Estimates

	$12 * \hat{\alpha}$	$\hat{b}$	$\hat{s}$	$\hat{h}$	$\hat{m}$	R^2
All (1,481)	-1.00%	0.96	0.03	0.01	-0.01	98.70%
	(0.02)	(0.00)	(0.08)	(0.52)	(0.58)	
Aggressive Growth (96)	-0.69%	1.04	0.30	-0.13	0.02	95.52%
	(0.40)	(0.00)	(0.00)	(0.00)	(0.26)	
Growth (849)	-1.01%	0.98	0.05	-0.03	0.00	98.51%
	(0.03)	(0.00)	(0.01)	(0.22)	(0.97)	
Growth & Income (430)	-1.03%	0.92	-0.05	0.10	-0.02	98.61%
	(0.01)	(0.00)	(0.01)	(0.00)	(0.05)	
Income (106)	-0.62%	0.90	-0.05	0.13	-0.03	97.90%
	(0.18)	(0.00)	(0.00)	(0.00)	(0.03)	

Panel B: Annualized Alpha for Five Year Subperiods

Time Period	1992-1997	1997-2002	2002-2007	2007-2012	2012-2017	2013-2017
All (Obs≥1)	-1.03	-1.15	-1.73	-1.83	-1.15	-1.46
All (Obs≥60)	-0.85	-0.48	-1.39	-1.51	-1.02	-1.32
Aggressive Growth (Obs≥1)	-1.13	-0.06	-1.85	-1.82	-1.15	-0.94
Aggressive Growth (Obs≥60)	-0.73	0.94	-1.30	-2.18	-0.82	-1.37
Growth (Obs≥1)	-1.06	-1.26	-1.69	-1.69	-1.10	-1.41
Growth (Obs≥60)	-0.69	-0.58	-1.36	-1.37	-0.94	-1.25
Growth and Income (Obs≥1)	-1.00	-1.25	-1.48	-2.05	-1.22	-1.53
Growth and Income (Obs≥60)	-1.23	-0.69	–1.49	-1.60	-1.15	-1.42
Income (Obs≥1)	0.37	-0.69	-3.32	-2.18	-1.30	-1.62
Income (Obs≥60)	0.90	0.77	-1.09	-1.67	-1.35	-1.57

Table III: Small Sample Properties with No Serial Dependence

This table shows estimates for the simulated cross-sectional skill distributions generated by Equation (3). Skill estimates are presented for the mutual fund selection methods: benchmark model (1), benchmark model with a Bonferroni correction (2), Kosowski et al. (2006) (3), Fama and French (2010) (4), Barras et al. (2010) (5), Ferson and Chen (2015) (6), and Harvey and Liu (2018) (7). We present estimates based on three different data generating processes for $N = 500$ funds across $T = 60$ months. For all results in this table return panels are generated with serially independent residuals. In Panel A we consider fixed pairwise correlation coefficients between all funds of $\rho = 0$. In Panel B we set $\rho = 0.3$, while Panel C presents results based on the empirical correlation coefficients to match actual cross-sectional data heterogeneity. The true weights of the $s = +, -, 0$ skill distributions are $\pi_+ = 11\%$, $\pi_- = 59\%$ and $\pi_0 = 30\%$. For $M = 100$ Monte Carlo simulation we calculate Bias and RMSE for individual skill distributions π_s. Bias and RMSE use the following definitions: $\text{Bias}_s = \frac{1}{M}\sum_{m=1}^{M}\hat{\pi}_{m,s} - \pi_s$ and $\text{RMSE}_s = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\pi}_{m,s} - \pi_s)^2}$. We also summarize the average absolute bias and the average RMSE across all skill distribution estimates where Avg. Abs. Bias $= \frac{1}{S}\sum_{s=1}^{S}|\text{Bias}_s|$ and Avg. RMSE $= \frac{1}{S}\sum_{s=1}^{S}\text{RMSE}_s$.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A					$\rho = 0$			
π_-	Bias	-38.13	-53.80	2.16	2.16	-29.27	-9.30	-33.28
	RMSE	38.16	53.81	2.92	2.92	29.47	10.24	46.39
π_0	Bias	41.24	63.76	-7.67	-1.29	30.36	2.06	-6.21
	RMSE	41.29	63.77	8.79	5.37	30.77	6.40	32.97
π_+	Bias	-3.11	-9.96	5.51	-0.87	-1.10	7.24	39.50
	RMSE	3.37	9.97	7.44	5.66	3.05	8.09	55.89
Average Absolute Bias		27.49	42.50	5.11	1.44	20.24	6.20	26.33
Average RMSE		27.61	42.51	6.38	4.65	21.09	8.24	45.08
Panel B					$\rho = 0.3$			
π_-	Bias	-37.60	-53.64	-2.44	-27.00	-29.23	-6.70	-32.13
	RMSE	38.69	53.72	20.58	35.23	32.62	22.83	46.62
π_0	Bias	40.15	63.50	-8.63	30.94	28.86	-5.38	4.13
	RMSE	40.56	63.55	10.77	36.31	30.24	12.55	39.13
π_+	Bias	-2.55	-9.86	11.07	-3.94	0.36	12.08	27.99
	RMSE	6.52	9.91	25.74	12.50	10.88	20.56	48.59
Average Absolute Bias		26.77	42.33	7.38	20.63	19.48	8.05	21.42
Average RMSE		28.59	42.39	19.03	28.01	24.58	18.65	44.78
Panel C					Data Dependent			
π_-	Bias	-22.70	-49.36	2.56	-5.83	-22.70	-11.86	-32.90
	RMSE	31.69	49.42	6.96	13.47	23.56	14.58	46.10
π_0	Bias	31.10	58.04	-15.55	6.01	18.84	4.05	-4.98
	RMSE	31.32	58.08	16.15	10.54	19.50	8.45	32.27
π_+	Bias	0.26	-8.67	12.99	-0.18	3.85	7.81	37.87
	RMSE	3.20	8.74	16.32	6.99	6.52	9.74	53.95
Average Absolute Bias		18.02	38.69	10.37	4.01	15.13	7.91	25.25
Average RMSE		22.07	38.75	13.14	10.33	16.53	10.92	44.11

Table IV: Small Sample Properties with First Order Serial Dependence

This table shows estimates for the simulated cross-sectional skill distributions generated by Equation (3). Skill estimates are presented for the mutual fund selection methods: benchmark model (1), benchmark model with a Bonferroni correction (2), Kosowski et al. (2006) (3), Fama and French (2010) (4), Barras et al. (2010) (5), Ferson and Chen (2015) (6) and Harvey and Liu (2018) (7). We present estimates based on three different data generating processes for $N = 500$ funds across $T = 60$ months. For all results in this table return panels are generated with first order serial dependence in residuals for $\gamma = 0.08$. In Panel A we consider fixed pairwise correlation coefficients between all funds of $\rho = 0$. In Panel B we set $\rho = 0.3$, while Panel C presents results based on the empirical correlation coefficients to match actual cross-sectional data heterogeneity. The true weights of the $s = +, -, 0$ skill distributions are $\pi_+ = 11\%$, $\pi_- = 59\%$ and $\pi_0 = 30\%$. For $M = 100$ Monte Carlo simulation we calculate Bias and RMSE for individual skill distributions π_s. Bias and RMSE use the following definitions: $\text{Bias}_s = \frac{1}{M}\sum_{m=1}^{M}\hat{\pi}_{m,s} - \pi_s$ and $\text{RMSE}_s = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\pi}_{m,s} - \pi_s)^2}$. We also summarize the average absolute bias and the average RMSE across all skill distribution estimates where Avg. Abs. Bias $= \frac{1}{S}\sum_{s=1}^{S}|\text{Bias}_s|$ and Avg. RMSE $= \frac{1}{S}\sum_{s=1}^{S}\text{RMSE}_s$.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A					$\rho = 0$			
π_-	Bias	-38.30	-54.19	1.96	1.96	-30.04	-10.25	-35.46
	RMSE	38.34	54.19	2.77	2.77	30.20	11.09	48.66
π_0	Bias	41.72	64.22	-6.74	-0.44	31.47	3.60	-2.55
	RMSE	41.78	64.23	8.05	4.68	31.87	7.36	36.23
π_+	Bias	-3.42	-10.03	4.78	-1.52	-1.42	6.65	38.01
	RMSE	3.62	10.04	6.94	5.17	3.48	7.62	55.25
Average Absolute Bias		27.82	42.81	4.49	1.31	20.98	6.83	25.34
Average RMSE		27.91	42.82	5.92	4.21	21.85	8.69	46.71
Panel B					$\rho = 0.3$			
π_-	Bias	-37.68	-53.89	-3.02	-25.25	-29.16	-6.50	-36.89
	RMSE	39.19	53.98	24.28	37.07	33.87	26.16	47.61
π_0	Bias	40.79	63.90	-8.81	30.56	29.75	-3.76	4.93
	RMSE	41.40	63.95	11.02	38.45	31.61	13.74	38.67
π_+	Bias	-3.12	-10.01	11.83	-5.31	-0.58	10.26	31.96
	RMSE	6.87	10.06	27.88	11.04	11.32	20.81	51.05
Average Absolute Bias		27.20	42.60	7.88	20.37	19.83	6.84	24.60
Average RMSE		29.15	42.66	21.06	28.85	25.60	20.24	45.77
Panel C					Data Dependent			
π_-	Bias	-31.80	-49.89	2.49	-5.40	-22.79	-12.19	-32.36
	RMSE	32.19	49.95	8.36	15.14	23.90	15.09	44.90
π_0	Bias	32.06	58.75	-13.97	5.98	19.98	5.08	-4.74
	RMSE	32.33	58.80	15.10	11.42	21.08	8.58	32.24
π_+	Bias	-0.27	-8.86	11.48	-0.58	2.81	7.11	37.09
	RMSE	4.09	8.94	17.15	8.71	7.34	10.66	52.50
Average Absolute Bias		21.38	39.17	9.31	3.99	15.19	8.13	24.73
Average RMSE		22.87	39.23	13.54	11.76	17.44	11.44	43.21

Table V: Large Sample Properties with No Serial Dependence

This table shows estimates for the simulated cross-sectional skill distributions generated by Equation (3). Skill estimates are presented for the mutual fund selection methods: benchmark model (1), benchmark model with a Bonferroni correction (2), Kosowski et al. (2006) (3), Fama and French (2010) (4), Barras et al. (2010) (5), Ferson and Chen (2015) (6) and Harvey and Liu (2018) (7). We present estimates based on three different data generating processes for $N = 500$ funds across $T = 307$ months. For all results in this table return panels are generated with serially independent residuals. In Panel A we consider fixed pairwise correlation coefficients between all funds of $\rho = 0$. In Panel B we set $\rho = 0.3$, while Panel C presents results based on the empirical correlation coefficients to match actual cross-sectional data heterogeneity. The true weights of the $s = +, -, 0$ skill distributions are $\pi_+ = 11\%$, $\pi_- = 59\%$ and $\pi_0 = 30\%$. For $M = 100$ Monte Carlo simulation we calculate Bias and RMSE for individual skill distributions π_s. Bias and RMSE use the following definitions: $\text{Bias}_s = \frac{1}{M}\sum_{m=1}^{M}\hat{\pi}_{m,s} - \pi_s$ and $\text{RMSE}_s = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\pi}_{m,s} - \pi_s)^2}$. We also summarize the average absolute bias and the average RMSE across all skill distribution estimates where Avg. Abs. Bias $= \frac{1}{S}\sum_{s=1}^{S}|\text{Bias}_s|$ and Avg. RMSE $= \frac{1}{S}\sum_{s=1}^{S}\text{RMSE}_s$.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A					$\rho = 0$			
π_-	Bias	-21.26	-37.95	3.95	3.95	-13.47	-11.35	-28.55
	RMSE	21.35	37.99	4.48	4.48	13.73	11.63	37.98
π_0	Bias	16.55	42.85	-21.56	-21.33	3.91	1.73	4.98
	RMSE	16.71	42.89	21.62	21.40	5.40	3.65	30.01
π_+	Bias	4.71	-4.89	17.60	17.38	9.57	9.62	23.57
	RMSE	4.93	5.00	17.84	17.63	9.90	9.87	36.99
Average Absolute Bias		14.17	28.57	14.37	14.22	8.98	7.57	19.04
Average RMSE		14.33	28.63	14.65	14.50	9.68	8.38	34.99
Panel B					$\rho = 0.3$			
π_-	Bias	-21.85	-38.35	2.72	-8.94	-14.05	-10.17	-30.18
	RMSE	23.09	38.67	9.28	14.47	16.55	13.90	40.30
π_0	Bias	16.32	42.88	-21.99	0.55	3.19	-2.53	4.12
	RMSE	16.76	43.04	22.21	5.34	5.53	5.11	30.76
π_+	Bias	5.53	-4.53	19.27	8.39	10.85	12.70	26.06
	RMSE	7.48	5.21	22.34	11.76	13.20	14.79	39.06
Average Absolute Bias		14.57	28.59	14.66	5.96	9.37	8.47	20.12
Average RMSE		15.78	28.97	17.94	10.52	11.76	11.26	36.71
Panel C					Data Dependent			
π_-	Bias	-14.21	-28.92	4.55	2.83	-7.68	-8.77	-21.91
	RMSE	14.63	29.10	5.51	5.03	8.51	9.42	30.85
π_0	Bias	5.08	29.58	-24.40	-17.27	-5.84	-3.57	4.08
	RMSE	5.71	29.72	24.44	17.40	6.71	4.38	25.61
π_+	Bias	9.12	-0.65	19.85	14.44	13.52	12.34	17.83
	RMSE	9.41	1.76	20.23	14.86	13.86	12.64	26.66
Average Absolute Bias		9.47	19.72	16.27	11.52	9.01	8.23	14.61
Average RMSE		9.92	20.19	16.73	12.43	9.69	8.81	27.70

Table VI: Full Sample Mutual Fund Skill Fractions

This table shows estimates for groups of unskilled (π_-), zero-alpha (π_0), and skilled (π_+) funds with at least 60 monthly return observations covering the sample period from June 1992 to December 2017. Skill estimates are presented for all investment styles (1,481 funds) in Panel A and for investment styles: Aggressive Growth (96 funds), Growth (849 funds), Growth and Income (430 funds), and Income (106 funds) in Panels B-E, respectively. The table presents results for the following mutual fund selection methods: benchmark model (1), benchmark model with a Bonferroni correction (2), Kosowski et al. (2006) (3), Fama and French (2010) (4), Barras et al. (2010) (5), Ferson and Chen (2015) (6), and Harvey and Liu (2018) (7).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: All Investment Styles							
π_-	23.70%	0.61%	79.27%	79.27%	35.85%	85.64%	100.00%
π_0	75.56%	99.39%	20.73%	20.73%	64.15%	14.36%	
π_+	0.74%	0.00%	0.00%	0.00%	0.00%	0.00%	
Panel B: Aggressive Growth Funds							
π_-	13.54%	0.00%	77.08%	76.04%	24.24%	68.88%	100.00%
π_0	86.46%	100.00%	22.92%	23.96%	75.76%	31.12%	
π_+	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
Panel C: Growth Funds							
π_-	22.85%	0.71%	79.86%	79.86%	37.97%	66.14%	100.00%
π_0	76.33%	99.29%	20.14%	20.14%	62.03%	33.86%	
π_+	0.82%	0.00%	0.00%	0.00%	0.00%	0.00%	
Panel D: Growth and Income Funds							
π_-	26.28%	1.63%	78.84%	78.84%	37.21%	80.39%	100.00%
π_0	73.26%	98.37%	21.16%	21.16%	62.79%	19.61%	
π_+	0.47%	0.00%	0.00%	0.00%	0.00%	0.00%	
Panel E: Income Funds							
π_-	29.25%	4.72%	77.36%	77.36%	43.40%	67.68%	100.00%
π_0	68.87%	95.28%	22.64%	22.64%	56.60%	32.32%	
π_+	1.89%	0.00%	0.00%	0.00%	0.00%	0.00%	

Table VII: Five-Year Mutual Fund Skill Fractions

This table shows estimates for groups of unskilled (π_-), zero-alpha (π_0), and skilled (π_+) funds across five-year subperiods from June 1992 to December 2017. For each five-year sub-period, a fund is included if it has at least 24 monthly return observations out of potentially 60 observations. Skill estimates are presented for the following mutual fund selection methods: benchmark model (1), benchmark model with a Bonferroni correction (2), Kosowski et al. (2006) (3), Fama and French (2010) (4), Barras et al. (2010) (5), Ferson and Chen (2015) (6) and Harvey and Liu (2018) (7).

Time Period		(1)	(2)	(3)	(4)	(5)	(6)	(7)
92-97	π_-	9.81%	0.38%	47.55%	0.75%	9.43%	68.24%	100.00%
	π_0	83.02%	99.62%	48.68%	99.25%	90.57%	12.06%	
	π_+	7.17%	0.00%	3.77%	0.00%	0.00%	19.70%	
97-02	π_-	6.29%	0.39%	47.74%	0.00%	10.43%	49.39%	100.00%
	π_0	88.21%	99.02%	48.53%	99.61%	86.88%	19.02%	
	π_+	5.50%	0.59%	3.73%	0.39%	2.69%	31.59%	
02-07	π_-	21.98%	1.22%	72.21%	71.91%	32.27%	66.68%	100.00%
	π_0	76.95%	98.63%	27.79%	28.09%	67.73%	33.32%	
	π_+	1.07%	0.15%	0.00%	0.00%	0.00%	0.00%	
07-12	π_-	24.25%	0.56%	78.26%	78.08%	34.14%	88.13%	100.00%
	π_0	75.00%	99.44%	21.74%	21.92%	65.86%	10.85%	
	π_+	0.75%	0.00%	0.00%	0.00%	0.00%	1.02%	
12-17	π_-	13.58%	0.57%	77.02%	77.02%	27.61%	63.79%	100.00%
	π_0	85.75%	99.43%	22.98%	22.98%	72.39%	36.21%	
	π_+	0.67%	0.00%	0.00%	0.00%	0.00%	0.00%	
13-17	π_-	19.88%	0.58%	81.85%	81.76%	39.19%	73.16%	100.00%
	π_0	79.73%	99.42%	18.15%	18.24%	60.81%	26.84%	
	π_+	0.39%	0.00%	0.00%	0.00%	0.00%	0.00%	

Research Papers 2018

2018-18: Giorgio Mirone: Cross-sectional noise reduction and more efficient estimation of Integrated Variance

2018-19: Kim Christensen, Martin Thyrsgaard and Bezirgen Veliyev: The realized empirical distribution function of stochastic variance with application to goodness-of-fit testing

2018-20: Ruijun Bu, Kaddour Hadri and Dennis Kristensen: Diffusion Copulas: Identification and Estimation

2018-21: Kim Christensen, Roel Oomen and Roberto Renò: The drift burst hypothesis

2018-22: Russell Davidson and Niels S. Grønborg: Time-varying parameters: New test tailored to applications in finance and macroeconomics

2018-23: Emilio Zanetti Chini: Forecasters' utility and forecast coherence

2018-24: Tom Engsted and Thomas Q. Pedersen: Disappearing money illusion

2018-25: Erik Christian Montes Schütte: In Search of a Job: Forecasting Employment Growth in the US using Google Trends

2018-26: Maxime Morariu-Patrichi and Mikko Pakkanen: State-dependent Hawkes processes and their application to limit order book modelling

2018-27: Tue Gørgens and Allan H. Würtz: Threshold regression with endogeneity for short panels

2018-28: Mark Podolskij, Bezirgen Veliyev and Nakahiro Yoshida: Edgeworth expansion for Euler approximation of continuous diffusion processes

2018-29: Isabel Casas, Jiti Gao and Shangyu Xie: Modelling Time-Varying Income Elasticities of Health Care Expenditure for the OECD

2018-30: Yukai Yang and Luc Bauwens: State-Space Models on the Stiefel Manifold with A New Approach to Nonlinear Filtering

2018-31: Stan Hurn, Nicholas Johnson, Annastiina Silvennoinen and Timo Teräsvirta: Transition from the Taylor rule to the zero lower bound

2018-32: Sebastian Ankargren, Måns Unosson and Yukai Yang: A mixed-frequency Bayesian vector autoregression with a steady-state prior

2018-33: Carlos Vladimir Rodríguez-Caballero and Massimiliano Caporin: A multilevel factor approach for the analysis of CDS commonality and risk contribution

2018-34: James G. MacKinnon, Morten Ørregaard Nielsen, David Roodman and Matthew D. Webb: Fast and Wild: Bootstrap Inference in Stata Using boottest

2018-35: Sepideh Dolatabadim, Paresh Kumar Narayan, Morten Ørregaard Nielsen and Ke Xu: Economic significance of commodity return forecasts from the fractionally cointegrated VAR model

2018-36: Charlotte Christiansen, Niels S. Grønborg and Ole L. Nielsen: Mutual Fund Selection for Realistically Short Samples