The Cross-Sectional Distribution of Fund Skill Measures

Laurent Barras, Patrick Gagliardini, and Olivier Scaillet*

This version, June 3, 2018

JEL Classification: G11, G12, C14, C33, C58.
Keywords: Mutual fund skill, nonparametric density estimation, large and unbalanced panel

*Barras is at McGill University (Desautels Faculty of Management), Gagliardini is at the University of Svizzera Italiana and at the Swiss Finance Institute (SFI), and Scaillet is at the University of Geneva (Geneva Finance Research Institute (GFRI) and at the SFI. We thank participants at the McGill/HEC 2018 Spring Workshop as well as seminar participants at McMaster University. The first author thank the Social Sciences and Humanities Research Council of Canada (SSHRC) for its financial support. The third author acknowledges financial support by the Geneva Finance Research Institute. He did part of this research when visiting the University of Cergy-Pontoise.

ABSTRACT

We develop a novel nonparametric approach to estimate the entire distribution of skill across mutual funds. Our approach is simple, fast, and immune to misspecification errors that plague traditional parametric approaches. As such, it provides a unified and consistent framework for jointly studying the two dimensions of skill, namely the ability to (i) detect profitable trades; (ii) mitigate capacity constraints. Our empirical analysis reveals that while 88.6% of the managers can detect profitable trades, 86.1% are also subject to capacity constraints. The two skill dimensions exhibit strong heterogeneity both within and across fund groups. Importantly, they are also negatively correlated. Combining them into a single skill measure—the value added—, we find that 70% of the funds create a positive value equal to $9 mio. per year on average.

I Introduction

Over the past 50 years, the academic literature on mutual funds has focused on performance. For instance, Jensen (1968), Elton et al. (1993), and Carhart (1997)) find that the average alpha received by investors is negative, while more recent studies show that the vast majority of funds deliver negative alphas (e.g., Barras, Scaillet, and Wermers (2010), Harvey and Liu (2018)). In constrast, far less attention has been given to the analysis of mutual fund skill—a point forcefully made by Berk and van Binsbergen (2015). Whereas skill and performance are typically used interchangeably, the two notions are very different. On the one hand, skill is defined from the viewpoint of fund managers—it measures their ability to create value through superior stock picking or market timing. On the other hand, performance is defined from the viewpoint of investors—it measures whether the value created by the funds, if any, is passed on to them.

In this paper, we develop a novel approach for estimating the entire distribution of mutual fund skill. We explicitly distinguish between the two dimensions of managerial skill using the model of Berk and Green (2004) in which the the gross alpha α_{it} of each fund i is a function of its lagged size $q_{i,t-1}$: $\alpha_{it} = a_i + b_i q_{i,t-1}$. The first dimension, which is measured by the first-dollar alpha a_i, determines the manager's ability to identify profitable trading opportunities. The second dimension, which is measured by the size coefficient b_i, captures the manager's ability to mitigate the impact of capacity constraints. We also measure skill using the value added va_i proposed by Berk and van Binsbergen (2015). This measure has a powerful economic interpretation because it determines the total profits from exploiting the two skill dimensions—as shown by Berk and Green (2004), the optimal value added is a function of both a_i and b_i. Finally, we examine the fund gross alpha α_i which is commonly used as the benchmark for measuring skill (e.g., Grinblatt and Titman (1989b), Jensen (1969)).

Our estimation approach is nonparametric. As such, it does not require any assumption on the shape of the cross-sectional skill density $\phi(m)$, where $m = \{\alpha, a, b, va\}$ is a general formulation that encompasses all four measures of skill. As such, it departs from the Bayesian or parametric approaches used in previous work which require a pre-specified distribution for $\phi(m)$—typically, a mixture of normal distributions (e.g., Harvey and Liu (2018), Jones and Shanken (2005)). The only required input for applying our approach is the estimated skill measure $\hat{m}_i$ of each fund. Our analysis reveals that using the estimated skill measures instead of the *true* ones ($\hat{m}_i$ instead of m_i) introduces an Error-in-Variable (EIV) bias that must be accounted for. This result is reminiscent of the well-known EIV bias for the two-pass regression (Shanken (1992)) and implies that the unadjusted estimators can be severely biased.

Our novel approach provides several advantages. First, it immune to misspecification errors because it is nonparametric. This is particularly important here because we simultaneously examine several skill measures—given that they are all theoretically closely related, specifying distributions for each skill measure is likely to introduce inconsistencies. Second, the implementation of our approach is simple and fast. Intuitively, the procedure is akin to computing an histogram using as only inputs the estimated skill measure $\hat{m}_i$ of each fund. This contrasts with the complex and lengthy Gibbs sampling and maximum likelihood methods used in Bayesian/parametric approaches. Third, our approach provides a unified framework to the different characterizations of the skill distribution. In addition to the density, we can use our approach to estimate (i) the distribution moments such as the mean, variance, skewness, and kurtosis, (ii) the cumulative distribution function (cdf), and (iii) the distribution quantile (such as the median).

The remainder of the paper is as follows. The next section outlines the problem of estimating the cross-sectional distribution of funds skill measures. It presents the different skill measures and the relationship between them before explaining how we can estimate them. It overviews our nonparametric approach to estimate the distribution of skill measures through kernel smoothing and how we can account for smoothing bias and EIV bias. It analyses the behaviour of the estimation bias before discussing several extensions. Section III describes the individual mutual fund and benchmark model data. Section IV contains the empirical results of the paper, while Section V concludes. A technical appendix contains the proofs of the theoretical results underlying our estimation approach.

II The Cross-Sectional Distribution of Skill Measures

A The Skill Measures

A.1 Overview of the Skill Measures

To begin the presentation of our approach, we define the different skill measures examined in the paper. While it is common to use skill and performance interchangeably, it is important to distinguish between the two notions. Skill is defined from the viewpoint of fund managers—it measures their ability to create value through superior stock picking or market timing. In contrast, performance is defined from the viewpoint of investors—it measures whether the value created by the funds, if any, is passed on to them.

Our first measure is the fund gross alpha α. This measure is commonly used in the previous literature and thus represents a natural candidate for measuring skill.[1] The gross alpha is defined as the average abnormal gross return measured as the return difference between the fund and its benchmark index. The gross alpha is useful for determining whether fund managers are skilled—under certain conditions, Grinblatt and Titman (1989) show that α must be positive if the manager has private information about future stock returns.

However, the gross alpha must be used with caution when measuring skill among managers because it does not control for the differences in fund size—α is typically negatively related to size because managers are subject to capacity constraints that increase the cost of active management. Therefore, it potentially fails to capture the two dimensions of managerial skill which are (i) the ability to generate profitable trading ideas and (ii) the ability to limit the impact of capacity constraints on the fund return. For one, manager 1 could be more skilled than manager 2 on every dimension and yet deliver the same gross alpha if he operates a larger fund.

We address this issue in two ways. To begin, we estimate the two skill dimensions separately. We measure the first dimension using the first-dollar (fd) alpha a proposed by Pastor, Stambaugh, and Taylor (2015). This measure allows us to identify managerial skill free of any capacity constraints by setting the fund size equal to zero.[2] It also provides a ranking of managers that remains unchanged regardless of whether they optimally choose the size of their funds. We measure the second dimension using the size coefficient b which determines how the gross alpha decreases with fund size ($b < 0$). The magnitude of b is larger if managers face higher execution costs for trading large orders (liquidity, price impact) and if managers are unable to generate investment ideas as the fund grows large (Berk and Green (2004), Perold and Salomon (1991)).

Next, we measure the value added va of Berk and van Binsbergen (2015; BvB hereafter) which is defined as the product between the gross alpha and the fund size. This measure has a powerful economic interpretation because it determines the total value (profits) created by managers from exploiting the two skill dimensions—as shown below, va is a function of both a and b. The value added differs from the gross alpha because it incorporates information about size—if manager 1 has the same gross alpha as manager

[1] A non-exhaustive list of studies that use gross alpha to measure skill includes Baks, Metrick, and Wachter (2001), Barras, Scaillet, and Wermers (2010), Carhart (1997), Grinblatt and Titman (1989b), Jensen (1969), Jones and Shanken (2005), Kosowski et al. (2006), Wermers (2000).

[2] Perold and Salomon (1991) refers to this measure as the paper return because it determines the fund return that is "unencumbered by the drag imposed by real world implementation".

2 but operates a larger fund, he creates more value. In other words, using the gross alpha is akin to measuring the rent of a monopolist with the markup price of the goods regardless of how many quantities are sold.

A.2 The Relations between the Skill Measures

To shed light on the relations between the skill measures, we need to understand how fund size is determined in equilibrium. A natural benchmark is the neoclassical model of Berk and Green (2004) and BvB in which skilled managers maximize profits and investors are rational and compete for performance so that the gross alpha equals fees, i.e., $\alpha = f_e$. Following the notation of BvB, we write the (benchmark-adjusted) total revenue r_i and total cost c_i of fund i from active management as $a_i q_i$ and $b_i q_i^2$, where a_i is the fd alpha, b_i is the size coefficient, and q_i is the fund size. Using the standard first order condition, the optimal size is $q_i^* = \frac{a_i}{2b_i}$ and the value added (profits) is given by $va_i = \alpha_i q_i^* = r_i^* - c_i^* = a_i q_i^* - b_i q_i^{*2} = \frac{a_i^2}{4b_i}$.

A key insight from this model is that the fees chosen by managers do not change the value added: (i) if managers choose low fees, they receive additional money from investors ($q_i - q_i^* > 0$) which is passively indexed to keep $r^* - c^*$ unchanged; (ii) if managers choose high fees, they can sell the index short and invest the proceeds in the fund ($q_i^* - q_i > 0$) to reach the optimal profits $r_i^* - c_i^*$. However, the choice of fees determine the fund q_i and thus the relations between the skill measures. To see this point, we examine four hypothetical compensation schemes in which managers set fees based on specific rules.[3]

Scheme I (optimal size). Managers set fees at $f_{e,i}^*$ such that the fund operates at the profit-mazimizing size q_i^*. The gross alpha is given by $\alpha_i = f_{e,i}^* = \frac{r_i^* - c_i^*}{q_i^*} = \frac{a_i q_i^* - b_i q_i^{*2}}{q_i^*} = \frac{a_i}{2} \div a_i$ and is proportional to the fd alpha. Therefore the gross alpha correctly measures the first skill dimension: the ability to detect profitable trading opportunities.

Scheme II (squared optimal size). Managers set fees at $f_{e,i}^b$ such that the fund operates at the squareed optimal size $q_i^b = q_i^{*2}$. We have $\alpha_i = f_{e,i}^b = \frac{r_i^* - c_i^*}{q_i^b} = \frac{a_1 q_1^* - b_1 q_1^{*2}}{q_i^b} = b_i$ which is proportional to the size coefficient. Under this compensation scheme, the gross alpha measures the second skill dimension: the ability to lower the impact of capacity constraints.

Scheme III (same size). Managers set fees at $f_{e,i}^q$ such that all funds have the same size q (equal to the median). The gross alpha becomes $\alpha_i = f_{e,i}^q = \frac{r_i^* - c_i^*}{q} = \frac{a_1 q_1^* - b_1 q_1^{*2}}{q} =$

[3] Our analysis largely builds on BvB who already discuss the relation between the gross alpha and the value added. Our contribution is to (i) include both the fd alpha and the size coefficient; (ii) systematically explain the relations between the skill measures under four different compensation schemes.

$\frac{a_i^2}{4qb_i} \div va_i$ and is proportional to the value added. Therefore, the gross alpha correctly measures the total profits earned by the manager.

Scheme IV (arbitrary size). Here, managers can choose fees $f_{e,i}^a$ arbitrarily at levels that potentially differ from $f_{e,i}^*$, $f_{e,i}^b$, and $f_{e,i}^q$. For instance, less skilled managers may charge high fees ($f_{e,i}^a > f_{e,i}^*$), while the opposite holds for more skilled managers ($f_{e,i}^a < f_{e,i}^*$).[4] In this case, the gross alpha becomes $\alpha_i = f_{e,i}^a = \frac{r_i^* - c_i^*}{q_i^a} = \frac{a_1 q_1^* - b_1 q_1^{*2}}{q_i^a} = \frac{a_i^2}{4q_i^a b_i}$ with $q_i^a(f_{e,i}^a) = \frac{a_i^2}{4f_{e,i}^a b_i}$. Therefore, the gross alpha is unrelated to the two dimensions of skill ($\alpha_i \neq a_i$, $\alpha_i \neq b_i$) and to the value added and ($\alpha_i \neq va_i$).

Our example in the previous section is a special case of Scheme IV where the less skilled manager 2 keeps fees unchanged ($f_{e,2} = f_{e,2}^*$) while the more skilled manager 1 lowers them to the same level ($f_{e,1} = f_{e,2}^*$). Figure 1 reveals that manager 2 is less skilled on every dimension ($a_2 < a_1$ and $b_2 > b_1$) and thus produces a lower value added ($va_2 < va_1$). Yet, a comparison based on the gross alpha fails to capture any skill difference, i.e., $\alpha_1 = \alpha_2 = f_{e,2}^*$.

Please insert Figure 1 here

Table I summarizes our analysis and reveals that the four compensation schemes yield different predictions for fund size and fees. Under Scheme I, we observe a moderate cross-fund variation in both fees and size as managers choose different fees to reach the fund-specific optimal size q_i^*. Under Scheme II, the chosen fees $f_{e,i}^b$ are tiny because they reflect the small magnitude of the size coefficient b_i—these low fees are required for receiving sufficient flows to reach the squared value of the optimal size q_i^{*2}. Under Scheme III, the fund size must be the same for all funds ($q_i = q$). Under Scheme IV, the cross-fund variation in size is large because this variable corrects for the arbitrary fee setting policy to reach the equilibrium (i.e., $f_{e,i}^q$ are potentially negatively correlated with skill). In the empirical analysis, we can build on these predictions to determine which compensation schemes are more realistic.

Please insert Table I here

[4] One possible justification for this fee structure is moral hazard. Habib and Johnson (2016) show that in a world with moral hazard $f_{e,i}^a$ are the minimum fees that guarantee that skilled managers exert effort. Therefore, $f_{e,i}^a$ captures the manager-specific probability of shirking and may be quite different from $f_{e,i}^*$, $f_{e,i}^b$, and $f_{e,i}^q$.

A.3 Formal Definition of the Skill Measures

To estimate the gross alpha, we run the following time-series regression for each fund i,

$$r_{i,t} = \alpha_i + \beta_i' f_t + \varepsilon_{i,t}, \tag{1}$$

where $r_{i,t}$ is the fund gross excess return (before fees) over the riskfree rate between time $t-1$ and t, f_t is a K-vector of benchmark excess returns, and $\varepsilon_{i,t}$ is the residual term. To capture the heterogeneity across funds, we interpret Equation (1) as a random coefficient model in which all the regression coefficients are random (e.g., Hsiao (2003))—in particular, the fund gross alpha is drawn from a common density function $\phi(\alpha)$. We do not treat the regression coefficients as fixed parameters but as random realizations from a continuum of funds.[5] Under this sampling scheme, we can invoke cross-sectional limits in order to conduct inference on the skill density $\phi(\alpha)$.

To estimate the two skill dimension of skill—the fd alpha and the size coefficient—we model the fund gross alpha as a linear function of its lagged size: $\alpha_{i,t} = a_i + b_i q_{i,t-1}$, where $q_{i,t-1}$ is the lagged fund size at time $t-1$.[6] The literature commonly assumes a common b for all funds (e.g., Chen et al. (2004), Pastor, Stambaugh, and Taylor (2015), Pollet and Wilson (2008)). Here, we follow Harvey and Liu (2017) and allow the size coefficient b_i to be fund-specific. We allow for heterogeneity across funds because (i) it is potentially a key driver of the empirical differences between skill measures, and (ii) there is a priori no reason why capacity constraints should be identical for all funds. Replacing α_i with $\alpha_{i,t}$ in Equation (1), we can estimate a_i and b_i for each fund i from the following time-series regression:

$$r_{i,t} = \alpha_{i,t} + b_i' f_t + \varepsilon_{i,t} = a_i + b_i q_{i,t-1} + \beta_i' f_t + \varepsilon_{i,t}, \tag{2}$$

Finally, we measure the value added of each fund as

$$va_i = \alpha_i \cdot \bar{q}_i, \tag{3}$$

where α_i is given by Equation (1). Similar to our analysis of the gross alphas, the fund fd alpha, size coefficient, and value added are drawn from the common density functions $\phi(a)$, $\phi(b)$, and $\phi(va)$ which can be estimated using our approach presented below.

[5] See Gagliardini, Ossola, and Scaillet (2016) for an arbitrage pricing theory consistent with such a framework.

[6] Note from Figure 1 that the relation between $\alpha_{i,t}$ and $q_{i,t-1}$ is not perfectly linear if the fund passively grows beyond the optimal size q_i^*. In the appendix, we account for potential non-linearities and find similar results.

B Overview of the Methodology

B.1 General Motivation

We now present our novel approach for estimating the cross-sectional distribution of skill among mutual funds. We denote the associated density (pdf) by $\phi(m)$ where the general formulation $m = \{\alpha, a, b, va\}$ encompasses all four measures of skill. The distinguishing feature of our approach is that it is non-parametric, i.e., it estimates $\phi(m)$ without imposing any parametric assumptions on its shape. This departs from the Bayesian/parametric approaches used in previous work which require a pre-specified distribution for $\phi(m)$—typically, a mixture of normal distributions (e.g., Harvey and Liu (2018), Jones and Shanken (2005))

Our approach provides several advantages. First, its non-parametric nature makes it immune to misspecification errors. In contrast, a Bayesian/parametric approach in which the distributions of α, a, b, and va are specified separately is likely to generate distribution inconsistencies because the skill measures are theoretically tightly related—that is, we have $\alpha_i = a_i + b_i\bar{q}_i$ and $\alpha_i = va_i/\bar{q}_i$. Whereas this issue could be solved by directly modelling their determinants (i.e., a_i, b_i, and $\bar{q}_i$), one faces the daunting task of specifying and estimating a joint trivariate distribution whose marginals are potentially mixtures of distributions. Second, the implementation of our approach is simple and fast. Intuitively, the procedure is akin to computing an histogram using as only inputs the estimated skill measure $\hat{m}_i$ of each fund. This contrasts with the complex and lengthy Gibbs sampling and maximum likelihood methods used in Bayesian/parametric approaches. Third, our approach provides a unified framework to the different characterizations of the skill distribution. In addition to $\phi(m)$, we can use our approach to estimate (i) the distribution moments such as the mean, variance, skewness, and kurtosis, (ii) the cumulative distribution function (cdf) $\Phi_m(x) = prob[m_i < x] = \int_{-\infty}^{x} m\phi(m)d_m$, (iii) the distribution quantile $Q(p) = \Phi_m^{-1}(p)$, where p denotes the chosen probability level (e.g, $p = 0.5$ for the median).

In the next section, we present our non-parametric approach in more detail. Our analysis reveals that using the estimated skill measures instead of the *true* ones ($\hat{m}_i$ instead of m_i) introduces an Error-in-Variable (EIV) bias that must be accounted for. This result is reminiscent of the well-known EIV bias for the two-pass regression (Shanken (1992)) and implies that the unadjusted estimators can be severely biased as shown below. For sake of brevity, we explain the procedure for estimating the skill density ϕ and relegate to the appendix the formal treatment of the three remaining measures (moment, cdf, quantiles) and the proofs of the different propositions.

B.2 Non-Parametric Density Estimation

To estimate the density $\phi(m)$, we begin by estimating the regression coefficients in Equation (2) for each of the n funds in the population. The $(K+2)$-vector of coefficients $\hat{\gamma}_i = \left(\hat{\alpha}_i^g, \hat{b}_i, \hat{\beta}_i'\right)'$ for fund i $(i = 1, ...n)$ is computed as

$$\hat{\gamma}_i = \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_{t=1}^{T} I_{i,t} x_{i,t} r_{i,t}, \tag{4}$$

where $I_{i,t}$ is an indicator variable equal to one if $r_{i,t}$ is observable (and zero otherwise), $T_i = \sum_{t=1}^{T} I_{i,t}$ is the total number of return observations, $x_{i,t} = [1, q_{i,t-1}, r_{b,t}']'$, is the $(K+2)$-vector of explanatory variables, $q_{i,t-1}$ is the lagged fund size, and $\hat{Q}_{x,i} = \frac{1}{T_i}\sum_{t=1}^{T} I_{i,t} x_{i,t} x_{i,t}'$ is the estimated matrix of the second moments of $x_{i,t}$. For the gross alpha and the value added, we only need to estimate the coefficients in Equation (1) in which case $\hat{\gamma}_i$ and $x_{i,t}$ are $(K+1)$-vectors defined as $\hat{\gamma}_i = \left(\hat{\alpha}_i, \hat{\beta}_i'\right)'$ and $x_t = [1, r_{b,t}']'$.

Given the unbalanced nature of the mutual fund panel, T_i can be very small for some funds. As a result, the inversion of the matrix $\hat{Q}_{x,i}$ can be numerically unstable and yield unreliable estimates of m_i. To address this issue, we follow Gagliardini, Ossola, and Scaillet (2016) and introduce a formal fund selection rule $\mathbf{1}_i^{\chi}$ equal to one if the following two conditions are met (and zero otherwise):

$$\mathbf{1}_i^{\chi} = \mathbf{1}\left\{CN\left(\hat{Q}_{x,i}\right) \leq \chi_{1,T}, \tau_{i,T} \leq \chi_{2,T}\right\}, \tag{5}$$

where $CN\left(\hat{Q}_{x,i}\right) = \sqrt{eig_{max}\left(\hat{Q}_{x,i}\right)/eig_{\min}\left(\hat{Q}_{x,i}\right)}$ is the condition number of $\hat{Q}_{x,i}$ defined as the ratio of the highest to lowest eigenvalues eig_{max} and eig_{min}, $\tau_{i,T} = T/T_i$ is the relative sample size, and $\chi_{1,T}$, $\chi_{2,T}$ denote the two threshold values. The first condition $\{CN\left(\hat{Q}_{x,i}\right) \leq \chi_{1,T}\}$ excludes funds for which the time series regression is poorly conditioned, i.e., a large value of $CN\left(\hat{Q}_{x,i}\right)$ indicates multicollinearity problems (Belsley, Kuh, and Welsch (2004), Greene (2008)). The second condition $\{\tau_{i,T} \leq \chi_{2,T}\}$ excludes funds for which the sample size is too small. Both thresholds $\chi_{1,T}$ and $\chi_{2,T}$ increase with the sample size T—with more return observations, the fund coefficients are estimated with greater accuracy which allows for a less stringent selection rule.

Next, we estimate the skill density function using a non-parametric approach (e.g., Silverman (1986)). The estimated density $\hat{\phi}$ at a given point m is computed as

$$\hat{\phi}(m) = \frac{1}{nh} \sum_{i=1}^{n} \mathbf{1}_i^{\chi} K\left(\frac{\hat{m}_i - m}{h}\right), \tag{6}$$

where $\hat{m}_i$ takes one of the following expressions,

$$\begin{aligned} \text{Gross alpha} &: \quad \hat{m}_i = \hat{\alpha}_i, \\ \text{Fd alpha} &: \quad \hat{m}_i = \hat{a}_i, \\ \text{Size coefficient} &: \quad \hat{m}_i = \hat{b}_i, \qquad (7) \\ \text{Value added} &: \quad \hat{m}_i = \hat{\alpha}_i \overline{q}_i, \qquad (8) \end{aligned}$$

and h is the vanishing smoothing bandwidth—similar to the length of histogram bars, the bandwidth h determines how many observations around point m we use for estimation. The function K is a symmetric kernel function that integrates to one. Because the choice of K is not a crucial aspect of nonparametric analysis, we use the standard Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})$ for simplicity (see Silverman (1986)).

The following proposition examines the asymptotic properties of $\hat{\phi}(m)$ as the size of the fund population n and the number of return observations T grow large for a vanishing bandwidth h.

Proposition II.1 *As $n, T \to \infty$ and $h \to 0$ such that $nh \to \infty$ and $\sqrt{nh}(h^3 + h^2 T + (1/T)^{\frac{3}{2}}) \to 0$, we have*

$$\sqrt{nh}\left(\hat{\phi}(m) - \phi(m) - bs(m)\right) \Rightarrow N\left(0, K_1 \phi(m)\right), \tag{9}$$

and the bias term $bs(m)$ is the sum of two components,

$$bs_1(m) = \frac{1}{2} h^2 K_2 \phi^{(2)}(m), \tag{10}$$

$$bs_2(m) = \frac{1}{2T} \psi^{(2)}(m), \tag{11}$$

where $K_1 = \int K(u)^2 du$, $K_2 = \int u^2 K(u) du$ (with a Gaussian kernel, $K_1 = \frac{1}{2\sqrt{\pi}}$ and $K_2 = 1$), $\phi^{(2)}(m)$ is the second derivative of the density $\phi(m)$ and $\psi^{(2)}(m)$ is the second derivative of the function $\psi(m) = \omega(m)\phi(m)$ with $\omega(m) = E(S_i | m_i = m)$. The term S_i is the asympotic variance of the estimated skill measure $\sqrt{T}\hat{m}_i$ equal to

$plim_{T\to\infty}(\frac{1}{T}\sum_{t,s} I_{i,t}I_{i,s}u_{i,t}u_{i,s})$. For each skill measure, the term $u_{i,t}$ is given by

$$\begin{aligned}
\textit{Gross alpha} &: \quad u_{i,t} = e_1' Q_x^{-1} x_{i,t}\varepsilon_{i,t}, \\
\textit{Fd alpha} &: \quad u_{i,t} = e_1' Q_x^{-1} x_{i,t}\varepsilon_{i,t}, \\
\textit{Size coefficient} &: \quad u_{i,t} = e_2' Q_x^{-1} x_{i,t}\varepsilon_{i,t}, \\
\textit{Value added} &: \quad u_{i,t} = \alpha_i q_{i,t-1} + \bar{q}_i e_1' Q_x^{-1} x_{i,t}\varepsilon_{i,t},
\end{aligned} \tag{12}$$

where e_1 (e_2) is a vector with one in the first (second) position and zeros elsewhere and $Q_x = E[x_t x_t']$.

Proof. See the appendix. ■

Proposition I.1 yields several important insights. First, it shows that the estimated density function $\hat{\phi}(m)$ is asymptotically normally distributed which facilitates the construction of confidence intervals. As shown in Equation (9), the width of this interval depends on the variance term $K_1\phi(m)$ which is higher in the peak of the density.

Second, $\hat{\phi}(m)$ is a biased estimator of $\phi(m)$—we can therefore improve the density estimation by adjusting for the bias term $bs(m)$. Equations (10)-(11) reveal that $bs(m)$ has two distinct components. The first component bs_1 is the standard smoothing bias in non-parametric density estimation (e.g., Silverman (1986), Wand and Jones (1995)). The second component bs_2, which is referred to as the EIV bias, is non-standard—it arises because we estimate ϕ using the estimated skill measure instead of the *true* one (i.e., $\hat{m}_i$ instead of m_i).

Finally, Proposition I.1 provides guidelines for the choice of the bandwidth. We show in the appendix that the choice of the optimal bandwidth—the one that minimizes the Asymptotic Mean Integrated Squared Error (AMISE) of $\hat{\phi}(m)$—depends on the relationship between T and n: (i) if T is small relative to n, h^* is proportional to $(nT)^{-\frac{1}{3}}$; (ii) if T is large relative to n, h^* is proportional to $n^{-\frac{1}{5}}$.[7] Our Monte-Carlo analysis reveals that given our actual sample size, the two bandwidth choices produce similar results with a slight advantage to the first case. Motivated by these results, we use the following bandwidth in our baseline specification:

$$h^* = \left(\frac{K_2}{K_1} e_1' Q_x^{-1} K_3 Q_x^{-1} e_1\right)^{-\frac{1}{3}} (nT)^{-\frac{1}{3}}, \tag{13}$$

where $K_3 = \int \phi^{(2)}(m)\psi^{(2)}(m)dm$.

[7] The AMISE is defined as the integrated sum of the leading terms of the asymptotic variance and squared bias of the estimated density $\hat{\phi}(m)$ (see Appendix A.3 for an analysis of its behaviour and a discussion of the optimal bandwidth).

B.3 The Bias-Adjusted Density Estimator

Building on the insights of Proposition I.1, we show how to compute the bias-adjusted estimator of the skill density $\phi(m)$. Our approach consists of estimating the two bias terms $bs_1(m)$, $bs_1(m)$ and the optimal bandwidth h^* using a Gaussian reference model in which the fund skill measure m_i and the log of the asymptotic variance $s_i = \log(S_i)$ are drawn from a bivariate normal distribution, i.e., $m_i \sim N(\mu_m, \sigma_m)$, $s_i \sim N(\mu_s, \sigma_s)$, and $corr(m_i, s_i) = \rho$.[8]

Using a simple reference model provides simplicity and estimation accuracy. The estimators of the bias and the bandwidth are simple to compute because they are all available in closed-form. They are also precisely estimated because they only rely on a few parameters. These benefits are not shared by the alternative approach in which the bias terms are directly inferred from Equations (10)-(11) via a non-parametric estimation of the second derivatives $\phi^{(2)}$ and $\psi^{(2)}$. Estimating these derivative terms is notoriously difficult and generally leads to large estimation errors (e.g., Wand and Jones (1995; Section 2.12)).[9] Finally, the reference model allows us to perform a comparative static analysis on the two bias components. In other words, we can examine how the different model parameters affect the shape of the bias.

The following proposition provides the closed-form expressions for the two bias components and the optimal bandwidth as the size of the fund population n and the number of return observations T grow large for a vanishing bandwidth h.

Proposition II.2 *As $n, T \to \infty$ and $h \to 0$ such that $nh \to \infty$ and $\sqrt{nh}(h^3 + h^2 T + (1/T)^{\frac{3}{2}}) \to 0$, the two bias components under the reference model are equal to*

$$bs_1^r(m) = \left[\frac{1}{2} K_2 h^2 \frac{1}{\sigma_m^2}(\bar{m}_1^2 - 1)\right] \frac{1}{\sigma_m}\varphi(\bar{m}_1), \tag{14}$$

$$bs_2^r(m) = \left[\frac{1}{2T} \exp\left(\mu_s + \frac{1}{2}\sigma_s^2\right) \frac{1}{\sigma_m^2}(\bar{m}_2^2 - 1)\right] \frac{1}{\sigma_m}\varphi(\bar{m}_2), \tag{15}$$

where $\bar{m}_1 = \frac{m-\mu_m}{\sigma_m}$, $\bar{m}_2 = \frac{m-\mu_m-\rho\sigma_m\sigma_s}{\sigma_m}$, $\varphi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$ is the density of the

[8] A Gaussian reference model underlies the celebrated Silverman rule of thumb for the choice of the bandwidth in standard non-parametric density estimation without EIV. This rule gives $h^* = 1.06\sigma n^{-\frac{1}{5}}$, where σ is the standard deviation of the observations (Silverman (1986)).

[9] We can estimate the rth-derivative of a density ϕ by kernel smoothing (Bhattacharya (1967)). Unfortunately, the rate of consistency of the derivative estimator equals $(nh^{2r+1})^{-\frac{1}{2}}$ and is much slower than the rate $(nh)^{-\frac{1}{2}}$ for the density estimator. In other words, the higher-order derivatives are imprecisely estimated because the rate of consistency decreases with the derivative order r.

standard normal distribution. In addition, the optimal bandwidth h^* is given by

$$h^* = \left[\frac{K_2}{K_1 2\sqrt{\pi}}\frac{3}{4\sigma_m^5}\left(\frac{\rho^4\sigma_s^4}{12}+\rho^2\sigma_s^2+1\right)\exp\left(\mu_s+\frac{1}{2}\sigma_s^2(1-\frac{\rho^2}{2})\right)\right]^{-\frac{1}{3}}(nT)^{-\frac{1}{3}}. \quad (16)$$

Proof. See the appendix ■

Building on Proposition I.2, we compute the bias-adjusted density $\hat{\phi}^*(m)$ using the following steps. First, we estimate the moments of the bivariate normal distribution in the reference model using the estimated quantities $\hat{m}_i$ and $\hat{s}_i$ $(i = 1, ..., n)$. To compute $\hat{s}_i = \log(\hat{S}_i)$, we use the standard variance estimator of Newey and West (1987):

$$\hat{S}_i = \frac{1}{T}\sum_{t=1}^{T} I_{i,t}\hat{u}_{i,t}^2 + 2\sum_{l_1=1}^{L}\left[\frac{1}{T}\sum_{t=1}^{T-l_1} I_{i,t}I_{i,t+l_1}\hat{u}_{i,t}\hat{u}_{i,t+l_1}\right], \quad (17)$$

where $\hat{u}_{i,t}$ is obtained by plugging the estimated quantities in Equation (12) for the chosen skill measure and L is the number of lags chosen to account for potential serial correlation. Second, we insert these estimated moments in Equations (13)-(15) to compute the bias terms $\widehat{bs}_1(m)$, $\widehat{bs}_2(m)$ and the optimal bandwidth h^*. Third, we remove the bias terms from the unadjusted density in Equation (7) to obtain the bias-adjusted density estimator $\hat{\phi}^*(m)$.

While our approach allows for a simple bias adjustment, an important question is whether the estimated bias terms are sufficiently accurate when the data do not follow our simple reference model. To address this issue, we perform an extensive Monte-Carlo analysis that replicates the salient features of the data. The results summarized in the appendix reveal that the bias-adjusted estimator captures the true distribution $\phi(m)$ with remarkable accuracy.[10]

C Analysis of the Density Bias

C.1 Magnitude of the Bias Components

To gain further insight into the bias adjustment mechanism, we now study the bias associated with the gross alpha density $\phi(\alpha)$. To begin, we compute the parameters of the reference model from the estimated monthly gross alphas and asympotic variances of all funds in our sample ($n = 2,772$). The mean μ_m and volatility σ_m of the *true* gross alpha are set equal to 0.06% and 0.18% per month. The mean μ_s and volatility

[10] This result is consistent with the previous analysis by Silverman (1986). He shows that the rule of thumb for the bandwidth choice which relies on a gaussian reference model is quite robust to departures from normality.

σ_s for the log of the asymptotic variance are equal to -7.3 and 1.1—these numbers yield typical values for the volatility of the estimated gross alpha between 0.11% and 0.34% per month.[11] For the parameter ρ, we obtain a value of 0.04 which implies that the gross alpha and the asympotic variance are essentially uncorrelated.[12] Inserting these parameters values in Equations (13)-(14), we can then estimate the two bias components $bs_1^r(\alpha)$ and $bs_2^r(\alpha)$.

In Figure 2, we plot the two bias components for different values of α ranging between -0.75% and 0.75% per month. To ease interpretation, we measure the bias relative to the average density value averaged across the grid of points. Panel A reveals that the smoothing bias is negligible given the large number of funds in the population. On the contrary, Panel B reveals that the EIV bias is large—it ranges between -129% and 58% of the average density. Adjusting for the EIV bias is therefore critical for improving the accuracy of the non-parametric density estimator.

The EIV bias is negative in the center of the distribution and positive in the tails. The intuition for this result is that the estimated alpha is a noisy version of the *true* alpha ($\hat{\alpha}_i = \alpha_i +$ estimation noise). Therefore, the density obtained with the estimated alphas overestimates the probability of observing extreme alphas (i.e., the unadjusted density estimate is too flat). Importantly, the shape of the EIV bias in Panel B is quite general and is not simply an artefact of our reference model. As shown in Proposition I.1, the *true* bias $bs_2(\alpha)$ is a function of the second-order derivative of the *true* density $\phi^{(2)}$. As long as this density peaks around its mean, $\phi^{(2)}$ takes negative (positive) values in the center (tails) of the distribution which implies that $bs_2(\alpha)$ follows the same shape as $bs_2^r(\alpha)$ obtained with the reference model.[13]

Please insert Figure 2 here

C.2 Comparative Statics

We can use the reference model to perform a comparative static analysis on the EIV bias. Using the closed-form expressions in Equation (14), we can examine the shape of

[11] A one-standard deviation change around the mean μ_s yields bounds for s_i equal to -8.4 and -6.2. To compute the equivalent bounds for the monthly volatility of the estimated alpha, we use $\sigma_\alpha = \frac{1}{\sqrt{T}}\sqrt{(\exp(s_i))}$, where T is set equal to the median number of observations (173).

[12] This finding resonates with the analysis of Jones and Shanken (2005) who note that it is not clear whether skilled managers can diversify their bets across a wide range of stocks or instead focus on particular industries/sectors. Given this uncertainty, they assume independent priors for α_i and s_i.

[13] The only case where $bs_2(\alpha)$ differs from $b^r s_2(\alpha)$ is if the *true* density $\phi(m)$ is a mixture of distributions with means extremely far away from one another such that we have a trough instead of a peak at the mean of $\phi(m)$.

$bs_2^r(\alpha)$ for different parameter values. In Panel A of Figure 3, we examine how the bias reponds to changes in the volatility σ_m of the *true* gross alpha. A higher value for σ_m makes the cross-sectional variation in the estimated alphas more aligned with that of the *true* alphas (i.e., the relative importance of α_i over the estimation noise increases). Therefore, the bias becomes less pronounced—if σ_m increases by 30% in relative terms, $bs_2^r(\alpha)$ only ranges between -58% and 26% of the average density (versus -129% and 58% in the baseline case).

In Panel B, we focus on the mean μ_s of the variance of the estimated gross alpha. Contrary to Panel A, a higher value for μ_s makes the cross-section of the estimated fund alphas more noisy (i.e., the relative importance of the estimation noise over α_i increases) which in turn amplifies the magnitude of the bias. For instance, if μ_s increases by 30%, the minimum and maximum values for $bs_2^r(\alpha)$ jump to -174% and 78% (we document the same results when increasing the volatility σ_s of the variance term).

Finally, we change the correlation ρ between the *true* alpha and the estimation uncertainty. A higher correlation implies that funds with higher alphas also exhibit higher estimation variance—this is the case if skilled managers take concentrated bets which increases the idiosyncratic variance of their funds. Panel C reveals that a rise in ρ shifts $bs_2^r(\alpha)$ shifts to the right. While the bias exhibits the same shape as in the baseline case, its mininum value is attained when the gross alpha is above the cross-sectional mean ($\alpha_i > \mu_m$).

Please insert Figure 3 here

D Extensions

D.1 Net Alpha

We conclude our presentation with several useful extensions. We can evaluate mutual fund performance using the fund net alpha α_i^n—that is, the abnormal average return earned by investors net of fees. Similar to the gross alpha, our non-parametric approach provides a robust, simple, and fast procedure to estimate the cross-sectional density $\phi(\alpha^n)$ of the net alpha. The estimation procedure remains exactly the same as in Section II except that we replace Equation (1) with

$$r_{i,t}^n = r_{i,t} - fe_{i,t} = \alpha_i^n + \beta_i' f_t + \varepsilon_{i,t}, \tag{18}$$

where $r_{i,t}^n$ is the fund net excess return defined as the difference between the fund gross excess return and the fees paid by investors.

D.2 Time-Varying Skill

We can also use our approach to examine the time-variation in fund returns at the individual fund level. This analysis can be used to extend recent studies which show the gross alpha at the aggregate level varies with size, turnover, and business cycle conditions (e.g., Chen et al. (2004), Kacperczyk, van Nieuwerburgh, and Veldkamp (2014), Pastor, Stambaugh, and Taylor (2015; 2017)). To estimate the density $\phi(b_l)$ associated with each alpha predictor l ($l = 1, ..., L$), we generalize of Equation (2):

$$\alpha_{i,t-1} = a_i + \sum_{l=1}^{L} b_{i,l} z_{l,i,t-1}, \tag{19}$$

where $z_{l,i,t-1}$ is predictor l and $b_{i,l}$ denotes its impact on the fund gross alpha. Then, we replace $\hat{m}_i$ with the estimated coefficient $\hat{b}_{i,l}$ in Equation (7) and modify $u_{i,t}$ accordingly to obtain the bias-adjusted density $\hat{\phi}^*(b_l)$.

D.3 Factor Beta

Finally, we can apply our approach to study the distribution $\phi(\beta_k)$ of the fund betas for each risk factor k ($k = 1, ..., K$). The estimation procedure remains essentially unchanged after replacing $\hat{m}_i$ with the estimated beta and modifying $u_{i,t}$ accordingly.

III Data Description

A Mutual Fund Data

We conduct our empirical analysis on the entire population of open-end US equity funds. We collect monthly data on net returns and size as well as annual data on fund fees and investment objectives from the CRSP Survivorship Bias Free Mutual Fund Database. We measure the monthly gross return of each fund as the sum of its monthly net return and fees. The net return is computed as a value-weighted average of the net returns across all shareclasses using their beginning-of-month total net asset values. Similarly, the monthly fees are defined as the value-weighted average of the most recently reported annual fees across shareclasses divided by 12.

To measure the size of the fund, we take the sum of the beginning-of-month net asset values across all shareclasses. Following BvB, we adjust the size for inflation by expressing all numbers in January 1, 2000 dollars. We also linearly interpolate any missing values between two observations. For the investment objectives, we rely on the classification provided by Lipper. If this information is missing, we use the objectives reported by Strategic Insight, Wiesenberger, and CRSP which are particularly useful for the earlier part of the sample (see the appendix for additional details).

To apply the fund selection rules in Equation (5), we follow Gagliardini, Ossola, and Scaillet (2016) and select funds for which (i) the condition number of the matrix $\hat{Q}_{x,i}$ is below 15 and (ii) the number of monthly return observations is at least equal to 60.[14] To avoid focusing on tiny funds, we eliminate return observations with lagged fund size below \$15 million (see Chen et al. (2004), Pastor, Stambaugh, and Taylor (2015)). We also eliminate fund observations if the fees are below 0.25% or above 10% per year. These selection criteria produce a final universe of 2,772 funds, as well as (i) 671 small- and 999 large-cap funds; (ii) 1,336 growth and 842 value funds; (iii) 1,133 low- and 1,052 high-expense funds.

B Benchmark Models

To estimate the different skill measures based on Equations (1)-(2), we use the four-factor models of Carhart (1997) and Cremers, Petajisto, and Zitzewitz (2012; CPZ hereafter). In both models, the vector of risk factors is defined as $f_t = [r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}]'$, where $r_{m,t}$, $r_{smb,t}$, $r_{hml,t}$, and $r_{mom,t}$ capture the excess returns of the market, size, value, and momentum factors.

In the Carhart model (1997), $r_{m,t}$ is proxied by the excess return of the CRSP index and $r_{smb,t}$, $r_{hml,t}$ and $r_{mom,t}$ are the returns of the Fama-French zero-investment portfolios sorted on size, book-to-market, and past annual returns. The CPZ model departs from the Carhart model in two respects: (i) $r_{m,t}$ is proxied by the excess return of the SP500 index, (ii) the size and value factors are index-based, i.e., $r_{smb,t}$ is measured as the return difference between the Russell 2000 and the SP500, while $r_{hml,t}$ is measured as the return difference between the Russell 3000 Value and the Russell 3000 Growth.

It is common for US equity managers to use as benchmarks the SP500 and the Russell 2000 which both cover about 85% of the total market capitalization. As noted by CPZ, the Carhart model fails to price these passive, well-diversified indices—for one, the Russell 2000 produces a Carhart-alpha of -2.41% per year over the period 1980-2005. This implies that small-cap managers would automatically be classified as unskilled if they use the Russell 2000 as benchmark. Motivated by these results, we use the CPZ model in our baseline specification and later examine how the results change when using the Carhart model.

[14] The monthly returns need not be contiguous. We delete the observation following any missing returns because CRSP reports the cumulated return since the last reported observation.

C Summary Statistics

Table II reports summary statistics for value-weighted portfolios of funds for the entire population and each group (small-cap, large-cap, growth, value, low-expense, high-expense). Our sample period starts in January 1979 and ends in December 2015 (for a total of 444 monthly return observations).[15] In Panel A, we report the first four moments of the portfolio gross excess returns. In the entire population, the portfolio achieves a risk-return tradeoff similar to that of the aggregate stock market with a mean and volatility equal to 7.69% and 15.06% per year. It also exhibit a negative skewness (-0.8) and a positive kurtosis (5.4). The results are similar across groups except for small-cap funds which produce higher mean and volatility.

In Panel B, we repeat the analysis for the estimated betas on the four factors in the CPZ model. Consistent with intuition, small-cap funds are heavily exposed to the size factor with a median beta of 0.82. We also find that growth funds are negatively exposed to the value factor (-0.27), whereas the opposite holds for value funds (0.19). Finally, high-expense funds heavily tilt towards small-cap, growth stocks (the size and value betas are equal to 0.42 and -0.27).

Please insert Table II here

IV Empirical Results

A The Two Dimensions of Skill

A.1 Skill at Detecting Trading Opportunities

We now examine the cross-sectional distributions of the two dimensions of skill. We begin our analysis with the fd alpha a which measures the ability of each manager to generate profitable trading ideas. To summarize the information contained in $\phi(\alpha)$, we report in Table III (i) the first four moments (mean, variance, skewness, kurtosis), (ii) the proportions of funds with negative and positive alphas denoted by π_{α}^{-} and π_{α}^{+}, and (iii) the distribution quantiles at 10% and 90%. As discussed in Section II.B, our estimation procedure adjusts for the EIV bias that arises because we use the estimated alphas instead of the *true* ones (see the appendix for the derivations). In other words, the estimated moments, proportions, and quantiles reflect the shape of the bias-adjusted density $\hat{\phi}^*$ which is markedly narrower than the unadjusted density $\hat{\phi}$ obtained with the estimated fd alphas (similar to Figure 2).

[15] The starting date correspond to the first month when the factors in the CPZ model are available.

To begin, we examine the proportions π_a^- and π_a^+ to determine whether managers are skilled at detecting trading opportunities. We find that the vast majority of managers have superior information. The proportion π_α^+ is equal to 88.1% in the population and ranges from 80.6% to 93.4% among the different groups.[16] The existence of 11.9% of funds with negative fd alphas is a priori surprising because managers always have the option to passively manage their funds. An natural explanation is that these managers are unskilled and yet decide to actively trade in order to hide their lack of skill from investors (e.g., Barras, Scaillet, Wermers (2010), Berk and van Binsbergen (2017)). For some funds, the positive alpha may simply reflect the compensation for providing diversification services—these services are included in our measure of gross alpha because the benchmarks are computed from net returns (see BvB). To examine this issue, we compute the proportion of funds with gross alphas above 30 bps per year (the typical expense ratio for the SP500 and Russell 2000 indices). This proportion remains high at 72.0% which implies that positive alphas primarily signal the presence of skill.

The ability of managers to identify profitable trades is not only widespread but economically large—the average fd alpha is reaches 3.48% per year in the population. Perhaps more strikingly, there is a wide heterogeneity among managers because the density $\phi(a)$ is both volatile (4.01% per year) and positively skewed (4.9). In other words, a few managers have stellar investment skills that are well above average. Similar to Pastor, Stambaugh, and Taylor (2015), we find unplausingly large values for the unadjusted quantiles. For instance, the quantile at 90% equals 13.1% which is 1.6 times larger than the one documented in Table III. This point reinforces the importance of adjusting the EIV bias when measuring the variation in fd alpha among managers.

Figure 4 reveals that the shape of the fd alpha distribution varies dramatically across fund groups. In Panel A, the small-cap group largely dominates the large-cap group—the average difference in fd alpha amounts to 3% per year (4.9-1.9). This sharp difference may be driven by competitive pressures given that a large number of small-cap stocks are untouched by mutual funds (e.g., Hong, Lim, and Stein (2000)). Panel B reveals a similar pattern for high- versus low-expense funds. The right tail of $\phi(a)$ for high-expense funds is significantly fatter, i.e., the difference in the 95%-quantile reaches 7.3% per year (12.8-5.5). Previous studies provide conflicting evidence on the level of skill among high-expense managers (e.g., Carhart (1997), Pastor, Stambaugh, and Taylor (2017)). Here, we show that when skill is defined as the ability to generate profitable

[16] These results resonate with the analysis by Berk and Green (2004) who shows that the vast majority of managers have skills at detecting trading opportunities. Calibrating their model on the observed survival rates and fund flows, they infer that 80% of managers have positive fd alphas.

trades, high-expense managers are unambiguously more skilled than low-expense funds.

Please insert Table III here

Please insert Figure 4 here

A.2 Skill at Alleviating Capacity Constraints

Next, we turn to the analysis of the second skill dimension—the ability of managers to limit the impact of capacity constraints. In Table IV, we report the bias-adjusted summary statistics for the density $\phi(b)$ of the fund size coefficient b_i. To ease interpretation, each size coefficient is standardized so that is corresponds to the annual change in the gross alpha for a one standard deviation change in fund size.

To begin we confirm that the size coefficient is negative on average. In the entire population, a one-standard deviation increase in fund size reduces the gross alpha by 1.41% per year—in other words, a $100 mio. increase in size lowers the alpha by 25 bps per year (the average standard deviation of size is $553 mio.).

In addition, our non-parametric estimation of the full density $\phi(b)$ brings new insights into the the impact of capacity constraints. First, the overwhelming majority of funds are subject to capacity constraints. We find that close to 90% of funds have negative size coefficients ($\pi_b^- = 86.8\%$), while the remaining funds all have coefficients close to zero (the quantile at 95% is a mere 0.76% per year). Second, we document a strong heterogeneity among funds. The distribution $\phi(b)$ is volatile (1.5% per year) and left-skewed (-7.7) which signals that a subset of managers face steep decreasing returns to scale. For instance, 5% of the managers see the fund gross alpha decrease by 2.92% per year after a one-standard deviation increase in size.

Similar to the fd alpha, the distribution $\phi(b)$ varies significantly across fund groups. As shown in Figure 5, the average size coefficient for small-cap funds is twice that of large-cap funds (-1.80% vs -0.92% per year). This is consistent with previous studies which relate capacity constraints to the illiquidity of small-cap stocks (e.g., Chen et al. (2004), Yan (2004)). Compared to value funds, some growth funds face tight capacity constraints while others are more skilled at scaling up their strategies. As a result, the distribution of growth funds is more spread out and exhibits thicker left and right tails. Finally, high-expense funds are severely impacted by capacity constraints—the average size coefficient is twice that of low-expense funds (-1.86% vs -0.83% per year).

An important insight from our analysis is that the two skill dimensions are negatively related. For instance, the pairwise correlation between the average fd alpha and size coefficient is nearly perfect across groups (-0.97). The small-cap group provides a clear illustration of this pattern. While they largely dominate their peers in the large-cap group in detecting profitable trades, they find it more difficult to scale up their investment strategies. To determine the overall effect of the two skill dimensions, we therefore need to measure the value added by these managers.

Please insert Table IV here

Please insert Figure 5 here

B The Value Added

B.1 Aggregating the Two Skill Dimensions

In this section, we examine the value added by fund managers. This measure is the total rent (profits) created by managers from their talent—in the neoclassical model, the value addes depends on both the fd alpha a and the size coefficient b: $va = \frac{a^2}{4b}$. Therefore, it aggregates the two dimensions of skill, namely the detection of trading opportunities and the reduction of capacity constraints. Table V shows the summary statistics for the density $\phi(va)$ computed from the estimated value added and adjusted for the EIV bias.

Consistent with the prevalence of skill inferred from the gross alpha, we find ample evidence that managers create value through their investment decisions. In the entire population, the value added is positive for 72% of the funds and its average is equal to $9.0 mio per year. Our non-parametric approach reveals that a large heterogeneity among funds, i.e. the volatility of the value added is equal to $24 mio. per year. It also shows that $\phi(va)$ is highly non-normal—it is right skewed (12.0) and exhibits extremely fat tails (44.1). Therefore, the population includes 5% of managers able to produce a value added in excess of $42.2 mio. per year.

Our analysis of the entire distribution of the value added also provides a more nuanced skill comparison across groups. Panel A of Figure 6 compares the small- and large-cap groups and sheds light on two opposite forces. On the one hand, a significant fraction of large-cap managers produce larger value added than small-cap managers. For one, the difference in the 95%-quantile is equal to $17.5 mio. per year (44.1-26.5). In other words, these large-cap managers exhibit lower fd alphas but are able to operate larger funds. On the other hand, some large-cap managers destroy more value that small-cap managers as they operate large funds with negative gross alphas. The fact that these negative alpha funds still attact large pools of money provides further empirical evidence that some investors fail to allocate funds optimally. We document the same pattern among low- and high-expense funds—the difference in volatility between the two groups is equal to $23.7 mio (38.0-14.3).

Please insert Table V here

Please insert Figure 6 here

B.2 Is the Value Added Optimized?

In progress

C The Gross Alpha

C.1 Prevalence of Skill among Managers

We conclude our analysis with the gross alpha α which is the most commonly-used measure in the literature. In Table VI, we report the summary statistics for the density $\phi(a)$ computed from the estimated fd alphas and adjusted for the EIV bias.

To begin, we examine the proportions π_α^- and π_α^+ of funds with negative and positive gross alphas. This analysis is motivated by the partial equilibrium model of Grinblatt and Titman (1989) which shows that a manager with superior stock picking abilities produces a positive gross alpha.[17] Consistent with our analysis of the fd alpha, we find that the vast majority of managers have positive gross alphas.[18] However, the proportion π_α^- is higher than that obtained with the fd alpha (27.4% versus 11.9%). This difference arises because some managers are skilled and yet deliver a negative gross alpha because they face capacity constraints (such constraints are absent from the model of Grinblatt and Titman (1989)). In a world where both investors and managers are uncertain about the *true* level of skill, the fund size may grow too large and drag the gross alpha in negative territory.

Please insert Table VI here

[17]This result requires that the manager (i) maximizes his utility function and (ii) does not time the market—a condition that is consistent with the empirical evidence documented in the literature (e.g., Ferson and Schadt (1996), Merton and Henrikson (1981), Treynor and Mazuy (1966)).

[18]These proportions are significantly larger than those obtained with the FDR approach (Barras, Scaillet, and Wermers (2010)). Intuitively, this approach estimates π_α^+ by counting funds with large alpha t-statistics and thus does not detect all the skilled funds. In contrast, our non-parametric approach directly estimates the gross alpha distribution using information from the entire cross-section of funds.

C.2 Fund Fees and Size

As shown in Table VI, the gross alphas vary significantly both across and within groups—for one, the annual average alpha is equal to 0.26% for large-cap funds compared to 1.52% for small-cap funds. In the context of the neoclassical model in Table I, this variation may or may not capture differences in managerial skill depending on the compensation scheme chosen by managers.

In Table VII, we report descriptive statistics for fees and size and find that they are not consistent with Schemes II and III—the average fees are not tiny (1.19% per year) and the size is not constant across funds. Scheme I under which the fund size is optimal seems more realistic because both fees and size vary across funds. However, the cross-fund dispersion in size is extremely large—for one, the difference between the quantiles at 90% and 10% reaches $1,504 mio. There is also a strong and negative correlation between the average fees and size across groups (-0.96). Both results resonates with the analysis by Habib and Johnson (2016) who argue that managers choose to maintain low fees and a larger fund size to mitigate several institutional constraints. The Investment Company Act imposes diversification rules on 75% of the portfolio which prevent managers from exhausting their trading opportunities if the fund is too small. In addition, holding a portion of the fund passively managed allow managers to hide their informed trades and obtain better prices.

Consistent with this analysis, the gross alpha is partly informative about the fd alpha. Figure 7 shows that the densities of the fd and gross alphas overlap in the center.[19] However, the gross alpha does not capture the large fd alphas produced by a fraction of the manager population. Figure 8 also reveals that the densities of the value added and the gross alpha are radically different—the value density is spread out, whereas the gross alpha density peaks around its mean.[20]. Because managers do not set fees to guarantee equal size across funds, the gross alpha fails to capture the large differences in scale at which managers implement their investment strategies. In short, the gross alpha correlate with the fd alpha but is informative about neither the size coefficient nor the value added. ross alpha than with the size coefficient and the value added.

Please insert Tables VII here

[19]To compare the two densities, we multiply the gross alpha by 2. because the neoclassical model predicts that $a_i = 2\alpha_i$ if managers choose the optimal size.

[20]To compare the two densities, we multiply the gross alpha by the median size q. As show in Section II.A, the neoclassical model predicts that $va_i = q\alpha_i$ if managers choose the optimal size.

Please insert Figures 7, 8 here

C.3 From Gross to Net Alphas

Finally, we apply our approach for estimating the density $\phi(\alpha^n)$ of the net alpha (as explained in Section II.D). This performance analysis reported in Table VIII determines whether investors receive any surplus alphas net of all expenses and provides information about the bargaining power of managers when setting fees.

The neoclassical model predicts that the net alpha of each fund equals zero—that is, skilled managers are in scarce supply and able to extract all the rent from their talent ($\alpha = f_e$). Consistent with this prediction, we find that the positive tail of $\phi(\alpha^n)$ is significantly thiner than that of $\phi(\alpha)$. In the whole population, positive alphas are only observed for one third of the funds ($\pi^+_{\alpha^n} = 35.4\%$) and are typically small (the quantile at 90% is a mere 1.04% per year). The same pattern emerges from Table IV because fund groups with higher gross alphas also charge higher fees (the pairwise correlation equals 0.57). The fact that $\pi^+_{\alpha^n}$ is not exactly zero can be driven by learning effects. While investors update their beliefs about the superior skill of the manager, the latter cannot extract the full rent from his talent. This learning phase can last a long time because investors must learn about both the fd alpha a and the impact of capacity constraints b (Pastor and Stambaugh (2012)). In addition, investors may have some bargaining power in specific segments of the market in which they are less willing to invest. We see that the highest proportion $\pi^+_{\alpha^n}$ is observed for small-cap funds (54.9%) which invest in stocks that are more subject to asymmetric information problems (see Merton (1987)).

One empirical limitation of the neoclassical model is that $\phi(\alpha^n)$ does not peak around zero: (i) the volatility remains nearly unchanged compared to the gross alpha; (ii) the skewness is negative in all but one group. To visualize these patterns, Figure 9 plots the gross and net alpha densities for each group. We see that $\phi(\alpha^n)$ is essentially a shifted version of $\phi(\alpha)$ with fatter left tails. In other words, some managers are particularly unskilled and yet able to charge hefty fees to investors (Christoffersen and Musto (2002), Gruber (1996)). Therefore, the behaviour of these apparently irrational investors drives a wedge between gross alphas and fees that is left unexplained by the neoclassical model.

Please insert Table VIII here

Please insert Figure 9 here

D Sensitivity Analysis

In progress

D.1 The Carhart Model

D.2 Industry-wide Capacity Constraints

V Conclusion

References

[1] Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, Should Investors Avoid all Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation, Journal of Finance 56, 45-85.

[2] Barras L., O. Scaillet, and R. Wermers, 2010, False discoveries in mutual fund performance: measuring luck in estimated alphas, Journal of Finance 65, 179-216.

[3] Berk J. B., and J. van Binsbergen, 2015, Measuring skill in the mutual fund industry, Journal of Financial Economics 118, 1-20.

[4] Berk, Jonathan B., and Richard C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, Journal of Political Economy 112, 1269-1295.

[5] Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, Journal of Finance 52, 57-82.

[6] Chen J., H. Hong, M. Huang, and J. D. Kubik, 2004, Does fund size erode mutual fund performance? The role of liquidity and organization, American Economic Review 94, 1276-1302.

[7] Christoffersen, Susan E. K., and David K. Musto, 2002, Demand Curves and the Pricing of Money Management, Review of Financial Studies 15, 1495-1524.

[8] Elton, Edwin J., Martin J. Gruber, and Jeffrey Busse, 2004, Are Investors Rational? Choices among Index Funds, Journal of Finance 59, 261-288.

[9] Gagliardini P., E. Ossola, and O. Scaillet, 2016, Time-varying risk premium in large cross-sectional equity datasets, Econometrica 84, 985-1056.

[10] Grinblatt, Mark, and Sheridan Titman, 1989, Portfolio Performance Evaluation: Old Issues and New Insights, Review of Financial Studies 2, 393-421.

[11] Harvey C., and Y. Liu, 2016, Rethinking Performance Evaluation, Review of Financial Studies, Forthcoming.

[12] Jensen, Michael C., 1968, The Performance of Mutual Funds in the Period 1945-1964, Journal of Finance 23, 389-416.

[13] Jones, Christopher S., and Jay Shanken, 2005, Mutual Fund Performance with Learning across Funds, Journal of Financial Economics 78, 507-552.

[14] Lynch, Anthony W., and David Musto, 2003, How Investors Interpret Past Fund Returns, Journal of Finance 58, 2033-2058.

[15] Newey, Whitney K., and Kenneth D. West, 1987, A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica 55, 703-708.

[16] Kacperczyk, M., S. van Nieuwerburgh, and L. Veldkamp, 2014, Time-Varying Fund Manager Skill, Journal of Finance 69, 1455-1484.

[17] Pastor, L., R. F. Stambaugh, and L. Taylor, 2015, Scale and Skill in Active Management, Journal of Financial Economics 116, 23-45.

[18] Pastor, L., R. F. Stambaugh, and L. Taylor, 2016, Do Funds Make More When They Trade More?, Journal of Finance, Forthcoming.

[19] Perold, A. F., and R. S. Salomon, Jr., 1991, The Right Amount of Assets under Management, Financial Analysts Journal 47, 31-39.

[20] Sirri, E., and P. Tufano, 1998, Costly Search and Mutual Fund Flows, Journal of Finance 53, 1589-1622.

VI Appendix

A.2 Estimation of the density of the skill measures

In this appendix, we focus on the proof of Proposition I.1 stated for the three performance measures. The proof for smoothing estimated slopes (betas) or other quantities estimated with parametric rates follows similar arguments. Let us first focus on the gross alpha $m_i = \alpha_i$. From the OLS estimation, we have:

$$\begin{aligned}\hat{m}_i &= e_1' \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_t I_{i,t} x_t r_{i,t} = m_i + e_1' \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_t I_{i,t} x_t \varepsilon_{i,t} \\ &= m_i + \frac{1}{\sqrt{T}} \tau_{i,T} e_1^{K+1\prime} \hat{Q}_{x,i}^{-1} \left(\frac{1}{\sqrt{T}} \sum_t I_{i,t} x_t \varepsilon_{i,t} \right) \equiv m_i + \frac{1}{\sqrt{T}} \hat{\eta}_{i,T}. \qquad (20)\end{aligned}$$

Moreover, let us write

$$\hat{\eta}_{i,T} = \eta_{i,T} + \frac{1}{\sqrt{T}} \hat{v}_{i,T}, \qquad (21)$$

where $\eta_{i,T} = \tau_i \frac{1}{\sqrt{T}} \sum_t I_{i,t} u_{i,t}$, $u_{i,t} = e_1' Q_x^{-1} x_t \varepsilon_{i,t}$, and $\tau_i = \text{plim}_{T \to \infty} \tau_{i,T}$. Hence, $\hat{\eta}_{i,T}/\sqrt{T}$ is the estimation error on $m_i = \alpha_i$. In particular, $\hat{v}_{i,T}/T$ is the component due to estimating the matrix Q_x and to the random sample size T_i. Then, we write $\hat{\phi}(m) - \phi(m) = I_1 + I_2 + I_3 + I_4$, where:

$$\begin{aligned}I_1 &= \frac{1}{h} E\left[K\left(\frac{m_i - m}{h} \right) \right] - \phi(m), \\ I_2 &= \frac{1}{h} E\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h} \right) \right] - \frac{1}{h} E\left[K\left(\frac{m_i - m}{h} \right) \right], \\ I_3 &= \frac{1}{nh} \sum_i \left\{ K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h} \right) - E\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h} \right) \right] \right\}, \\ I_4 &= \frac{1}{nh} \sum_i \left[\mathbf{1}_i^{\chi} K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} + \hat{v}_{i,T}/T - m}{h} \right) - K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h} \right) \right].\end{aligned}$$

The first term I_1 is the smoothing bias, the second term I_2 is the Error-in-Variable (EIV) bias, I_3 is the main stochastic term, and I_4 is a remainder term. We characterise the first three terms in the following.

(i) From standard results in kernel density estimation, the smoothing bias is such that $I_1 = \frac{1}{2} \phi^{(2)}(m) K_2 h^2 + O(h^3)$, with $K_2 = \int u^2 K(u) du$.

(ii) By a Taylor expansion of the kernel function K we have:

$$I_2 = \sum_{j=1}^{\infty} \frac{1}{j!T^{j/2}h^{j+1}} E\left[K^{(j)}\left(\frac{m_i - m}{h}\right)\eta_{i,T}^j\right].$$

Moreover, by applying j times partial integration and a change of variable:

$$\begin{aligned} \frac{1}{h^{j+1}} E\left[K^{(j)}\left(\frac{m_i - m}{h}\right)\eta_{i,T}^j\right] &= \frac{1}{h^{j+1}} \int K^{(j)}\left(\frac{u-m}{h}\right)\psi_{T,j}(u)du \\ &= (-1)^j \frac{1}{h} \int K\left(\frac{u-m}{h}\right)\psi_{T,j}^{(j)}(u)du \\ &= (-1)^j \int K(u)\,\psi_{T,j}^{(j)}(m+hu)du, \end{aligned}$$

where $\psi_{T,j}(m) = E[\eta_{i,T}^j | m_i = m]\phi(m)$ for $j = 1, 2....$ We have $\psi_{T,1}(m) = 0$ and $\psi_{T,2}(m) = E[S_i | m_i = m]\phi(m) \equiv \psi(m)$ where $S_i = \text{plim}_{T\to\infty} \frac{1}{T}\sum_{t,s} I_{i,t}I_{i,s}u_{i,t}u_{i,s}$. By weak serial dependence of the error terms, functions $\psi_{T,j}(m)$ for $j > 2$ are bounded with respect to T. Thus, we get: $I_2 = \frac{1}{2T}\psi^{(2)}(m) + O(1/T^{3/2} + h^2/T)$.

(iii) Let us now consider term I_3. For expository purpose, let us assume that the error terms are cross-sectionally independent, and the factor values x_t are treated as given constants. Then:

$$V[I_3] = \frac{1}{nh^2} V\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right].$$

From the above arguments, we have $\frac{1}{h}E\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right] = \phi(m) + o(1)$ and

$$\begin{aligned} \frac{1}{h}E\left[K\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)^2\right] &= \int K(u)^2 du \frac{1}{h} E\left[\bar{K}\left(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h}\right)\right] \\ &= \phi(m)\int K(u)^2 du + o(1), \end{aligned}$$

where $\bar{K}(u) = K(u)^2 / \int K(u)^2 du$. Therefore:

$$V[I_3] = \frac{1}{nh}\phi(m)\int K(u)^2 du + o\left(\frac{1}{nh}\right).$$

Under regularity conditions, the application of an appropriate Central Limit Theorem (CLT) implies $\sqrt{nh}I_3 \Rightarrow N\left(0, \phi(m)K_1\right)$, with $K_1 = \int K(u)^2 du$.

For the a (fd dollar) and va (value added) measures, we can proceed similarly by using the corresponding definition for $u_{i,t}$ listed in the statement of Proposition I.1.

A.3 Asymptotic Mean Integrated Squared Error

From the previous subsection, we get the asymptotic expansion of the bias and variance of the estimator $\hat{\phi}(m)$ with leading terms: $bs(m) = bs_1(m) + bs_1(m)$, and $\sigma^2(m) = \frac{1}{nh}\phi(m)K_1$. The AMISE (asymptotic mean integrated squared error) is given by:

$$\begin{aligned} AMISE(h) &= \int [\sigma^2(u) + bs(u)^2] du \\ &= \frac{1}{nh}K_1 + \frac{h^4 K_2^2}{4}\int [\phi^{(2)}(u)]^2 du \\ &\quad + \frac{h^2 K_2}{2T}\int \phi^{(2)}(u)\psi^{(2)}(u)du + \frac{1}{4T^2}\int \left[\psi^{(2)}(u)\right]^2 du. \end{aligned}$$

The optimal bandwidth h^* is the minimizer of the AMISE, and solves the equation:

$$\begin{aligned} & -\frac{1}{nh^2} + c_1 h^3 + c_2\frac{h}{T} = 0 \\ \iff \quad & 1 = c_1 n h^5 + c_2 \frac{nh^3}{T}. \end{aligned} \tag{22}$$

where $c_1 = K_2^2 \int [\phi^{(2)}(u)]^2 du / K_1$ and $c_2 = K_2 \int \phi^{(2)}(u)\psi^{(2)}(u)du/K_1$.

Let us investigate the speed of convergence to 0 of the optimal bandwidth h^* as a function of n and T. We assume that $c_2 > 0$. There are three possible cases:

(i) The optimal bandwidth is such that nh^5 tends to a nonzero constant and $nh^3/T \to 0$. Then, we have $h^* \sim c_1^{-1/5} n^{-1/5}$, that is the Silverman rule. This solution is admissible, i.e., satisfies $nh^3/T \to 0$, if the sample sizes n and T are such that $n^{2/5}/T \to 0$, i.e., T is sufficiently large.

(ii) The optimal bandwidth is such that nh^3/T tends to a nonzero constant and $nh^5 \to 0$. This case is possible only if $n/T \to \infty$. Then, we have $h^* \sim c_2^{-1/3}(n/T)^{-1/3}$. This solution is admissible, i.e., satisfies $nh^5 \to 0$, if the sample sizes are such that $n^{2/5}/T \to \infty$.

(iii) When $n^{2/5}/T \to \rho$, with $\rho > 0$, the two rates of convergence $n^{-1/5}$ and $(n/T)^{-1/3}$

coincide. Then, equation (22) has a solution such that $h^* \sim \bar{c}^{1/5} n^{-1/5}$, where $\bar{c}$ solves the equation $1 = c_1 \bar{c} + c_2 \rho \bar{c}^{3/5}$.

Let us now consider the asymptotic distribution of estimator $\hat{\phi}(m)$ for a generic bandwidth sequence h shrinking to zero such that $nh \to \infty$. From the above analysis, we have:

$$\sqrt{nh}\left(\hat{\phi}(m) - \phi(m) - bs(m)\right) \Rightarrow N\left(0, \phi(m)K_1\right).$$

For some bandwidth sequences, the asymptotic bias is negligible. (i) If $Th^2 \to \infty$, the dominant component in the asymptotic bias is due to smoothing and is of order $O(h^2)$. Then, the asymptotic bias is negligible if $nh^5 \to 0$. This condition is compatible with the condition $Th^2 \to \infty$ if $n/T^{5/2} \to 0$. (ii) If $Th^2 \to 0$, the dominant component in the asymptotic bias is due to the EIV problem and is of order $O(1/T)$. The asymptotic bias is negligible if $nh/T^2 \to 0$.

Let us now consider the asymptotic distribution when $h = h^*$ is the optimal bandwidth. We can check that $\sqrt{nh^*}(h^{*3} + h^{*2}/T + 1/T^{3/2}) = o(1)$ if $n/T^4 \to 0$. Then, we can replace the bias component $bs(m)$ by its asymptotic approximation to get:

$$\sqrt{nh^*}\left(\hat{\phi}(m) - \phi(m) - \frac{1}{2}\phi^{(2)}(m)K_2 h^{*2} - \frac{1}{2T}\psi^{(2)}(m)\right) \Rightarrow N\left(0, \phi(m)K_1\right). \tag{23}$$

If $n^{2/5}/T \to \infty$, we have $Th^{*2} \to 0$, and the smoothing bias is negligible. If $n^{2/5}/T \to 0$, we have $Th^{*2} \to \infty$, and the EIV bias is negligible.

A.4 A simple plug-in method

In this section we develop a simple plug-in method to implement the optimal bandwidth that is solution of equation (22). We focus on a Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-u^2/2\right)$. Then, the kernel constants are $\int K(u)^2 du = \frac{1}{2\sqrt{\pi}}$ and $K_2 = \int u^2 K(u) du = 1$. To compute constants c_1 and c_2, we rely on a reference model which assumes a bivariate Gaussian distribution for m_i and $s_i = \log(S_i)$ with mean parameters μ_m, μ_s, variance parameters σ_m^2, σ_s^2, and correlation parameter ρ. The Gaussian marginal density of m_i implies that our reference model nests the one underlying the derivation of the Silverman rule for kernel smoothing. The constants c_1 and c_2 are given by:

$$
\begin{aligned}
c_1 &= 2\sqrt{\pi}\int [\phi^{(2)}(u)]^2 du, \\
c_2 &= 2\sqrt{\pi}\int \phi^{(2)}(u)\psi^{(2)}(u)du = 2\sqrt{\pi}\int \phi^{(4)}(u)\psi(u)du,
\end{aligned}
$$

where we have used twice partial integration in c_2. Let us now compute the two integrals appearing in these formulas.

(i) We have $\phi(m) = \frac{1}{\sigma_m}\varphi\left(\frac{m-\mu_m}{\sigma_m}\right)$, where $\varphi(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-z^2/2\right)$ is the standard Gaussian density. We get: $\phi^{(1)}(m) = -\frac{1}{\sigma_m}\left(\frac{m-\mu_m}{\sigma_m}\right)\phi(m)$, and $\phi^{(2)}(m) = \frac{1}{\sigma_m^2}\left(\left(\frac{m-\mu_m}{\sigma_m}\right)^2 - 1\right)\phi(m)$. Then:

$$
\begin{aligned}
\int [\phi^{(2)}(u)]^2 du &= \frac{1}{\sigma_m^5}\int (z^2-1)^2 \frac{1}{2\pi}\exp(-z^2)dz = \frac{1}{2\sqrt{\pi}\sigma_m^5}\int (v^2/2-1)^2\varphi(v)dv \\
&= \frac{3}{8\sqrt{\pi}\sigma_m^5},
\end{aligned}
$$

with the changes of variables from u to $z = (u-\mu_m)/\sigma_m$, and from z to $v = \sqrt{2}z$.

(ii) We have $E[\exp(s_i)|m_i = m] = \exp\left(\mu_s + \rho\sigma_s\left(\frac{m-\mu_m}{\sigma_m}\right) + \frac{1}{2}\sigma_s^2(1-\rho^2)\right)$, so that $\psi(m) = \exp\left(\mu_s + \rho\sigma_s\left(\frac{m-\mu_m}{\sigma_m}\right) + \frac{1}{2}\sigma_s^2(1-\rho^2)\right)\phi(m)$ and

$$
\begin{aligned}
\psi^{(2)}(m) &= \exp\left(\mu_s + \rho\sigma_s\left(\frac{m-\mu_m}{\sigma_m}\right) + \frac{1}{2}\sigma_s^2(1-\rho^2)\right)\phi(m) \\
&\quad \times \left\{\left(\frac{\sigma_s\rho}{\sigma_m}\right)^2 - 2\frac{\sigma_s\rho}{\sigma_m^2}\left(\frac{m-\mu_m}{\sigma_m}\right) + \frac{1}{\sigma_m^2}\left[\left(\frac{m-\mu_m}{\sigma_m}\right)^2 - 1\right]\right\} \\
&= \exp\left(\mu_s + \frac{1}{2}\sigma_s^2\right)\frac{1}{\sigma_m^2}\left(\left(\frac{m-\mu_m-\rho\sigma_s\sigma_m}{\sigma_m}\right)^2 - 1\right)\frac{1}{\sigma_m}\varphi\left(\frac{m-\mu_m-\rho\sigma_s\sigma_m}{\sigma_m}\right).
\end{aligned}
$$

We can also derive directly the last line by rewriting $\psi(m) = \omega(m)\phi(m)$ in the bivariate Gaussian reference model as a recentered Gaussian density up to a multiplicative constant:

$$
\psi(m) = \exp\left(\mu_s + \frac{1}{2}\sigma_s^2\right)\frac{1}{\sigma_m}\varphi\left(\frac{m-\mu_m-\rho\sigma_s\sigma_m}{\sigma_m}\right),
$$

and then differentiating twice that expression. Besides, we have:

$$\int \phi^{(4)}(a)E[\exp(s_i)|m_i = m]\phi(m)dm = \frac{\exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1-\rho^2)\right)}{\sigma_m^5} \int \varphi^{(4)}(z)\exp(\rho\sigma_s z)\varphi(z)dz$$
$$= \frac{\exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1-\rho^2)\right)}{2\sqrt{\pi}\sigma_m^5} \int (v^4/4 - 3v^2 + 3)\exp(\lambda v)\varphi(v)dv,$$

where $\lambda = \rho\sigma_s/\sqrt{2}$, by using the same changes of variables as above and $\varphi^{(4)}(z) = (z^4 - 6z^2 + 3)\varphi(z)$. Moreover, we have $\int z^k \exp(\lambda z)\varphi(z)dz = E[Z^k \exp(\lambda Z)] = \frac{\partial^k}{\partial \lambda^k} E[\exp(\lambda Z)]$ with $E[\exp(\lambda Z)] = \exp(\lambda^2/2)$. This yields

$$\begin{aligned}\int (v^4/4 - 3v^2 + 3)\exp(\lambda v)\varphi(v)dv &= \left(\frac{1}{4}\frac{\partial^4}{\partial\lambda^4} - 3\frac{\partial^2}{\partial\lambda^2} + 3\right)\exp(\lambda^2/2) \\ &= \frac{1}{4}(\lambda^4 - 6\lambda^2 + 3)\exp(\lambda^2/2).\end{aligned}$$

Thus, we get:

$$\int \phi^{(4)}(m)E[\exp(s_i)|m_i = m]\phi(m)dm = \frac{3\exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1-\rho^2/2)\right)}{8\sqrt{\pi}\sigma_m^5}(\rho^4\sigma_s^4/12 - \rho^2\sigma_s^2 + 1).$$

The optimal bandwidth h^* is obtained by solving equation (22) with coefficients c_1 and c_2 given by:

$$c_1 = \frac{3}{4\sigma_m^5}, \qquad c_2 = \frac{3}{4\sigma_m^5}(\rho^4\sigma_s^4/12 - \rho^2\sigma_s^2 + 1)\exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1-\rho^2/2)\right).$$

We have $c_2 \geq 0$ when either $\rho^2\sigma_s^2 \leq 6 - 2\sqrt{6}$, or $\rho^2\sigma_s^2 \geq 6 + 2\sqrt{6}$. In our implementation, the parameters σ_m, μ_s, σ_s and ρ are estimated by the sample moments of $\hat{m}_i$ and $\hat{s}_i = \log \hat{S}_i$.

A.5 Estimation of the moments of the skill measures

Let us consider the estimation of the cross-sectional expectation $E[g(m_i)]$, where g is a given smooth function. We investigate the convergence properties of the cross-sectional estimator $\frac{1}{n}\sum_{i=1}^{n} g(\hat{m}_i)\mathbf{1}_i^{\chi}$ based on the OLS estimates $\hat{m}_i$ of the non-trimmed assets only. Proposition VI.1 states the asymptotic normality under the double asymptotics "large

n, small T" in unbalanced panel when we assume the linear factor model (2).

Proposition VI.1 *As $n, T \to \infty$, such that $n = o(T^3)$,*

$$\sqrt{n}\left(\frac{1}{n}\sum_i g(\hat{m}_i)\mathbf{1}_i^{\chi} - E\left[g(m_i)\right] - \mathcal{B}_T\right) \Rightarrow N\left(0, V\left[g(m_i)\right]\right), \tag{24}$$

with $\mathcal{B}_T = \dfrac{1}{2T}E[g^{(2)}(m_i)S_i]$.

Here we have an asymptotic bias $\mathcal{B}_T$ of order $1/T$ where S_i is the same as in Proposition I.1. It comes from the estimation error in $\hat{m}_i$ (EIV contribution). The asymptotic variance is simply the cross-sectional variance $V\left[g(m_i)\right]$. The condition $n = o(T^3)$ is used to control the remainder term in the Taylor expansion of function g, and the bias term. For the mean of the skill measures, we have $g(m) = m$ and the asymptotic bias is zero from $g^{(1)}(m) = g^{(2)}(m) = 0$. For that particular case, we do not need the condition $n = o(T^3)$ for Proposition VI.1 to hold. For the second moment of the skill measures, we have $g(m) = m^2$, and the asymptotic bias is not zero from $g^{(2)}(m) = 2$. We need to estimate the asymptotic bias and asymptotic variance to get a feasible asymptotic normality result and to build asymptotic confidence intervals. We can achieve that by replacing the unknown moments by consistent estimators based on empirical averages as done earlier in the paper. For the variance of the skill measures, the asymptotic bias is the same as the one for the second moment, but the asymptotic variance is $V\left[(m_i - E\left[m_i\right])^2\right]$. The asymptotic bias and variance for the skewness and kurtosis can be obtained via the delta method. For example, for the skewness $Sk = E\left[(m_i - E\left[m_i\right])^3\right]/V\left[(m_i - E\left[m_i\right])^2\right]^{3/2}$, its asymptotic bias is

$$\mathcal{B}_T(Sk) = (\nabla_3 Sk)\mathcal{B}_T(E[m_i^3]) + (\nabla_2 Sk)\mathcal{B}_T(E[m_i^2]),$$

with

$$\begin{aligned}
\nabla_3 Sk &= V[m_i]^{-3/2}, \\
\nabla_2 Sk &= -3E[m_i]V[m_i]^{-3/2} + \left\{E[m_i^3] - 3E[m_i^2]E[m_i] + 2E[m_i^3]\right\}(\frac{-3}{2})V[m_i]^{-5/2}
\end{aligned}$$

where we denote the derivative of Sk w.r.t. $E[m_i^j]$ as $\nabla_j Sk$. Similarly, for the kurtosis $Ku = E\left[(m_i - E\left[m_i\right])^4\right]/V\left[(m_i - E\left[m_i\right])^2\right]^2$, its asymptotic bias is

$$\mathcal{B}_T(Ku) = (\nabla_4 Ku)\mathcal{B}_T(E[m_i^4]) + (\nabla_3 Ku)\mathcal{B}_T(E[m_i^3]) + (\nabla_2 Ku)\mathcal{B}_T(E[m_i^2]),$$

with

$$\nabla_4 Ku = V[m_i]^{-2}, \qquad \nabla_3 Ku = -4E[m_i]V[m_i]^{-2},$$
$$\nabla_2 Ku = 6E[m_i]^2 V[m_i]^{-2} + \left\{E[m_i^4] - 4E[m_i^3]E[m_i] + 6E[m_i^2]E[m_i]^2 - 3E[m_i]^4\right\}(-2)V[m_i]^{-3}.$$

Proof of Proposition VI.1 : Equation (20) yields the mean value expansion

$$g(\hat{m}_i) = g(m_i) + g^{(1)}(\bar{m}_i)\frac{1}{\sqrt{T}}\hat{\eta}_{i,T} + g^{(2)}(\bar{m}_i)\frac{1}{2T}\hat{\eta}_{i,T}\hat{\eta}_{i,T}',$$

where $\bar{m}_i$ lies between $\hat{m}_i$ and m_i. Then we get

$$\begin{aligned}
& \sqrt{n}\left(\frac{1}{n}\sum_i g(\hat{m}_i)\mathbf{1}_i^{\chi} - E\left[g(m_i)\right] - \mathcal{B}_T\right) \\
= & \frac{1}{\sqrt{n}}\sum_i \left(g(m_i) - E\left[g(m_i)\right]\right) - \frac{1}{\sqrt{n}}\sum_i g(m_i)(1 - \mathbf{1}_i^{\chi}) + \frac{1}{\sqrt{nT}}\sum_i \mathbf{1}_i^{\chi} g^{(1)}(\bar{m}_i)\hat{\eta}_{i,T} \\
& + \frac{1}{2T}\frac{1}{\sqrt{n}}\sum_i \mathbf{1}_i^{\chi}\left(g^{(2)}(\bar{m}_i)\hat{\eta}_{i,T}\hat{\eta}_{i,T}' - E\left[g^{(2)}(m_i)S_i\right]\right) \\
\equiv & I_{71} + I_{72} + I_{73} + I_{74}.
\end{aligned}$$

We have $I_{71} \Rightarrow N(0, V[g(m_i)])$ from the standard CLT. We also have $I_{72} = o_p(1)$. The bound $I_{73} = O_p(1/\sqrt{T}) = o_p(1)$ follows from similar arguments as in Lemma 2 of GOS. Then, the asymptotic distribution (24) follows from the remainder term $I_{74} = O_p(\sqrt{n/T^3} + \sqrt{n}/T^2 + 1/T)$, which gives $I_{74} = o_p(1)$ if $n = o(T^3)$.

A.6 Estimation of the cumulative distribution function and quantiles of the skill measures

We consider the estimation of the cumulative distribution function (cdf) $\Phi(m) = P[m_i \leq m]$ of the skill measures, for any given real argument m, and of the associated quantile function $Q(u) = \Phi^{-1}(u)$, for any given percentile level $u \in (0,1)$, in the linear factor model (2). Building on the previous section, the estimator of the cdf is the cross-sectional average of the indicator function $g(\hat{m}_i) = \mathbf{1}\{\hat{m}_i \leq m\}$ based on the OLS estimates $\hat{m}_i$ for the non-trimmed assets, namely $\hat{\Phi}(m) = \frac{1}{n}\sum_i \mathbf{1}\{\hat{m}_i \leq m\}\mathbf{1}_i^{\chi}$. The quantile estimator is the inverse function $\hat{Q}(u) = \hat{\Phi}^{-1}(u)$.

The next proposition extends Proposition VI.1 to the estimation of the cdf and quantile function of the skill measures in the linear factor model (2).

Proposition VI.2 *As $n, T \to \infty$, such that $n = o(T^3)$,*

$$\sqrt{n}\left(\hat{\Phi}(m) - \Phi(m) - \mathcal{B}_T(m)\right) \Rightarrow N\left(0, \Phi(m)(1-\Phi(m))\right),$$
$$\sqrt{n}\left(\hat{Q}(u) - Q(u) + \frac{\mathcal{B}_T(Q(u))}{\phi(Q(u))}\right) \Rightarrow N\left(0, \frac{u(1-u)}{\phi(Q(u))^2}\right),$$

with $\mathcal{B}_T(m) = \dfrac{1}{2T}\psi^{(1)}(m)$, where $\psi(m) = E[S_i|m_i = m]\phi(m)$.

As in the previous section, we can approximate the asymptotic bias through a reference model and estimate the asymptotic variance to get feasible results. With our bivariate Gaussian reference model, we get:

$$\begin{aligned}
\psi^{(1)}(m) &= \exp\left(\mu_s + \rho\sigma_s\left(\frac{m-\mu_m}{\sigma_m}\right) + \frac{1}{2}\sigma_s^2(1-\rho^2)\right)\phi(m)\left(\frac{\sigma_s\rho}{\sigma_m} - \frac{m-\mu_m}{\sigma_m^2}\right) \\
&= \exp\left(\mu_s + \frac{1}{2}\sigma_s^2\right)\frac{-1}{\sigma_m}\left(\frac{m-\mu_m-\rho\sigma_s\sigma_m}{\sigma_m}\right)\frac{1}{\sigma_m}\varphi\left(\frac{m-\mu_m-\rho\sigma_s\sigma_m}{\sigma_m}\right).
\end{aligned}$$

We can proceed similarly for the quantile case.

Proof of Proposition VI.2 : From (20), we have: $E\left[1\{\hat{m}_i \leq m\}\right] = P\left[m_i + \frac{1}{\sqrt{T}}\hat{\eta}_{i,T} \leq m\right]$. By using the results in Gourieroux, Laurent, and Scaillet (2000), Martin, and Wilde (2001), Gordy (2003), Gagliardini and Gourieroux (2011), we get:

$$\begin{aligned}
P\left[m_i + \frac{1}{\sqrt{T}}\hat{\eta}_{i,T} \leq m\right] &= \Phi(m) - \frac{1}{\sqrt{T}}\phi(m)E[\hat{\eta}_{i,T}|m_i = m] \\
&\quad + \frac{1}{2T}\frac{d}{dm}\left(\phi(m)E[\hat{\eta}_{i,T}|m_i = m]\right) + o(1/T).
\end{aligned}$$

From (21), the bias expansion is such that: $E[\hat{\Phi}(m)] - \Phi(m) = \mathcal{B}_T(m) + E\left[1\{\hat{m}_i \leq m\}(1 - \mathbf{1}_i^{\chi})\right] + o(1/T)$. We deduce the asymptotic normality of the cdf estimator by controlling the different terms and applying a CLT.

We deduce the asymptotic normality of the quantile estimator by using the Bahadur expansion for the quantile estimator at level $u \in (0,1)$: $\hat{Q}(u) - Q(u) = -\frac{1}{\phi(Q(u))}\left(\hat{\Phi}(Q(u)) - u\right)$.

Table I
Relations Between the Skill Measures

Legend to be added

Compensation Scheme	Scheme I (optimal size)	Scheme II (squared optimal size)	Scheme III (same size)	Scheme IV (arbitrary size)
Fees are set such that	Managers choose the optimal size	Managers choose the squared optimal size	Managers choose the same average size	Funds choose an arbitrary size
Predictions for Size/Fees	Moderate cross-fund variation in fees/size	Huge size Tiny fees	Same Size for all funds	Large cross-fund variation in size
Does the Gross Alpha Measure Skill?	YES	YES	YES	NO
	First-Dollar Alpha (1st skill dimension)	Size Coefficient (2nd skill dimension)	Value Added	

Table II
Summary Statistics for the Value-Weighted Portfolio of Funds

Legend to be added

Panel A: Gross Excess Return

	Average Nb. Funds	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis
All Funds	1191	7.69	15.06	-0.8	5.4
Small-cap Funds	257	8.80	19.32	-0.6	5.0
Large-cap Funds	459	7.78	14.79	-0.7	5.3
Growth Funds	529	8.14	16.29	-0.8	5.3
Value Funds	309	7.45	13.88	-0.7	5.5
Low-Expense Funds	401	7.72	14.66	-0.8	5.3
High-Expense Funds	314	8.03	16.43	-0.8	5.2

Panel B: Estimated Fund Betas

	Market	Size	Value	Momentum	Adj. R2
All Funds	0.93	0.27	-0.09	0.01	0.98
Small-cap Funds	0.99	0.82	-0.27	0.04	0.97
Large-cap Funds	0.95	0.15	-0.05	0.01	0.99
Growth Funds	0.95	0.33	-0.27	0.03	0.97
Value Funds	0.92	0.14	0.19	-0.01	0.98
Low-Expense Funds	0.93	0.21	-0.03	0.00	0.98
High-Expense Funds	0.93	0.42	-0.27	0.01	0.97

Table III
Cross-Sectional Distribution of The First-Dollar Alpha

Legend to be added

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	5%	95%
All Funds	3.38	4.01	4.9	6.5	11.9	88.1	-2.19	8.17
Small-cap Funds	4.92	3.85	8.6	2.2	6.6	93.4	-1.44	9.91
Large-cap Funds	1.95	2.69	4.7	4.3	19.4	80.6	-2.57	6.03
Growth Funds	3.72	5.17	6.1	14.3	17.3	82.7	-3.73	11.23
Value Funds	3.49	5.02	7.2	25.0	12.5	87.5	-2.01	8.18
Low-Expense Funds	2.33	3.87	5.4	8.0	17.1	82.9	-2.13	5.50
High-Expense Funds	4.95	5.56	6.4	5.7	14.0	86.0	-2.91	12.81

Table IV
Cross-Sectional Distribution of The Size Coefficient

Legend to be added

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	5%	95%
All Funds	-1.42	1.50	-7.7	7.4	86.8	13.2	-2.93	0.76
Small-cap Funds	-1.83	1.50	-9.1	3.0	89.9	10.1	-3.33	0.68
Large-cap Funds	-0.90	0.98	-6.1	3.1	81.2	18.8	-2.29	0.59
Growth Funds	-1.58	1.85	-7.0	5.4	82.0	18.0	-4.26	1.14
Value Funds	-1.23	1.49	-6.2	5.6	86.6	13.4	-2.02	0.61
Low-Expense Funds	-0.83	1.21	-6.1	9.6	82.4	17.6	-1.66	0.57
High-Expense Funds	-1.86	2.07	-6.9	4.8	85.4	14.6	-4.78	1.09

Table V
Cross-Sectional Distribution of The Value Added

Legend to be added

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	5%	95%
All Funds	9.0	24.2	12.0	44.1	28.0	72.0	-1.2	42.2
Small-cap Funds	8.8	16.3	7.7	13.3	16.3	83.7	-0.6	26.5
Large-cap Funds	8.5	26.9	11.4	26.7	34.9	65.1	-2.1	44.1
Growth Funds	10.7	31.6	13.8	68.3	31.9	68.1	-3.8	53.9
Value Funds	9.6	28.1	13.9	57.3	24.4	75.6	-0.9	47.1
Low-Expense Funds	11.5	38.0	12.7	42.3	27.0	73.0	-2.3	45.5
High-Expense Funds	6.0	14.3	11.5	33.8	25.2	74.8	-2.1	24.0

Table VI
Cross-Sectional Distribution of Gross Alpha

Legend to be added

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skew.	Kurt.	Negative Skill	Positive Skill	5%	95%
All Funds	0.77	1.39	1.3	7.6	27.4	72.6	-1.41	2.82
Small-cap Funds	1.55	1.33	5.1	-0.4	12.9	87.1	-0.58	3.82
Large-cap Funds	0.30	0.94	0.6	3.9	37.5	62.5	-1.27	1.75
Growth Funds	0.78	1.85	1.9	11.2	32.0	68.0	-2.09	3.65
Value Funds	0.95	1.63	4.7	5.3	23.1	76.9	-1.27	2.75
Low-Expense Funds	0.49	1.34	3.1	3.8	32.1	67.9	-1.36	2.10
High-Expense Funds	1.12	1.95	2.5	5.2	26.6	73.4	-1.97	4.01

Table VII
Fund Fees and Fund Size

Legend to be added

	Mean		Quantiles 75%-25%		Quantiles 95%-5%	
	Fees	Size	Fees	Size	Fees	Size
All Funds	1.19	216	0.53	522	1.52	2665
Small-cap Funds	1.30	197	0.46	372	1.52	1527
Large-cap Funds	1.08	258	0.49	708	1.39	3716
Growth Funds	1.24	231	0.52	554	1.35	3118
Value Funds	1.12	282	0.52	668	1.48	3787
Low-Expense Funds	0.81	357	0.27	904	0.67	4871
High-Expense Funds	1.66	143	0.34	294	0.85	1183
Correlation Alpha	0.61	-0.60				
Correlation Size	-0.96					

Table VIII
Cross-Sectional Distribution of Net Alpha

Legend to be added

	Moments				Proportions (%)		Quantiles (Ann.)	
	Mean (Ann.)	Volatility (Ann.)	Skewness	Kurtosis	Negative Skill	Positive Skill	5%	95%
All Funds	-0.44	1.44	-4.2	8.7	62.4	37.6	-2.65	1.74
Small-cap Funds	0.24	1.36	-1.3	2.8	44.4	55.6	-1.88	2.43
Large-cap Funds	-0.84	1.04	-6.1	3.9	78.7	21.3	-2.59	0.73
Growth Funds	-0.51	1.89	-2.5	10.7	62.5	37.5	-3.46	2.44
Value Funds	-0.18	1.63	0.7	5.1	56.2	43.8	-2.35	1.80
Low-Expense Funds	-0.27	1.36	-0.1	3.3	58.7	41.3	-2.14	1.48
High-Expense Funds	-0.60	2.02	-2.4	5.7	62.1	37.9	-3.87	2.41