The Cross-Sectional Distribution of Fund Skill Measures

Laurent Barras, Patrick Gagliardini, and Olivier Scaillet∗

This version, October 1, 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 Università della Svizzera italiana (USI Lugano) and the Swiss Finance Institute (SFI), and Scaillet is at the University of Geneva (Geneva Finance Research Institute (GFRI) and the SFI. We thank participants at the McGill/HEC 2018 Spring Workshop, the 2018 SFI Research Days, 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 simple approach for estimating the entire distribution of skill across mutual funds. Our approach is non-parametric—as such, it avoids the challenge of correctly specifying the skill distribution. It also allows for a joint analysis of multiple measures—a key requirement for examining skill. Our empirical analysis reveals that most funds are skilled at detecting profitable trades, but unskilled at resisting capacity constraints. These two skill dimensions exhibit strong heterogeneity both within and across fund groups. In addition, they are strongly correlated. Aggregating them using the value added reveals that 75% of the funds earn profits for a total of 7.8 mio. per year on average.
I Introduction

Over the past 50 years, the academic literature on mutual funds has largely focused on mutual fund performance. For instance, Carhart (1997), Elton et al. (1993), and Jensen (1968) find that the aggregate alpha net of fees and trading costs is negative, while recent studies further show that the majority of funds deliver negative net alphas (e.g., Barras, Scaillet, and Wermers (2010), Harvey and Liu (2018a)). Far less attention has been devoted to the analysis of mutual fund skill.1 Whereas these two notions are often used interchangeably, they differ in important ways. Skill is defined from the viewpoint of funds—it measures their ability to create value through their investment and trading decisions. 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 cross-sectional distribution of mutual fund skill. Our approach is non-parametric—it does not impose any structure on the shape of the distribution. As such, it brings two key advantages for examining skill. First, it avoids the challenge of correctly specifying the skill distribution. For one, theory offers little guidance to specify the mean, dispersion, or asymmetry of skill across funds. Data cannot be used to guide specification either—a simple histogram of the estimated skill measures is plagued by estimation errors and thus biased.

Second, our approach accommodates the analysis of multiple skill measures. This joint analysis is necessary because funds are skilled along several dimensions. In a world with decreasing returns to scale, Berk and Green (2004; BG hereafter) show that funds can be skilled (i) at detecting profitable trading ideas, or (ii) at mitigating the negative impact of capacity constraints. Furthermore, it is natural to aggregate both dimensions to infer the overall skill level of each fund. This step requires an aggregate skill measure, such as the value added proposed by Berk and van Binsbergen (2015).

In addition to its flexibility, our novel approach is simple, widely applicable, and supported by econometric theory. It involves simple manipulations of the fund estimated skill measures. It provides a unified framework that is applicable to all the descriptive statistics of the skill distribution, including its density, cumulative function, quantiles, and moments (e.g., mean, skewness). Finally, it rests on solid theoretical foundations that allow us to infer the asymptotic properties of the different estimators.

Our non-parametric approach departs from standard parametric/Bayesian approaches that require a full parametric specification of the distribution. In the context of skill

---

1 Notable exceptions include Berk and van Binsbergen (2015), Grinblatt and Titman (1989b), Jones and Shanken (2005), and Pastor, Stambaugh, and Taylor (2015).
evaluation, these approaches are less appealing because they are prone to misspecification errors and cannot easily handle the analysis of multiple skill measures. While a joint specification across all measures is subject to the curse of dimensionality, a separate specification is prone to inconsistencies because the skill measures are theoretically linked together. Another limitation of these approaches is that they rely on sophisticated, computer-intensive estimation methods (e.g., Gibbs sampling).

A key challenge in estimating the skill distribution is to control for bias. Given that the true skill measures are unobservable, we can only use as inputs the estimated skill measures. This introduces an Error-in-Variable (EIV) bias that is reminiscent of the well-known EIV bias in the two-pass regression approach (Shanken (1992)). To address this issue, we derive a closed-form expression of the EIV bias that can be easily computed and interpreted. In essence, the bias adjustment consists of removing probability mass from the tails to reduce the likelihood of observing extreme skill levels. We further validate this procedure through an extensive Monte-Carlo analysis.

Our empirical analysis is based on monthly return data for all actively managed US equity funds between January 1979 and December 2015. We examine the skill distribution for the entire population, as well as different groups sorted on investment styles and fund characteristics. Our analysis covers four skill measures. We measure the two skill dimensions using the BG model in which the fund gross alpha $\alpha_{it}$ depends on its lagged size $q_{it-1}$: $\alpha_{it} = a_i - b_i q_{it-1}$. The first skill dimension—the first dollar (fd) alpha $a_i$—determines the fund ability to identify profitable trades. The second dimension—the size coefficient $b_i$—captures the fund ability to resist capacity constraints. The third measure is the value added $\nu a_i = a_i \cdot q_i$, where $a_i$ and $q_i$ respectively denote the average gross alpha and size. This measure has a powerful economic interpretation because it determines the total profit from exploiting the two skill dimensions. The fourth measure is the gross alpha $\alpha_i$, which commonly serves as a benchmark for measuring skill.

To begin, we find that the ability of funds to detect profitable trades is both widespread and economically significant. The fd alpha is positive for 87.9% of the funds and reaches 3.4% per year on average. In other words, most funds are able to detect undervalued stocks and correct for mispricing caused by noise trading (e.g., Stambaugh (2014)). At the same time, only a handful of funds have the ability to resist capacity constraints. The size coefficient is positive for 87.1% of the funds and implies an average 1.5% decrease in annual alpha following a one standard deviation increase in fund size. Taken together, these fund-level results provides strong support to the BG model in which investment skills and capacity constraints represent central features of the mutual fund industry.
The distributions of the two skill dimensions exhibit several common features. First, they reveal a substantial heterogeneity across funds. For instance, the volatility of the size coefficient is as large as its average (1.3% vs 1.5%). Therefore, the standard approach of imposing a constant size coefficient across funds ignores the diverse impact of capacity constraints. Second, both distributions are highly non-normal and positively skewed. This implies that a minority of funds exhibit stellar investment skills—for 10% of the population, the fd alpha is above 7.5% per year. Third, adjusting for the EIV bias is essential for estimating the two skill distributions. A simple histogram based on the estimated coefficients $\hat{a}_i$ and $\hat{b}_i$ is heavily biased—it largely overestimates the probability mass in the tails and completely fails to capture the asymmetry in skill.

Our joint analysis shows that the two skill dimensions are positively correlated. Small cap funds exhibit both higher fd alphas and higher size coefficients than large cap funds. We observe the same pattern for funds sorted on expense ratio—high expense funds are skilled at detecting profitable trades, but unskilled at mitigating capacity constraints. These results help reconcile previous studies that interpret high expenses as a signal of both superior and inferior skill (e.g., Elton et al. (1993), Pastor, Stambaugh, and Taylor (2017)). They also imply that examining each dimension separately is not sufficient for determining the overall skill level.

Next, we turn to the value added $\nu a_i$ which allows us to aggregate the two skill dimensions. Overall, mutual funds earn large profits—the average value added equals 7.8 mio. per year once funds reach their average size. Among the minority of funds that destroy value (30%), around half of them lack trading ideas, while the others grow too large to maintain positive alphas. Comparing the different fund groups, we find that the value added distributions are similar among small and large cap funds. Whereas both groups differ radically along the two skill dimensions, these differences largely offset each other. In contrast, we observe strong differences between groups sorted on expense ratio and turnover—on average, low expense/turnover funds produce an incremental value added of up to 5 mio. per year. Therefore, these funds achieve better combinations of fd alphas and size coefficients and are rewarded by investors via larger money inflows.

Our results further show that the fund value added $\nu a_i$ is relatively close to its optimal value $\nu a_i^\ast$. In the BG model, skilled funds are in scarce supply and can thus choose the amount of money that is actively managed to maximize the value added: $\max_{q} \nu a_i = \nu a_i^\ast = \frac{a_i^2}{4b_i}$. Consistent with this prediction, we find the ratio $\frac{\nu a_i}{\nu a_i^\ast}$ equals 70.4% in the entire population, and increases up to 76.8% among small-cap funds. The biggest gap is observed for high turnover funds where the ratio is equal to 55.0%. Therefore, these funds produce low profits partly because they fail to exploit their skill potential.
Finally, we find that the gross alpha $\alpha_i$ contains limited information about individual fund skill. Our theoretical analysis reveals that the relations between $\alpha_i$ and the three skill measures, $a_i$, $b_i$, $va_i$, depend on the fee structure chosen by the funds. For instance, $\alpha_i$ is uninformative about skill if funds have a preference for managing a larger asset base—in this case, a fund could be more skilled than its peers, and yet deliver the same gross alpha if it trades off lower fees for a larger size. Overall, the empirical evidence documented here is consistent with this fee setting behavior. Therefore, the fund gross alpha is weakly related to the fd alpha, and contains no information about the size coefficient and the value added.

Our work is related to several strands of the literature. Recent papers use parametric/Bayesian approaches to infer the entire distribution of fund performance (e.g., Chen, Cliff, and Zhao (2017), Jones and Shanken (2005), Harvey and Liu (2018a)). Here, we apply a non-parametric approach to examine mutual fund skill. Several studies apply the False Discovery Rate approach to measure the proportions of funds with negative/positive performance (e.g., Avramov, Barras, and Kosowski (2013), Barras, Scaillet, and Wemers (2010), Ferson and Chen (2018)). This paper estimates the entire skill distribution and thus encompasses the analysis of the proportions. Berk and van Binsbergen (2015) and Pastor, Stambaugh, and Taylor (2015) discuss the advantages of using the value added and the fd alpha. We build on these papers to estimate mutual fund skill. Finally, several studies provide evidence of decreasing return to scale at the aggregate level (e.g., Chen et al. (2004), Pastor, Stambaugh, and Taylor (2015)). Here, we examine the impact of capacity constraints at the individual fund level.

The remainder of the paper is as follows. Section II presents the different skill measures and describes our non-parametric approach. Section III describes the mutual fund data set and provides summary statistics. Section IV contains the empirical analysis of the different skill measures, and Section V concludes. The appendix contains the proofs of the different propositions, as well as additional empirical results.

II The Cross-Sectional Distribution of Skill Measures

A Measuring Fund Skill

A.1 Definition of the Skill Measures

The Two Dimensions of Skill We begin our presentation with the two dimensions of skill that funds potentially exhibit, namely (i) their ability to generate profitable trading ideas, and (ii) their ability to mitigate the impact of capacity constraints. To this end,
we build on the BG model and express the gross alpha of each fund $i$ as a linear function of its lagged size $q_{i,t-1}$, i.e., $\alpha_{i,t} = a_i - b_i q_{i,t-1}$.\footnote{This linear specification is used, among others, by Berk and van Binsbergen (2015), Harvey and Liu (2018b), Pastor, Stambaugh, and Taylor (2015), and Pollet and Wilson (2008). In the appendix, we consider alternative specifications which leave our results largely unchanged.}

We measure the first skill dimension using the first dollar (fd) alpha $a_i$, which is defined as the gross alpha on the first invested dollar.\footnote{Here, we follow the terminology proposed by Pastor, Stambaugh, and Taylor (2015).} The fd alpha captures the fund investment skills. In other words, it measures the fund abnormal return unencumbered by the drag of real world implementation (Perold and Salomon (1991)).

We measure the second skill dimension using the size coefficient $b_i$, which captures the fund sensitivity to capacity constraints. As discussed by BG, $b_i$ is larger if the fund faces high execution costs for trading large orders (liquidity, price impact). A distinguishing feature of our specification is that $b_i$ is fund-specific. This contrasts with the previous literature which commonly assumes a common coefficient $\bar{b}$ for all funds (e.g., Chen et al. (2004), Pastor, Stambaugh, and Taylor (2015)). Capturing this heterogeneity is potentially important because there is a priori no reason why the impact of capacity constraints should be identical across funds.

To estimate the two skill dimensions, we use the following time-series regression:

$$r_{i,t} = a_i - b_i q_{i,t-1} + \beta_i f_t + \varepsilon_{i,t},$$

where $r_{i,t}$ is the fund gross excess return (before fees) over the riskfree rate, $f_t$ is a $K$-vector of benchmark excess returns, and $\varepsilon_{i,t}$ is the residual term.\footnote{The estimation of the two skill dimensions requires a benchmark model. In our baseline specification, we use the four-factor model of Cremers, Petajisto, and Zitzewitz (2012). We formally test that this model is correctly specified using the procedure proposed by Gagliardini, Ossola, and Scaillet (2018).} 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 other words, we do not treat the coefficients $a_i$, $b_i$, and $\beta_i$ as fixed parameters, but as random realizations from a continuum of funds. Under this sampling scheme, we can invoke cross-sectional limits in order to infer the common density function $\phi(a)$ and $\phi(b)$ from which the fd alpha and size coefficient of each fund are drawn.\footnote{Gagliardini, Ossola, and Scaillet (2016) use a similar sampling scheme to develop testable applications of the arbitrage pricing theory in a large cross-section of assets.}

**Aggregating the Skill Dimensions**  Our main aggregate skill measure is the value added $va_i$ proposed by Berk and van Binsbergen (2015). This measure has a powerful economic interpretation—similar to the monopolist rent, the value added captures the
total profits from exploiting the two skill dimensions. For one, the maximized value added $va_i^*$ in the BG model is a function of both $a_i$ and $b_i$: $va_i^* = \frac{a_i^2}{2b_i}$.

We also consider the gross alpha $\alpha_i$ because it is commonly used in the literature on mutual fund skill. As discussed below, the gross alpha may provide limited information about the overall skill level because it does not control for the impact of size. Intuitively, using the gross alpha alone is akin to measuring the monopolist rent with the markup price of the goods, regardless of how much quantity is sold.

Formally, we define the value added as the total profits earned by fund $i$ once it reaches its average size $q_i$:

$$va_i = \alpha_i : q_i,$$

where the gross alpha is estimated using the following time-series regression:

$$r_{i,t} = \alpha_i + \beta_i q_i + \varepsilon_{i,t}.$$  \hspace{1cm} (3)

Similar to the other skill measures, we interpret Equation (3) as a random coefficient model so that $va_i$ and $\alpha_i$ are drawn from the common density functions $\phi(va)$ and $\phi(\alpha)$.

### A.2 The Relations between the Skill Measures

To shed light on the relations between the different skill measures, we need to understand how fund size is determined in equilibrium. A natural benchmark is the neoclassical BG model in which skilled managers maximize profit and rational investors compete for performance such that the gross alpha equals fees, i.e., $\alpha_i = f_{i,i}$. Following the notation of Berk and van Binsbergen (2015), we write the (benchmark-adjusted) fund total revenue and cost from active management as $r_i = a_i q_i$ and $c_i = b_i q_i^2$. Using the standard first order condition, we obtain the optimal size $q_i^* = \frac{a_i}{2b_i}$, and the optimal value added $va_i^* = r_i^* - c_i^* = a_i q_i^* - b_i q_i^{*2} = \frac{a_i^2}{2b_i}$. \hspace{1cm} (7)

A key insight from this model is that fund fees do not change the value added: (i) if the fund chooses low fees, it receives additional money from investors ($q_i - q_i^* > 0$) which is passively indexed to keep $r_i^* - c_i^*$ unchanged; (ii) if the fund chooses high fees, it can sell the index short and invest the proceeds in the fund ($q_i^* - q_i > 0$) to obtain

---


7 Our analysis largely builds on that of Berk and van Binsbergen (2015) which already discusses the relation between the gross alpha and the value added. Our contribution is to include all four skill measures (fd alpha, size coefficient, value added, gross alpha) and examine their relations under different compensation schemes.
However, the choice of fees determine the equilibrium size $q_i$ 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.

**Scheme I (fd alpha).** Funds set fees at $f_{e,i}^a$ such that they operate at the profit-maximizing size $q_i^*$. We have $\alpha_i = f_{e,i}^a = \frac{r_i^* - c_i^*}{q_i} = \frac{aq_i^*-b_iq_i^*}{q_i} \div a_i$. Therefore, the gross alpha captures the fd alpha (first skill dimension).

**Scheme II (size coefficient).** Funds set fees at $f_{e,i}^b$ such that they operate at the squared optimal size $q_i^{*2}$. We have $\alpha_i = f_{e,i}^b = \frac{r_i^*-c_i^*}{q_i^{*2}} = \frac{aq_i^*-b_iq_i^{*2}}{q_i^{*2}} = b_i$. Therefore, the gross alpha captures the size coefficient (the second skill dimension).

**Scheme III (value added).** Funds set fees at $f_{e,i}^v$ such that the size remains constant across all funds at $\bar{q}$. The gross alpha is given by $\alpha_i = f_{e,i}^v = \frac{r_i^*-c_i^*}{\bar{q}} = \frac{aq_i^*-b_i\bar{q}^2}{\bar{q}} = \frac{a\bar{q}^2}{\bar{q}^2} \div \nu a_i$. Therefore, the gross alpha captures the value added.

**Scheme IV (no information).** Here, funds choose fees $f_{e,i}^{ra}$ at levels that differ from $f_{e,i}^a$, $f_{e,i}^b$, and $f_{e,i}^{va}$.

We have $\alpha_i = f_{e,i}^{ra} = \frac{r_i^*-c_i^*}{q_i^{ra}} = \frac{aq_i^*-b_iq_i^{ra^2}}{q_i^{ra^2}} = \frac{a^2}{q_i^{ra^2}}$, where $q_i^{ra}(f_{e,i}^{ra}) = \frac{q_i^2}{q_i^{ra^2}}$. Therefore, the gross alpha is uninformative about skill (i.e., $\alpha_i \neq a_i, \alpha_i \neq b_i, \alpha_i \neq \nu a_i$). For instance, suppose that fund A is more skilled than fund B on every dimension, but chooses the same level of fees ($f_{e,i}^{ra_A} = f_{e,i}^{ra_B} = \tilde{f}_e$). A comparison based on the gross alpha fails to capture any skill difference ($\alpha_A = \alpha_B = \tilde{f}_e$).

Table I summarizes the above analysis and reveals that the four compensation schemes yield different predictions for fund size and fees. These predictions can be examined empirically to shed light on the information content of the gross alpha. Under Scheme I, we observe a moderate cross-sectional variation in fees and size as funds choose different fees to reach the fund-specific optimal size. Under Scheme II, the fees must be tiny so that funds grow large and reach the squared optimal size. Under Scheme III, the fund size is the same for all funds. Finally, Scheme IV predicts a cross-sectional variation in size because it corrects for the arbitrary fees set by funds.

Please insert Table I here

---

8 For instance, funds may choose fees to mitigate moral hazard (Habib and Johnson (2016)). In this case, fees are set at the minimum level such that skilled managers exert effort. As fees are driven by the manager-specific probability of shirking, they may be quite different from $f_{e,i}^a$, $f_{e,i}^b$, and $f_{e,i}^{va}$. 7
B Overview of the Non-Parametric Approach

B.1 General Motivation

In this section, we describe the approach for estimating the cross-sectional distribution of fund skill. For simplicity, we denote the skill measure by $\mu$, where $\mu \in \{a, b, va, \alpha\}$ encompasses all four measures presented above. Our approach is non-parametric, i.e., it estimates the skill distribution without imposing any structure on its shape. As such, it provides several key advantages.

First, our approach is immune to misspecification errors. This is not the case for standard Bayesian/parametric approaches because they require to specify the shape of the distribution. In the context of skill, choosing the correct specification is challenging. For one, theory can be used to anchor the distribution of net alpha around zero, but offers no such guidance for skill. Whereas flexibility can be gained by specifying a mixture of normals (e.g., Chen, Cliff, and Zhao (2017), Harvey and Liu (2018a)), determining the number of mixture components is technically difficult.\(^9\)

Second, it allows for a joint analysis of all four skill measures, which is extremely challenging with Bayesian/parametric approaches. On the one hand, specifying each skill distribution separately is likely to generate inconsistencies because the skill measures are theoretically related, i.e., we have $\alpha_i = a_i - b_iq_i = va_i/q_i$. On the other, jointly specifying all skill distributions involves the daunting task of estimating a multivariate distribution whose marginals are potentially mixtures of distributions.

Third, the implementation of the non-parametric approach is simple and fast. Intuitively, it is akin to computing an histogram using as inputs the estimated skill measure $\hat{m}_i$ of each fund. In contrast, Bayesian/parametric approaches require sophisticated and computer-intensive Gibbs sampling and Expectation Maximum (EM) methods (e.g., Chen, Cliff, and Zhao (2017), Jones and Shanken (2005)).

Fourth, it provides a unified framework for estimating the different characterizations of the skill distribution: (i) the density function $\phi(m)$, (ii) the moments (e.g., mean, variance), (iii) the cumulative function $\Phi(x) = \text{prob}[m_i \leq x] = \int_{-\infty}^{x} \phi(m)dm$, and (iv) the distribution quantile $q(p) = \Phi^{-1}(p)$, where $p\%$ denotes the probability level.

Last but not least, it comes with a full-fledged econometric theory. For each estimators that characterize the skill distribution, we derive its asymptotic distribution as the number of funds $n$ and the number of return observations $T$ grow large ($n, T \to \infty$).

\(^9\)Due to their non-regularity, there are many technical challenges concerning inference problems on various aspects of finite mixture models (see the review paper of Chen (2017)). For example, the classical distribution theory of the log likelihood test statistic does not hold for testing for no mixture (homogeneity) against mixture alternatives (Ghosh and Sen (1985)).
B.2 Non-Parametric Estimation

We now explain the main steps of our non-parametric approach. For sake of brevity, we describe the procedure for estimating the skill density $\phi(m)$ and relegate to the appendix the formal treatment of the three remaining measures (moment, cdf, quantiles), as well as the proofs of the different propositions.

To begin, we estimate the regression coefficients in Equation (1) for each of the $n$ funds in the population. The vector of coefficients $\hat{\gamma}_i = (\hat{\alpha}_i, \hat{b}_i, \hat{\beta}_i)'$ for fund $i$ ($i = 1, ..., n$) is computed as

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

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, $x_{i,t} = [1, -q_{i,t-1}, f_{i,t}']'$, is the vector of explanatory variables, and $\hat{Q}_{x,i} = \frac{1}{T_i} \sum_{t=1}^{T_i} I_{i,t} x_{i,t} x_{i,t}'$ is the estimated matrix of the second moments of $x_{i,t}$. We then repeat the procedure for Equation (3) except that $\hat{\gamma}_i = (\hat{\alpha}_i, \hat{\beta}_i)'$ and $x_{i,t} = [1, f_{i,t}']'$. Using the estimated coefficients along with the average size $\bar{q}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} I_{i,t} q_{i,t-1}$, we can then compute each of the four skill measures as

- Fd alpha : $\hat{m}_i = \hat{a}_i$,
- Size coefficient : $\hat{m}_i = \hat{b}_i$,
- Value added : $\hat{m}_i = \hat{\alpha}_i \bar{q}_i$,
- Gross alpha : $\hat{m}_i = \hat{\alpha}_i$.

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 an unreliable estimate $\hat{m}_i$. To address this issue, we follow Gagliardini, Ossola, and Scaillet (2016) and introduce a formal fund selection rule $1_i^X$ equal to one if the following two conditions are met (and zero otherwise):

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

where $CN \left( \hat{Q}_{x,i} \right) = \sqrt{\frac{eig_{max} \left( \hat{Q}_{x,i} \right)}{eig_{min} \left( \hat{Q}_{x,i} \right)}}$ is the condition number of the matrix $\hat{Q}_{x,i}$ defined as the ratio of the largest to smallest eigenvalues $eig_{max}$ and $eig_{min}$, $\tau_{i,T} = T/T_i$ is the inverse of relative sample size $T_i/T$, 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 $CN(Q_{x,i})$ 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 the standard non-parametric approach. The estimated density $\hat{\phi}$ at a given point $m$ is computed as

$$\hat{\phi}(m) = \frac{1}{nh} \sum_{i=1}^{n} 1_{i} K\left(\frac{\bar{m}_{i} - m}{h}\right), \quad (6)$$

where $h$ 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 $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$ 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 $\sqrt{nh}(h^3 + h^2 T + (1/T)^{\frac{1}{2}}) \to 0$, we have

$$\sqrt{nh} \left(\hat{\phi}(m) - \phi(m) - bs(m)\right) \Rightarrow N(0, K_1 \phi(m)), \quad (7)$$

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

$$bs_1(m) = \frac{1}{2} h^2 K_2 \phi^{(2)}(m), \quad (8)$$

$$bs_2(m) = \frac{1}{2T} \psi^{(2)}(m), \quad (9)$$

where $K_1 = \int K(u)^2 du$, $K_2 = \int u^2 K(u) du$, $\phi^{(2)}(m)$ 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 asymptotic variance of the estimated

---

10Similar to Equation (7), Okui and Yanagi (2018) consider a kernel estimator for the density of the mean and autocorrelation of random variables. However, their distributional results differ from those derived in our regression context.
centered skill measure $\sqrt{T} (\tilde{m}_i - m_i)$ equal to \( \text{plim}_{T \to \infty} \left( \frac{1}{T} \sum_{t,s} I_{i,t} I_{i,s} u_{i,t} u_{i,s} \right) \). For each skill measure, the term \( u_{i,t} \) is given by

\begin{align*}
\text{Fd alpha} & : u_{i,t} = e_1' Q_{x,t}^{-1} x_{i,t} \varepsilon_{i,t}, \\
\text{Size coefficient} & : u_{i,t} = e_2' Q_{x,t}^{-1} x_{i,t} \varepsilon_{i,t}, \\
\text{Value added} & : u_{i,t} = \alpha_i (q_{i,t-1} - \tilde{q}_i) + \tilde{q}_i e_1' Q_x^{-1} x_t \varepsilon_{i,t}, \quad (10) \\
\text{Gross alpha} & : u_{i,t} = e_1' Q^{-1}_x x_t \varepsilon_{i,t}, \quad (11)
\end{align*}

where \( e_1 (e_2) \) is a vector with one in the first (second) position and zeros elsewhere while \( Q_x = E[x_i x'_i] \) and \( Q_{x,t} = E[x_{i,t} x'_{i,t}] \). Under a Gaussian kernel, the two constants \( K_1 \) and \( K_2 \) are equal to \( \frac{1}{\sqrt{\pi}} \) and 1, respectively.

**Proof.** See the appendix. □

Proposition II.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 \( \phi(m) \). Therefore, we can 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 smoothing bias, which is standard in non-parametric density estimation (e.g., Silverman (1986), Wand and Jones (1995)). The second component \( bs_2 \), which is referred to as the error-in-variable (EIV) bias, is non-standard—it arises because we estimate \( \phi \) using the estimated skill measure instead of the true one (i.e., \( \hat{m}_i \) instead of \( m_i \)).

Finally, Proposition II.1 provides guidelines for the choice of the bandwidth. We show in the appendix that the choice of the optimal bandwidth \( h^* \)—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 \( n \) \( \left( n^{2/5}/T \to \infty \right) \), \( h^* \) is proportional to \( (nT)^{-\frac{1}{5}} \); (ii) if \( T \) is not small relative to \( n \) \( \left( n^{2/5}/T \to 0 \right) \), \( h^* \) is proportional to \( n^{-\frac{1}{7}} \). 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.

\[ \text{AMISE} \text{ is defined as the integrated sum of the leading terms of the asymptotic variance and squared bias of the estimated density } \hat{\phi}(m). \]
Motivated by these results, we use the following bandwidth in our baseline specification:

\[ h^* = \left( \frac{K_2}{K_1} \int \phi^{(2)}(m)\psi^{(2)}(m) \, dm \right)^{-\frac{1}{2}} (n/T)^{-\frac{1}{5}}. \] (12)

B.3 Adjusting for the Bias

Building on the insights of Proposition II.1, we show how to compute the bias-adjusted density estimator. Our approach consists of estimating the two bias terms \( b_{s_1}(m) \), \( b_{s_2}(m) \), 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: \( m_i \sim N(\mu_m, \sigma_m) \), \( s_i \sim N(\mu_s, \sigma_s) \), and \( corr(m_i, s_i) = \rho \).\(^{12}\)

This simple reference model has several appealing properties. First, the estimators of the bias and the bandwidth are simple to compute because they are all available in closed form. Second, they are precisely estimated because they only depend on the five parameters of the normal distribution. Finally, we can perform a comparative static analysis on the two bias components, i.e., we can determine how their shapes change with the different parameters.

These benefits are not shared by the alternative approach in which the bias terms are directly inferred from Equations (9)-(10) via a non-parametric estimation of the second-order 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)).\(^{13}\)

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 \( \sqrt{nh}(h^3 + h^2T + \ldots) \to 0 \) and \( nh^{-\frac{1}{5}} \to 0 \),

\[^{12}\text{A Gaussian reference model underlies the celebrated Silverman rule of thumb for the choice of the bandwidth in standard non-parametric density estimation without the EIV problem. This rule gives } h^* = 1.06\sigma n^{-\frac{1}{5}}, \text{ where } \sigma \text{ is the standard deviation of the observations (Silverman (1986)).} \]

\[^{13}\text{We can estimate the } r\text{th-derivative of a density } \phi \text{ by kernel smoothing (Bhattacharya (1967)). The rate of consistency of the derivative estimator equals } (nh^{2r+1})^{-\frac{1}{2}} \text{ and is much slower than the rate } (nh)^{-\frac{1}{5}} \text{ for the density estimator. In other words, the higher-order derivatives are imprecisely estimated because the rate of consistency decreases with the derivative order } r. \]
\[(1/T)^{\frac{3}{2}} \rightarrow 0, \text{ the two bias components under the reference model are equal to} \]

\[bs_1^m(m) = \left[ \frac{1}{2} K_2 h^2 \frac{1}{\sigma_m} (\bar{m}_1^2 - 1) \right] \frac{1}{\sigma_m} \varphi(\bar{m}_1), \quad (13)\]

\[bs_2^m(m) = \left[ \frac{1}{2T} \exp \left( \mu_s + \frac{1}{2} \sigma_s^2 \right) \frac{1}{\sigma_m} (\bar{m}_2^2 - 1) \right] \frac{1}{\sigma_m} \varphi(\bar{m}_2), \quad (14)\]

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

\[h^* = \left[ \frac{K_2}{K_1 2 \sqrt{\pi}} \frac{3}{4 \sigma_m^3} \left( \frac{\rho^4 \sigma_s^4}{12} - \rho^2 \sigma_s^2 + 1 \right) \exp \left( \mu_s + \frac{1}{2} \sigma_s^2 (1 - \frac{\rho^2}{2}) \right) \right]^{-\frac{1}{3}} (nT)^{-\frac{1}{6}}. \quad (15)\]

**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 \(\hat{\mu}_1\) and \(\hat{\sigma}_1(m)\). To compute \(\hat{\sigma}_1(m) = \log(\hat{\Sigma})\), we use the standard variance estimator of Newey and West (1987):

\[\hat{\Sigma} = \left[ \frac{T}{T} \sum_{t=1}^{T} I_{i,t} \hat{u}_{i,t}^2 + 2 \sum_{l=1}^{L} \left( 1 - \frac{l}{L+1} \right) \left[ \frac{T}{T} \sum_{t=1}^{T-l} I_{i,t+l} \hat{u}_{i,t+l} \hat{u}_{i,t+l}^T \right] \right], \quad (16)\]

Second, we insert these estimated moments in Equations (13)-(15) to compute the bias terms \(\hat{b}_1^s(m), \hat{b}_2^s(m)\) 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

\[\hat{\phi}^*(m) = \hat{\phi}(m) - \hat{b}_1^s(m) - \hat{b}_2^s(m). \quad (17)\]

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.\(^{15}\)

\(^{14}\) In the baseline specification, we set \(L = 3\).

\(^{15}\) Our results resonate with those reported by Silverman (1986) for the standard non-parametric density estimation without the EIV problem. He shows that the rule of thumb for the bandwidth...
C Analysis of the Density Bias

C.1 The Shape of the two Bias Components

To shed light on the bias adjustment mechanism, we now study the shape of the two bias components in Equations (13)-(14). A key feature of the smoothing bias $b_{d1}(m)$ is that it depends on the total number of funds $n$. As $n$ increases, $h^*$ shrinks towards zero, which reduces the magnitude of $b_{d1}(m)$. With a population of several thousand funds, the contribution of $b_{d1}(m)$ thus becomes negligible. In contrast, the EIV bias $b_{d2}(m)$ depends on the number of observations $T$ because it arises from the gap between $\hat{m}_i$ and $m_i$. Therefore, it can remain significant even if the fund population is large.

The sign of the EIV bias is driven by the term $(\hat{m}_2^2 - 1)$ which is negative when $m$ is close to the average $\mu_m$ and positive otherwise. As a result, $b_{d2}(m)$ is negative in the center of the distribution and positive in the tails. The intuition for this result is that the estimated skill measure is a noisy version of the true one ($\hat{m}_i = m_i +$ estimation noise). Therefore, the density obtained with $\hat{m}_i$ overestimates the probability of observing extreme skill levels, i.e., the unadjusted density $\tilde{\phi}(m)$ is too flat.

Importantly, the shape of the EIV bias obtained with the reference model is quite general. As shown in Proposition II.1, the true bias $b_{d2}(m)$ is a function of the second-order derivative of the true density $\phi^{(2)}(m)$. As long as this density peaks around its mean, $\phi^{(2)}(m)$ takes negative (positive) values in the center (tails) of the distribution—just like the bias obtained with the reference model.\(^{16}\)

To illustrate, Figure 1 plots the two bias components $b_{d1}(m)$ and $b_{d2}(m)$ for the gross alpha obtained with our sample ($n = 2,504$). The estimated values for the mean $\mu_m$ and volatility $\sigma_m$ of the true gross alpha distribution are equal to 0.06% and 0.18% per month. The mean $\mu_s$ and volatility $\sigma_s$ for the log of the asymptotic variance are equal to -7.3 and 1.1. These numbers yield an average asymptotic variance $\mu_S = \exp(\mu_s + 0.5\sigma_s^2)$ of 0.0012, which translates into a monthly volatility for $\hat{\sigma}_i$ equal to 0.26% per month.\(^{17}\) Finally, we obtain a value of 0.04 for the correlation $\rho$.

There are two main insights from Figure 1. First, the smoothing bias is close to zero. In other words, the bias is entirely driven by the EIV bias. Second, the impact of the EIV bias is large. Integrating $b_{d2}(m)$ over the interval $[\mu_m - \sigma_m, \mu_m + \sigma_m]$, we find that choice, which relies on a Gaussian reference model, is quite robust to departures from normality.

\(^{16}\)The only case where $b_{d2}(m)$ differs from $b_{d2}(m)$ is if the true density $\phi(m)$ is a mixture of distributions whose components have means extremely far away from one another such that we have a trough instead of a peak at the mean of $\phi(m)$.

\(^{17}\)We have $\sigma_{\hat{\sigma}_i} \approx \sqrt{\frac{2\pi}{T}} = 0.26\%$ when replacing $T$ with the median number of observations (173).
the EIV bias represents a probability mass of 25% in the center of the distribution.

Please insert Figure 1 here

C.2 Comparative Statics for the EIV Bias

We can also use the reference model to perform a comparative static analysis on the EIV bias. There are three key parameters that determine the shape of $bs_2^*(m)$: (i) the volatility $\sigma_m$ of the skill measure, (ii) the average level of the estimation variance $\mu_S$, and (iii) the correlation $\rho$ between the skill measure and the estimation variance.

A higher value for $\sigma_m$ reduces the magnitude of $bs_2^*(m)$ because it makes the cross-sectional variation in the estimated measure more aligned with that of the true one (i.e., the relative importance of $m_i$ over noise increases). A higher value for $\mu_S$ increases the magnitude of $bs_2^*(m)$ because the estimated skill measure becomes more volatile (i.e., the relative importance of noise over $m_i$ increases). Finally, a higher value for $\rho$ moves $bs_2^*(m)$ to the right because funds with higher skill are also more likely to exhibit higher estimation variance.

To illustrate, Figure 2 examines the magnitude of these changes for the gross alpha. Consistent with the previous analysis, an increase in $\sigma_m$ leads to a sharp decrease in the magnitude of $bs_2^*(\alpha)$, while an increase in $\mu_S$ has the opposite effect. Finally, an increase in $\rho$ shifts $bs_2^*(\alpha)$ to the right, i.e., the minimum value of $bs_2^*(m)$ is obtained at $m = \mu_m + \rho \sigma_m \sigma_s > \mu_m$.

Please insert Figure 2 here

D Extensions

D.1 Net Alpha

We conclude our presentation with several useful extensions. We can evaluate mutual fund performance using the net alpha $\alpha_i^n$, which measures the abnormal average return earned by investors. Similar to the gross alpha, our non-parametric approach provides a simple, robust, and fast procedure to infer the cross-sectional density of the net alpha $\phi(\alpha^n)$. The estimation procedure remains exactly the same as for the gross alpha after replacing Equation (3) with

$$r_{i,t}^n = r_{i,t} - f e_{i,t} = \alpha_i^n + \beta_i f_t + \epsilon_{i,t},$$

where $r_{i,t}^n$ is the difference between the fund gross excess return and fees.
D.2 Time-Varying Skill

We can also use our approach to examine the time-variation in individual fund returns. This analysis provides an extension of recent studies that examine how 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 Equation (1) as:

$$r_{i,t} = \alpha_{i,t-1} + \beta_l f_t + \varepsilon_{i,t} = \bar{a}_i + \sum_{l=1}^{L} b_{i,l} z_{i,t-1} + \beta_l f_t + \varepsilon_{i,t},$$

(19)

where $z_{i,t-1}$ 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.

D.3 Factor Beta

Finally, we can apply our approach to study the fund betas on the different risk factors. To compute the distribution $\phi(\beta_k)$ associated with each risk factor $k$ ($k = 1, ..., K$), we simply need to replace $\hat{m}_i$ with the estimated beta $\hat{\beta}_{i,k}$ and to modify $u_{i,t}$ accordingly.

III Data Description

A Mutual Fund Data

We conduct our analysis on the entire population of actively managed US equity funds. We collect monthly data on net returns and size, as well as annual data on fees, turnover, and investment objectives from the CRSP database. We measure the monthly gross return $r_{i,t}$ as the sum of the fund monthly net return and fees. The net return $r^m_{i,t}$ is computed as a value-weighted average of the net returns across all shareclasses using their beginning-of-month total net asset values. The monthly fees $f e_{i,t}$ are defined as the value-weighted average of the most recently reported annual fees across shareclasses divided by 12. We measure the fund size $q_{i,t-1}$, by taking the sum of the beginning-of-month net asset values across all shareclasses. Following Berk and van Binsbergen (2015), we adjust size for inflation by expressing all numbers in January 1, 2000 dollars.

Our initial sample includes all funds with a valid equity investment objective from Lipper, Strategic Insight, Wiesenberger, or CRSP. We exclude index funds and eliminate return observations of tiny funds by imposing a minimum size of $15$ million (see Chen...
et al. (2004), Pastor, Stambaugh, and Taylor (2015)). To apply the fund selection rules in Equation (6), we follow Gagliardini, Ossola, and Scaillet (2016) and select funds for which the condition number (CN) of the matrix $\hat{Q}_{x,i}$ is below 15 and the number of monthly observations $T_i$ is above 60.\textsuperscript{18} These selection criteria produce a final universe of more than 2,000 funds.

We further group funds according to their investment styles (small cap, large cap, growth, and value). At the start of each month, we classify each fund using the style information provided by Lipper. If this information is missing, we use the investment objectives reported by Strategic Insight, Wiesenberger, and CRSP (see the appendix for additional details). A fund is included in a given group if its style corresponds to that of the group for a sufficiently long period such that the two selection rules are satisfied (i.e., CN≤15 and $T_i \geq 60$). We follow the same procedure to form terciles of funds based on their characteristics (expense ratios and turnover), where the monthly turnover is defined as the most recently observed ratio of \text{min}(\text{buys}, \text{sells}) on fund size (see Pastor, Stambaugh, and Taylor (2017)).

B Benchmark Models

To estimate the skill measures in Equations (1)-(3), we primarily use the benchmark model of Cremers, Petajisto, and Zitzewitz (2012; CPZ hereafter). Similar to the model of Carhart (1997), the CPZ model includes four factors, i.e., $f_t$ is defined as $[r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}]'$, 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. However, the CPZ model departs from the Carhart model in two respects: (i) $r_{m,t}$ is proxied by the excess return of the S&P500 index (instead of the CRSP market index), and (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 S&P500, 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 S&P500 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 instance, the Russell 2000 produces a Carhart alpha of -2.41% per year over the period 1980-2005. This implies that small-cap managers are likely be classified as unskilled if

\textsuperscript{18}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. To mitigate the impact of outliers on the estimated parameters in the reference model, we also exclude the values for $\hat{m}_i$ and $\hat{s}_i$ that are four standard deviation away from the mean.
they use the Russell 2000 as benchmark. Motivated by these results, we use the CPZ model in our baseline specification, and later repeat the analysis with the Carhart model.

C Summary Statistics

Table II reports summary statistics for value weighted portfolios of funds for the entire population and the different groups. Our sample period starts in January 1979 and ends in December 2015 (for a total of 444 monthly return observations).\textsuperscript{19} 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 exhibits 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 and high turnover funds tilt toward small cap, growth stocks.

Please insert Table II here

IV Empirical Results

A The Two Dimensions of Skill

A.1 The First Dollar Alpha

We begin our empirical analysis with the fd alpha $\alpha$, which measures the fund ability to detect profitable trades. After estimating $\alpha_i$ for each fund using Equation (1), we apply our non-parametric approach described in Section II.B to infer the cross-sectional distribution $\phi(a)$. To describe the properties of the distribution, we compute the bias-adjusted (i) moments (mean, variance, skewness, kurtosis), (ii) proportions of funds with negative and positive fd alphas, $\pi^-_\alpha$ and $\pi^+_\alpha$, and (iii) distribution quantiles at 10\% and 90\%, $q(10)$ and $q(90)$.

Table III reveals that the fund investment skills are both economically large and widespread—on average, the fd alpha equals 3.4\% per year and around 90\% of the funds in the population are skilled. These results resonate with the calibration exercise in BG\textsuperscript{19}\textsuperscript{19}The starting date correspond to the first month when the factors in the CPZ model are available.
that yields a value for $\pi_+^a$ around 80%. Another striking feature of the population is the wide heterogeneity in skill as captured by the 7.8%-difference between $q(10)$ and $q(90)$. This heterogeneity is mostly visible in the right tail—$\phi^* (a)$ is heavily skewed (6.4) which implies that a minority of funds exhibit stellar investment skills. An open question is whether these funds diversify their bets across a wide range of stocks or instead hold a concentrated portfolio. Consistent with the second interpretation, funds with higher fd alphas also have higher idiosyncratic risk (i.e., the correlation $\rho$ equals 0.26).

Adjusting for the EIV bias is essential to infer the true distribution $\phi(a)$. As discussed in Section II.C, this adjustment requires that we remove the probability mass from the tails, and transfer it onto the center—slightly to the right because $\rho$ is positive (i.e., $\mu_m + \rho \sigma_m \sigma_s > \mu_m$). Both changes are economically significant and explain why all the estimators directly obtained from $\hat{\alpha}_i$ are biased. For one, the unadjusted distribution $\hat{\phi}(a)$ produces an implausibly large difference of 11.2% between $q(10)$ and $q(90)$, and fails to capture the asymmetry in skill (the skewness is a mere 0.4).

Whereas the skill distributions are similar for growth and value funds, we document sharp differences between the small and large cap groups (Table III and Figure 4). Small cap funds produce a higher average annual fd alpha (5.0% vs 2.0%) and a higher skill proportion $\pi_+^a$ (96.0% vs 81.9%). There are two potential explanations for this skill gap. First, small cap stocks have higher idiosyncratic volatility and thus a larger flow of company specific information. Therefore, they provide more opportunities for stock picking activities (e.g., Duan, Yu, and McLean (2009)). Second, their mispricing is more persistent because there is less competitive pressure among funds.20

Finally, we measure skill among funds with different expense ratios and turnover. This analysis is motivated by the previous literature that commonly uses both characteristics as predictors of mutual fund performance.21 Table III show that high expense funds unambiguously dominate low expense funds, i.e., $\hat{\phi}^* (a)$ exhibits both a higher average (4.6% vs 2.0%) and a fatter right tail. In contrast, trading activity is weakly related to the fd alpha as the proportions $\pi_+^a$ and $\pi_+^+ \alpha$ remain largely unchanged. This result suggests that some funds in the population are unskilled ($\pi_+^a =12.1\%$) not because their managers are overconfident and trade excessively—instead, they may simply take active positions to hide their lack of skill from investors (Berk and van Binsbergen (2018)).

20 Small cap stocks are prone to idiosyncratic fluctuations and asymmetric information problems and are thus largely untouched by mutual funds (e.g., Grompers and Metrick (2001), Hong, Lim, and Stein (2000)).
21 An non-exhaustive list include Elton et al. (93)), Gil-Bazo and Verdu (2009), Grinblatt and Titman (1989b), Pastor, Stambaugh, and Taylor (2015), and Wermers (2000).
A.2 The Size Coefficient

We now turn to the analysis of the size coefficient $b$, which measures the fund sensitivity to capacity constraints. To ease interpretation, we compute the standardized size coefficient $\hat{b}_i$ of each fund in Equation (1) so that it corresponds to the annual change in gross alpha for a one standard deviation change in fund size. Because we estimate the entire cross-sectional distribution $\phi(b)$, we contribute to the previous literature that focuses on the average size coefficient $\bar{b}$.

Table IV shows that the overwhelming majority of funds are subject to capacity constraints. We find that close to 90% of the funds in the population have a positive size coefficient. The magnitude of the size coefficient is typically large—on average, a one standard deviation increase in fund size reduces the gross alpha by 1.5% per year.\(^{22}\) These results provide strong support to previous theoretical work that emphasizes the importance of capacity constraints for the fund industry (e.g., BG, Pastor and Strambaugh (2012)).

Similar to the fd alpha, we observe a strong heterogeneity in the size coefficient—the difference between $q(90)$ and $q(10)$ reaches 3.4% per year. Therefore, assuming a constant $\bar{b}$ provides a poor summary of the diverse impact of capacity constraints across funds. Harvey and Liu (2018b) reach a similar conclusion using a parametric approach in which $\phi(b)$ is assumed to be normal. Our non-parametric approach suggests that this assumption is too stringent because $\hat{\phi}^*(b)$ is heavily skewed (8.0). In other words, a normal distribution fails to capture the asymmetric nature of capacity constraints—wheras we expect $b_i$ to be close to zero for unconstrained funds, it can rise dramatically for those subject to capacity constraints ($q(90)$ equals 3.3%).

Finally, the impact of capacity constraints varies significantly across groups (Table IV and Figure 4). The density $\hat{\phi}^*(b)$ for small cap funds exhibits a higher average (2.0% vs 1.0%) and a fatter right tail than that of large cap funds. This finding resonates with previous studies that establish a link between the tightness of capacity constraints

\(^{22}\text{This estimate is similar to that reported in the recent study by Harvey and Liu (2018b). They find that a }$100 \text{ mio. increase in size reduces the gross alpha by 0.17% per year. In our case, the implied reduction is equal to } \frac{\bar{b}}{\bar{\sigma}} \times 100 \text{ mio} = \frac{0.23}{0.15} = 0.23\% \text{ per year, where } \bar{b}, \bar{\sigma} \text{ are the average size coefficient and standard deviation of fund size equal to 1.5% and }$650 \text{ mio., respectively.}
and the illiquidity of small-cap stocks (e.g., Chen et al. (2004), Yan (2004)). We also document a similar pattern for groups sorted on characteristics—funds with high expense ratios and high turnover tend to follow strategies that are difficult to scale up. Part of this difficulty is due to the fact that these funds tilt their portfolios towards small cap stocks (see Table II).

Please insert Table IV here

Please insert Figure 4 here

A.3 The Correlation between the two Skill Dimensions

An important insight from our joint analysis of the two skill dimensions is that they are positively correlated. A clear example is provided by the two size groups. We find that small cap funds exhibit both higher alpha alphas and higher size coefficients than large cap funds. We observe the same empirical regularity among high and low expense funds.

The implications of this positive correlation are twofold. First and foremost, we need to combine the two skill dimensions to determine the overall skill level. For instance, large cap funds may well dominate small cap funds if they are more able to scale up their less profitable trading ideas. To address this issue, we need to compute the value added as it allows us to aggregate skill into a single measure.

Second, it helps us reconcile the literature on the predictive content of expense ratios. On the one hand, some studies argue that high expense ratios signal superior skill (e.g., Kacperczyk, van Nieuwerburgh, and Veldkamp (2014), Pastor, Stambaugh, and Taylor (2017)). On the other hand, previous work provides empirical evidence that favors the opposite interpretation (e.g., Elton et al. (93), Gil-Bazo and Verdu (2009)). Our results show that both conclusions hold depending on which skill dimension is examined—high expense funds are skilled at generating profitable ideas, but unskilled at resisting capacity constraints.

To check that our results are not driven by the omitted exposure of small cap funds to aggregate liquidity risk, we use an alternative benchmark model that includes the traded liquidity factor of Pastor and Stambaugh (2003). Our results reported in the appendix remain unchanged.

In the entire population, a simple correlation calculation based on the estimated coefficients $\hat{a}_i$ and $\hat{b}_i$ yields a coefficient of 0.84.
B The Value Added

B.1 Aggregating the Two Skill Dimensions

We now examine the value added \( va \), which measures the total fund profits. The value added depends on both skill dimensions and thus provides an economically-motivated approach for aggregating them. The summary statistics for the bias-adjusted density \( \hat{\phi} (va) \) are reported in Table V and use as inputs the estimated \( \hat{va} \) of each fund obtained from Equation (2).

Overall, mutual funds earn large profits through their investment and trading decisions. The average value added in the entire population equals 7.8 mio. per year and more than 70% of the funds are able to create value. This proportion is lower than the one obtained with the fd alpha \( \pi_a^+ > \pi_{va}^+ \) because some funds grow too large to maintain a positive gross alpha. Using this information, we can therefore split the set of funds with negative value added in two groups: half of them have poor investment ideas \( (\pi_a - \pi_{va}) = 12.1 \div 24.7 = 49\% \), while the other half are initially skilled but become too large \( (\pi_{va} - \pi_a) = 12.6 \div 24.7 = 51\% \).

The average value added is higher than the one reported by BvB (3.2 mio. per year) because of methodological differences. BvB measure \( va \) as \( E(\alpha_{i,t}q_{i,t-1}) \) to assess the profits earned by the fund each period. In contrast, we measure \( va \) as \( \alpha_{i,t}E(q_{i,t-1}) \) to assess the profits once the fund reaches its average size. The difference, which is equal to the covariance between \( q_{i,t-1} \) and \( \alpha_{i,t} \), is strongly negative in the presence of capacity constraints.25

The distribution \( \hat{\phi} (va) \) maintains the positive skewness observed for the two skill dimensions. As a result, funds in the right tail reap staggering profits from exploiting their skills, i.e., \( q(90) \) reaches 28.2 mio. per year. However, the positive correlation between the two skill dimensions largely reduces the heterogeneity in the value added. To see this point, suppose that we draw \( a_i \) and \( b_i \) independently from their respective distributions \( \hat{\phi} (a) \) and \( \hat{\phi} (b) \), and then compute the implied optimal value added as \( \frac{a_i^2}{b_i} \). The resulting volatility of the value added equals to 80 mio., which is around 6 times larger than the number reported in Table V.

We find that the different style groups produce similar levels of value added (Table V and Figure 5). This result is particularly striking for small cap and large cap funds given their strong differences along the two skill dimensions. For these two groups, the

25To evaluate the magnitude of the covariance term, we can replace it with its upper bound: \( \text{cov}(\alpha_{i,t}, q_{i,t-1}) = \rho_{\alpha,q}\sigma_{\alpha}\sigma_{q_i} < \sigma_{\alpha}\sigma_{q_i} = b_i\sigma_{q_i}^2 \), where \( \rho_{\alpha,q} \) is the correlation between the gross alpha and the size, and \( \sigma_{\alpha}, \sigma_{q_i} \) denote their respective volatilities. In our sample, the average value of \( b_i\sigma_{q_i}^2 \) amounts to 5.2 mio. per year.
effects of the fd alpha and the size coefficient offset each other, leaving the value added distribution $\hat{\phi}^*(va)$ largely unchanged.

A clear pattern also emerges for turnover-sorted funds. We find that high turnover funds creates significantly less value than low turnover funds. In addition to their lower average (6.6 mio. vs 11.7 mio.), the high turnover group contains a higher proportion of funds that destroy value ($\pi_{va} = 23\%$ vs 10\%). For these two groups, Figure 5 reveals strong departures from normality—while $\hat{\phi}^*(va)$ is heavily peaked for high turnover funds, the opposite holds for low turnover funds. This observation emphasizes the benefit of using a non-parametric approach for measuring skill in the cross-section.

Finally, the comparison of groups sorted on expense ratios is more subtle. Overall, low expense funds dominate high expense funds—both the average and $q(90)$ are around two times larger (9.4 mio. vs 5.4 mio. and 41.5 mio vs 19.5). However, funds in the left tail destroy more value in the low expense group (i.e., $q(5)$ is equal to -2.1 vs -1.1). This negative outcome points toward two sources of inefficiencies: (i) investors allocate too much money to these funds, (ii) which then fail to eliminate the negative impact of this additional money on the value added.

Please insert Table V here

Please insert Figure 6 here

B.2 Is the Value Added Optimized?

A key prediction of the neoclassical BG model is that skilled funds are in scarce supply and thus able to maximize the profits they earn. In practice, the value added may differ from its optimal value for several reasons. An intuitive one is the presence of learning effects. Because investors do not observe the fund skill dimensions $a_i$ and $b_i$, they must learn about them using past data (Pastor and Stambaugh (2012)). During this learning phase, they may therefore invest less than the optimal amount in specific funds. Another possibility is that some managers simply fail to correctly evaluate the impact of each skill dimension on the value added.

To address this issue, we use our non-parametric approach to estimate the distribution of the optimal value added $\phi(va^*)$. Using the estimated coefficients $\hat{a}_i$ and $\hat{b}_i$ from Equation (2), we compute the estimated optimal value added for each fund $i$ as

$$\hat{m}_i = \hat{va}_i^* = \frac{\hat{a}_i^2}{4\hat{b}_i}.$$ (20)
and apply the delta method to obtain the error term $u_{i,t}$:

$$u_{i,t} = \frac{2a_i}{4b_i} e_1^i Q_{z_i}^{-1} x_{i,t} \varepsilon_{i,t} - \frac{a_i^2}{4b_i^2} e_2^i Q_{z_i}^{-1} x_{i,t} \varepsilon_{i,t}. \tag{21}$$

Then, we apply the procedure in Section II.B to compute the bias-adjusted density $\hat{\phi}^*(\nu \alpha^*)$. The equilibrium value added requires both $\hat{\alpha}$ and $\hat{\beta}$ to be positive. Therefore, we conduct our analysis on funds for which this condition is satisfied (75% of the sample).

The empirical evidence in Table VI is broadly supportive of the predictions of the model. The average values for the actual and optimal profits are equal to 8.0 mio. and 11.3 mio. per year. This implies that funds extract 70% of the profit they would produce if they fully optimize.\textsuperscript{26} Consistent with our previous results, we find that low expense and low turnover funds produce the highest optimal value added (the average equal to 15.0 and 17.5 mio.). In other words, both groups achieve the most profitable combinations of the two skill dimensions. The biggest gap (in relative terms) between actual and optimal value added is observed among high-turnover funds. Therefore, one reason why these funds achieve low profits is that they only exploit 55.0% of their skill potential.

Please insert Table VI here.

\section*{C The Gross Alpha}

\subsection*{C.1 The Information Content of the Gross Alpha}

In this section, we examine the skill information contained in the gross alpha $\alpha$. Our previous analysis in Section II.A shows that this information depends on the compensation scheme chosen by the funds. To shed light on this choice, we compute summary statistics for fees and size and compare them with the predictions shown in Table I.

The overall evidence suggests that the gross alpha contains limited information about the fund skill levels. Consistent with scheme IV (no information), Table VII shows that fees and size are negatively correlated (-0.19 across funds and -0.96 across fund groups). In other words, some funds are willing to charge low fees in order to manage a large asset base and, possibly, mitigate several institutional constraints (see Habib and Johnson (2016)).\textsuperscript{27} In addition, funds do not set fees according to the schemes II (size

\textsuperscript{26}In addition, the correlation between the estimated actual and optimal value added equals (0.94), which suggests that funds with the highest skill potential do create more value.

\textsuperscript{27}For instance, the Investment Company Act imposes diversification rules on 75% of the portfolio that 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...
coefficient) and III (value added)—the fees are not tiny and the size is not constant across funds. Therefore, the gross alpha is unrelated to the size coefficient and the value added (the pairwise correlations are equal to 0.07 and 0.05, respectively).

Only scheme I (fd alpha) under which the gross and fd alphas are related is partly supported by the data, as both fees and size vary across funds. Consistent with this analysis, the distribution of the gross alpha $\hat{\phi}^*(\alpha)$ examined in Table VIII bears some similarities with that of the fd alpha. However, the large difference between fund groups are largely muted compared to those reported in Table III.

Please insert Table VII here

Please insert Table VIII here

C.2 From Gross to Net Alphas

Finally, we apply the non-parametric approach to the net alpha $\alpha^n$. This performance analysis reported in Table IX determines whether investors receive any surplus alphas and provides information about the bargaining power of funds when setting fees.

The neoclassical BG model predicts that the net alphas equal zero as skilled funds are able to extract all the profits they generate (i.e., $\alpha_i = f_{c, i}$). Consistent with this prediction, the proportion of funds with positive alphas drops from 72.6% to 35.4% as move from gross to net alphas.\textsuperscript{28,29} In addition, fund groups with higher gross alphas also tend to charge higher fees (the pairwise correlation equals 0.57).

However, we find that the left tail of the net alpha distribution does not shrink towards zero—$\hat{\hat{\phi}}^*(\alpha^n)$ is essentially a shifted version of $\hat{\phi}(\alpha)$ with a similar volatility and a fatter left tail. In other words, some funds are able to charge excessively high fees to investors (Christoffersen and Musto (2002), Gruber (1996)). The behaviour of these apparently irrational investors drives a wedge between gross alphas and fees that is left unexplained by the neoclassical model. The same behavior also explains why some unskilled funds are allowed to sufficiently large and destroy a sizeable amount of value (Table V).

\textsuperscript{28}Both proportions are significantly larger than those obtained with the False Discovery Rate approach (Barras, Scaillet, and Wermers (2010)). Intuitively, this approach estimates $\pi^+_{\alpha}$ and $\pi^+_{\alpha^n}$ by counting funds with large alpha t-statistics and thus does not detect all the funds with positive gross/net alphas (see Barras (2018)). In contrast, our non-parametric approach directly estimates the gross alpha distribution using information from the entire cross-section of funds.

\textsuperscript{29}The fact that some investors receive positive net alphas can be rationalized by the presence of search costs for which they need to be compensated (Garleanu and Pedersen (2018)).
D Sensitivity Analysis

D.1 Alternative Asset Pricing Models

(Carhart model). We now summarize the additional results reported in the appendix. First, we use the Carhart model to estimate fund skill. Overall, our analysis reveals that the distributions of the two skill dimensions remain largely unchanged. The fd alpha is equal to 2.6% per year and is positive for 83.3% of the funds (vs 3.4% and 88.0% for the CPZ model). For the size coefficient, the similarity is even more striking—on average, it is equal to 1.4% per year and 83.8% of the funds have a positive coefficient (vs 1.5% and 87.1%). The main difference is observed for the small-cap group in which the annual fd alpha drops 5.0% to 3.2% on average. This sharp reduction arises because the Carhart model assigns a negative alpha to the Russell 2000 index.

(Five-factor model). We also measure fund skill using the five-factor model of Fama and French (2015) which includes the market, size, value, profitability, and investment factors. Using this model leaves the distribution of the size coefficient largely unchanged. However, it lowers the proportion of funds with a fd alpha from 88.0% to 76.4%. In other words, around 12% of the funds achieve a positive fd alpha because they implement profitability- and investment-based strategies. This tilt is particularly pronounced for low expense and low turnover funds as their average fd alphas respectively drop by 0.9% and 1.3% per year.

(CPZ-liquidity model). Finally, we check that the strong differences in skill between small cap and large cap funds are not simply due to an omitted liquidity risk factor. To address this issue, we add the traded liquidity factor of Pastor and Stambaugh (2003) to the CPZ model. We find that the distributions of the two skill dimensions remain largely unchanged. Therefore, the superior fd alphas produced by small cap funds is not driven by an aggregate liquidity risk premium.

D.2 Small Sample Bias

We also use an alternative estimation procedure to address the small sample bias in the estimated fund size coefficient $\hat{\beta}_i$. This bias stems from the positive correlation between the return residual $\varepsilon_{i,t}$ and the change in size $\varepsilon_{q,t}$, as positive return surprises increase fund size (e.g., Stambaugh (1999), Pastor, Stambaugh, and Taylor (2015)). Whereas this bias vanishes asymptotically, it can potentially affect the estimated coefficients of
funds with short return histories. To address this issue, we use the approach of Amihud and Hurvich (2004) that adds a proxy for the size innovation \( \hat{\varepsilon}_{qi,t} \) to the set of regressors in Equation (1) (see the appendix for details).\(^{30}\) The empirical results reveal the bias adjusted densities for the two skill dimensions remain largely unchanged.

D.3 Alternative Specifications for Size

(Industry-adjusted size). We consider different specifications to model the relation between the fund gross alpha \( \alpha_{i,t} \) and its size \( q_{i,t-1} \). Following Harvey and Liu (2018b), we replace \( q_{i,t-1} \) with \( q_{i,t-1}^{\text{act}} \), defined as the ratio between the fund size and the size of the active fund industry. This choice is motivated by the possibility that capacity constraints depend on the fund size relative to that of the entire industry. We find that the bias-adjusted distribution \( \hat{\phi}^* (b) \) remains largely unchanged.

(Log size). Next, we replace \( q_{i,t-1} \) with \( \log(q_{i,t-1}) \) and find qualitatively similar results to those reported in Table IV for the size coefficient. Using a log specification is useful when imposing a constant size coefficient, as fund alphas are more likely to respond similarly to relative changes in size (as opposed to absolute changes). However, this specification seems less important here because we allow for fund-specific size coefficients.

(Fixed costs). We also test whether the relation between \( \alpha_{i,t} \) and \( q_{i,t-1} \) is subject to non-linearities that may bias the estimated size coefficient \( \hat{b}_t \). One main source of non-linearities is the existence of fixed operating costs \( F_i \). In this case, we have: \( \alpha_{i,t} = a_i - b_i q_{i,t-1} - \frac{F_i}{q_{i,t-1}} = a_i - b_i q_{i,t-1} - c_i \frac{1}{q_{i,t-1}} \). Estimating this equation for each fund, we find that the distribution \( \hat{\phi}^* (b) \) remains largely unchanged.

(Industry-wide capacity constraints). Finally, we re-estimate the Equation (1) after replacing the fund size with the ratio of the industry size on the total market capitalization. Consistent with Pastor, Stambaugh, and Taylor (2015), we find evidence that individual fund returns vary with industry wide capacity constraints. However, the estimated coefficients are less precisely estimated. Because the regression condition numbers increase significantly, the sample size drops to 486 funds (as per Equation (5)).

V Conclusion

In this paper, we apply a new approach for measuring the entire skill distribution across mutual funds. Our approach is non-parametric and thus particularly suited to the

\(^{30}\) The inclusion of the estimated variable \( \hat{\varepsilon}_{qi,t} \) does not change the properties of our non-parametric density estimator. In particular, the smoothing and EIV biases in Equations (9)-(10) remain unchanged because \( \hat{\varepsilon}_{qi,t} \) only affects the higher order terms beyond \( T^{-1} \).
analysis of skill. For one, it avoids the challenge of specifying the skill distribution. It also allows for the joint analysis of multiple skill measures, including the two skill dimensions—the fd alpha and size coefficient—, the value added, and the gross alpha. In addition to its flexibility, our approach brings several advantages. It is simple to implement, applicable to all the descriptive statistics of the skill distribution, and supported by econometric theory.

Our empirical results yield several insights into the skill level in the fund industry. First, it reveals that most funds are able to detect profitable trades—around 90% of them produce positive fd alphas. However, most funds are also heavily exposed to capacity constraints—around 90% of them have a positive size coefficient. Second, the distributions of the skill dimensions (fd alpha and size coefficient) are highly non-normal and positively skewed. Therefore, taking simple averages across all funds fails to capture the large heterogeneity in skill. Third, both skill dimensions are positively correlated and need to be combined for measuring the overall fund skill. For one, small cap funds produce higher fd alphas than large cap funds, but also are also more exposed to capacity constraints. Fourth, the value added analysis shows that funds earn substantial profits from exploiting the two skill dimensions—around 75% of them generate positive profits for a total of 7.8 mio. per year on average.

Whereas our paper focuses on skill, our non-parametric approach has potentially wide applications in finance. It can be used to estimate the cross-sectional distribution of any coefficient of interest in a random coefficient model. This is the case, for instance, the case in asset pricing when we want to capture the heterogeneity across stocks (e.g., risk exposure, commonality in liquidity), or in corporate finance when we want to capture the heterogeneity across firms (e.g., investment and financing decisions).
References


VI Appendix

A.1 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:

$$ \hat{m}_i = e'_i \hat{Q}^{-1}_{x,i} \frac{1}{T_i} \sum_{t} I_{i,t} x_{t} r_{i,t} = m_i + e'_i \hat{Q}^{-1}_{x,i} \frac{1}{T_i} \sum_{t} I_{i,t} x_{t} \varepsilon_{i,t} $$

$$ = m_i + \frac{1}{\sqrt{T}} \tau_{i,T} e'_i \hat{Q}^{-1}_{x,i} \left( \frac{1}{\sqrt{T}} \sum_{t} I_{i,t} x_{t} \varepsilon_{i,t} \right) \equiv m_i + \frac{1}{\sqrt{T}} \hat{\eta}_{i,T}. \quad (22) $$

Moreover, let us write

$$ \hat{\eta}_{i,T} = \eta_{i,T} + \frac{1}{\sqrt{T}} \hat{\varepsilon}_{i,T}, \quad (23) $$

where $\eta_{i,T} = \tau_{i,T} \frac{1}{\sqrt{T}} \sum_{t} I_{i,t} u_{i,t}$, $u_{i,t} = e'_i Q^{-1}_{x,i} x_{t} \varepsilon_{i,t}$, and $\tau_{i,T} = \plim_{T \to \infty} \tau_{i,T}$. Hence, $\hat{\eta}_{i,T}/\sqrt{T}$ is the estimation error on $m_i = \alpha_i$. In particular, $\hat{\varepsilon}_{i,T}/T$ 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:

$$ I_1 = \frac{1}{h} E \left[ K \left( \frac{m_i - m}{h} \right) \right] - \phi(m), $$

$$ I_2 = \frac{1}{h} E \left[ K \left( \frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h} \right) \right] - \frac{1}{h} E \left[ K \left( \frac{m_i - m}{h} \right) \right], $$

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

$$ I_4 = \frac{1}{nh} \sum_{i} \left[ 1^X K \left( \frac{m_i + \eta_{i,T}/\sqrt{T} + \hat{\varepsilon}_{i,T}/T - m}{h} \right) - K \left( \frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h} \right) \right]. $$

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, which is negligible w.r.t. the others. We characterise the first three dominating 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 $K_2 = \int u^2 K(u) du$. 

31
(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:

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

where $\psi_{T,j}(m) = E[\eta_{i,T}^j | m_i = m] \phi(m)$ for $j = 1, 2, \ldots$. We have $\psi_{T,1}(m) = 0$ and

$$\lim_{T \to \infty} \psi_{T,2}(m) = E[S_i | m_i = m] \phi(m) \equiv \psi(m) \text{ where } S_i = \tau_i \overline{\lim}_{T \to \infty} \frac{1}{T} \sum_{i,s} I_i I_s u_i u_s.$$

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: $I_2 = \frac{1}{2T} \psi^{(2)}(m) + O(1/T^3/h^2/T)$.

(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 $x_i$ 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

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

where $\tilde{K}(u) = K(u)^2 / \int K(u)^2 du$. 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 \( \sqrt{n} I_3 \Rightarrow N(0, \phi(m) K_1) \), with \( K_1 = \int K(u)^2 du \).

For the \( \alpha \) (fd dollar), \( \beta \) (size coefficient), and \( \nu \alpha \) (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.2 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: \( bs(m) = bs_1(m) + bs_2(m) \), and \( \sigma^2(m) = \frac{1}{nh} \phi(m) K_1 \). The AMISE (asymptotic mean integrated squared error) is given by:

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

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

\[
-\frac{1}{nh^2} + c_1 h^3 + c_2 \frac{h}{T} = 0
\]

\[
\iff 1 = c_1 nh^5 + c_2 \frac{nh^3}{T}.
\] (24)

where \( c_1 = K_2 \int [\phi^{(2)}(u)]^2 du / K_1 \) 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 (24) 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:

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

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 \( \sqrt{nh^*}(h^* + T + 1/T^3/2) = o(1) \) if \( n/T^4 \to 0 \). Then, we can replace the bias component \( bs(m) \) by its asymptotic approximation to get:

\[
\sqrt{nh^*} \left( \hat{\phi}(m) - \phi(m) - \frac{1}{2} \phi(2)(m)K_2h^* - \frac{1}{2T} \phi(2)(m) \right) \Rightarrow N(0, \phi(m)K_1) . \quad (25)
\]

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

### A.3 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 (24). We focus on a Gaussian kernel \( K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) \). Then, the kernel constants are \( K_1 = K(u)^2du = \frac{1}{2\sqrt{\pi}} \) and \( K_2 = \int u^2K(u)du = 1 \). 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 \( \mu_m, \mu_s \), variance parameters \( \sigma_m^2, \sigma_s^2 \), and correlation parameter \( \rho \). 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:

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

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

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

\[
\int [\phi^{(2)}(u)]^2 du = \frac{1}{\sigma_m^5} \int (z^2 - 1)^2 \frac{1}{2\pi} \exp(-z^2) dz = \frac{1}{2\sqrt{\pi}\sigma_m^2} \int (v^2/2 - 1)^2 \varphi(v) dv
\]

= \frac{3}{8\sqrt{\pi}\sigma_m^5},

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

\[
\psi^{(2)}(m) = \exp\left(\mu_s + \rho \sigma_s \left(\frac{m - \mu_m}{\sigma_m}\right) + \frac{1}{2} \sigma_s^2 (1 - \rho^2)\right) \varphi(m)
\]
\[
\times \left\{\left(\frac{\sigma_s \rho}{\sigma_m}\right)^2 - 2 \frac{\sigma_s \rho}{\sigma_m} \left(\frac{m - \mu_m}{\sigma_m}\right) + \frac{1}{2} \left(\frac{m - \mu_m}{\sigma_m}\right)^2 - 1\right\}
\]
\[
= \exp\left(\mu_s + \frac{1}{2} \sigma_s^2\right) \left\{\frac{1}{\sigma_m^2} \left(\left(\frac{m - \mu_m - \rho \sigma_s \sigma_m}{\sigma_m}\right)^2 - 1\right) \frac{1}{\sigma_m} \varphi\left(\frac{m - \mu_m - \rho \sigma_s \sigma_m}{\sigma_m}\right)\right\}.
\]

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

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

35
and then differentiating twice that expression. Besides, we have:

\[
\int \phi^{(4)}(m)\psi(m)dm = \frac{\exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1 - \rho^2)\right)}{\sigma_m^3} \int \varphi^{(4)}(z) \exp(\rho \sigma_s z) \varphi(z) dz
\]

\[
= \frac{\exp\left(\mu_s + \frac{1}{2}\sigma_s^2(1 - \rho^2)\right)}{2\sqrt{\pi} \sigma_m^3} \int (v^4 / 4 - 3v^2 + 3) \exp(\lambda v) \varphi(v) dv,
\]

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 \( E[\exp(\lambda Z)] = \exp(\lambda^2 / 2) \) for a standard Gaussian variable \( Z \). This yields

\[
\int (v^4 / 4 - 3v^2 + 3) \exp(\lambda v) \varphi(v) dv = \left( \frac{1}{4} \frac{\partial^4}{\partial \lambda^4} - 3 \frac{\partial^2}{\partial \lambda^2} + 3 \right) \exp(\lambda^2 / 2)
\]

\[
= \frac{1}{4}(\lambda^4 - 6\lambda^2 + 3) \exp(\lambda^2 / 2).
\]

Thus, we get:

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

The optimal bandwidth \( h^* \) is obtained by solving Equation (24) with coefficients \( c_1 \) and \( c_2 \) given by:

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

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 \( \sigma_m \), \( \mu_s \), \( \sigma_s \) and \( \rho \) are estimated by the sample moments of \( \hat{m}_i \) and \( \hat{s}_i = \log \hat{S}_i \).

**A.4 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 \( \frac{1}{n} \sum_{i=1}^{n} g(\hat{m}_i) \mathbf{1}_i^X \) 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
Proposition VI.1. As $n, T \to \infty$, such that $n = o(T^3)$, we have
\[
\sqrt{n} \left( \frac{1}{2} \sum_{i=1}^{T} g(m_i) \right) - \mathbb{E}[g(m_i)] - B_r \to N(0, \mathbb{V}[g(m_i)]).
\]
with
\[
\nabla_4 K u = V[m_i]^{-2}, \quad \nabla_3 K u = -4E[m_i]V[m_i]^{-2}, \\
\nabla_2 K u = 6E[m_i]^2V[m_i]^{-2} + \{E[m_i^4] - 4E[m_i^3]E[m_i] + 6E[m_i^2]E[m_i]^2 - 3E[m_i]^4\}(-2)V[m_i]^{-3}. \\
\]

**Proof of Proposition VI.1:** Equation (22) yields the mean value expansion
\[
g(\hat{m}_i) = g(m_i) + g^{(1)}(\hat{m}_i) \frac{1}{\sqrt{T}} \hat{\eta}_{i,T} + g^{(2)}(\hat{m}_i) \frac{1}{2T} \hat{\eta}_{i,T}'^T, \\
\]
where \(\hat{m}_i\) lies between \(\hat{m}_i\) and \(m_i\). Then we get
\[
\sqrt{n} \left( \frac{1}{n} \sum_i g(\hat{m}_i)1^\chi_i - E[g(m_i)] - B_T \right) \\
= \frac{1}{\sqrt{n}} \sum_i (g(m_i) - E[g(m_i)]) - \frac{1}{\sqrt{n}} \sum_i g(m_i)(1 - 1^\chi_i) + \frac{1}{\sqrt{nT}} \sum_i 1^\chi_i g^{(1)}(\hat{m}_i) \hat{\eta}_{i,T} \\
+ \frac{1}{2T} \sqrt{n} \sum_i \left( 1^\chi_i g^{(2)}(\hat{m}_i) \hat{\eta}_{i,T}'^T - E\left[g^{(2)}(m_i)S_i\right] \right) \\
\equiv I_{71} + I_{72} + I_{73} + I_{74}. \\
\]

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 (26) 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.5 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 \(g(\hat{m}_i) = 1\{\hat{m}_i \leq m\}\) based on the OLS estimates \(\hat{m}_i\) for the non-trimmed assets, namely \(\hat{\Phi}(m) = \frac{1}{n} \sum_i 1\{\hat{m}_i \leq m\}1^\chi_i\). 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 \( n = o(T^3) \),

\[
\sqrt{n} \left( \Phi(m) - \Phi(m) - B_T(m) \right) \Rightarrow N \left( 0, \Phi(m)(1 - \Phi(m)) \right),
\]

\[
\sqrt{n} \left( \hat{Q}(u) - Q(u) + \frac{B_T(Q(u))}{\phi(Q(u))} \right) \Rightarrow N \left( 0, \frac{u(1-u)}{\phi(Q(u))^2} \right),
\]

with \( B_T(m) = \frac{1}{2T} \psi^{(1)}(m) \), 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:

\[
\psi^{(1)}(m) = \exp \left( \mu_s + \rho \sigma_s \left( \frac{m - \mu_m}{\sigma_m} \right) + \frac{1}{2} \sigma_s^2 (1 - \rho^2) \right) \phi(m) \left( \frac{\sigma_s \rho}{\sigma_m} - \frac{m - \mu_m}{\sigma_m^2} \right)
\]

\[
= \exp \left( \mu_s + \frac{1}{2} \sigma_s^2 \right) \left( \frac{m - \mu_m - \rho \sigma_s \sigma_m}{\sigma_m} \right) \left( \frac{1}{\sigma_m} \frac{m - \mu_m - \rho \sigma_s \sigma_m}{\sigma_m} \right).
\]

We can proceed similarly for the quantile case.

**Proof of Proposition VI.2:** From (22), we have: \( E \left[ 1 \{ \hat{m}_i \leq m \} \right] = P \left[ m_i + \frac{1}{\sqrt{T}} \hat{n}_{i,T} \leq m \right] \).

By using the results in Gourieroux, Laurent, and Scaillet (2000), Martin, and Wilde (2001), Gordy (2003), Gagliardini and Gourieroux (2011), we get:

\[
P \left[ m_i + \frac{1}{\sqrt{T}} \hat{n}_{i,T} \leq m \right] = \Phi(m) - \frac{1}{\sqrt{T}} \phi(m) E[\hat{n}_{i,T} | m_i = m]
\]

\[
+ \frac{1}{2T} \frac{d}{dm} \left( \phi(m) E[\hat{n}_{i,T}^2 | m_i = m] \right) + o(1/T).
\]

From (23), the bias expansion is such that: \( E \left[ \frac{\Phi(m)}{\Phi(m)} - \Phi(m) \right] = B_T(m) + E \left[ 1 \{ \hat{m}_i \leq m \} (1 - 1^\chi) \right] + o(1/T) \). 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 \( u \in (0, 1) \): \( \hat{Q}(u) - Q(u) = -\frac{1}{\phi(Q(u))} \left( \Phi(Q(u)) - u \right) \).

39
### Table I

**Relations Between the Skill Measures**

This table describes different compensation schemes chosen by the funds when setting their level of fees. Each compensation scheme yields specific predictions regarding the cross-sectional distributions of fund fees and size, and the relation between the gross alpha and the three other skill measures (first-dollar alpha, size coefficient, and value added).

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

Panel A reports the average number of funds and the first four moments of the portfolio gross excess return for all funds in the population, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). Panel B reports the estimated portfolio betas on the market, size, value, and momentum factors, as well as the adjusted \( R^2 \) using the Cremers, Petajisto, and Zitzewitz benchmark model. All statistics are computed using monthly data between January 1979 and December 2015.

### Panel A: Gross Excess Return

<table>
<thead>
<tr>
<th></th>
<th>Average Nb. Funds</th>
<th>Mean (Ann.)</th>
<th>Volatility (Ann.)</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Funds</td>
<td>1103</td>
<td>7.68</td>
<td>15.01</td>
<td>-0.8</td>
<td>5.4</td>
</tr>
<tr>
<td><strong>Investment Styles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>236</td>
<td>8.82</td>
<td>19.30</td>
<td>-0.6</td>
<td>4.9</td>
</tr>
<tr>
<td>Large-cap</td>
<td>436</td>
<td>7.77</td>
<td>14.77</td>
<td>-0.7</td>
<td>5.3</td>
</tr>
<tr>
<td>Growth</td>
<td>511</td>
<td>8.14</td>
<td>16.26</td>
<td>-0.8</td>
<td>5.3</td>
</tr>
<tr>
<td>Value</td>
<td>279</td>
<td>7.45</td>
<td>13.81</td>
<td>-0.7</td>
<td>5.5</td>
</tr>
<tr>
<td><strong>Fund Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>373</td>
<td>7.71</td>
<td>14.58</td>
<td>-0.8</td>
<td>5.3</td>
</tr>
<tr>
<td>High Expense</td>
<td>289</td>
<td>8.03</td>
<td>16.40</td>
<td>-0.8</td>
<td>5.2</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>274</td>
<td>7.49</td>
<td>14.31</td>
<td>-0.8</td>
<td>5.5</td>
</tr>
<tr>
<td>High Turnover</td>
<td>260</td>
<td>8.58</td>
<td>16.59</td>
<td>-0.7</td>
<td>5.0</td>
</tr>
</tbody>
</table>

### Panel B: Estimated Betas

<table>
<thead>
<tr>
<th></th>
<th>Market</th>
<th>Size</th>
<th>Value</th>
<th>Momentum</th>
<th>Adj. R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Funds</td>
<td>0.93</td>
<td>0.26</td>
<td>-0.10</td>
<td>0.01</td>
<td>0.98</td>
</tr>
<tr>
<td><strong>Investment Styles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>0.98</td>
<td>0.81</td>
<td>-0.29</td>
<td>0.05</td>
<td>0.97</td>
</tr>
<tr>
<td>Large-cap</td>
<td>0.94</td>
<td>0.15</td>
<td>-0.06</td>
<td>0.01</td>
<td>0.99</td>
</tr>
<tr>
<td>Growth</td>
<td>0.95</td>
<td>0.33</td>
<td>-0.27</td>
<td>0.03</td>
<td>0.97</td>
</tr>
<tr>
<td>Value</td>
<td>0.92</td>
<td>0.13</td>
<td>0.19</td>
<td>-0.01</td>
<td>0.98</td>
</tr>
<tr>
<td><strong>Fund Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>0.92</td>
<td>0.21</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.98</td>
</tr>
<tr>
<td>High Expense</td>
<td>0.93</td>
<td>0.42</td>
<td>-0.27</td>
<td>0.01</td>
<td>0.97</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>0.91</td>
<td>0.20</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.98</td>
</tr>
<tr>
<td>High Turnover</td>
<td>0.95</td>
<td>0.37</td>
<td>-0.28</td>
<td>0.08</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Table III
Cross-Sectional Distribution of the First Dollar Alpha

The table contains the summary statistics on the cross-sectional distribution of the first dollar (fd) alpha for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive fd alpha, and the distribution quantiles a 10% and 90%. All of the estimated statistics are adjusted for bias (smoothing and EIV) using our non-parametric approach.

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>Proportions (%)</th>
<th>Quantiles (Ann.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Ann.)</td>
<td>Volatility (Ann.)</td>
<td>Skewness</td>
</tr>
<tr>
<td>All Funds</td>
<td>3.4</td>
<td>3.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Investment Styles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>5.0</td>
<td>3.3</td>
<td>10.1</td>
</tr>
<tr>
<td>Large-cap</td>
<td>2.0</td>
<td>2.4</td>
<td>5.0</td>
</tr>
<tr>
<td>Growth</td>
<td>3.5</td>
<td>3.9</td>
<td>5.5</td>
</tr>
<tr>
<td>Value</td>
<td>3.3</td>
<td>3.8</td>
<td>5.7</td>
</tr>
<tr>
<td>Fund Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>2.0</td>
<td>3.2</td>
<td>3.4</td>
</tr>
<tr>
<td>High Expense</td>
<td>4.6</td>
<td>4.0</td>
<td>6.3</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>3.2</td>
<td>3.8</td>
<td>4.6</td>
</tr>
<tr>
<td>High Turnover</td>
<td>3.7</td>
<td>4.6</td>
<td>5.2</td>
</tr>
</tbody>
</table>
Table IV
Cross-Sectional Distribution of the Size Coefficient

The table contains the summary statistics on the cross-sectional distribution of the size coefficient for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive size coefficient, and the distribution quantiles a 10% and 90%. All of the estimated statistics are adjusted for bias (smoothing and EIV) using our non-parametric approach.

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>Proportions (%)</th>
<th>Quantiles (Ann.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Ann.)</td>
<td>Volatility (Ann.)</td>
<td>Skewness</td>
</tr>
<tr>
<td>All Funds</td>
<td>1.5</td>
<td>1.3</td>
<td>8.0</td>
</tr>
<tr>
<td>Investment Styles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>2.0</td>
<td>1.4</td>
<td>9.5</td>
</tr>
<tr>
<td>Large-cap</td>
<td>1.0</td>
<td>0.9</td>
<td>6.9</td>
</tr>
<tr>
<td>Growth</td>
<td>1.6</td>
<td>1.7</td>
<td>6.8</td>
</tr>
<tr>
<td>Value</td>
<td>1.3</td>
<td>1.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Fund Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>0.9</td>
<td>1.1</td>
<td>6.1</td>
</tr>
<tr>
<td>High Expense</td>
<td>1.9</td>
<td>1.9</td>
<td>7.0</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>1.3</td>
<td>1.6</td>
<td>6.4</td>
</tr>
<tr>
<td>High Turnover</td>
<td>1.7</td>
<td>2.1</td>
<td>4.7</td>
</tr>
</tbody>
</table>
The table contains the summary statistics on the cross-sectional distribution of the value added for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive value added, and the distribution quantiles a 10% and 90%. All of the estimated statistics are adjusted for bias (smoothing and EIV) using our non-parametric approach.

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>Proportions (%)</th>
<th>Quantiles (Ann.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Ann.)</td>
<td>Volatility (Ann.)</td>
<td>Skewness</td>
</tr>
<tr>
<td>All Funds</td>
<td>7.8</td>
<td>14.2</td>
<td>10.7</td>
</tr>
<tr>
<td><strong>Investment Styles</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>9.7</td>
<td>15.4</td>
<td>9.0</td>
</tr>
<tr>
<td>Large-cap</td>
<td>6.9</td>
<td>16.2</td>
<td>12.1</td>
</tr>
<tr>
<td>Growth</td>
<td>9.0</td>
<td>16.7</td>
<td>10.4</td>
</tr>
<tr>
<td>Value</td>
<td>7.6</td>
<td>14.3</td>
<td>9.4</td>
</tr>
<tr>
<td><strong>Fund Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>9.4</td>
<td>23.5</td>
<td>8.7</td>
</tr>
<tr>
<td>High Expense</td>
<td>5.4</td>
<td>8.3</td>
<td>10.3</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>11.7</td>
<td>22.5</td>
<td>7.9</td>
</tr>
<tr>
<td>High Turnover</td>
<td>6.6</td>
<td>12.1</td>
<td>10.9</td>
</tr>
</tbody>
</table>
Table VI

Actual versus Optimal Value Added

The table contains the summary statistics on the cross-sectional distribution of the optimal and actual value added for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic-sorted groups (low expense, high expense, low turnover, high turnover). It reports the mean, the distribution quantiles at 10% and 90%. All of the estimated statistics are adjusted for bias (smoothing and EIV) using our non-parametric approach.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Quantile 10%</th>
<th>Quantile 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Funds</td>
<td>11.3</td>
<td>8.0</td>
<td>3.4</td>
</tr>
<tr>
<td>Investment Styles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>11.0</td>
<td>8.4</td>
<td>2.5</td>
</tr>
<tr>
<td>Large-cap</td>
<td>9.6</td>
<td>6.1</td>
<td>3.6</td>
</tr>
<tr>
<td>Growth</td>
<td>12.2</td>
<td>8.2</td>
<td>4.0</td>
</tr>
<tr>
<td>Value</td>
<td>12.2</td>
<td>8.0</td>
<td>4.2</td>
</tr>
<tr>
<td>Fund Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>15.0</td>
<td>9.3</td>
<td>5.7</td>
</tr>
<tr>
<td>High Expense</td>
<td>9.1</td>
<td>6.0</td>
<td>3.1</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>17.5</td>
<td>11.9</td>
<td>5.7</td>
</tr>
<tr>
<td>High Turnover</td>
<td>12.3</td>
<td>6.8</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Table VII
Fund Fees and Size

The table contains the summary statistics on the cross-sectional distribution of fund fees and size for all funds, four styles groups (small cap, large cap, growth, value), and four characteristic.sorted groups (low expense, high expense, low turnover, high turnover). It reports the mean, volatility, and the distribution quantiles at 10% and 90%.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Volatility</th>
<th>Quantile 10%</th>
<th>Quantile 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fees</td>
<td>Size</td>
<td>Fees</td>
<td>Size</td>
</tr>
<tr>
<td>All Funds</td>
<td>1.25</td>
<td>753</td>
<td>0.41</td>
<td>2076</td>
</tr>
<tr>
<td>Investment Styles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>1.37</td>
<td>410</td>
<td>0.36</td>
<td>673</td>
</tr>
<tr>
<td>Large-cap</td>
<td>1.16</td>
<td>1068</td>
<td>0.39</td>
<td>2953</td>
</tr>
<tr>
<td>Growth</td>
<td>1.30</td>
<td>769</td>
<td>0.41</td>
<td>2059</td>
</tr>
<tr>
<td>Value</td>
<td>1.18</td>
<td>980</td>
<td>0.38</td>
<td>2690</td>
</tr>
<tr>
<td>Fund Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>0.82</td>
<td>1411</td>
<td>0.18</td>
<td>3563</td>
</tr>
<tr>
<td>High Expense</td>
<td>1.73</td>
<td>344</td>
<td>0.28</td>
<td>727</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>1.14</td>
<td>1475</td>
<td>0.37</td>
<td>3980</td>
</tr>
<tr>
<td>High Turnover</td>
<td>1.30</td>
<td>650</td>
<td>0.37</td>
<td>1272</td>
</tr>
</tbody>
</table>
Table VIII
Cross-Sectional Distribution of Gross Alpha

The table contains the summary statistics on the cross-sectional distribution of the gross alpha for all funds, four styles groups (small-cap, large-cap, growth, value), and four characteristic-sorted groups (low-expense, high-expense, low-turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive gross alpha, and the distribution quantiles a 10% and 90%. All of the estimated statistics are adjusted for bias (smoothing and EIV) using our non-parametric approach.

<table>
<thead>
<tr>
<th></th>
<th>Mean (Ann.)</th>
<th>Volatility (Ann.)</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Negative Skill</th>
<th>Positive Skill</th>
<th>10%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Funds</td>
<td>0.8</td>
<td>1.3</td>
<td>3.2</td>
<td>1.2</td>
<td>27.8</td>
<td>72.2</td>
<td>-0.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Investment Styles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>1.6</td>
<td>1.2</td>
<td>6.4</td>
<td>-1.9</td>
<td>12.3</td>
<td>87.7</td>
<td>-0.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Large-cap</td>
<td>0.3</td>
<td>0.9</td>
<td>1.6</td>
<td>3.1</td>
<td>36.0</td>
<td>64.0</td>
<td>-0.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Growth</td>
<td>0.8</td>
<td>1.5</td>
<td>2.6</td>
<td>1.7</td>
<td>31.6</td>
<td>68.4</td>
<td>-1.3</td>
<td>2.9</td>
</tr>
<tr>
<td>Value</td>
<td>1.0</td>
<td>1.6</td>
<td>4.2</td>
<td>2.1</td>
<td>26.5</td>
<td>73.5</td>
<td>-1.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Fund Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>0.4</td>
<td>1.4</td>
<td>2.4</td>
<td>2.7</td>
<td>38.8</td>
<td>61.2</td>
<td>-1.3</td>
<td>2.0</td>
</tr>
<tr>
<td>High Expense</td>
<td>1.1</td>
<td>1.8</td>
<td>2.9</td>
<td>3.3</td>
<td>25.2</td>
<td>74.8</td>
<td>-1.2</td>
<td>3.6</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>0.9</td>
<td>1.5</td>
<td>2.8</td>
<td>0.8</td>
<td>27.8</td>
<td>72.2</td>
<td>-1.1</td>
<td>3.0</td>
</tr>
<tr>
<td>High Turnover</td>
<td>0.7</td>
<td>1.9</td>
<td>2.8</td>
<td>3.1</td>
<td>34.8</td>
<td>65.2</td>
<td>-1.6</td>
<td>3.1</td>
</tr>
</tbody>
</table>
The table contains the summary statistics on the cross-sectional distribution of the net alpha for all funds, four styles groups (small-cap, large-cap, growth, value), and four characteristic-sorted groups (low-expense, high-expense, low-turnover, high turnover). It reports the first four moments, the proportions of funds with a negative and positive net alpha, and the distribution quantiles a 10% and 90%. All of the estimated statistics are adjusted for bias (smoothing and EIV) using our non-parametric approach.

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>Proportions (%)</th>
<th>Quantiles (Ann.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Ann.)</td>
<td>Volatility (Ann.)</td>
<td>Skewness</td>
</tr>
<tr>
<td>All Funds</td>
<td>-0.5</td>
<td>1.3</td>
<td>-2.8</td>
</tr>
<tr>
<td>Investment Styles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>0.2</td>
<td>1.3</td>
<td>-0.6</td>
</tr>
<tr>
<td>Large-cap</td>
<td>-0.8</td>
<td>0.9</td>
<td>-5.3</td>
</tr>
<tr>
<td>Growth</td>
<td>-0.5</td>
<td>1.6</td>
<td>-2.8</td>
</tr>
<tr>
<td>Value</td>
<td>-0.2</td>
<td>1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Fund Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Expense</td>
<td>-0.4</td>
<td>1.3</td>
<td>-1.0</td>
</tr>
<tr>
<td>High Expense</td>
<td>-0.6</td>
<td>1.9</td>
<td>-2.5</td>
</tr>
<tr>
<td>Low Turnover</td>
<td>-0.3</td>
<td>1.5</td>
<td>-1.7</td>
</tr>
<tr>
<td>High Turnover</td>
<td>-0.6</td>
<td>2.0</td>
<td>-1.3</td>
</tr>
</tbody>
</table>
Figure 1
The Two Bias Components

This figure plots the two bias components (smoothing and Error-in-Variable (EIV)) for the cross-sectional distribution of the gross alpha. Each component is obtained from the normal reference model whose five parameters are estimated using the estimated gross alpha and its asymptotic variance for all funds in the population.
This figure shows how the shape of the Error-in-Variable (EIV) bias for the cross-sectional distribution of the gross alpha changes for different parameter values. In Panel A, we increase the volatility \( \sigma_m \) of the true gross alpha by 30% (in relative terms). In Panel B, we increase the average level \( \mu_\alpha \) of the asymptotic variance of the estimated gross alpha by 30% (in relative terms). In Panel C, we increase the correlation \( \rho \) between the fund true alpha and the asymptotic variance by 0.3.
Figure 3
Cross-sectional Distribution of the First Dollar Alpha: Analysis across Fund Groups

Panel A plots the cross-sectional densities of the first dollar alpha for small cap and large cap funds. Panel B compares growth and value funds. Panel C compares low expense and high expense funds. Finally, Panel D compares low turnover and high turnover funds. All the estimated densities are adjusted for bias (smoothing and EIV) using our non-parametric approach.
Figure 4
Cross-sectional Distribution of the Size Coefficient:
Analysis across Fund Groups

Panel A plots the cross-sectional densities of the size coefficient for small cap and large cap funds. Panel B compares growth and value funds. Panel C compares low expense and high expense funds. Finally, Panel D compares low turnover and high turnover funds. All the estimated densities are adjusted for bias (smoothing and EIV) using our non-parametric approach.
Figure 5  
Cross-sectional Distribution of the Value Added: Analysis across Fund Groups

Panel A plots the cross-sectional densities of the value added for small cap and large cap funds. Panel B compares growth and value funds. Panel C compares low expense and high expense funds. Finally, Panel D compares low turnover and high turnover funds. All the estimated densities are adjusted for bias (smoothing and EIV) using our non-parametric approach.