False Discoveries in Mutual Fund Performance:
Measuring Luck in Estimated Alphas[*]

Laurent Barras[†], Olivier Scaillet[‡] and Russ Wermers[§]

First version, September 2005; This version, April 2008

ABSTRACT

This paper uses a new approach to determine the frequency of truly skilled managers among the universe of U.S. domestic-equity mutual funds. Specifically, we develop a simple technique that properly accounts for "false discoveries," or mutual funds which exhibit significant alphas by luck alone. We use this technique to precisely separate actively managed funds into those having (1) unskilled, (2) zero-alpha, and (3) skilled fund managers, even with cross-fund dependencies in estimated alphas. This precise separation into skill groups allows several new insights. First, we find that the majority of funds (75.4%) pick stocks well enough to cover their trading costs and other expenses, consistent with the equilibrium model of Berk and Green (2004). Further, we find a significant number of skilled funds prior to 1995, but almost none by 2006, accompanied by a large increase in unskilled fund managers–due both to a large reduction in the proportion of fund managers with stockpicking skills and to a persistent level of fees that exceed the value generated by these managers. Finally, we show that controlling for false discoveries substantially improves the ability to find funds with persistent performance.

Investors and academic researchers have long searched for outperforming mutual fund managers. Although several researchers document negative average fund alphas, net of expenses and trading costs (e.g., Jensen (1968), Lehman and Modest (1987), Elton et al. (1993), and Carhart (1997)), recent papers show that some fund managers have stock-selection skills. For instance, Kosowski, Timmermann, Wermers, and White (2006; KTTW) use a bootstrap technique to document outperformance by some funds, while Baks, Metrick, and Wachter (2001), Pastor and Stambaugh (2002b), and Avramov and Wermers (2007) illustrate the benefits of investing in actively-managed funds from a Bayesian perspective. While these papers are useful in uncovering whether, on the margin, outperforming mutual funds exist, they are not particularly informative regarding their prevalence in the entire fund population. For instance, it is natural to wonder how many fund managers possess true stockpicking skills (according to a chosen asset pricing model), and where these funds are located in the cross-sectional *estimated* alpha distribution. From an investment perspective, precisely locating skilled funds maximizes our chances of achieving persistent outperformance.[1]

Of course, we cannot observe the *true* alphas of each fund in the population. Therefore, a seemingly reasonable way to estimate the prevalence of skilled fund managers is to simply count the number of funds with sufficiently high estimated alphas, $\widehat{\alpha}$. In implementing such a procedure, we are actually conducting a multiple (hypothesis) test, since we simultaneously examine the performance of several funds in the population (instead of just one fund).[2] However, a simple count of significant alphas does not properly adjust for luck in such a multiple test setting–many of the funds have significant estimated alphas purely by luck (i.e., their true alphas are zero). To illustrate, consider a population of funds with skills just sufficient to cover trading costs and expenses (zero-alpha funds). With the usual chosen significance level of 5%, we expect that 5% of these zero-alpha funds will have significant estimated alphas–some of them will be unlucky ($\widehat{\alpha} < 0$) while others are lucky ($\widehat{\alpha} > 0$), but all will be "false discoveries"–funds with significant *estimated* alphas, but zero *true* alphas.

This paper provides a new approach to controlling for false discoveries in such a multiple fund setting. Our approach much more accurately identifies (1) the proportions of unskilled and skilled funds in the population (those with *truly* negative and positive

---

[1] From an investor's perspective, "skill" is manager talent, in selecting stocks, that is sufficient to generate a positive alpha, net of trading costs and fund expenses.

[2] This multiple test should not be confused with the joint hypothesis test with the null that all fund alphas are equal to zero in a sample. This test, which is employed by several past research papers (e.g., Grinblatt and Titman (1989, 1993)), addresses only whether at least one fund has a non-zero alpha among several funds, but is silent on the prevalence of these non-zero alpha funds.

alphas, respectively), and (2) their respective locations in the left and right tails of the cross-sectional *estimated* alpha (or estimated alpha t-statistic) distribution. A main virtue of our approach is its simplicity–to determine the proportions of unlucky and lucky funds, the only parameter needed is the proportion of zero-alpha funds in the population, π_0. Rather than arbitrarily imposing a prior assumption on π_0, we estimate it with a straightforward computation that uses the p-values of the individual fund estimated alphas–no further econometric tests are necessary. A second advantage of our approach is its accuracy. Using a simple Monte-Carlo experiment, we demonstrate that our approach provides much more accurate estimates of the proportion of unskilled and skilled funds in the population than previous approaches that impose an a priori assumption about the proportion of zero-alpha funds in the population.[3]

Another important advantage of our approach to multiple testing (as shown in further Monte-Carlo tests) is its robustness to cross-sectional dependencies among fund estimated alphas. Prior literature has indicated that such dependencies likely exist due to herding and other correlated trading behaviors (e.g., Wermers (1999)), which result in funds holding similar stocks or industries. While, for prior approaches, this dependence greatly complicates performance measurement, this is not the case with our approach, since it only requires the (alpha) p-value for each fund in the population, estimated in isolation–and not the estimation of the cross-fund covariance matrix.

We apply our new approach to the monthly returns of 2,076 U.S. open-end, domestic equity mutual funds, and revisit several important themes examined in the previous literature. We start with an examination of the long-term performance (net of trading costs and expenses) of the mutual fund industry during 1975 to 2006. Our decomposition of the population reveals that 75.4% are zero-alpha funds–funds having managers with some stockpicking abilities, but that extract all of the rents of these abilities. Among remaining funds, only 0.6% are skilled (true $\alpha > 0$), while 24.0% are unskilled (true $\alpha < 0$). While our finding of a majority of zero-alpha funds–those that capture the entire surplus created by their managers (through fees)–is supportive of the long-run equilibrium of Berk and Green (2004), it is surprising that we find so many truly negative-alpha funds–those that overcharge relative to the skills of their managers. Indeed, we find that such unskilled funds are not particularly new, indicating that investors have had some time to evaluate and identify them as underperformers.

---

[3]The reader should note the difference between our approach and that of KTWW (2006). Our approach simultaneously estimates the prevalence and location of outperforming funds in a group, while KTWW test for the skills of a single fund that is chosen from the universe of alpha-ranked funds. As such, our approach examines fund performance from a different perspective, with a richer set of information about active fund manager skills.

We also find some notable time trends in our study. Examining the evolution of each skill group between 1990 and 2006, we observe that the proportion of skilled funds dramatically decreases from 14.4% to 0.6%, while the proportion of unskilled funds increases sharply from 9.2% to 26.6%. Thus, although the number of actively managed funds has substantially increased, skilled managers (those capable of picking stocks well enough to overcome their trading costs and expenses) have become increasingly rare.

Motivated by the possibility that funds may outperform over the short-run, before investors compete away their performance with inflows, we conduct further tests over five-year subintervals–treating each five-year fund record as a separate "fund." Here, we find that the proportion of skilled funds equals 2.4%, implying that a small number of managers have "hot hands." These skilled funds are located in the extreme right tail of the cross-sectional estimated alpha distribution, which indicates that a *very* low p-value is an accurate signal of short-run fund manager skill (relative to pure luck). Across the investment subgroups, Aggressive Growth funds have the highest proportion of managers with short-term skills, while Growth & Income funds exhibit no skills.

The concentration of skilled funds in the extreme right tail of the estimated alpha distribution suggests a natural way to choose funds in seeking out-of-sample persistent performance. Specifically, we form portfolios of right-tail funds that condition on the frequency of "false discoveries"–during years when our tests indicate higher proportions of lucky, zero-alpha funds in the right tail, we move further to the extreme tail to decrease false discoveries. Forming such a false-discovery controlled portfolio at the beginning of each year from January 1980 to 2006, we find an equal-weighted four-factor alpha of 1.45% per year, which is statistically significant. Notably, this luck-controlled strategy outperforms prior persistence strategies used by Carhart (1997) and others, where constant top-decile portfolios of funds are chosen with no control for luck.

Our final tests examine the performance of fund managers before expenses (but after trading costs) are subtracted. Specifically, while fund managers may be able to pick stocks well enough to cover their trading costs, they do not usually exert direct control over the level of fund expenses and fees–management companies set these expenses, with the approval of fund directors. We find evidence that indicates a very large impact of fund fees and other expenses–on a pre-expense basis, we find a much higher incidence of skilled funds with positive alphas–9.6%, compared to our prior finding of 0.6% after expenses. Thus, almost all outperforming funds appear to capture (or waste through operational inefficiencies) the entire surplus created by their portfolio managers. It is also noteworthy that the proportion of skilled managers (before expenses) declines substantially over time, again indicating that portfolio managers with skills have become

3

increasingly rare. We also observe a large reduction in the proportion of unskilled funds when we move from net alphas to pre-expense alphas (from 26.6% to 4.5%), indicating a big role for excessive fees (relative to manager stockpicking skills) in underperforming funds. It remains a major puzzle why investors do not use their flows to compete such fees away, given that there are numerous available funds in the population that do not underperform.

The remainder of the paper is as follows. The next section explains our approach to separating luck from skill in measuring the performance of asset managers. Section 2 presents the performance measures, and describes the mutual fund data. Section 3 contains the results of the paper. An appendix contains the details of the estimation procedure, as well as an extensive Monte-Carlo study on the accuracy of our estimators under cross-sectional independence and dependence.

I   The Impact of Luck on Mutual Fund Performance

A   Overview of the Setting

A.1   Luck in a Multiple Performance Setting

Our objective is to develop a framework to precisely estimate the proportion of a group of mutual funds that truly outperform their benchmarks. To begin, suppose that a population of M actively managed mutual funds is composed of three distinct performance categories, where performance is due to stock-selection skills. We define such performance as the ability of fund managers to generate superior model alphas, net of trading costs as well as all fees and other expenses (except loads and taxes). We define our performance categories as follows:

  • **Unskilled funds**: funds having managers with stockpicking skills insufficient to recover their trading costs and expenses, such that there is an "alpha shortfall" ($\alpha < 0$)

  • **Zero-alpha funds**: funds having managers with stockpicking skills sufficient to just recover trading costs and expenses ($\alpha = 0$)

  • **Skilled funds**: funds having managers with stockpicking skills sufficient to provide an "alpha surplus," beyond simply recovering their trading costs and expenses ($\alpha > 0$).

Note that our above definition of skill is one that is relative to expenses, and not in an absolute sense. This definition is driven by the idea that consumers look for mutual fund that delivers surplus alpha, net of all expenses.[4]

---

[4] However, perhaps a manager exhibits skill sufficient to more than compensate for trading costs,

Of course, we cannot observe the true alphas of each fund in the population. There-fore, how do we best infer the prevalence of each of the above skill groups from perfor-mance measures for individual funds? First, we use the t-statistic, $\widehat{t}_i = \widehat{\alpha}_i/\widehat{\sigma}_{\widehat{\alpha}_i}$, where $\widehat{\alpha}_i$ is the estimated alpha for fund i, and $\widehat{\sigma}_{\widehat{\alpha}_i}$ is its estimated standard deviation. KTTW (2006) show that performance measured with the t-statistic has superior properties rel-ative to alpha, since alpha estimates have differing precision across funds with varying lives and portfolio volatilities. Second, after choosing a significance level, γ (e.g., 10%), we observe whether $\widehat{t}_i$ lies outside the thresholds implied by γ (denoted by t_γ^- and t_γ^+), and label it "significant" if it is such an outlier. It is important to note that this proce-dure, simultaneously applied across all funds, is a multiple-hypothesis test:

$$
\begin{aligned}
H_{0,1} &: \alpha_1 = 0, \quad H_{A,1} : \alpha_1 \neq 0, \\
\dots &: \dots \\
H_{0,M} &: \alpha_M = 0, \quad H_{A,M} : \alpha_M \neq 0.
\end{aligned}
\tag{1}
$$

To illustrate the difficulty of this multiple test, Figure 1 presents a hypothetical example that borrows from our empirical findings over the last five years of our sample period (to be presented later). In Panel A, we show the distribution of fund t-statistics for each of the three skill categories described above.[5] Specifically, the mean t is fixed at -2.5 for unskilled and 3.0 for skilled funds, which corresponds to an annual four-factor alpha of -3.2% and 3.8%, respectively (the choice of these values is explained in Appendix B). The t-distribution shown in Panel B is the cross-section that (hypothetically) would be observed by a researcher. This distribution is a mixture of the three distributions in Panel A, where the weight on each distribution is equal to the proportion of zero-alpha, unskilled, and skilled funds that we estimate from our empirical tests to come later (specifically, $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$).

Please insert Figure 1 here

To illustrate further, suppose that we choose a significance level, γ, of 10% (correspond-ing to $t_\gamma^- = -1.65$ and $t_\gamma^+ = 1.65$). With the test shown in Equation (1), the researcher

but the fund management company overcharges fees or inefficiently generates other services (such as administrative services, such as record-keeping)–costs that the manager usually has little control over. In a later section of this paper, we redefine stockpicking skill in an absolute sense (net of trading costs only) and revisit some of our basic tests to be described.

[5] For simplicity, this figure illustrates the (hypothetical) distributions of the fund t-statistics as being normal, even though the distribution is non-normal for most U.S. domestic equity funds (KTTW (2006)). In our empirical section to follow, we use a bootstrap approach to more accurately determine the distribution of t-statistics for each fund (and their associated p-values).

expects to find 5.4% of funds with a positive and significant t-statistic.[6] This proportion, denoted by $E(S_\gamma^+)$, is represented by the shaded region in the right tail of the cross-sectional t-distribution (Panel B). Does this area consist merely of skilled funds, as defined above? Clearly not, because some funds are just lucky; as shown in the shaded region of the right tail in Panel A, zero-alpha funds can be lucky and exhibit positive and significant estimated t's. By the same token, the proportion of funds with a negative and significant t-statistic (the shaded region in the left-tail of Panel B) overestimates the proportion of unskilled funds, because it includes some unlucky zero-alpha funds (the shaded region in the left tail in Panel A).[7]

The message conveyed by Figure 1 is that we measure performance with a limited sample of data, therefore, unskilled and skilled funds cannot easily be distinguished from zero-alpha funds. This issue is compounded by the small magnitude of alphas, relative to the sampling error from asset pricing models. To proceed, we must account for "false discoveries," i.e., funds that falsely exhibit significant estimated alphas (i.e., their true alphas are zero).

A.2 Measuring Luck

We next describe our approach to estimating the proportion of funds that lie in the tails of the cross-sectional t-distribution purely by (good or bad) luck. At a given significance level γ, the probability that a zero-alpha fund (as defined in the last section) exhibits luck equals $\gamma/2$ (shown as the dark shaded region in Panel A of Figure 1)). If the proportion of zero-alpha funds in the population is π_0, the expected proportion of "lucky funds" (zero-alpha funds with positive and significant t-statistics) is equal to

$$E(F_\gamma^+) = \pi_0 \cdot \gamma/2. \tag{2}$$

Now, to determine the expected proportion of skilled funds, $E(T_\gamma^+)$, we simply adjust $E(S_\gamma^+)$ for the presence of these lucky funds:

$$E(T_\gamma^+) = E(S_\gamma^+) - E(F_\gamma^+) = E(S_\gamma^+) - \pi_0 \cdot \gamma/2. \tag{3}$$

---

[6]From Panel A, the probability that the observed t-statistic is greater than $t_\gamma^+ = 1.65$ equals 5% for a zero-alpha fund and 84% for a skilled fund. Multiplying these two probabilities by the respective proportions respresented by their categories gives 5.4%.

[7]Note that we have not considered the possibility that skilled funds could be *very* unlucky, and exhibit a negative and significant t-statistic. In our example of Figure 1, the probability that the estimated t-statistic of a skilled fund is lower than $t_\gamma^- = -1.65$ is less than 0.001%. This probability is negligible, so we ignore this pathological case. The same comment applies to unskilled funds who are very lucky.

Since the probability of being unlucky is also equal to $\gamma/2$ (i.e., the grey and black areas in Panel A of Figure 1 are identical), $E(F_\gamma^-)$, the expected proportion of "unlucky funds," is equal to $E(F_\gamma^+)$. As a result, the expected proportion of unskilled funds, $E(T_\gamma^-)$, is similarly given by

$$E(T_\gamma^-) = E(S_\gamma^-) - E(F_\gamma^-) = E(S_\gamma^-) - \pi_0 \cdot \gamma/2. \qquad (4)$$

What is the role played by the significance level, γ, chosen by the researcher? By defining the significance thresholds t_γ^- and t_γ^+, γ determines the portion of the right (or left) tail which is examined for lucky vs. skilled funds (unlucky vs. unskilled), as described by Equations (3) and (4). By varying γ, we can measure the proportion of skilled funds in any segment of the cross-section of funds (ranked by their t-statistics).

This flexibility in choosing γ provides us with important opportunities for insights into the merits of active fund management. First, by choosing a larger γ (i.e., lower t_γ^- and t_γ^+, in absolute value), we can estimate the proportions of unskilled and skilled funds in a larger portion of the left and right tails of the cross-sectional t-distribution, respectively–thus, giving us an appreciation of the prevalence of unskilled and skilled funds in the entire population. That is, as we increase γ, $E(T_\gamma^-)$ and $E(T_\gamma^+)$ converge to π_A^- and π_A^+, thus minimizing Type II error (failing to locate truly unskilled or skilled funds). Alternatively, by reducing γ, we can determine the precise location of unskilled or skilled funds in the extreme tails of the t-distribution. For instance, choosing a very low γ (i.e., very large t_γ^- and t_γ^+, in absolute value) allows us to determine whether extreme tail funds are simply lucky (unlucky), or skilled (unskilled)–information that is quite useful to investors trying to locate skilled managers.

A.3 Estimation Procedure

The key to our approach to measuring luck in a group setting, as shown in Equation (2), is the estimator of the proportion, π_0, of zero-alpha funds in the population. Here, we turn to a novel estimation approach developed by Storey (2002)–called the "False Discovery Rate" (FDR) approach. The FDR approach is very straightforward, as its sole inputs are the p-values (two-sided) associated with the (alpha) t-statistics for each of the M funds. By definition, zero-alpha funds satisfy the null hypothesis, $H_{0,i} : \alpha_i = 0$, and, therefore, have p-values that are uniformly distributed over the interval $[0,1]$.[8] On the contrary, p-values of unskilled and skilled funds tend to be very small because their

---

[8]To see this, let us denote by t and p the t-statistic and p-value of the zero-alpha fund. We have $p = 1 - F(|t|)$, where $F(t) = prob(|\hat{t}_i| < |t| \,|\, \alpha_i = 0)$. The p-value is uniformly distributed over $[0, 1]$ since its cdf, $G(p) = prob(\hat{p}_i < p) = prob(1 - F(|t_{\hat{p}_i}|) < p) = prob(|t_{\hat{p}_i}| > F^{-1}(1-p)) = 1 - F(F^{-1}(1-p)) = p$.

estimated t's tend to be far from zero (see Panel A of Figure 1). We can exploit this information to estimate π_0 without knowing the exact distribution of the p-values of the unskilled and skilled funds.

Thus, a key intuition of the FDR approach is that it uses information from the center of the cross-sectional t-distribution (where estimated t-statistics are close to zero) to draw inferences about the tails (where t-statistics are very large, in absolute value). To illustrate this procedure, consider a population of 2,076 funds (the number of funds in our study, which we will introduce shortly). Suppose, for each fund, we draw its t-statistic from one of the three t-distributions in Panel A of Figure 1, with probability according to our estimate of the proportion of unskilled, zero-alpha, and skilled funds in the population, $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$. From these t-statistics, we compute two-sided p-values, $\widehat{p}_i$, for each of the 2,076 funds, then plot them in Figure 2.

Please insert Figure 2 here

The darkest grey zone near zero represents the proportion of p-values in the population that correspond to the unskilled and skilled funds ($\pi_A^- + \pi_A^+ = 25\%$). The area below the horizontal line at 0.075 equals the true (unknown) proportion, π_0, of zero-alpha funds (75%), since zero-alpha funds have uniformly distributed p-values. Since this quantity is unknown, we must estimate it from the histogram of observed p-values as follows. If we take a sufficiently high threshold λ^* (e.g., $\lambda^* = 0.6$), we know that the vast majority of p-values higher than λ^* come from zero-alpha funds. Thus, we first estimate the area covered by the four lightest grey bars on the right of λ^*, $\widehat{W}(\lambda^*)/M$ (where $\widehat{W}(\lambda^*)$ denotes the number of funds having p-values exceeding λ^*), then simply extrapolate this area over the entire interval $[0,1]$ by multiplying by $1/(1-\lambda^*)$ (if $\lambda^* = 0.6$, the area is multiplied by 2.5):[9]

$$\widehat{\pi}_0(\lambda^*) = \frac{\widehat{W}(\lambda^*)}{M} \cdot \frac{1}{(1-\lambda^*)}. \tag{5}$$

To select λ^*, we use the data-driven approach suggested by Storey (2002) and explained in detail in Appendix A.

Substituting the estimate $\widehat{\pi}_0$ in Equations (2), (3), and replacing $E(S_\gamma^+)$ with the observed proportion of significant funds in the right tail, $\widehat{S}_\gamma^+$, we can easily estimate $E(F_\gamma^+)$, and $E(T_\gamma^+)$ corresponding to any chosen significance level, γ. The same approach

---

[9]This estimation procedure cannot be used in a one-sided multiple test, since the null hypothesis is tested under the least favorable configuration (LFC). For instance, consider the following null hypothesis $H_{0,i} : \alpha_i \leq 0$. Under the LFC, it is replaced with $H_{0,i} : \alpha_i = 0$. Therefore, all funds with $\alpha_i \leq 0$ (i.e., drawn from the null) have inflated p-values which are not uniformly distributed over $[0,1]$.

can be used in the left tail by replacing $E(S_\gamma^-)$ with the observed proportion of significant funds in the left tail, $\widehat{S}_\gamma^-$. This implies the following estimates of the proportions of unlucky and lucky funds:

$$\widehat{F}_\gamma^- = \widehat{F}_\gamma^+ = \widehat{\pi}_0 \cdot \gamma/2. \tag{6}$$

Using Equation (6), the estimated proportion of unskilled and skilled funds (at the chosen significance level, γ) are, respectively, equal to

$$\begin{aligned} \widehat{T}_\gamma^- &= \widehat{S}_\gamma^- - \widehat{F}_\gamma^- = \widehat{S}_\gamma^- - \widehat{\pi}_0 \cdot \gamma/2, \\ \widehat{T}_\gamma^+ &= \widehat{S}_\gamma^+ - \widehat{F}_\gamma^+ = \widehat{S}_\gamma^+ - \widehat{\pi}_0 \cdot \gamma/2. \end{aligned} \tag{7}$$

Finally, we estimate the proportions of unskilled and skilled funds in the population as

$$\widehat{\pi}_A^- = \widehat{T}_{\gamma^*}^-, \qquad \widehat{\pi}_A^+ = \widehat{T}_{\gamma^*}^+, \tag{8}$$

where γ^* is a sufficiently high significance level–we choose γ^* with a simple data-driven method explained in Appendix A.

B   Comparison of Our Approach with Existing Methods

The previous literature has followed two alternative approaches when estimating the proportion of unskilled and skilled funds. The "full luck" approach proposed by Jensen (1968) and Ferson and Qian (2004) assumes, a priori, that all funds in the population have zero alphas, $\pi_0 = 1$. Thus, for a given significance level, γ, this approach implies an estimate of the proportions of unlucky and lucky funds equal to $\gamma/2$.[10] At the other extreme, the "no luck" approach reports the observed number of significant funds (for instance, Ferson and Schadt (1996)) without making a correction for luck.

What are the errors committed by assuming, a priori, that π_0 equals 0 or 1, when it does not accurately describe the population? To address this question, we compare the bias produced by these two approaches relative to our FDR approach across the possible range of values for π_0 ($\pi_0 \in [0, 1]$) using our simple framework of Figure 1. Our procedure consists of three steps. First, for a chosen value of π_0, we create a simulated sample of 2,076 fund t-statistics (corresponding to our fund sample size) by randomly drawing from the three distributions in Panel A of Figure 1 in the proportions π_0, π_A^-, and π_A^+.[11] Second, we use these sampled t-statistics to estimate the proportion of

---

[10] Jensen (1968) summarizes the "full luck" approach as follows: "...if all the funds had a true α equal to zero, we would expect (merely by random chance) to find 5% of them having t values 'significant' at the 5% level."

[11] For each chosen π_0, the values for π_A^- and π_A^+ are fixed such that their relative proportions remain

unlucky ($\alpha = 0$, $\widehat{\alpha} < 0$), lucky ($\alpha = 0$, $\widehat{\alpha} > 0$), unskilled ($\alpha < 0$, $\widehat{\alpha} < 0$), and skilled ($\alpha > 0$, $\widehat{\alpha} > 0$) funds under each of the three approaches–the "no luck," "full luck," and FDR techniques.[12] Third, under each approach, we repeat these first two steps 1,000 times to compare the average value of each estimator with its true population value.

Please insert Figure 3 here

Specifically, Panel A of Figure 3 compares the three estimators of the expected proportion of unlucky funds. The true population value, $E(F_\gamma^-)$, is an increasing function of π_0 by construction, as indicated by Equation (2). While the average value of the FDR estimator closely tracks $E(F_\gamma^-)$, this is not the case for the other two approaches. Note that, by assuming that $\pi_0 = 0$, the "no luck" approach consistently underestimates $E(F_\gamma^-)$ when $\pi_0 > 0$. Conversely, the "full luck" approach, which assumes that $\pi_0 = 1$, overestimates $E(F_\gamma^-)$ for $\pi_0 < 1$. To illustrate the extent of the bias, consider the case where $\pi_0 = 75\%$. While the "no luck" approach substantially underestimates $E(F_\gamma^-)$ (0% instead of its true value of 7.5%), the "full luck" approach overestimates $E(F_\gamma^-)$ (10% instead of its true 7.5%). The biases for estimates of lucky funds $E(F_\gamma^+)$ shown in Panel B are exactly the same, since $E(F_\gamma^+) = E(F_\gamma^-)$.

The estimators of the expected proportions of unskilled and skilled funds ($E(T_\gamma^-)$ and $E(T_\gamma^+)$) provided by the three estimators are shown in Panels C and D, respectively. As we choose a higher proportion of zero-alpha funds (a higher value of π_0), the true proportions of unskilled and skilled funds, $E(T_\gamma^-)$ and $E(T_\gamma^+)$, decrease by construction. In both panels, our FDR estimator accurately captures this feature, while the other approaches do not fare well due to their fallacious assumptions about the prevalence of luck. For instance, when $\pi_0 = 75\%$, the "no luck" approach exhibits a large upward bias in its estimates of the total proportion of unskilled and skilled funds, $E(T_\gamma^-) + E(T_\gamma^+)$ (37.3% rather than the correct value of 22.3%). At the other extreme, the "full luck" approach underestimates $E(T_\gamma^-) + E(T_\gamma^+)$ (17.8% instead of 22.3%).

When the true proportion of skilled funds in the population is low, the "no luck" and "full luck" approaches exhibit a nonsensical positive relation between π_0 and $E(T_\gamma^+)$. With an increased proportion of zero-alpha funds, the proportion of lucky funds in the right tail increases by construction, yet both approaches erroneously exhibit an increase in the proportion of skilled funds (panel D). While the "no luck" approach wrongly considers these lucky funds as skilled, the "full luck" approach makes an excessive luck

unchanged ($\pi_A^-/\pi_A^+ = 0.23/0.02$).

[12]We choose $\gamma = 0.20$ to examine a large portion of the tails of the cross-sectional t-distribution, although other values for γ provide similar results.

10

adjustment and actually produces negative estimates of $E(T_\gamma^+)$.

In addition to the bias properties of our FDR estimators, their variability is low because of the large cross-section of funds ($M = 2,076$). To understand this, consider our main estimator $\widehat{\pi}_0$ (the same arguments apply to the other estimators). Since $\widehat{\pi}_0$ is a proportion estimator that depends on the proportion of $\widehat{p}_i > \lambda^*$, the Law of Large Numbers drives it close to its true value with our large sample size. For instance, taking $\lambda^* = 0.6$ and $\pi_0 = 75\%$, we find that σ_{π_0} is as low as 2.5% with independent p-values ($1/30^{th}$ the magnitude of π_0).[13] In the appendix, we provide further evidence of the remarkable accuracy of our estimators based on Monte-Carlo simulations.

C   Estimation under Cross-Sectional Dependence among Funds

Mutual funds can have correlated residuals if they "herd" in their stockholdings (Wermers (1999)) or hold similar industry allocations. In general, cross-sectional dependence in fund estimated alphas greatly complicates performance measurement. Any inference test with dependencies becomes quickly intractable as M rises, since this requires the estimation and inversion of an $M \times M$ residual covariance matrix. In a Bayesian framework, Jones and Shanken (2005) show that performance measurement requires intensive numerical methods when investor prior beliefs about fund alphas include cross-fund dependencies. Further, KTWW (2006) show that a complicated bootstrap is necessary to test the significance of performance of a fund located at a particular alpha rank, since this test depends on the joint distribution of all fund estimated alphas–cross-correlated fund residuals must be bootstrapped simultaneously.

An important advantage of our approach is that we estimate the p-value of each fund in isolation–avoiding the complications that arise because of the dependence structure of fund residuals. However, high cross-sectional dependencies could potentially bias our estimators. To illustrate this point with an extreme case, suppose that all funds are zero-alpha ($\pi_0 = 100\%$), and that fund residuals are perfectly correlated (perfect herding). In this case, all fund p-values would be the same, and the p-value histogram would not converge to the uniform distribution, as shown in Figure 2 Clearly, we would make serious errors no matter where we set λ^*.

In our sample, we are not very concerned with dependencies, since we find that the average correlation between four-factor model residuals of pairs of funds is only 0.08.

---

[13]Specifically, $\widehat{\pi}_0 = (1 - \lambda^*)^{-1} \cdot 1/M \sum_{i=1}^{M} x_i$, where x_i follows a binomial distribution with probability of success $P_{\lambda^*} = prob(\widehat{p}_i > \lambda^*) = 0.30$ (i.e., the rectangle area delimited by the horizontal black line and the vertical line at $\lambda^* = 0.6$ in Figure 2). Therefore, we have $\sigma_x = (P_{\lambda^*}(1 - P_{\lambda^*}))^{\frac{1}{2}} = 0.46$, and $\sigma_{\pi_0} = (1 - \lambda^*)^{-1} \cdot \sigma_x/\sqrt{M} = 2.5$.

Further, many of our funds do not have highly overlapping return data, thus, ruling out highly correlated residuals by construction. Specifically, we find that 15% of the funds pairs do not have a single monthly return observation in common; on average, only 55% of the return observations of fund pairs is overlapping. As a result, we believe that cross-sectional dependencies are sufficiently low to produce consistent estimators (i.e., mutual fund residuals satisfy the ergodicity conditions discussed in Storey, Taylor, and Siegmund (2004)).

However, in order to explicitly verify the properties of our estimators, we run a Monte-Carlo simulation. In order to closely reproduce the actual pairwise correlations between funds in our dataset, we estimate the residual covariance matrix directly from the data, then use these dependencies in our simulations. In further simulations, we impose other types of dependencies, such as residual block correlations or residual factor dependencies, as in Jones and Shanken (2005). In all simulations, we find both that average estimates (for all of our estimators) are very close to their true values, and that confidence intervals for estimates are comparable to those that result from simulations where independent residuals are assumed. For brevity, these results, as well as further details on the simulation experiment are discussed in Appendix B.

II   Performance Measurement and Data Description

A   Asset Pricing Models

To compute fund performance, our baseline asset pricing model is the four-factor model proposed by Carhart (1997):

$$r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + \varepsilon_{i,t}, \qquad (9)$$

where $r_{i,t}$ is the month t excess return of fund i over the riskfree rate (proxied by the monthly T-bill rate); $r_{m,t}$ is the month t excess return on the value-weighted market portfolio; and $r_{smb,t}$, $r_{hml,t}$, and $r_{mom,t}$ are the month t returns on zero-investment factor-mimicking portfolios for size, book-to-market, and momentum obtained from Kenneth French's website.

We also implement a conditional four-factor model to account for time-varying exposure to the market portfolio (Ferson and Schadt (1996)),

$$r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + B^{'}\left(z_{t-1} \cdot r_{m,t}\right) + \varepsilon_{i,t}, \quad (10)$$

where z_{t-1} denotes the $J \times 1$ vector of centered predictive variables, and B is the $J \times 1$ vector of coefficients. The four predictive variables are the one-month T-bill rate; the dividend yield of the CRSP value-weighted NYSE/AMEX stock index; the term spread, proxied by the difference between yields on 10-year Treasurys and three-month T-bills; and the default spread, proxied by the yield difference between Moody's Baa-rated and Aaa-rated corporate bonds. We have also computed fund alphas using the CAPM and the Fama-French (1993) models. These results are summarized in Section 3.5.

To compute each fund t-statistic, we use the Newey-West (1987) heteroscedasticity and autocorrelation consistent standard deviation, $\widehat{\sigma}_{\widehat{\alpha}_i}$. Further, KTWW (2006) find that the finite-sample distribution of $\widehat{t}$ is non-normal for approximately half of the funds. Therefore, we use a bootstrap procedure (instead of asymptotic theory) to compute fund p-values. In order to approximate the distribution of $\widehat{t}_i$ for each fund i under the null hypothesis $\alpha_i = 0$, we use a residual-only bootstrap procedure, which draws with replacement from the regression estimated residuals $\{\widehat{\varepsilon}_{i,t}\}$.[14] For each fund, we implement 1,000 bootstrap iterations. The reader is referred to KTWW (2006) for details on this bootstrap procedure.

B Mutual Fund Data

We use monthly mutual fund return data provided by the Center for Research in Security Prices (CRSP) between January 1975 and December 2006 to estimate fund alphas. Each monthly fund return is computed by weighting the net return of its component shareclasses by their beginning-of-month total net asset values. The CRSP database is matched with the Thomson/CDA database using the MFLINKs product of Wharton Research Data Services (WRDS) in order to use Thomson fund investment-objective information, which is more consistent over time. Wermers (2000) provides a description of how an earlier version of MFLINKS was created. Our original sample is free of survivorship bias, but we further select only funds having at least 60 monthly return observations in order to obtain precise four-factor alpha estimates. These monthly returns need not be contiguous. However, when we observe a missing return, we delete the following-month return, since CRSP fills this with the cumulated return since the last non-missing return. In unreported results, we find that reducing the minimum fund

---

[14]To determine whether assuming homoscedasticity and temporal independence in individual fund residuals is appropriate, we have checked for heteroscedasticity (White test), autocorrelation (Ljung-Box test), and Arch effects (Engle test). We have found that only a few funds present such regularities. We have also implemented a block bootstrap methodology with a block length equal to $T^{\frac{1}{5}}$ (proposed by Hall, Horowitz, and Jing (1995)), where T denotes the length of the fund return time-series. All of our results to be presented remain unchanged.

return requirement to 36 months has no material impact on our main results, thus, we believe that any biases introduced from the 60-month requirement are minimal.

Our final universe has 2,076 open-end, domestic equity mutual funds existing for at least 60 months between 1975 and 2006. Funds are classified into three investment categories: Growth (1,304 funds), Aggressive Growth (388 funds), and Growth & Income (384 funds). If an investment objective is missing, the prior non-missing objective is carried forward. A fund is included in a given investment category if its objective corresponds to the investment category for at least 60 months.

Table I shows the estimated annualized alpha as well as factor loadings of equally-weighted portfolios within each category of funds. The portfolio is rebalanced each month to include all funds existing at the beginning of that month. Results using the unconditional and conditional four-factor models are shown in Panels A and B, respectively.

<center>Please insert Table I here</center>

Similar to results previously documented in the literature, we find that unconditional estimated alphas for each category is negative, ranging from -0.45% to -0.60% per annum. Aggressive Growth funds tilt towards small capitalization, low book-to-market, and momentum stocks, while the opposite holds for Growth & Income funds. Introducing time-varying market betas provides similar results (Panel B). In tests available upon request from the authors, we find that all results to be discussed in the next section are qualitatively similar whether we use the unconditional or conditional version of the four-factor model. For brevity, we present only results from the unconditional four-factor model.

III   Empirical Results

A   Impact of Luck on Long-Term Performance

We begin our empirical analysis by measuring the impact of luck on long-term mutual fund performance. To this end, we estimate the performance of each fund during its entire life (during the period 1975-2006) using the four-factor model of Equation (9). Panel A of Table II shows estimated proportions of zero-alpha, unskilled, and skilled funds in the population ($\widehat{\pi}_0$, $\widehat{\pi}_A^-$, and $\widehat{\pi}_A^+$), as defined in Section I.A.1, with standard deviations of estimates in parentheses. These point estimates are computed using the procedure described in Section I.A.3, while standard deviations are computed using the method proposed by Genovese and Wasserman (2004)–which is described in the

<center>14</center>

appendix.

Among the 2,076 funds, we estimate that a large majority–75.4%–are zero-alpha funds. Managers of these funds exhibit stockpicking skills just sufficient to cover their trading costs and other expenses (including fees). These funds, therefore, capture all of the economic rents that they generate–consistent with the long-run predictions of Berk and Green (2004).

Further, it is quite surprising that the estimated proportion of skilled funds is statistically indistinguishable from zero (see "Skilled" column). This result may seem surprising in light of some prior studies, such as Ferson and Schadt (1996), which find that a small group of top mutual fund managers appear to outperform their benchmarks, net of costs. However, a closer examination–in Panel B–shows that our adjustment for luck is key in understanding the difference between our study and those of prior researchers.

To be specific, Panel B shows the proportion of significant alpha funds in the left and right tails ($\widehat{S}_\gamma^-$ and $\widehat{S}_\gamma^+$, respectively) at four different significance levels ($\gamma = 0.05$, 0.10, 0.15, 0.20). Similar to past research, there are many significant alpha funds in the right tail–$\widehat{S}_\gamma^+$ peaks at 8.2% of the total population (170 funds) at $\gamma = 0.20$. However, of course, "significant alpha" does not always mean "skilled fund manager." Illustrating this point, the right side of Panel B decomposes these significant funds into the proportions of lucky zero-alpha funds and truly skilled funds ($\widehat{F}_\gamma^+$ and $\widehat{T}_\gamma^+$, respectively). Clearly, we cannot reject that all of the right tail funds are merely lucky outcomes among the large number of zero-alpha funds (1,565), and that none of these right-tail funds have truly skilled fund managers.

The bottom of Panel B presents characteristics of the average fund in each segment of the tails. Although the average estimated alpha of right-tail funds is somewhat high (between 4.8% and 6.5% per year), this is simply due to very lucky outcomes for a small proportion of the 1,565 zero-alpha funds in the population. It is also interesting that expense ratios are higher for left-tail funds, which likely explains some of the underperformance of these funds (we will revisit this issue when we examine pre-expense returns in a later section). Turnover does not vary systematically among the various tail segments, but left-tail funds are much smaller than right-tail funds, presumably due to the combined effects of flows and investment returns. Results for the three investment-objective subgroups (Aggressive Growth, Growth, and Growth & Income) are similar–these results are available upon request from the authors.

Please insert Table II here

15

It is interesting that Panel A shows that 24% of the population (499 funds) are unskilled fund managers–unable to pick stocks well enough to recover their trading costs and other expenses.[15] While the majority of funds in our sample are zero-alpha, consistent with the long-run predicted equilibrium of Berk and Green (2004), it is puzzling that a substantial number of negative alpha funds can survive. In fact, in untabulated results, we find that left-tail funds, which are overwhelmingly comprised of unskilled (and not merely unlucky) funds, have a relatively long fund life–12.6 years, on average. Perhaps, as discussed by Elton, Gruber, and Busse (2003), such funds exist if they are able to attract a sufficient number of unsophisticated investors, who are also charged higher fees (Christoffersen and Musto (2002)).

As mentioned earlier, the universe of U.S. domestic equity mutual funds has expanded substantially since 1990. Accordingly, we next examine the evolution of the proportion of unskilled and skilled funds over time. To accomplish this, at the end of each year from 1989 to 2006, we estimate the proportion of unskilled and skilled funds using the entire return history for each fund up to that point in time. For instance, our initial estimates, on December 31, 1989, cover the first 15 years of the sample (1975-89), while our final estimates, on December 31, 2006, are based on the entire period 1975-2006 (i.e., these are the estimates shown in Panel A of Table II).[16] The results in Panel A of Figure 4 show that the proportion of funds with non-zero alphas (equal to the sum of the proportion of skilled plus unskilled funds) remains fairly constant over time. However, there are dramatic changes in the relative proportions of unskilled and skilled funds: from 1989 to 2006, the proportion of skilled funds declines from 14.4% to 0.6%, while the proportion of unskilled rises from 9.2% to 24.0% of the entire universe of funds. These changes are also reflected in the population average estimated alpha (shown in Panel B), which drops from 0.16% to -0.97% per year over the same period.

Please insert Figure 4 here

Panel B also displays the yearly count of funds that we analyze. From 1996 to 2005, there are more than 100 new actively managed domestic-equity mutual funds per year.[17] Interestingly, this coincides with the time trend in unskilled and skilled funds

---

[15]This minority of funds is the driving force explaining the negative average estimated alpha that is widely documented in the literature (e.g., Jensen (1968), Carhart (1997), Elton et al. (1993), and Pastor and Stambaugh (2002a)).

[16]To be included in a given year, a fund must have at least 60 monthly return observations, although these observations need not be contiguous.

[17]Since we require 60 monthly observations to measure the fund performance, this rise reflects the massive entry of new funds over the period 1993-2001.

shown in Panel A. Thus, the huge increase in numbers of actively managed mutual funds has resulted in a much larger proportion of unskilled funds, at the expense of skilled funds. Either the growth of the fund industry has coincided with greater levels of stock market efficiency, making stockpicking a more difficult and costly endeavor, or the large number of new managers simply have inadequate skills. It is also interesting that during our period of analysis, many fund managers with good track records left the sample to manage hedge funds, and that indexed investing increased substantially.

B Impact of Luck on Short-Term Performance

Our above results indicate that funds do not appear to provide superior long-term performance, but that they may provide outperformance over the short-run. This prompts us to more formally examine mutual fund performance over short time horizons. Specifically, we partition our data into six non-overlapping subperiods of five years, beginning with 1977-1981 and ending with 2002-2006. For each subperiod, we select all funds having 60 monthly return observations, then compute their respective alpha p-values–in other words, we treat each fund during each five-year period as a separate fund.[18] After pooling all of these subperiods together, we obtain a total of 3,311 p-values from which we compute our different estimators. Results for the entire population (All Funds) are shown in Table III, while results for Growth, Aggressive Growth, and Growth & Income funds are displayed in Panels A, B, and C of Table IV, respectively. In each panel, the rightmost columns examine the right tail of the cross-sectional t-distribution. For each significance level ($\gamma = 0.05, 0.10, 0.15, 0.20$), we show the proportion of significant funds $\left(\widehat{S}_\gamma^+\right)$, and its decomposition into lucky and skilled funds ($\widehat{F}_\gamma^+$, $\widehat{T}_\gamma^+$, respectively; standard deviations of the point estimates are shown in parentheses). The leftmost columns repeat this analysis for the left tail.

<center>Please insert Table III here</center>

First, note that a small fraction of funds exhibit skill over the short-run. Specifically, Panel A of Table III shows that 2.4% of funds exhibit such skills (which is statistically significant, since the standard deviation is 0.7%). Thus, short-term superior performance is rare, but does exist, as opposed to long-term performance. Second, these skilled funds are located in the extreme right tail of the cross-sectional t-distribution. Panel B of Table III shows that, with a γ of only 10%, we capture almost all skilled funds, as $\widehat{T}_\gamma^+$ reaches

---

[18]Note that reducing the number of observations comes at a cost: it greatly increases the standard deviation of the estimated alphas, making the p-values of non-zero alpha funds harder to distinguish from those of zero-alpha funds.

2.3% (close to its maximum value of 2.4%). Thus, proceeding toward the center of the distribution (by increasing γ to 0.15 and 0.20) produces mainly additional zero-alpha funds that are lucky. This finding indicates that skilled fund managers, while rare, may be somewhat easy to find, since they have extremely high t-statistics (extremely low p-values)–we will use this finding in our next section, where we attempt to find funds with out-of-sample skills. It is notable that we find evidence of short-term outperformance of some funds here, but no evidence of long-term outperformance in the last section of this paper. These findings are consistent with the model of Berk and Green (2004), where outperforming funds exist only until investors are successfully able to locate them.

In the left tail, we observe that the great majority of funds are unskilled, and not merely unlucky zero-alpha funds. For instance, in the extreme left tail (at $\gamma = 0.05$), the proportion of unskilled funds, $\widehat{T}_\gamma^-$, is roughly five times the proportion of unlucky funds, $\widehat{F}_\gamma^-$ (9.4% versus 1.8%). Here, the short-term results are similar to the prior-discussed long-term results–the great majority of left-tail funds are truly unskilled. It is also interesting that true skills seem to be inversely related to turnover, as indicated by the substantially higher levels of turnover of left-tail funds (which are mainly composed of unskilled funds). Unskilled managers apparently trade frequently to appear skilled, which ultimately hurts their performance.

Table IV shows results for investment-objective subgroups. Panel A shows that the proportion of skilled Growth funds in the right-tail is similar to that of the entire universe (discussed above). However, Aggressive-Growth funds (Panel B) exhibit somewhat higher skills. For instance, at $\gamma = 0.05$, 73% of significant Aggressive-Growth funds are truly skilled (3.1/4.9). On the other hand, Panel C shows that no Growth & Income funds are truly skilled, but that a substantial proportion of them are unskilled. The long-term existence of this category of actively-managed funds, which includes "value funds" and "core funds" is remarkable in light of these poor results.

Please insert Table IV here

C Performance Persistence

Our previous analysis reveals that only 2.4% of the funds are skilled over the short-term. Can we detect these skilled funds over time, in order to capture their superior alphas? Ideally, we would like to form a portfolio containing only the truly skilled funds in the right tail; however, since we only know which segment of the tails in which they lie, but not their identity, such an approach is not feasible.

Nevertheless, the reader should recall from the last section that skilled funds are

located in the extreme right tail. By forming portfolios containing all funds in this extreme tail, we have a greater chance of capturing the superior alphas of the truly skilled ones. For instance, Panel A of Table III shows that, at $\gamma = 0.05$, the proportion of skilled funds among all significant funds, $\widehat{T}_\gamma^+ / \widehat{S}_\gamma^+$, is about 50%, which is much higher than the proportion of skilled funds among the entire universe, 2.4%.

To select a portfolio of funds, we use the False Discovery Rate in the right tail, FDR^+. At a given significance level, γ, the FDR^+ is defined as the expected proportion of lucky funds among the significant funds in this tail:

$$FDR_\gamma^+ = E\left(\frac{F_\gamma^+}{S_\gamma^+}\right). \tag{11}$$

The FDR_γ^+ provides a simple portfolio formation rule.[19] When we set a low FDR^+ target, we only allow a small proportion of lucky funds ("false discoveries") in the chosen portfolio. Specifically, we set a sufficiently low significance level, γ, so as to include skilled funds along with a small number of zero-alpha funds that are extremely lucky. Conversely, increasing the FDR^+ target has two opposite effects on a portfolio. First, it decreases the portfolio's expected future performance, since the proportion of lucky funds in the portfolio is higher. However, it also increases its diversification, since more funds are selected. Accordingly, we examine five FDR^+ target levels in our persistence test: 10%, 30%, 50%, 70%, and 90%.

The construction of the portfolios proceeds as follows. At the end of each year, we estimate the alpha p-values of each existing fund using the previous five-year period. Using these p-values, we estimate the FDR_γ^+ over a range of chosen significance levels ($\gamma = 0.01, 0.02,..., 0.60$). Following Storey (2002) and Storey and Tibshirani (2003), we propose the following straightforward estimator of the FDR_γ^+ :

$$\widehat{FDR}_\gamma^+ = \frac{\widehat{F}_\gamma^+}{\widehat{S}_\gamma^+} = \frac{\widehat{\pi}_0 \cdot \gamma/2}{\widehat{S}_\gamma^+}, \tag{12}$$

where $\widehat{\pi}_0$ is the estimator of the proportion of zero-alpha funds described in Section I.A.3. For each FDR^+ target level, we determine the significance level, γ^P, that provides an $\widehat{FDR}_{\gamma^P}^+$ as close as possible to this target. Then, only funds with p-values smaller than γ^P are included in an equally-weighted portfolio. This portfolio is held for one year,

---

[19]Our new measure, FDR_γ^+, is an extension of the traditional FDR introduced in the statistical literature (e.g., Benjamini and Hochberg (1995), Storey (2002)), since the latter does not distinguish between bad and good luck. The traditional measure is $FDR_\gamma = E\left(F_\gamma / S_\gamma\right)$, where $F_\gamma = F_\gamma^+ + F_\gamma^-$, $S_\gamma = S_\gamma^+ + S_\gamma^-$.

after which the selection procedure is repeated. If a selected fund does not survive after a given month, its weight is reallocated to the remaining funds during the rest of the year to mitigate survival bias. The first portfolio formation date is December 31, 1979 (after the first five years), while the last is December 31, 2005.

In Panel A of Table V, we show the FDR level ($\widehat{FDR}_{\gamma P}^{+}$) of the five portfolios, as well as the proportion of funds in the population that they include ($\widehat{S}_{\gamma P}^{+}$) during the five-year formation period, averaged over the 27 formation dates (from 1979 to 2005), as well as their respective distributions. First, we observe (as expected) that the achieved FDR increases with the FDR target assigned to a portfolio. However, the average $\widehat{FDR}_{\gamma P}^{+}$ does not always match its target. For instance, $FDR10\%$ achieves an average of 41.5%, instead of the targeted 10%. The reason for this shortfall is simple. During some periods, the proportion of skilled funds in the population is too low to achieve a 10% FDR target.[20] Of course, we also observe a positive relation between the FDR target and the proportion of funds included in a portfolio, since our selection becomes less restrictive.

In Panel B, we present the average out-of-sample performance (over the following year) of these five false-discovery controlled portfolios, starting January 1, 1980 and ending December 31, 2006. We compute the estimated annualized alpha, $\widehat{\alpha}$, along with its bootstrapped p-value; annualized residual standard deviation, $\widehat{\sigma}_{\varepsilon}$; information ratio, IR= $\widehat{\alpha}/\widehat{\sigma}_{\varepsilon}$; four-factor model loadings; annualized mean return (minus T-bills); and annualized time-series standard deviation of monthly returns. The results reveal that our FDR portfolios successfully detect funds with short-term skills. For example, the portfolios $FDR10\%$ and 30% produce out-of-sample alphas (net of expenses) of 1.45% and 1.15% per year (significant at the 5% level). As the FDR target rises to 90%, the proportion of funds in the portfolio increases, which improves diversification ($\widehat{\sigma}_{\varepsilon}$ falls from 4.0% to 2.7%). However, we also observe a sharp decrease in the alpha (from 1.45% to 0.39%), reflecting the large proportion of lucky funds contained in the $FDR90\%$ portfolio.

<center>Please insert Table V here</center>

Finally, Panel C examines portfolio turnover. For each portfolio, we determine the proportion of funds which are still selected 1, 2, 3, 4, and 5 years after their initial inclusion. The results sharply illustrate the short-term nature of truly outperforming funds. After 1 year, 40% or fewer of funds remain in portfolios $FDR10\%$ and 30%, while

---

[20]For instance, the minimum achievable FDR at the end of 2003 and 2004 is equal to 47.0% and 39.1%, respectively. If we look at the $\widehat{FDR}_{\gamma P}^{+}$ distribution for the portfolio FDR 10% in Panel A, we observe that in 6 years out of 27, the $\widehat{FDR}_{\gamma P}^{+}$ is higher than 70%.

after 3 years, these percentages are lower than 6%.

Finally, we examine, in Figure 5, how the estimated alpha of the portfolio $FDR10\%$ evolves over time. Our initial estimate, on December 31, 1989, covers the first 10 years of the portfolio return history (1980-89), while our final estimate, on December 31, 2006, is based on the entire period 1980-2006 (i.e., this is the estimated alpha shown in Panel B of Table V). The similarity with Figure 4 is striking. While the alpha accruing to the $FDR10\%$ portfolio is impressive at the beginning of the 1990s, it consistently declines thereafter. As the proportion, π_A^+, of skilled funds falls, the FDR approach moves much further to the extreme right tail of the cross-sectional t-distribution (from 5.7% of all funds in 1990 to 0.9% in 2006) in search of skilled funds. However, this change is not sufficient to prevent the performance of $FDR10\%$ from dropping substantially.

Please insert Figure 5 here

It is important to note the differences between our approach to persistence and that of the previous literature (e.g., Hendricks, Patel, and Zeckhauser (1993), Elton, Gruber, and Blake (1996), Carhart (1997)). These prior papers generally classify funds into fractile portfolios based on their past performance (past returns, estimated alpha, or alpha t-statistic) over a previous ranking period (one to three years). The size of fractile portfolios (e.g., deciles) are held fixed, with no regard to the changing proportion of lucky funds within these fixed fractiles. As a result, the signal used to form portfolios is likely to be noisier than our FDR approach. To compare these approaches with ours, Figure 5 displays the performance evolution of two top decile portfolios which are formed based on ranking funds by their alpha t-statistic, estimated over the previous one and three years, respectively.[21] Over most years, the FDR approach performs much better, consistent with the idea that it much more precisely detects skilled funds. However, this performance advantage declines during later years, when the proportion of skilled funds decreases substantially, making them much tougher to locate. Therefore, we find that the superior performance of the FDR portfolio is tightly linked to the prevalence of skilled funds in the population.

---

[21]We use the t-statistic to be consistent with the rest of our paper, but the results are qualitatively similar when we rank on the estimated alpha.

D   Additional Results

D.1   Performance Measured with Pre-Expense Returns

In our baseline framework described previously, we define a fund as skilled if it generates a positive alpha net of trading costs, fees, and other expenses. Alternatively, skill could be defined, in an absolute sense, as the manager's ability to produce a positive alpha before expenses are deducted. Measuring performance on a pre-expense basis allows one to disentangle the manager's stockpicking skills, net of trading costs but before expenses, from the fund's expense policy–which may be out of the control of the fund manager. To address this issue, we add monthly expenses (1/12 times the most recent reported annual expense ratio) to net returns for each fund, then revisit the long-term performance of the mutual fund industry.[22]

Panel A of Table VI contains the estimated proportions of zero-alpha, unskilled, and skilled funds in the population ($\widehat{\pi}_0$, $\widehat{\pi}_A^-$, and $\widehat{\pi}_A^+$), on a pre-expense basis. Comparing these estimates with those shown in Table II, we observe a striking reduction in the proportion of unskilled funds–from 26.6% to 4.5%. This result indicates that only a small fraction of fund managers have stockpicking skills that are insufficient to at least compensate for their trading costs. Instead, mutual funds produce negative net-of-expense alphas chiefly because they charge excessive fees, in relation to the selection abilities of their managers. In Panel B, we further find that the average expense ratio across funds in the left tail is lower when performance is measured prior to expenses (1.4% versus 1.6% per year), indicating that high fees (potentially charged to unsophisticated investors) are a chief reason why funds end up in the extreme left tail, net of expenses. In addition, turnover seems only weakly correlated with performance, before expenses, as opposed to our finding in the last section that turnover is negatively correlated, after expenses. Perhaps poor governance of some funds explains why they end up in the left tail–they overexpend on both trading costs (through high turnover) and other expenses relative to their skills.

<div align="center">Please insert Table VI here</div>

In the right tail, we find that 9.6% of fund managers have stockpicking skills sufficient to more than compensate for trading costs (Panel A). Consistent with Berk and Green (2004), the rents stemming from their skills are extracted through fees and other expenses, driving the proportion of net-expense skilled funds to zero.

Since 75.4% of funds produce zero net-expense alphas, it seems surprising that that

---

[22]We discard funds which do not have at least 60 pre-expense return observations over the period 1975-2006. This leads to a small reduction in our sample from 2,076 to 1,836 funds.

we do not find more pre-expense skilled funds. However, this is due to the relatively small impact of expense ratios on funds in the center of the cross-sectional t-distribution. Adding back these expenses leads only to a marginal increase in the alpha t-statistic, making the power of the tests rather low.[23]

Finally, in untabulated tests, we find that the proportion of skilled funds in the population decreases from 27.5% to 10% between 1996 and 2006. This implies that the decline in net-expense skills noted in Figure 4 is mostly driven by the reduction in stockpicking skills over time (as opposed to an increase in expenses for (pre-expense) skilled funds).

On the contrary, the proportion of pre-expense unskilled funds remains equal to zero until the end of 2003. Thus, poor stock-picking skills cannot expain the large increase in the proportion of unskilled funds (net of expenses) from 1996 onwards. Therefore, this increase is likely to be due to rising expenses charged by funds with no particular selection abilities, or the introduction of new funds with high expense ratios.

D.2 Performance Measured with Other Asset Pricing Models

Our estimation of the proportions of unskilled and skilled funds, ($\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$) obviously depends on the choice of the asset pricing model. To examine the sensitivity of our results, we repeat the long-term performance analysis using the (unconditional) CAPM and Fama-French models. Based on the CAPM, we find that $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$ are equal to 14.3% and 8.6% respectively, which is much more supportive of active management skills, compared to Section III.A.1. However, this result may be due to the omission of the size, book-to-market, and momentum factors. This conjecture is confirmed in Panel A of Table VII: the funds located in the right tail (according to the CAPM) have substantial loadings on the size and the book-to-market factors, which carry positive risk premia over our sample period (3.7% and 5.4% per year, respectively).

Please insert Table VII here

Turning to the Fama-French model, we find that $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$ amount to 25.0% and 1.7%, respectively. These proportions are very close to those obtained with the four-factor model, since only one factor is omitted. As expected, the 1.1% difference in

---

[23] The average expense ratio across funds with $|\widehat{\alpha}_i| < 1\%$ is approximately 10 bp per month. Adding back these expenses to a fund with zero net-expense alpha only increases its t-statistic mean from 0 to 0.9 (based on $T^{\frac{1}{2}}\alpha_A/\sigma_\varepsilon$, with $T = 384$, and $\sigma_\varepsilon = 0.021$). It implies that the null and alternative t-statistic distributions are extremely difficult to distinguish (for a fund with a pre-expense mean t of 0.9, the probability of observing a negative pre-expense t-statistic is equal to 18%!).

the estimated proportion of skilled funds between the two models (1.7%-0.6%) can be explained by the momentum factor. As shown in Panel B, the funds located in the right tail (according to the Fama-French model) have substantial loadings on the momentum factor, which carries a positive risk premium over the period (9.4% per year).

D.3 Bayesian Interpretation

Although we operate in a classical frequentist framework, our new FDR measure, FDR^+, also has a natural Bayesian interpretation.[24] To see this, we denote, by H_i, a random variable which takes the value of -1 if fund i is unskilled, 0 if it has zero alpha, and +1 if it is skilled. The prior probabilities for the three possible values (-1, 0, +1) are given by the proportion of each category in the population, π_A^-, π_0, and π_A^+. The Bayesian version of our FDR^+ measure, denoted by fdr_γ^+, is defined as the posterior probability that the fund i has a zero alpha given that its t-statistic, $\widehat{t_i}$, is positive and significant: $fdr_\gamma^+ = prob\left(H_i = 0 \middle| \widehat{t_i} \in \Gamma_A^+(\gamma)\right)$, where $\Gamma_A^+(\gamma) = \left(t_\gamma^+, +\infty\right)$. Using Bayes theorem, we have:

$$fdr_\gamma^+ = \frac{prob\left(\widehat{t_i} \in \Gamma_A^+(\gamma)\middle| H_i = 0\right) \cdot prob(H_i = 0)}{prob\left(\widehat{t_i} \in \Gamma_A^+(\gamma)\right)} = \frac{\gamma/2 \cdot \pi_0}{E(S_\gamma^+)}. \tag{13}$$

Stated differently, the fdr_γ^+ indicates how the investor changes his prior probability that fund i has a zero alpha ($H_i = 0$) after observing that its t-statistic is significant. In light of Equation (13), our estimator $\widehat{FDR}_\gamma^+ = (\gamma/2 \cdot \widehat{\pi}_0)/\widehat{S}_\gamma^+$ can therefore be interpreted as an empirical Bayes estimator of fdr_γ^+, where π_0 and $E(S_\gamma^+)$ are directly estimated from the data.[25]

In the recent Bayesian literature on mutual fund performance (e.g., Baks, Metrick, and Wachter (2001) and Pastor and Stambaugh (2002a)), attention is given to the posterior distribution of the fund alpha, α_i, as opposed to the posterior distribution of H_i. Interestingly, our approach also provides some relevant information for modeling the fund alpha prior distribution in an empirical Bayes setting. The parameters of the prior can be specified based on the relative frequency of the three fund categories (zero-alpha, unskilled, and skilled funds). In light of our estimates, an empirically-based alpha prior

---

[24]Our demonstration follows from the arguments used by Efron and Tibshirani (2002) and Storey (2003) for the traditional FDR, defined as $FDR_\gamma = E(F_\gamma/S_\gamma)$, where $F_\gamma = F_\gamma^+ + F_\gamma^-$, $S_\gamma = S_\gamma^+ + S_\gamma^-$.

[25]A full Bayesian estimation of fdr_γ^+ requires to posit prior distributions for the proportions π_0, π_A^-, and π_A^+, and for the distribution parameters of $\widehat{t_i}$ for each category. This method based on additional assumptions (including independent p-values) as well as intensive numerical methods is illustrated by Tang, Ghosal, and Roy (2007) in the case of the traditional FDR.

distribution is characterized by a point mass at $\alpha = 0$, reflecting the fact that 75.4% of the funds yield zero alphas, net of expenses. Since $\widehat{\pi}_A^-$ is higher than $\widehat{\pi}_A^+$, the prior probability of observing a negative alpha is higher than that of observing a positive alpha. These empirical constraints yield an asymmetric prior distribution. A tractable way to model the left and right parts of this distribution is to exploit two truncated normal distributions in the same spirit as in Baks, Metrick, and Wachter (2001). Further, we estimate that 9.6% of the funds have an alpha greater than zero, before expenses. While Baks, Metrick, and Wachter (2001) set this probability to 1% in order to examine the portfolio decision made by a skeptical investor, our analysis reveals that this level represents an overly skeptical belief.

IV   Conclusion

In this paper, we apply a new method for measuring the skills of fund managers in a group setting. Specifically, the "False Discovery Rate" (FDR) approach provides a simple and straightforward method to estimate the proportion of funds within a population that have stockpicking skills. In Monte Carlo simulations, we show that the FDR provides very accurate estimates of the proportion of skilled funds (those providing a positive alpha, net of trading costs and expenses), zero-alpha funds, and unskilled funds (those providing a negative alpha) in the entire population. Further, the FDR uses these estimates to provide accurate counts of skilled funds within various intervals in the right tail of the cross-sectional (estimated) alpha distribution, as well as unskilled funds within segments of the left tail.

We also apply the FDR technique to show that the proportion of skilled fund managers has diminished rapidly over the past 20 years. On the contrary, unskilled fund managers have increased substantially in the population over this period. Further analysis of pre-expense alphas shows that the increase in unskilled fund managers (net of expenses) is due to an increase in the number of funds who charge high fees while possessing no particular stockpicking skills.

Our paper focuses the long-standing puzzle of actively managed mutual fund underperformance on the minority of truly underperforming funds. Most actively managed funds provide either positive or zero net-of-expense alphas, putting them at least on par with passive funds. Still, it is puzzling why investors seem to increasingly tolerate the existence of a large minority of funds that produce negative alphas, when an increasing array of passively managed funds have become available (such as ETFs). Perhaps a class of unsophisticated or inattentive investors remain shareholders in funds after they

have clearly demonstrated (over time) their inferior returns. Or, as Elton, Gruber, and Blake (2007; EGB) document, maybe investors are forced to make constrained rational decisions–since EGB document that many 401(k) plans offer inefficient choices of mutual funds.

While our paper focuses on mutual fund performance, our approach has potentially wide applications in finance. It can be used in any setting in which a multiple hypothesis test is run and a large sample is available. We list two illustrative examples. First, technical trading can be implemented with a myriad of trading rules (e.g., Sullivan, Timmermann, and White (1999)). Our estimators can be used to determine the impact of luck on the performance of all these trading rules simultaneously. Second, testing the presence of commonality in liquidity boils down to regressing an individual stock liquidity measure on the market liquidity measure (e.g., Chordia, Roll, and Subrahmanyam (2000)). Since this regression is run for each individual stock, we are dealing with multiple testing. As a result, a correct measurement of commonality in liquidity necessitates a proper adjustment for luck. Because our approach only requires the estimation of π_0, controlling for luck in multiple testing is trivial: the only input required is a vector of p-values, one for each stock.

V   Appendix

A   Estimation Procedure

A.1   Determining the Value for λ^* from the Data

We use the bootstrap procedure proposed by Storey (2002) and Storey, Taylor, and Siegmund (2004). This resampling approach chooses λ such that an estimate of the Mean-Squared Error (MSE) of $\widehat{\pi}_0(\lambda)$ is minimized. First, we compute $\widehat{\pi}_0(\lambda)$ using Equation (5) across a range of λ values ($\lambda = 0.30, 0.35, ..., 0.70$). Second, for each possible value of λ, we form 1,000 bootstrap versions of $\widehat{\pi}_0(\lambda)$ by drawing with replacement from the $M \times 1$ vector of fund p-values. These are denoted by $\widehat{\pi}_0^b(\lambda)$, for $b = 1, ..., 1,000$. Third, we compute the estimated MSE for each possible value of λ:

$$\widehat{MSE}(\lambda) = \frac{1}{1,000} \sum_{b=1}^{1,000} \left[\widehat{\pi}_0^b(\lambda) - \min_\lambda \widehat{\pi}_0(\lambda) \right]^2. \qquad (14)$$

We choose λ^* such that $\lambda^* = \arg\min_\lambda \widehat{MSE}(\lambda)$. In unreported results (available upon request), we find that fixing λ to 0.5 or 0.6 yields similar results to those obtained

with the bootstrap procedure (see also Storey (2002)). Still, the main advantage of the bootstrap approach is that it is entirely data-driven.

A.2 Determining the Value for γ^* from the Data

Similar to the approach used to determine λ^*, we use a bootstrap procedure which minimizes the estimated MSE of $\widehat{\pi}_A^-(\gamma)$ and $\widehat{\pi}_A^+(\gamma)$.[26] First, we compute $\widehat{\pi}_A^-(\gamma)$ using Equation (8) across a range of γ ($\gamma = 0.10, 0.15, ..., 0.40$). Second, we form 1,000 bootstrap versions of $\widehat{\pi}_A^-(\gamma)$ for each possible value of γ. These are denoted by $\widehat{\pi}_A^{b-}(\gamma)$, for $b = 1, ..., 1,000$. Third, we compute the estimated MSE for each possible value of γ:

$$\widehat{MSE}^-(\gamma) = \frac{1}{1,000} \sum_{b=1}^{1,000} \left[\widehat{\pi}_A^{b-}(\gamma) - \max_{\gamma} \widehat{\pi}_A^-(\gamma) \right]^2. \tag{15}$$

We choose γ^- such that $\gamma^- = \arg\min_{\gamma} \widehat{MSE}^-(\gamma)$. We use the same data-driven procedure for $\widehat{\pi}_A^+(\gamma)$ to determine $\gamma^+ = \arg\min_{\gamma} \widehat{MSE}^+(\gamma)$. If $\min_{\gamma} \widehat{MSE}^-(\gamma) < \min_{\gamma} \widehat{MSE}^+(\gamma)$, we set $\widehat{\pi}_A^-(\gamma^*) = \widehat{\pi}_A^-(\gamma^-)$. To preserve the equality $1 = \pi_0 + \pi_A^+ + \pi_A^-$, we set $\widehat{\pi}_A^+(\gamma^*) = (1 - \widehat{\pi}_0) - \widehat{\pi}_A^-(\gamma^*)$. Otherwise, we set $\widehat{\pi}_A^+(\gamma^*) = \widehat{\pi}_A^+(\gamma^+)$ and $\widehat{\pi}_A^-(\gamma^*) = (1 - \widehat{\pi}_0) - \widehat{\pi}_A^+(\gamma^*)$.

A.3 Determining the Standard Deviation of the Estimators

We rely on the large-sample theory proposed by Genovese and Wasserman (2004). The essential idea is to recognize that the estimators $\widehat{\pi}_0(\lambda^*)$, $\widehat{S}_\gamma^+$, $\widehat{F}_\gamma^+$, $\widehat{T}_\gamma^+$, $\widehat{S}_\gamma^-$, $\widehat{F}_\gamma^-$, and $\widehat{T}_\gamma^-$ are all stochastic processes indexed by λ^* or γ which converge to a Gaussian process when the number of funds, M, goes to infinity. Proposition 3.2 of Genovese and Wasserman (2004) shows that $\widehat{\pi}_0(\lambda^*)$ is asymptotically normally distributed when $M \to \infty$, with standard deviation $\widehat{\sigma}_{\widehat{\pi}_0(\lambda^*)} = \left(\frac{\widehat{W}(\lambda^*)(M - \widehat{W}(\lambda^*))}{M^3(1-\lambda^*)^2} \right)^{\frac{1}{2}}$, where $\widehat{W}(\lambda^*)$ denotes the number of funds having p-values exceeding λ^*. Similarly, we have $\widehat{\sigma}_{\widehat{F}_\gamma^+} = (\gamma/2)\widehat{\sigma}_{\widehat{\pi}_0(\lambda^*)}$, $\widehat{\sigma}_{\widehat{S}_\gamma^+} = \left(\frac{\widehat{S}_\gamma^+(1 - \widehat{S}_\gamma^+)}{M} \right)^{\frac{1}{2}}$, and $\widehat{\sigma}_{\widehat{T}_\gamma^+} = \left(\widehat{\sigma}_{\widehat{S}_\gamma^+}^2 + (\gamma/2)^2 \widehat{\sigma}_{\widehat{\pi}_0(\lambda^*)}^2 + 2\frac{(\gamma/2)}{1-\lambda^*} \widehat{S}_\gamma^+ \frac{\widehat{W}(\lambda^*)}{M} \right)^{\frac{1}{2}}$ (using the equality $\widehat{S}_\gamma^+ = \widehat{F}_\gamma^+ + \widehat{T}_\gamma^+$). Standard deviations for the estimators in the left tail $\left(\widehat{S}_\gamma^-, \widehat{F}_\gamma^-, \widehat{T}_\gamma^- \right)$ are obtained by simply replacing $\widehat{S}_\gamma^+$ with $\widehat{S}_\gamma^-$ in the above formulas.

Finally, if $\gamma^* = \gamma^+$, the standard deviation of $\widehat{\pi}_A^+$ and $\widehat{\pi}_A^-$ are respectively given by $\widehat{\sigma}_{\widehat{\pi}_A^+} = \widehat{\sigma}_{\widehat{T}_{\gamma^*}^+}$, and $\widehat{\sigma}_{\widehat{\pi}_A^-} = \left(\widehat{\sigma}_{\widehat{\pi}_A^+}^2 + \widehat{\sigma}_{\widehat{\pi}_0(\lambda^*)}^2 - 2\left(\frac{1}{1-\lambda^*} \right) \widehat{S}_{\gamma^*}^+ \frac{\widehat{W}(\lambda^*)}{M^2} - (\gamma^*/2)\widehat{\sigma}_{\widehat{\pi}_0(\lambda^*)}^2 \right)^{\frac{1}{2}}$ (using

---

[26] We make the standard assumption that the density of the t-statistic of the unskilled and skilled funds decreases monotonically as γ rises.

27

the equality $\widehat{\pi}_A^+ = 1 - \widehat{\pi}_0^+ - \widehat{\pi}_A^-$). Otherwise if $\gamma^* = \gamma^-$, we just reverse the signs $+/-$ in the two formulas above.

B Monte-Carlo Analysis

B.1 Under Cross-Sectional Independence

We use Monte-Carlo simulations to examine the performance of all estimators used in the paper: $\widehat{\pi}_0$, $\widehat{\pi}_A^-$, $\widehat{\pi}_A^+$, $\widehat{S}_\gamma^-$, $\widehat{F}_\gamma^-$, $\widehat{T}_\gamma^-$, and $\widehat{S}_\gamma^+$, $\widehat{F}_\gamma^+$, $\widehat{T}_\gamma^+$. We generate the $M \times 1$ vector of fund monthly excess returns, r_t, according to the four-factor model (market, size, book-to-market, and momentum factors):

$$
\begin{aligned}
r_t &= \alpha + \beta F_t + \varepsilon_t, \qquad t = 1, ..., T, \\
F_t &\sim N(0, \Sigma_F), \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2 I),
\end{aligned}
\tag{16}
$$

where α denotes the $M \times 1$ vector of fund alphas, and β is the $M \times 4$ matrix of factor loadings. The 4×1 vector of factor excess returns, F_t, is normally distributed with covariance matrix Σ_F. ε_t is the $M \times 1$ vector of normally distributed residuals. We initially assume that the residuals are cross-sectionally independent and have the same variance σ_ε^2, so that the covariance matrix of ε_t can simply be written as $\sigma_\varepsilon^2 I$ (where I is the $M \times M$ identity matrix).

Our estimators are compared with their respective true population values defined as follows. The parameters π_0, π_A^-, and π_A^+ denote the true proportions of zero-alpha, unskilled, and skilled funds. The expected proportions of unlucky and lucky funds, $E(F_\gamma^-)$ and $E(F_\gamma^+)$, are both equal to $\pi_0 \cdot \gamma/2$. To determine the expected proportions of unskilled and skilled funds, $E(T_\gamma^-)$ and $E(T_\gamma^+)$, we use the fact that, under the alternative hypothesis $\alpha_i \neq 0$, the fund t-statistic follows a non-central student distribution with $T - 5$ degrees of freedom and a noncentrality parameter equal to $T^{\frac{1}{2}}\alpha_A/\sigma_\varepsilon$ (Davidson and MacKinnon (2004), p. 169):

$$
\begin{aligned}
E(T_\gamma^-) &= \pi_A^- \cdot prob\left(t < t_{T-5,\gamma/2} \,|\, H_A, \alpha_A < 0\right), \\
E(T_\gamma^+) &= \pi_A^+ \cdot prob\left(t > t_{T-5,1-\gamma/2} \,|\, H_A, \alpha_A > 0\right),
\end{aligned}
\tag{17}
$$

where $t_{T-2,\gamma/2}$ and $t_{T-2,1-\gamma/2}$ denote the quantiles of probability level $\gamma/2$ and $1 - \gamma/2$, respectively (these quantiles correspond to the thresholds t_γ^- and t_γ^+ used in the text). Finally, we have $E(S_\gamma^-) = E(F_\gamma^-) + E(T_\gamma^-)$, and $E(S_\gamma^+) = E(F_\gamma^+) + E(T_\gamma^+)$.

To compute these population values, we need to set values for the (true) proportions π_0, π_A^-, π_A^+, as well as for the t means (required to compute Equation (17)). In order

to set realistic values, we estimate π_0, π_A^-, and π_A^+ at the end of each of the final five years of our sample (2002-2006) using the entire return history for each fund up to that point in time. These estimates are then averaged to produce values that reflect the recent trend observed in Figure 4: $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$. To determine the t means of the unskilled and skilled funds, we use a simple calibration method. We first compute the average $\widehat{T}_\gamma^-$ and $\widehat{T}_\gamma^+$ (at $\gamma = 0.20$) over the final 5 years of our sample (2002-2006). Inserting these values along with $\pi_A^- = 23\%$ and $\pi_A^+ = 2\%$ in Equation (17), we can determine what are the t-means which satisfy both equalities. The resulting values are -2.5 and 3, and correspond to an annual four-factor alpha of -3.2% and 3.8%, respectively (using the equality $t_A = T^{\frac{1}{2}} \alpha_A / \sigma_\varepsilon$).

The total number of funds, M, used in the simulation is equal to 1,400.[27] The input for β is equal to the empirical loadings of a random draw of 1,400 funds (among the total population of 2,076 funds). Consistent with our database, we set $T = 384$ (months), $\sigma_\varepsilon = 0.021$ (equal to the empirical average across the 1,400 funds), and proxy Σ_F by its empirical counterpart. To build the vector of fund alphas, α, we need to determine the identity of the unskilled and skilled funds. This is done by randomly choosing 322 funds (i.e., 23% of the entire population) to which we assign a negative alpha (-3.2% per year), and 28 funds (2% of the population), to which we assign a positive alpha (3.8% per year).

After randomly drawing F_t and ε_t ($t = 1, ..., 384$), we construct the fund return time-series according to Equation (14), and compute their t-statistics by regressing the fund returns on the four-factor model. To determine the alpha p-values, we use the fact that the fund t-statistic follows a Student distribution with $T - 5$ degrees of freedom under the null hypothesis $\alpha_i = 0$. Then, we compute $\widehat{\pi}_0$, $\widehat{\pi}_A^-$, and $\widehat{\pi}_A^+$ using Equations (5) and (8). $\widehat{S}_\gamma^-$ and $\widehat{S}_\gamma^+$ correspond to the observed number of significant funds with negative and positive alphas, respectively. $\widehat{F}_\gamma^-$ and $\widehat{F}_\gamma^+$ are computed with Equation (6). $\widehat{T}_\gamma^-$ and $\widehat{T}_\gamma^+$, are given in Equation (7). We repeat this procedure 1,000 times.

In Table AI, we compare the average value of each estimator (over the 1,000 replications) with the true values. The figures in parentheses denote the lower and upper bounds of the estimator 90%-confidence interval. We set γ equal to 0.05 and 0.20. In all cases, the simulation results reveal that the average values of our estimators closely match the true values, and that their 90%-confidence intervals are narrow. This result

---

[27]We use this sample size to allow for comparison with the dependence case (described hereafter), which uses a sample of 1,400 correlated fund returns. Since our original sample of funds is larger than 1,400 ($M = 2,076$), our assessment of the precision of the estimators in this section is conservative.

is not surprising in light of the large cross-section of funds available in our sample.

Please insert Table AI here

B.2 Under Cross-Sectional Dependence

The return-generating process is the same as the one shown in Equation (16), except that the fund residuals are cross-correlated:

$$\varepsilon_t \sim N(0, \Sigma), \tag{18}$$

where Σ denotes the $M \times M$ residual covariance matrix. The main constraint imposed on Σ is that it must be positive semi-definite. To achieve this, we proceed as follows. We select all funds with 60 valid return observations over the final five years (2002-2006), which is the period over which we have the largest possible cross-section of funds existing simultaneously–898 funds, whose covariance matrix, Σ_1, is directly estimated from the data.[28] To assess the precision of our estimators, we also need to account for the non-overlapping returns observed in the long-term fund data due to funds that do not exist at the same time. To address this issue, we introduce 502 uncorrelated funds, and write the covariance matrix for the resulting 1,400 funds as follows:[29]

$$\Sigma = \left(\begin{array}{cc} \Sigma_1 & 0 \\ 0 & \sigma_\varepsilon^2 I \end{array} \right). \tag{19}$$

An an input for β, we use the empirical factor loadings of the 898 funds, along with the loadings of a random draw of the 502 remaining funds. The vector of fund alphas, α, is built by randomly choosing the identity of the unskilled and skilled funds among the 1,400 funds. The results in Table AII indicate that all estimators remain nearly unbiased ($\widehat{\pi}_0$, $\widehat{\pi}_A^-$, and $\widehat{\pi}_A^+$ exhibit small biases). Looking at the 90% confidence intervals, we logically observe that the dispersion of the estimators widens under cross-sectional dependence. However, the performance of the estimators is still very good.

Please insert Table AII here

[28] The 25%, 50%, and 75% pairwise correlation quantiles are -0.09, 0.05, and 0.19, respectively.
[29] The total number of fund pairs, P, is given by $M(M-1)/2$, where $M = 1,400$. If there are X uncorrelated funds in the population, the total number of uncorrelated fund pairs, I, equals $X \cdot (M - X) + X(X-1)/2$. In our data, 15% of the funds pairs do not have any return observations in common, and 55% of the observations are common to the remaining pairs (85%). Therefore, we estimate that the proportion of uncorrelated pairs is equal to 53% ($15\% + 85\% \cdot 45\%$). With 502 uncorrelated funds, I/P amounts to 58%, and is very close to the ratio observed in the data.

Apart from this baseline dependence scenario, we also examine three other cases (the results are available upon request). First, we introduce correlation by block among each skill group (zero-alpha, unskilled, and skilled funds) to account for their possible similar bets. Inside each block (representing 10% of each skill group), we set the pairwise correlation equal to 0.15 or 0.30. Second, we use the residual factor specification proposed by Jones and Shanken (2005) in order to capture the role of non-priced factors. We assume that all fund residuals depend on a common residual factor, and that the unskilled and skilled funds are affected by specific residual factors. In the two cases, the results show that the precision of the estimators remain very close to those obtained under the independence case. Finally, we consider the extreme dependence case where the fund population only consists of the 898 correlated funds. We find that all estimators remain unbiased as shown in Tables AI and AII. But unsurprisingly, the confidence intervals widen a bit compared to those shown in Table AII (on average 2% are added on each side of the interval).

References

[1] Avramov, Doron, and Russ Wermers, 2006, Investing in Mutual Funds When Return are Predictable, Journal of Financial Economics 81, 339-377.

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

[3] Benjamini, Yoav, and Yosef Hochberg, 1995, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society 57, 289-300.

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

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

[6] Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in Liquidity, Journal of Financial Economics 56, 3-28.

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

[8] Davidson, Russell, and James G. MacKinnon, 2004, Econometric Theory and Methods, Oxford University Press, New-York.

[9] Efron, Bradley, and Robert Tibshirani, 2002, Empirical Bayes Methods and False Discovery Rates for Microarrays, Genetic Epidemiology 23, 70-86.

[10] Elton, Edwin J., Martin J. Gruber, Sanjiv Das, and Matthew Hlavka, 1993, Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios, Review of Financial Studies 6, 1-22.

[11] Elton, Edwin J., Martin J. Gruber, Christopher R. Blake, 1996, The Persistence of Risk-Adjusted Mutual Fund Performance, Journal of Business 69, 133-157.

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

[13] Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2007, Participant Reaction and the Performance of Funds offered by 401(k) Plans, Journal of Financial Intermediation, 16, 249-271.

[14] Fama, Eugene F., and Kenneth R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, Journal of Financial Economics 33, 3-56.

[15] Ferson, Wayne E., and Rudi W. Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions, Journal of Finance 51, 425-461.

[16] Ferson, Wayne E., and Meijun Qian, 2004, Conditional Performance Evaluation Revisited, in Research Foundation Monograph of the CFA Institute.

[17] Genovese, Christopher, and Larry Wasserman, 2004, A Stochastic Process Approach to False Discovery Control, Annals of Statistics, 32, 1035-1061.

[18] Grinblatt, Mark, and Sheridan Titman, 1989, Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings, Journal of Business 62, 393-416.

[19] Grinblatt, Mark, and Sheridan Titman, 1993, Performance Measurement without Benchmarks: An Examination of Mutual Fund Returns, Journal of Business 66, 47-68.

[20] Gruber, Martin J., 1996, Another Puzzle: The Growth of Actively Managed Mutual Funds, Journal of Finance 51, 783-810.

[21] Hall Peter, Joel L. Horowitz, and Bing-Yi Jing, 1995, On Blocking Rules for the Bootstrap with Dependent Data, Biometrika 82, 561-574.

[22] Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot Hands in Mutual Funds: The Persistence of Performance, 1974-88, Journal of Finance 48, 93-130.

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

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

[25] Kosowski, Robert, Allan Timmermann, Halbert White, and Russ Wermers, 2006, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, Journal of Finance 61, 2551-2595.

[26] Lehmann, Bruce N., and David M. Modest, 1987, Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons, Journal of Finance 42, 233-265.

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

[28] Pastor, Lubos, and Robert F. Stambaugh, 2002a, Mutual Fund Performance and Seemingly Unrelated Assets, Journal of Financial Economics 63, 315-349.

[29] Pastor, Lubos, and Robert F. Stambaugh, 2002b, Investing in Equity Mutual Funds, Journal of Financial Economics 63, 351-380.

[30] Storey, John D., 2002, A Direct Approach to False Discovery Rates, Journal of the Royal Statistical Society 64, 479–498.

[31] Storey, John D., 2003, The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value, Annals of Statistics 31, 2013-2035.

[32] Storey, John D., and Robert Tibshirani, 2003, Statistical Significance for Genomewide Studies, Proceedings of the National Academy of Sciences 100, 9440-9445.

[33] Storey, John D., Jonathan E. Taylor, and David Siegmund, 2004, Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach, Journal of the Royal Statistical Society 66, 187-205.

[34] Sullivan Ryan, Allan Timmermann, and Halbert White, 1999, Data-Snooping, Technical Trading Rule Performance and the Bootstrap, Journal of Finance 54, 1647-1691.

[35] Tang, Yongqiang, Subhashis Ghosal, and Anindya Roy, 2007, Nonparametric Bayesian Estimation of Positive False Discovery Rates, Biometrics 63, 1126-1134.

[36] Wermers, Russ, 1999, Mutual Fund Herding and the Impact on Stock Prices, Journal of Finance 55, 581-622

[37] Wermers, Russ, 2000, Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transaction Costs, and Expenses, Journal of Finance 55, 1655-1695.

Table I

Performance of the Equally-Weighted Portfolio of Funds

Results for the unconditional and conditional four-factor models are shown in Panels A and B for the entire fund population (All funds), as well as Growth, Aggressive Growth, and Growth & Income funds. The regressions are based on monthly data between January 1975 and December 2006. Each Panel contains the estimated annualized alpha $(\widehat{\alpha})$, the estimated exposures to the market $(\widehat{b}_m)$, size $(\widehat{b}_{smb})$, book-to-market $(\widehat{b}_{hml})$, and momentum factors $(\widehat{b}_{mom})$, as well as the adjusted R^2 of an equally-weighted portfolio that includes all funds that exist at the beginning of each month. Figures in parentheses denote the Newey-West (1987) heteroskedasticity-consistent p-values, under the null hypothesis that the regression parameters are equal to zero.

Panel A Unconditional Four-Factor Model

	$\widehat{\alpha}$	$\widehat{b}_m$	$\widehat{b}_{smb}$	$\widehat{b}_{hml}$	$\widehat{b}_{mom}$	R^2
All (2,076)	-0.48%	0.95	0.17	-0.01	0.02	98.0%
	(0.12)	(0.00)	(0.00)	(0.38)	(0.09)	
Growth (1,304)	-0.45%	0.95	0.16	-0.03	0.02	98.0%
	(0.16)	(0.00)	(0.00)	(0.15)	(0.07)	
Aggressive Growth (388)	-0.53%	1.04	0.43	-0.17	0.09	95.8%
	(0.22)	(0.00)	(0.00)	(0.00)	(0.00)	
Growth & Income (384)	-0.47%	0.87	-0.04	0.17	-0.03	98.2%
	(0.09)	(0.00)	(0.02)	(0.00)	(0.01)	

Panel B Conditional Four-Factor Model

	$\widehat{\alpha}$	$\widehat{b}_m$	$\widehat{b}_{smb}$	$\widehat{b}_{hml}$	$\widehat{b}_{mom}$	R^2
All (2,076)	-0.60%	0.96	0.17	-0.02	0.02	98.2%
	(0.09)	(0.00)	(0.00)	(0.23)	(0.08)	
Growth (1,304)	-0.59%	0.96	0.16	-0.03	0.03	98.2%
	(0.10)	(0.00)	(0.00)	(0.08)	(0.05)	
Aggressive Growth (388)	-0.49%	1.05	0.43	-0.19	0.08	96.2%
	(0.24)	(0.00)	(0.00)	(0.00)	(0.00)	
Growth & Income (384)	-0.58%	0.87	-0.04	0.16	-0.03	98.3%
	(0.05)	(0.00)	(0.02)	(0.00)	(0.02)	

35

Table II

Impact of Luck on Long-Term Performance

Panel A displays the estimated proportions of zero-alpha, unskilled, and skilled funds in the entire fund population (2,076 funds). We measure fund performance with the unconditional four-factor model over the entire period 1975-2006. Panel B counts the proportions of significant funds in the left and right tails of the cross-sectional t-statistic distribution $(\widehat{S}_\gamma^-, \widehat{S}_\gamma^+)$ at four significance levels (γ=0.05, 0.10, 0.15, 0.20). In the leftmost columns, the significant group in the left tail, $\widehat{S}_\gamma^-$, is decomposed into unlucky and unskilled funds $(\widehat{F}_\gamma^-, \widehat{T}_\gamma^-)$. In the rightmost columns , the significant group in the right tail, $\widehat{S}_\gamma^+$, is decomposed into lucky and skilled funds $(\widehat{F}_\gamma^+, \widehat{T}_\gamma^+)$. Figures in parentheses denote the standard deviation of the different estimators. The bottom of Panel B also presents the characteristics of each significant group $(\widehat{S}_\gamma^-, \widehat{S}_\gamma^+)$: the average estimated alpha (% per year), expense ratio (% per year), turnover (% per year), and median size measured by the total net asset under management (millions USD).

Panel A Proportion of Unskilled and Skilled Funds

	Zero alpha($\widehat{\pi}_0$)	Non-zero alpha	Unskilled($\widehat{\pi}_A^-$)	Skilled($\widehat{\pi}_A^+$)
Proportion	75.4 (2.5)	24.6	24.0 (2.3)	0.6 (0.8)
Number	1,565	511	499	12

Panel B Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$(%)	11.6	17.2	21.5	25.4	8.2	6.0	4.2	2.2	Signif. $\widehat{S}_\gamma^+$(%)
	(0.7)	(0.8)	(0.9)	(0.9)	(0.6)	(0.5)	(0.4)	(0.3)	
Unlucky $\widehat{F}_\gamma^-$(%)	1.9	3.8	5.6	7.6	7.6	5.6	3.8	1.9	Lucky $\widehat{F}_\gamma^+$(%)
	(0.0)	(0.1)	(0.2)	(0.3)	(0.3)	(0.2)	(0.1)	(0.0)	
Unskilled $\widehat{T}_\gamma^-$(%)	9.8	13.6	16.1	18.2	0.6	0.4	0.4	0.3	Skilled $\widehat{T}_\gamma^+$(%)
	(0.7)	(0.9)	(1.0)	(1.1)	(0.7)	(0.6)	(0.5)	(0.3)	
Alpha(% year)	-5.5	-5.0	-4.7	-4.6	4.8	5.2	5.6	6.5	Alpha(% year)
Exp.(% year)	1.6	1.5	1.5	1.5	1.2	1.2	1.2	1.2	Exp.(% year)
Turn.(% year)	100	99	98	96	126	95	95	105	Turn.(% year)
Size(million $)	81	88	86	84	731	985	888	745	Size(million $)

Table III

Impact of Luck on Short-Term Performance

Panel A displays the estimated proportions of zero-alpha, unskilled, and skilled funds in the entire fund population (3,311 funds). We measure fund performance with the unconditional four-factor model over non-overlapping 5-year periods between 1977-2006. Panel B counts the proportions of significant funds in the left and right tails of the cross-sectional t-statistic distribution ($\widehat{S}_\gamma^-$, $\widehat{S}_\gamma^+$) at four significance levels (γ=0.05, 0.10, 0.15, 0.20). In the leftmost columns, the significant group in the left tail, $\widehat{S}_\gamma^-$, is decomposed into unlucky and unskilled funds ($\widehat{F}_\gamma^-$, $\widehat{T}_\gamma^-$). In the rightmost columns , the significant group in the right tail, $\widehat{S}_\gamma^+$, is decomposed into lucky and skilled funds ($\widehat{F}_\gamma^+$, $\widehat{T}_\gamma^+$). Figures in parentheses denote the standard deviation of the different estimators. The bottom of Panel B also presents the characteristics of each significant group ($\widehat{S}_\gamma^-$, $\widehat{S}_\gamma^+$): the average estimated alpha (% per year), expense ratio (% per year), turnover (% per year), and median size measured by the total net asset under management (millions USD).

Panel A Proportion of Unskilled and Skilled Funds

	Zero alpha($\widehat{\pi}_0$)	Non-zero alpha	Unskilled($\widehat{\pi}_A^-$)	Skilled($\widehat{\pi}_A^+$)
Proportion	72.2 (2.0)	27.8	25.4 (1.7)	2.4 (0.7)
Number	2,390	921	841	80

Panel B Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$(%)	11.2	16.8	21.4	24.9	9.6	7.8	5.9	3.5	Signif. $\widehat{S}_\gamma^+$(%)
	(0.5)	(0.6)	(0.7)	(0.8)	(0.5)	(0.5)	(0.4)	(0.3)	
Unlucky $\widehat{F}_\gamma^-$(%)	1.8	3.6	5.4	7.2	7.2	5.4	3.6	1.8	Lucky $\widehat{F}_\gamma^+$(%)
	(0.0)	(0.0)	(0.1)	(0.2)	(0.2)	(0.1)	(0.0)	(0.0)	
Unskilled $\widehat{T}_\gamma^-$(%)	9.4	13.2	16.0	17.7	2.4	2.4	2.3	1.7	Skilled $\widehat{T}_\gamma^+$(%)
	(0.6)	(0.7)	(0.8)	(0.8)	(0.6)	(0.5)	(0.4)	(0.3)	
Alpha(% year)	-6.5	-5.9	-5.5	-5.3	6.7	7.0	7.2	7.5	Alpha(% year)
Exp.(% year)	1.4	1.3	1.3	1.3	1.2	1.2	1.2	1.2	Exp.(% year)
Turn.(% year)	98	105	100	99	80	81	83	81	Turn.(% year)
Size(million $)	242	244	252	244	623	628	664	749	Size(million $)

Table IV

Short-Term Performance across Investment Categories

The impact of luck in the left and right tails for three investment categories (Growth, Aggressive Growth, and Growth & Income funds) is presented in Panels A, B, and C, respectively. We measure fund performance with the unconditional four-factor model over non-overlapping 5-year periods between 1977-2006. For each panel, we count the proportions of significant funds in the left and right tails of the cross-sectional t-statistic distribution ($\widehat{S}_\gamma^-$, $\widehat{S}_\gamma^+$) at four significance levels (γ=0.05, 0.10, 0.15, 0.20). In the leftmost columns, the significant group in the left tail, $\widehat{S}_\gamma^-$, is decomposed into unlucky and unskilled funds ($\widehat{F}_\gamma^-$, $\widehat{T}_\gamma^-$). In the rightmost columns, the significant group in the right tail, $\widehat{S}_\gamma^+$, is decomposed into lucky and skilled funds ($\widehat{F}_\gamma^+$, $\widehat{T}_\gamma^+$). Figures in parentheses denote the standard deviation of the different estimators. The bottom of Panel B also presents the characteristics of each significant group ($\widehat{S}_\gamma^-$, $\widehat{S}_\gamma^+$): the average estimated alpha (% per year), expense ratio (% per year), turnover (% per year), and median size measured by the total net asset under management (millions USD).

Panel A Growth funds

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$(%)	11.3	16.6	21.4	25.2	9.9	8.1	5.9	3.5	Signif. $\widehat{S}_\gamma^+$(%)
	(0.7)	(0.8)	(0.9)	(1.0)	(0.7)	(0.6)	(0.5)	(0.4)	
Unlucky $\widehat{F}_\gamma^-$(%)	1.8	3.6	5.5	7.3	7.3	5.5	3.6	1.8	Lucky $\widehat{F}_\gamma^+$ (%)
	(0.0)	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.0)	
Unskilled $\widehat{T}_\gamma^-$(%)	9.5	13.0	15.9	17.9	2.6	2.6	2.3	1.7	Skilled $\widehat{T}_\gamma^+$ (%)
	(0.7)	(0.9)	(1.0)	(1.1)	(0.8)	(0.7)	(0.6)	(0.4)	
Alpha(% year)	-6.0	-5.6	-5.2	-5.1	6.8	6.8	6.8	7.3	Alpha(% year)
Exp.(% year)	1.4	1.3	1.3	1.3	1.2	1.2	1.2	1.2	Exp.(% year)
Turn.(% year)	101	113	110	107	79	79	79	75	Turn.(% year)
Size(million $)	253	253	253	252	589	593	621	580	Size(million $)

Table IV

Panel B Aggressive Growth funds

		Left Tail				Right Tail			
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$ (%)	12.0	16.0	19.4	22.2	11.2	9.4	7.1	4.9	Signif. $\widehat{S}_\gamma^+$ (%)
	(1.1)	(1.4)	(1.6)	(1.8)	(1.3)	(1.1)	(1.0)	(0.8)	
Unlucky $\widehat{F}_\gamma^-$ (%)	1.8	3.6	5.4	7.2	7.2	5.4	3.6	1.8	Lucky $\widehat{F}_\gamma^+$ (%)
	(0.1)	(0.2)	(0.3)	(0.4)	(0.4)	(0.3)	(0.2)	(0.1)	
Unskilled $\widehat{T}_\gamma^-$ (%)	10.2	12.3	14.0	15.0	4.0	4.0	3.5	3.1	Skilled $\widehat{T}_\gamma^+$ (%)
	(1.3)	(1.6)	(1.7)	(1.9)	(1.4)	(1.2)	(1.1)	(0.9)	
Alpha(% year)	-9.3	-8.6	-8.1	-7.6	8.5	8.8	9.7	9.7	Alpha(% year)
Exp.(% year)	1.5	1.5	1.5	1.5	1.3	1.3	1.3	1.3	Exp.(% year)
Turn.(% year)	127	123	119	117	105	104	107	104	Turn.(% year)
Size(million $)	154	187	160	192	1,014	949	1,021	1,073	Size(million $)

Panel C Growth & Income funds

		Left Tail				Right Tail			
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$ (%)	11.5	17.4	22.5	26.8	7.3	5.5	3.7	1.8	Signif. $\widehat{S}_\gamma^+$ (%)
	(1.1)	(1.4)	(1.5)	(1.6)	(1.0)	(0.8)	(0.7)	(0.5)	
Unlucky $\widehat{F}_\gamma^-$ (%)	1.8	3.7	5.5	7.3	7.3	5.5	3.7	1.8	Lucky $\widehat{F}_\gamma^+$ (%)
	(0.1)	(0.2)	(0.3)	(0.4)	(0.4)	(0.3)	(0.2)	(0.1)	
Unskilled $\widehat{T}_\gamma^-$ (%)	9.8	13.7	17.0	19.5	0.0	0.0	0.0	0.0	Skilled $\widehat{T}_\gamma^+$ (%)
	(1.2)	(1.5)	(1.7)	(1.8)	(1.1)	(0.9)	(0.8)	(0.5)	
Alpha(% year)	-4.9	-4.5	-4.2	-4.0	4.9	5.3	5.1	4.9	Alpha(% year)
Exp.(% year)	1.3	1.2	1.2	1.2	1.1	1.1	1.0	1.0	Exp.(% year)
Turn.(% year)	69	68	66	64	56	57	54	45	Turn.(% year)
Size(million $)	295	346	348	337	492	482	473	1,787	Size(million $)

Table V

Performance Persistence Based on the False Discovery Rate

For each of the five FDR targets (10%, 30%, 50%, 70%, and 90%), Panel A contains descriptive statistics on the FDR level $(\widehat{FDR}_{\gamma P}^{+})$ achieved by each portfolio, as well as the proportion of funds in the population that they include $(\widehat{S}_{\gamma P}^{+})$. the panel shows the average values of $\widehat{FDR}_{\gamma P}^{+}$ and $\widehat{S}_{\gamma P}^{+}$ over the 27 annual formation dates (from December 1979 to 2005), as well as respective distributions. Panel B displays the performance of each portfolio over the period 1980-2006. We estimate the annual four-factor alpha $(\widehat{\alpha})$ with its bootstrap p-value, its annual residual standard deviation $(\widehat{\sigma}_{\varepsilon})$, its annual information ratio (IR=$\widehat{\alpha}/\widehat{\sigma}_{\varepsilon}$), its loadings on the market $(\widehat{b}_m)$, size $(\widehat{b}_{smb})$, book-to-market $(\widehat{b}_{hml})$, and momentum factors $(\widehat{b}_{mom})$, and its annual excess mean, and standard deviation. In Panel C, we examine the turnover of each portfolio. We compute the proportion of funds that are still included in the portfolio 1, 2, 3, 4, and 5 years after their initial selection.

Panel A Portfolio Statistics

| | Achieved False Discovery Rate $\left(\widehat{FDR}_{\gamma P}^{+}\right)$ | | | | | Included proportion of funds$(\widehat{S}_{\gamma P}^{+})$ | | | | |
	Mean	10-30	30-50	50-70	>70%	Mean	0-6	6-12	12-24	>24%
FDR10%	41.5%	14	6	1	6	3.0%	25	2	0	0
FDR30%	47.5%	8	12	1	6	8.2%	15	7	3	2
FDR50%	60.4%	0	14	7	6	20.9%	5	7	4	11
FDR70%	71.3%	0	4	12	11	29.7%	1	5	5	16
FDR90%	75.0%	0	4	9	14	33.7%	0	3	4	20

Panel B Performance Analysis

	$\widehat{\alpha}$(p-value)	$\widehat{\sigma}_{\varepsilon}$	IR	$\widehat{b}_m$	$\widehat{b}_{smb}$	$\widehat{b}_{hml}$	$\widehat{b}_{mom}$	Mean	Std dev
FDR10%	1.45%(0.04)	4.0%	0.36	0.93	0.16	-0.04	-0.02	8.3%	15.4%
FDR30%	1.15%(0.05)	3.3%	0.35	0.94	0.17	-0.02	-0.03	8.1%	15.4%
FDR50%	0.95%(0.10)	2.9%	0.33	0.96	0.20	-0.06	-0.01	8.1%	16.1%
FDR70%	0.68%(0.15)	2.7%	0.25	0.97	0.19	-0.06	-0.01	7.9%	16.1%
FDR90%	0.39%(0.30)	2.7%	0.14	0.97	0.19	-0.05	-0.00	7.8%	16.0%

Panel C Portfolio Turnover

| | Proportion of funds | | | | |
	After 1 year	After 2 years	After 3 years	After 4 years	After 5 years
FDR10%	36.7	12.8	3.4	0.8	0.0
FDR30%	40.0	14.7	5.1	1.7	1.3
FDR50%	48.8	23.5	12.3	4.7	2.6
FDR70%	52.2	29.0	17.4	9.5	6.3
FDR90%	55.9	33.8	20.4	13.0	8.5

Table VI

Impact of Luck on Long-Term Pre-Expense Performance

Panel A displays the estimated proportions of zero-alpha, unskilled, and skilled funds in the entire fund population on a pre-expense basis (1,836 funds). We add the monthly expenses to net return of each fund, and measure performance with the unconditional four-factor model over the entire period 1975-2006. Panel B counts the proportions of significant funds in the left and right tails of the cross-sectional t-statistic distribution ($\widehat{S}_\gamma^-$, $\widehat{S}_\gamma^+$) at four significance levels (γ=0.05, 0.10, 0.15, 0.20). In the leftmost columns, the significant group in the left tail, $\widehat{S}_\gamma^-$, is decomposed into unlucky and unskilled funds ($\widehat{F}_\gamma^-$, $\widehat{T}_\gamma^-$). In the rightmost columns , the significant group in the right tail, $\widehat{S}_\gamma^+$, is decomposed into lucky and skilled funds ($\widehat{F}_\gamma^+$, $\widehat{T}_\gamma^+$). Figures in parentheses denote the standard deviation of the different estimators. The bottom of Panel B also presents the characteristics of each significant group ($\widehat{S}_\gamma^-$, $\widehat{S}_\gamma^+$): the average estimated alpha prior to expenses (in % per year), expense ratio (in % per year), turnover (in % per year), and median size measured by the total net asset under management (in millions of dollars).

Panel A Proportion of Unskilled and Skilled Funds

	Zero alpha($\widehat{\pi}_0$)	Non-zero alpha	Unskilled($\widehat{\pi}_A^-$)	Skilled($\widehat{\pi}_A^+$)
Proportion	85.9 (2.7)	14.1	4.5 (1.0)	9.6 (1.5)
Number	1,577	259	176	83

Panel B Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$(%)	4.3	7.5	10.2	12.8	17.3	13.1	9.3	5.8	Signif. $\widehat{S}_\gamma^+$(%)
	(0.5)	(0.6)	(0.7)	(0.8)	(0.9)	(0.8)	(0.7)	(0.5)	
Unlucky $\widehat{F}_\gamma^-$(%)	2.1	4.3	6.4	8.6	8.6	6.4	4.3	2.1	Lucky $\widehat{F}_\gamma^+$(%)
	(0.0)	(0.1)	(0.1)	(0.2)	(0.2)	(0.1)	(0.1)	(0.0)	
Unskilled $\widehat{T}_\gamma^-$(%)	2.2	3.2	3.8	4.2	8.7	6.6	5.0	3.6	Skilled $\widehat{T}_\gamma^+$(%)
	(0.5)	(0.6)	(0.8)	(0.9)	(1.0)	(0.9)	(0.7)	(0.5)	
Pre Expense									Pre Expense
Alpha(% year)	-5.9	-5.2	-4.8	-4.5	4.4	4.8	5.0	5.3	Alpha(% year)
Exp.(% year)	1.5	1.4	1.3	1.3	1.3	1.3	1.3	1.3	Exp.(% year)
Turn.(% year)	105	108	108	108	106	111	122	84	Turn.(% year)
Size(million $)	81	90	93	90	430	578	948	1,000	Size(million $)

Table VII
Loadings on Omitted Factors

We determine the proportions of significant funds in the left and right tails $(\widehat{S}_\gamma^-, \widehat{S}_\gamma^+)$ at four significance levels (γ=0.05, 0.10, 0.15, 0.20) according to each asset-pricing model over the period 1975-2006. For each of these significant groups, we compute their average loadings on the omitted factors from the four-factor model: size $(\widehat{b}_{smb})$, book-to-market $(\widehat{b}_{hml})$, and momentum $(\widehat{b}_{mom})$. Panel A shows the results obtained with the unconditional CAPM, while Panel B repeats the same procedure with the unconditional Fama-French model.

Panel A Unconditional CAPM

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Size($\widehat{b}_{smb}$)	0.06	0.07	0.09	0.09	0.27	0.28	0.28	0.36	Size($\widehat{b}_{smb}$)
BTM($\widehat{b}_{hml}$)	-0.14	-0.14	-0.13	-0.14	0.34	0.35	0.36	0.37	BTM($\widehat{b}_{hml}$)
Mom.($\widehat{b}_{mom}$)	0.00	0.00	0.00	0.01	-0.01	-0.01	-0.02	-0.01	Mom.($\widehat{b}_{mom}$)

Panel B Unconditional Fama-French model

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Mom.($\widehat{b}_{mom}$)	-0.02	-0.03	-0.02	-0.03	0.09	0.10	0.11	0.12	Mom.($\widehat{b}_{mom}$)

Table AI

Monte-Carlo Analysis under Cross-Sectional Independence

We examine the average value and the 90%-confidence interval (in parentheses) of the different estimators based on 1,000 replications. For each replication, we generate monthly fund returns for 1,400 funds and 384 periods using the four-factor model (market, size, book-to-market, and momentum factors). Fund residuals are independent from one another. The true parameter values for the proportions of zero-alpha, unskilled, and skilled funds (π_0, π_A^-, and π_A^+) are set to 75%, 23%, and 2%. The mean of the t-statistic of the unskilled and skilled funds are set to -2.5 and 3 (corresponding to an annual four-factor alpha of -3.2% and 3.8%, respectively). In each tail (left and right), we assess the precision of the different estimators at two significance levels (γ=0.05 and 0.20).

Fund Proportion	True	Estimator (90% interval)		
Zero-alpha funds (π_0)	75.0	75.1 (71.7,78.6)		
Unskilled funds (π_A^-)	23.0	22.9 (19.7,25.9)		
Skilled funds (π_A^+)	2.0	2.0 (0.3,3.8)		

	Significance level $\gamma = 0.05$		Significance level $\gamma = 0.20$	
Left Tail	True	Estimator (90% interval)	True	Estimator (90% interval)
Significant funds $E\left(S_\gamma^-\right)$	18.1	18.1 (16.4,19.7)	27.9	27.9 (26.1,30.0)
Unlucky funds $E\left(F_\gamma^-\right)$	1.8	1.8 (1.8,1.9)	7.5	7.5 (7.1,7.9)
Unskilled funds $E\left(T_\gamma^-\right)$	16.2	16.2 (14.6,17.9)	20.4	20.4 (18.2,22.7)

	Significance level $\gamma = 0.05$		Significance level $\gamma = 0.20$	
Right Tail	True	Estimator (90% interval)	True	Estimator (90% interval)
Significant funds $E\left(S_\gamma^+\right)$	3.6	3.6 (2.8,4.4)	9.4	9.4 (8.2,10.8)
Lucky funds $E\left(F_\gamma^+\right)$	1.8	1.8 (1.8,1.9)	7.5	7.5 (7.1,7.9)
Skilled funds $E\left(T_\gamma^+\right)$	1.7	1.7 (0.9,2.5)	1.9	1.9 (0.5,3.3)

Table AII

Monte-Carlo Analysis under Cross-Sectional Dependence

We examine the average value and the 90%-confidence interval (in parentheses) of the different estimators based on 1,000 replications. For each replication, we generate monthly fund returns for 1,400 funds and 384 periods using the four-factor model (market, size, book-to-market, and momentum factors). We assume that funds are cross-sectionally correlated. In order to closely reproduce the relations across all funds, we use the emprical covariance matrix of the fund residuals as the true covariance matrix. The true parameter values for the proportions of zero-alpha, unskilled, and skilled funds (π_0, π_A^-, and π_A^+) are set to 75%, 23%, and 2%. The mean of the t-statistic of the unskilled and skilled funds are set to -2.5 and 3 (corresponding to an annual four-factor alpha of -3.2% and 3.8%, respectively). In each tail (left and right), we assess the precision of the different estimators at two significance levels (γ=0.05 and 0.20).

Fund Proportion	True	Estimator (90% interval)		
Zero-alpha funds (π_0)	75.0	75.2 (69.5,80.8)		
Unskilled funds (π_A^-)	23.0	22.8 (17.0,28.9)		
Skilled funds (π_A^+)	2.0	1.9 (0.0,6.5)		

	Significance level $\gamma = 0.05$		Significance level $\gamma = 0.20$	
Left Tail	True	Estimator (90% interval)	True	Estimator (90% interval)
Significant funds $E\left(S_\gamma^-\right)$	18.1	18.1 (15.3,20.7)	27.9	27.9 (24.3,32.3)
Unlucky funds $E\left(F_\gamma^-\right)$	1.8	1.8 (1.6,2.1)	7.5	7.6 (6.6,8.3)
Unskilled funds $E\left(T_\gamma^-\right)$	16.2	16.2 (13.4,19.1)	20.4	20.4 (16.3,24.6)

	Significance level $\gamma = 0.05$		Significance level $\gamma = 0.20$	
Right Tail	True	Estimator (90% interval)	True	Estimator (90% interval)
Significant funds $E\left(S_\gamma^+\right)$	3.5	3.6 (2.4,5.3)	9.4	9.4 (6.6,12.6)
Lucky funds $E\left(F_\gamma^+\right)$	1.8	1.8 (1.6,2.1)	7.5	7.6 (6.6,8.3)
Skilled funds $E\left(T_\gamma^+\right)$	1.7	1.7 (0.5,3.8)	1.9	1.9 (0.1,5.6)

Figure 1

Figure 1

Outcome of the Multiple Performance Test

Panel A shows the distribution of the fund t-statistic across the three skill categories (zero-alpha, unskilled, and skilled funds). We set the mean of the t-statistic (mean t) equal to -2.5 for the unskilled funds and +3 for the skilled ones. Panel B displays the cross-sectional t-statistic distribution. It is a mixture of the three distributions in Panel A, where the weight on each distribution depends on the proportion of zero-alpha, unskilled, and skilled funds in the population (π_0, π_A^-, and π_A^+). In this example, we set $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$ to match our average estimated values over the final 5 years of our sample.

PANEL A: T-STATISTIC DISTRIBUTION

PANEL B: CROSS-SECTIONAL T-STATISTIC DISTRIBUTION

45
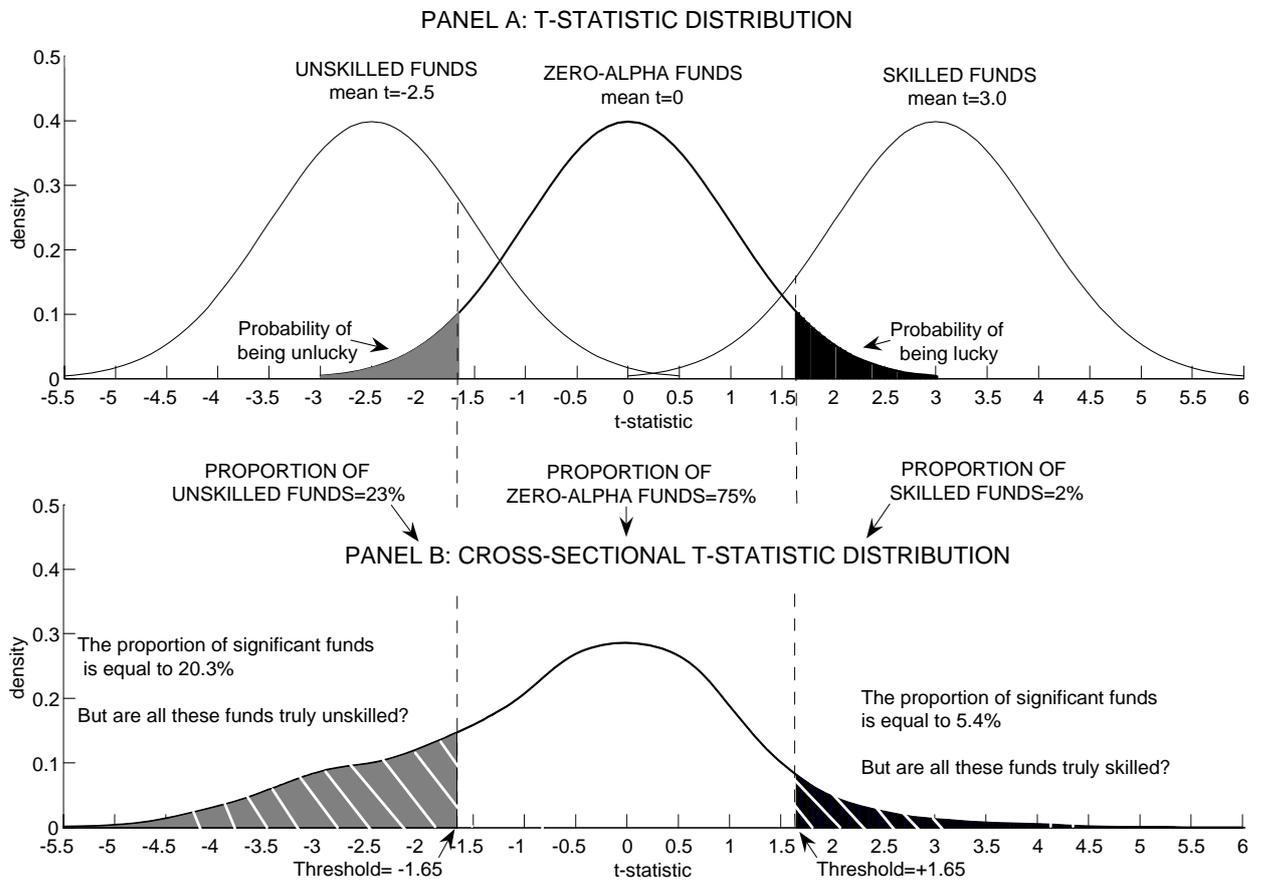
Figure 2

Histogram of Fund p-values

This figure represents the p-value histogram of 2,076 funds (as in our database). For each fund, we draw its t-statistic from one of the distributions in Figure 1 (Panel A) according to the proportion of zero-alpha, unskilled, and skilled funds in the population (π_0, π_A^-, and π_A^+). In this example, we set $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$ to match our average estimated values over the final 5 years of our sample. Then, we compute the two-sided p-values of each fund from its respective t-statistic, and plot them in the histogram.
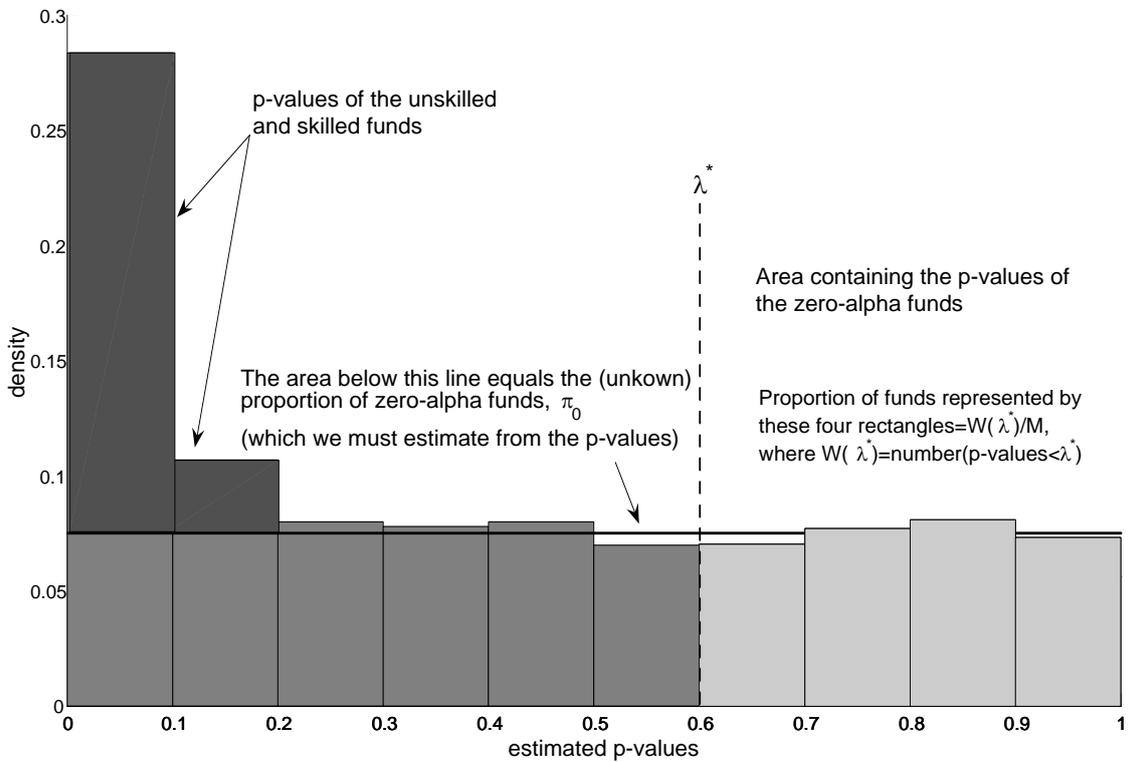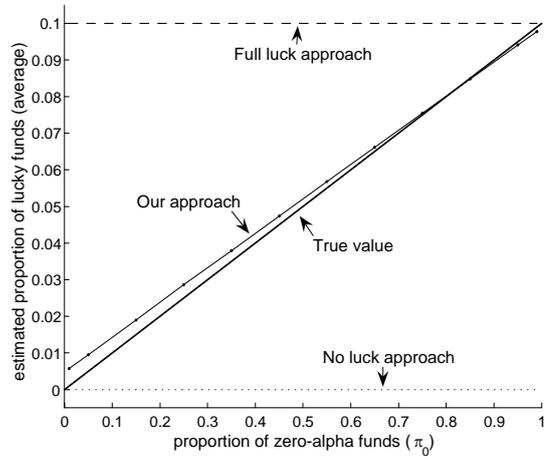
Figure 3

## Measuring Luck: Comparison with Existing Approaches

This figure examines the bias of different estimators produced by the three approaches as a function of the proportion of zero-alpha funds, π_0. We examine the estimators of the proportions of unlucky, lucky, unskilled, and skilled funds in Panel A, B, C, and D, respectively. The "no luck" approach assumes that $\pi_0 = 0$, the "full luck" approach assumes that $\pi_0 = 1$, while "our approach" estimates π_0 directly from the data. For each approach, we compare the average estimator value (over 1,000 replications) with the true population value. For each replication, we draw the t-statistic for each fund i ($i = 1,...,2,076$) from one of the distributions in Figure 1 (Panel A) according to the weights π_0, π_A^-, and π_A^+, and compute the different estimators at the significance level $\gamma = 0.20$. For each π_0, the ratio π_A^- over π_A^+ is held fixed to 11.5 (0.23/0.02) as in Figure 1.
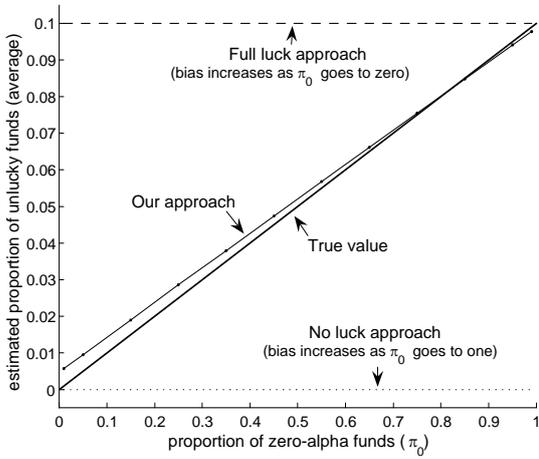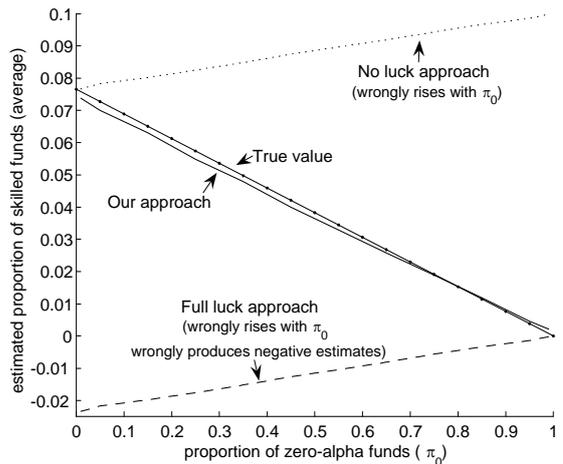
(A) Unlucky funds (left tail)

(B) Lucky funds (right tail)
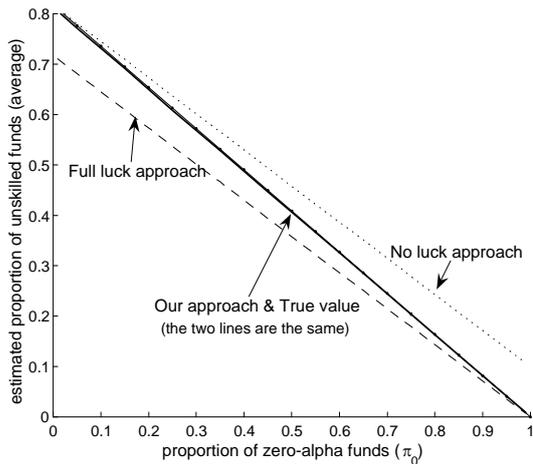
(C) Unskilled funds (left tail)

(D) Skilled funds (right tail)

47

Figure 4

Evolution of Mutual Fund Performance over Time

Panel A plots the evolution of the estimated proportions of unskilled and skilled funds ($\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$) between 1989 and 2006. At the end of each year, we measure $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$ using the entire fund return history up to that point. The initial estimates at the end of 1989 cover the period 1975-1989, while the last ones in 2006 use the period 1975-2006. The performance of each fund is measured with the unconditional four-factor model. Panel B displays the growth in the mutual fund industry (proxied by the total number of funds used to compute $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$ over time), as well as its average alpha (in % per year).

PANEL A: PROPORTIONS OF UNSKILLED AND SKILLED FUNDS
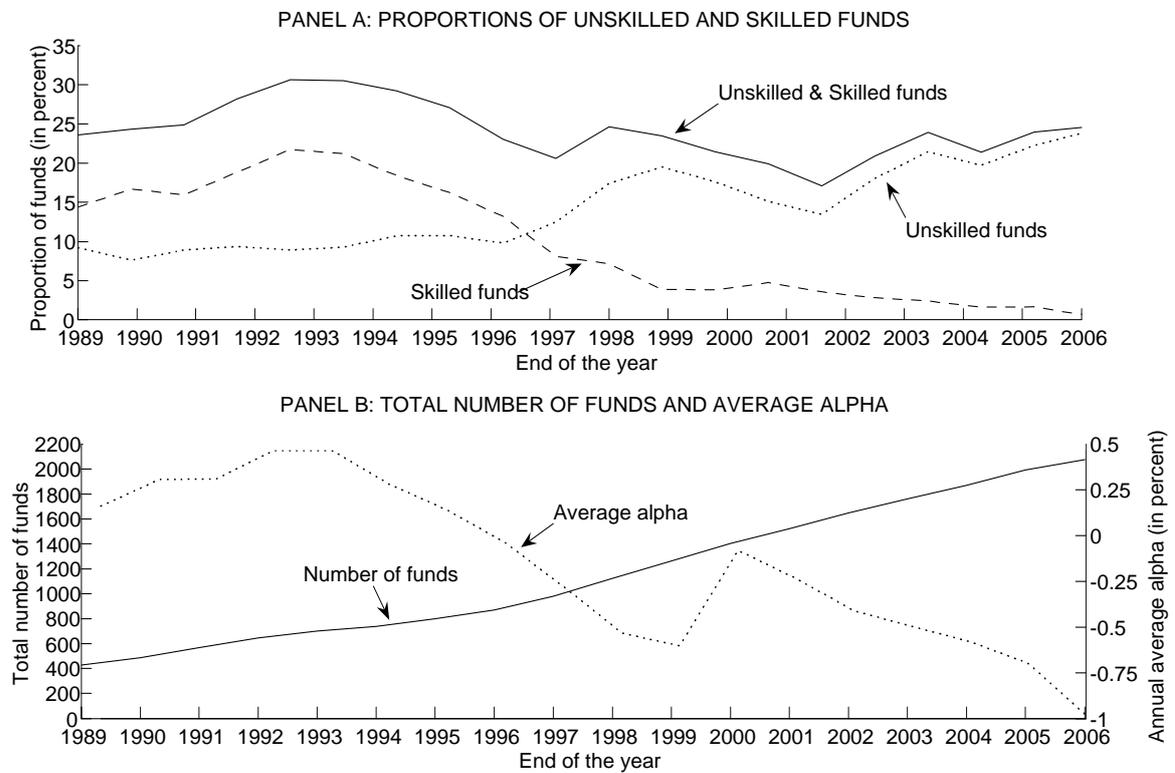

PANEL B: TOTAL NUMBER OF FUNDS AND AVERAGE ALPHA

Figure 5

Performance of the Portfolios FDR10% over Time

The graph plots the evolution of the estimated annual four-factor alpha of the portfolio $FDR10\%$. This portfolio contains funds located in the right tail of the cross-sectional t distribution such that the targeted proportion of lucky funds included in the portfolio is equal to 10%. At the end of each year from 1989 to 2006, the portfolio's alpha is estimated using the portfolio return history up to that point. The initial estimates cover the period 1980-1989 (the first five years are used for the initial portfolio formation on December 31, 1979), while the last ones use the entire portfolio history from 1980 up to 2006. For comparison purposes, we also show the performance of top decile portfolios formed according to a t-statistic ranking, where the t-statistic is estimated over the prior one and three years, respectively.