THE JOURNAL OF FINANCE • VOL. LXV, NO. 1 • FEBRUARY 2010

False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas

LAURENT BARRAS, OLIVIER SCAILLET, and RUSS WERMERS*

ABSTRACT

This paper develops a simple technique that controls for "false discoveries," or mutual funds that exhibit significant alphas by luck alone. Our approach precisely separates funds into (1) unskilled, (2) zero-alpha, and (3) skilled funds, even with dependencies in cross-fund estimated alphas. We find that 75% of funds exhibit zero alpha (net of expenses), consistent with the Berk and Green equilibrium. Further, we find a significant proportion of skilled (positive alpha) funds prior to 1996, but almost none by 2006. We also show that controlling for false discoveries substantially improves the ability to find the few 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), Elton et al. (1993), and Carhart (1997)), recent papers indicate that some fund managers have stock selection skills. For instance, Kosowski et al. (2006;

*Barras is at the Desautels Faculty of Management at McGill University, Scaillet is at the Swiss Finance Institute at HEC-University of Geneva, and Wermers is at the Robert H. Smith School of Business at the University of Maryland at College Park. We thank Stephen Brown, Bernard Dumas, Amit Goyal, Mark Grinblatt, Mark Huson, Andrew Metrick, Lars Pedersen, Elvezio Ronchetti, René Stulz, Sheridan Titman, Maria-Pia Victoria-Feser, and Michael Wolf, as well as seminar participants at Banque Cantonale de Genève, BNP Paribas, Bilgi University, CREST, Greqam, Imperial College, INSEAD, London School of Economics, Maastricht University, MIT, Princeton University, Queen Mary, Solvay Business School, NYU (Stern School), UBP Geneva, Universita della Svizzera Italiana, University of Geneva, University of Georgia, University of Indiana, University of Missouri, University of Notre-Dame, University of Pennsylvania, Vienna University of Economics and Business Administration, University of Virginia (Darden), the Swiss Doctoral Workshop (2005), the Research and Knowledge Transfer Conference (2006), the Zeuthen Financial Econometrics Workshop (2006), the Professional Asset Management Conference at RSM Erasmus University (2008), the Joint University of Alberta/Calgary Finance Conference (2008), the 2005 European Conference of the Econom[etr]ics Community, 2006 Econometric Society European Meeting, 2006 European Conference on Operational Research, 2006 International Congress of Actuaries, 2006 French Finance Association Meeting, 2006 Swiss Society for Financial Market Research Meeting, and 2007 Campus for Finance Meeting (Otto Beisheim School of Management) for their comments. We are also grateful to Campbell Harvey (the editor), an associate editor, and the referee (anonymous) for numerous helpful insights. This paper won the 2008 Banque Privée Espírito Santo Prize for best paper of the Swiss Finance Institute. The first and second authors acknowledge financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK). Part of this research was done while the second author was visiting the Centre Emile Bernheim (ULB).

KTWW) use a bootstrap technique to document outperformance by some funds, while Baks, Metrick, and Wachter (2001), Pástor and Stambaugh (2002b), and Avramov and Wermers (2006) illustrate the benefits of investing in actively managed funds from a Bayesian perspective. Although 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 stock-picking skills, 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* alpha 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, $\hat{\alpha}$. In implementing such a procedure, we are actually conducting a multiple hypothesis test, because we simultaneously examine the performance of all funds in the population (instead of just one fund).[2] However, a simple count of significant-alpha funds does not properly adjust for luck in such a multiple test setting—many of the funds will have significant estimated alphas by luck alone (i.e., their true alphas are zero). To illustrate, consider a population of funds with skills just sufficient to cover trading costs and expenses (truly zero-alpha funds). With the usual significance level of 5%, we should expect that 5% of these zero-alpha funds will have significant estimated alphas—some of them will be unlucky (significant with $\hat{\alpha} < 0$) while others will be lucky (significant with $\hat{\alpha} > 0$), but all will be "false discoveries"—funds with significant *estimated* alphas, but zero *true* alphas.

This paper implements a new approach to controlling for false discoveries in such a multiple fund setting. Our approach much more precisely estimates (1) the proportions of unskilled and skilled funds in the population (those with *truly* negative and positive 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. One main virtue of our approach is its simplicity: to determine the frequency of false discoveries, the only parameter needed is the proportion of zero-alpha funds in the population, π_0. Rather than arbitrarily impose a prior assumption on π_0, as in past studies, our approach estimates it with a straightforward computation that uses the p-values of individual fund

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

[2] This multiple test should not be confused with the joint test of the null hypothesis that all fund alphas are equal to zero in a sample (e.g., Grinblatt and Titman (1989)) or to the KTWW test of single-fund performance. The first test addresses 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. The second test examines the skills of a single fund that is chosen from the universe of alpha-ranked funds. In contrast, our approach simultaneously estimates the prevalence and location of multiple outperforming funds in a group. As such, our approach examines fund performance from a more general perspective, with a richer set of information about active fund manager skills.

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 a much more accurate partition of the universe of mutual funds into zero-alpha, unskilled, and skilled funds than previous approaches that impose an a priori assumption about the proportion of zero-alpha funds in the population.

Another important advantage of our approach to multiple testing is its robustness to cross-sectional dependencies among fund estimated alphas. Prior literature indicates that such dependencies, which exist due to herding and other correlated trading behaviors (e.g., Wermers (1999)), greatly complicate performance measurement in a group setting. With our approach, the computation of the proportions of unskilled and skilled funds only requires the (alpha) p-value for each fund in the population, and not the estimation of the cross-fund covariance matrix. Indeed, the large cross-section of funds in our database makes these estimated proportions very accurate estimators of the true values, even when funds are cross-sectionally correlated. We confirm, with Monte Carlo simulations, that our simple approach is quite robust to cross-fund dependencies.

We apply our novel approach to the monthly returns of 2,076 actively managed U.S. open-end, domestic equity mutual funds that exist at any time between 1975 and 2006 (inclusive), and revisit several important themes examined in the previous literature. We start with an examination of the long-term (lifetime) performance of these funds, net of trading costs and expenses. Our decomposition of the population reveals that 75.4% are zero-alpha funds—funds that have managers with some stock-picking ability, but that extract all of the rents generated by these abilities through fees. Further, 24.0% of the funds are unskilled (true $\alpha < 0$), while only 0.6% are skilled (true $\alpha > 0$)—the latter being statistically indistinguishable from zero. Although our empirical finding that the majority are zero-alpha funds is supportive of the long-run equilibrium theory of Berk and Green (2004; BG), 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 underperform for long time periods, indicating that investors have had some time to evaluate and identify them as underperformers. Across the investment subgroups, aggressive growth funds have the highest proportion of skilled managers, while none of the growth and income funds exhibit skills.

We also uncover some notable time trends in our study. Specifically, we observe that the proportion of skilled funds decreases from 14.4% in early 1990 to 0.6% in late 2006, while the proportion of unskilled funds increases from 9.2% to 24.0%. Thus, although the number of actively managed funds dramatically increases over this period, skilled managers (those capable of picking stocks well enough, over the long-run, to overcome their trading costs and expenses) have become exceptionally rare.

Motivated by the possibility that funds may outperform over the short run, before investors compete away their performance with inflows (as modeled by BG), we conduct further tests over 5-year subintervals, treating each 5-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" over short time periods. These skilled funds are concentrated 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). Further analysis indicates that larger and older funds consist of far more unskilled funds than smaller and newer funds, and that high inflow funds exhibit the highest proportion of skilled funds (18%) during the 5 years ending with the flow year, but the largest reduction in skilled funds during the 5 years subsequent to the flow year (from 18% to 2.4%). Conversely, funds in the lowest flow quintile exhibit high proportions of unskilled funds prior to the measured flows, but much lower proportions afterwards (perhaps due to a change in strategy or portfolio manager in response to the outflows; Lynch and Musto (2003)). These results are generally consistent with the predictions of the BG model.

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 such false discoveries. Forming this false discovery controlled portfolio at the beginning of each year from January 1980 to 2006, we find a four-factor alpha of 1.45% per year, which is statistically significant. Notably, we show that 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. Although fund managers may be able to pick stocks well enough to cover their trading costs, they usually do not exert direct control over the level of fund expenses and fees—management companies set these expenses, with the approval of fund directors. We find, on a pre-expense basis, a much higher incidence of funds with positive alphas—9.6%, compared to our above-mentioned 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 noteworthy that the proportion of skilled managers (before expenses) declines substantially over time, again indicating that skilled portfolio managers have become 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 24.0% to 4.5%), indicating a big role for excessive fees (relative to manager stock-picking skills in excess of trading costs) in underperforming funds. Although industry sources argue that competition among funds has reduced fees and expenses substantially since 1980 (Rea and Reid (1998)), our study indicates that a large subgroup of investors are either unaware that they are being overcharged (Christoffersen and Musto (2002)), or constrained to invest in high-expense funds (Elton, Gruber, and Blake (2007)).

The remainder of this paper is organized as follows. Section I explains our approach to separating luck from skill in measuring the performance of asset managers. Section II presents the performance measures, and describes the mutual fund data. Section III contains the results of the paper, while Section IV concludes.

I. The Impact of Luck on Mutual Fund Performance

A. Overview of the Approach

A.1. Luck in a Multiple Fund Setting

Our objective is to develop a framework to precisely estimate the fraction 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). Our performance categories are defined as follows:

- Unskilled funds: funds that have managers with stock-picking skills insufficient to recover their trading costs and expenses, creating an "alpha shortfall" ($\alpha < 0$),
- Zero-alpha funds: funds that have managers with stock-picking skills sufficient to just recover trading costs and expenses ($\alpha = 0$), and
- Skilled funds: funds that have managers with stock-picking skills sufficient to provide an "alpha surplus," beyond simply recovering trading costs and expenses ($\alpha > 0$).

Note that our above definition of skill is one that captures performance in excess of expenses, and not in an absolute sense. This definition is driven by the idea that consumers search for actively managed mutual funds that deliver surplus alpha, net of all expenses.[3]

Of course, we cannot observe the true alphas of each fund in the population. So, how do we best infer the prevalence of each of the above skill groups from performance estimates for individual funds? First, we use the t-statistic $\hat{t}_i = \hat{\alpha}_i/\hat{\sigma}_{\hat{\alpha}_i}$ as our performance measure, where $\hat{\alpha}_i$ is the estimated alpha for fund i and $\hat{\sigma}_{\hat{\alpha}_i}$ is its estimated standard deviation—KTWW show that the t-statistic has superior statistical properties relative to alpha because alpha estimates have differing precision across funds with varying lives and portfolio

[3] However, perhaps a manager exhibits skill sufficient to more than compensate for trading costs, but the fund management company overcharges fees or inefficiently generates other services (such as administrative services, e.g., record-keeping)—costs that the manager usually has little control over. In a later section (III.D.1), we redefine stock-picking skill in an absolute sense (net of trading costs only) and revisit some of our basic tests to be described.

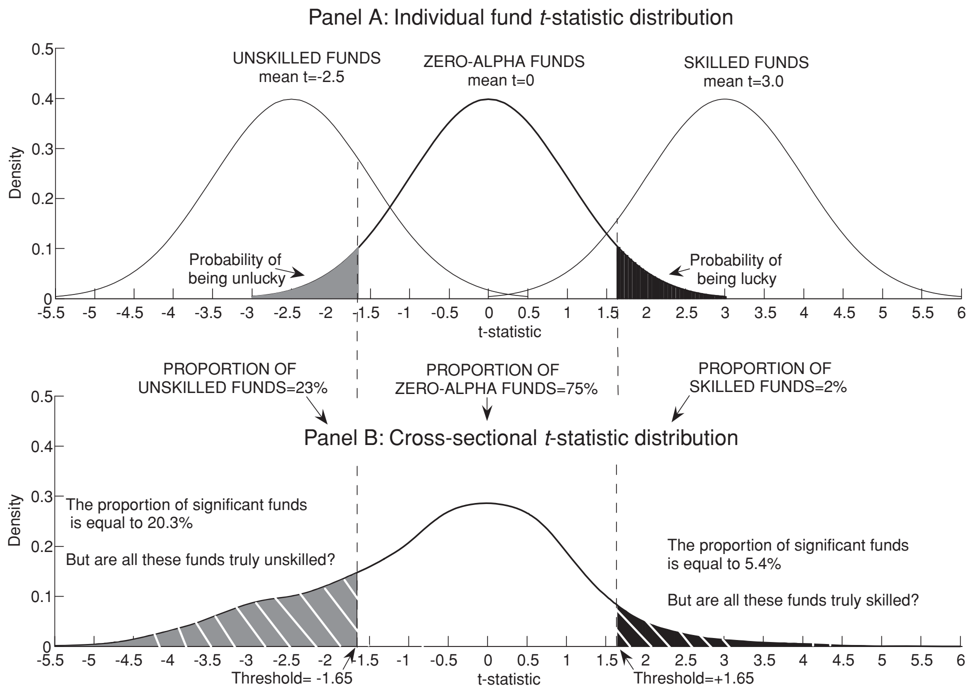

Figure 1. Outcome of the multiple performance test. Panel A shows the distribution of the fund t-statistic across the three skill groups (zero-alpha, unskilled, and skilled funds). We set the true four-factor alpha equal to -3.2% and +3.8% per year for the unskilled and skilled funds (implying that the t-statistic distributions are centered at -2.5 and +3). 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.

volatilities. Second, after choosing a significance level, γ (e.g., 10%), we observe whether $\hat{t}_i$ lies outside the thresholds implied by γ (denoted by t_γ^- and t_γ^+) and label it "significant" if it is such an outlier. This procedure, simultaneously applied across all funds, is a multiple hypothesis test (for several null hypotheses, $H_{0,i}$, and alternative hypotheses, $H_{A,i}$, $i = 1, \ldots, M$):

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

To illustrate the difficulty of controlling for luck in this multiple test setting, Figure 1 presents a simplified hypothetical example that borrows from our empirical findings (to be presented later) over the last 5 years of our sample period. In Panel A, individual funds within the three skill groups—unskilled, zero alpha, and skilled—are assumed to have true annual four-factor alphas of

-3.2%, 0%, and 3.8%, respectively (the choice of these values is explained in the Internet Appendix).[4] The individual fund t-statistic distributions shown in the panel are assumed to be normal for simplicity, and are centered at -2.5, 0, and 3.0 (which correspond to the prior-mentioned assumed true alphas; see the Internet Appendix).[5] 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 skill group distributions in Panel A, where the weight on each distribution is equal to the proportion of zero-alpha, unskilled, and skilled funds in the population, denoted by π_0, π_A^-, and π_A^+, respectively (specifically, $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$; see the Internet Appendix).

To illustrate further, suppose that we choose a significance level, γ, of 10% (corresponding to $t_\gamma^- = -1.65$ and $t_\gamma^+ = 1.65$). With the test shown in expression (1), the researcher would expect to find 5.6% 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 of Panel A, zero-alpha funds can exhibit positive and significant estimated t-statistics. 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 of Panel A). 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 applies to unskilled funds that are very lucky.

The message conveyed by Figure 1 is that we measure performance with a limited sample of data, and therefore unskilled and skilled funds cannot easily be distinguished from zero-alpha funds. This problem can be worse if the cross-section of actual skill levels has a complex distribution (and not all fixed at the same levels, as assumed by our simplified example), and is further compounded if a substantial proportion of skilled fund managers have low levels of skill, relative to the error in estimating their t-statistics. To proceed, we must employ a procedure that is able to precisely account for false discoveries, that is, zero-alpha funds that falsely exhibit significant estimated alphas in the face of these complexities.

[4] Individual funds within a given skill group are assumed to have identical true alphas in this illustration. In our empirical section, our approach makes no such assumption. An Internet Appendix for this article is online in the "Supplements and Datasets" section at http//www.afajof.org/supplements.asp.

[5] The actual t-statistic distributions for individual funds are nonnormal for most U.S. domestic equity funds (KTWW). Accordingly, in our empirical section, we use a bootstrap approach to more accurately estimate the distribution of t-statistics for each fund (and their associated p-values).

[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 91% for a skilled fund. Multiplying these two probabilities by the respective proportions represented by their categories (π_0 and π_A^+) gives 5.6%.

A.2. Measuring Luck

How do we measure the frequency of false discoveries in the tails of the cross-sectional (alpha) t-distribution? At a given significance level γ, it is clear that 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) equals

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

To illustrate, if we take our previous example with $\pi_0 = 75\%$ and $\gamma = 0.10$, we find using equation (2) that $E(F_\gamma^+) = 3.75\%$. 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:

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

From Figure 1, we see that $E(S_\gamma^+) = 5.6\%$ (the shaded region in the right tail of Panel B). By subtracting $E(F_\gamma^+) = 3.75\%$, the expected proportion of skilled funds, $E(T_\gamma^+)$, amounts to 1.85%.

Because the probability of a zero-alpha fund 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

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

The significance level, γ, chosen by the researcher determines the segment of the tail examined for lucky versus skilled (or unlucky versus unskilled) mutual funds, as described by equations (3) and (4). This flexibility in choosing γ provides us with opportunities to gain important insights into the merits of active fund management. One objective of this paper—estimating the proportions of unskilled and skilled funds in the entire population, π_A^- and π_A^+—is achieved only by choosing an appropriately large value for γ. Ultimately, 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).

Another objective of this paper—determining the *location* of truly skilled (or unskilled) funds in the tails of the cross-sectional t-distribution—can only be achieved by evaluating equations (3) and (4) at several different values of γ. For instance, if the majority of skilled funds lie in the extreme right tail, then increasing the value of γ from 0.10 to 0.20 in equation (3) would result in a very small increase in $E(T_\gamma^+)$, the proportion of truly skilled funds, because most of the additional significant funds, $E(S_\gamma^+)$, would be lucky funds. Alternatively, if skilled funds are dispersed throughout the right tail, then increases in γ would result in larger increases in $E(T_\gamma^+)$.

To illustrate the impact of fund location, consider two different fund populations (A and B) identical to the one shown in Figure 1 (with $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$), except that the (true) annual alpha of the skilled funds is equal to 3.8% in A (t-mean of 3.0) and 1.9% in B (t-mean of 1.5). Although these two populations have the same proportion of skilled funds ($\pi_A^+ = 2\%$), their locations differ because the skilled funds in A are more concentrated in the extreme right tail. This information is useful for investors trying to form portfolios with skilled managers, because, in population A, the skilled funds can be more easily distinguished from the zero-alpha funds. For instance, by forming a portfolio of the significant funds in A at $\gamma = 0.05$ ($t_\gamma^+ = 1.96$), the investor would obtain an expected alpha of 1.8% per year, as opposed to only 45 basis points in population B.[7] Our approach to fund selection presented later (in Section III.C) explicitly accounts for fund location in order to choose the significance level γ used to construct the portfolio.

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 of zero-alpha funds in the population, π_0. Here, we turn to a recent 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 (two-sided) p-values associated with the (alpha) t-statistics of 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 other hand, p-values of unskilled and skilled funds tend to be very small because their estimated t-statistics 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.

To explain further, a key intuition of the FDR approach is that it uses information from the center of the cross-sectional t-distribution (which is dominated by zero-alpha funds) to correct for luck in the tails. To illustrate the FDR procedure, suppose we randomly draw 2,076 t-statistics (the number of funds in our study), each from one of the three t-distributions in Panel A of Figure 1—with probability according to our estimates of the proportion of unskilled, zero-alpha, and skilled funds in the population, $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$, respectively. Thus, our draw of t-statistics comes from a known frequency of each type (75%, 23%, and 2%, respectively). Next, we apply the

[7] From Figure 1 (Panel A), the probability of including a zero-alpha fund (skilled fund) in the portfolio equals 2.5% (85%) in population A. This gives $E(T_\gamma^+) = \pi_A^+ \cdot 85\% = 1.7\%$, $E(F_\gamma^+) = \pi_0 \cdot 2.5\% = 1.8\%$, $E(S_\gamma^+) = 3.5\%$, and an expected alpha of $(E(T_\gamma^+)/E(S_\gamma^+)) \cdot 3.8\% = 1.8\%$ per year.

[8] To see this, we denote by T_i and P_i the t-statistic and p-value of the zero-alpha fund, $\hat{t}_i$ and $\hat{p}_i$ their estimated values, and $T_i(P_i)$ the t-statistic associated with the p-value, P_i. We have $\hat{p}_i = 1 - F(|\hat{t}_i|)$, where $F(|\hat{t}_i|) = \text{prob}(|T_i| < |\hat{t}_i| | \alpha_i = 0)$. The p-value P_i is uniformly distributed over $[0, 1]$ because its cdf, $\text{prob}(P_i < \hat{p}_i) = \text{prob}(1 - F(|T_i(P_i)|) < \hat{p}_i) = \text{prob}(|T_i(P_i)| > F^{-1}(1 - \hat{p}_i)) = 1 - F(F^{-1}(1 - \hat{p}_i)) = \hat{p}_i$.

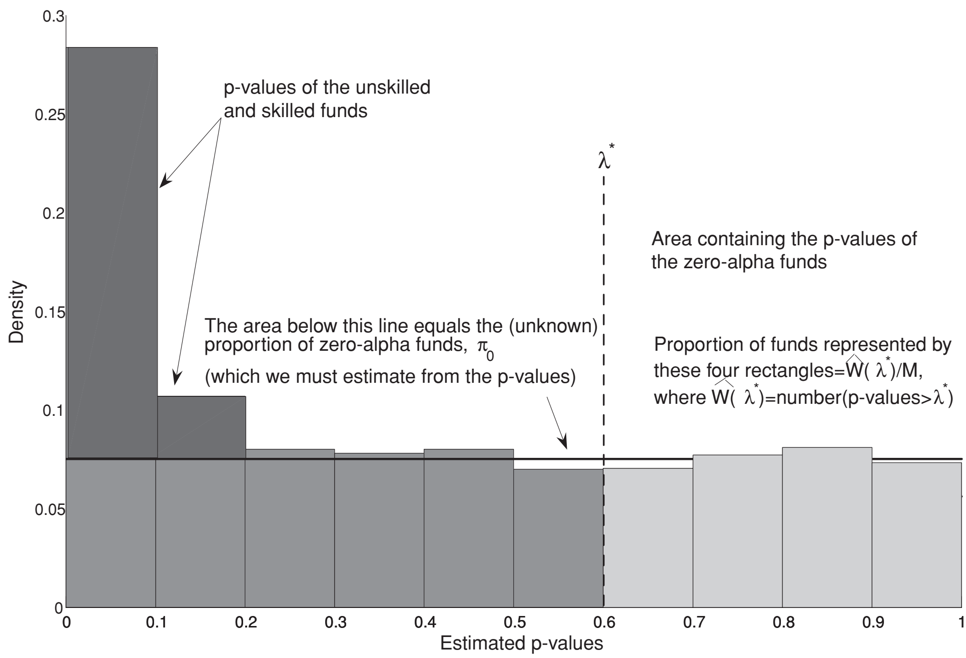

Figure 2. Histogram of fund *p*-values. This figure represents the p-value histogram of $M =$ 2,076 funds (as in our database). For each fund, we randomly draw its t-statistic from one of the distributions in Figure 1 (Panel A) according to the proportion of zero-alpha, unskilled, and skilled funds in 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 all funds from their respective sampled t-statistics and plot them in the histogram.

FDR technique to estimate these frequencies: from the *sampled t*-statistics, we compute two-sided p-values for each of the 2,076 funds, then plot them in Figure 2.

Given the sampled p-values, we estimate π_0 as follows. First, we know that the vast majority of p-values larger than a sufficiently high threshold, λ^* (e.g., $\lambda^* = 0.6$, as shown in the figure), come from zero-alpha funds. Accordingly, after choosing λ^*, we measure the proportion of the total area that is covered by the four lightest grey bars to the right of λ^*, $\hat{W}(\lambda^*)/M$ (where $\hat{W}(\lambda^*)$ equals the number of funds with p-values exceeding λ^*). Note the nearly uniform mass of sampled p-values in intervals between 0.6 and 1—each interval has a mass close to 0.075. Extrapolating this area over the entire region between zero and one, we have

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

which indicates that our estimate of the proportion of zero-alpha funds, $\hat{\pi}_0(\lambda^*)$, is close to 75%, which is the true (but unknown to the researcher) value of π_0

(because the 75% proportion of zero-alpha funds have uniformly distributed p-values).[9]

To select λ^*, we apply a simple bootstrap procedure introduced by Storey (2002), which minimizes the estimated mean squared error (MSE) of $\hat{\pi}_0(\lambda)$ (see the Internet Appendix).[10] Although the main advantage of this procedure is that it is entirely data driven, we find that $\hat{\pi}_0(\lambda^*)$ is not overly sensitive to the choice of λ^*. For instance, a simple approach that fixes the value of λ^* to intermediate levels (such as 0.5 or 0.6) produces estimates similar to the MSE approach (see the Internet Appendix).

Substituting the resulting estimate, $\hat{\pi}_0$, in equations (2) and (3), and proxying $E(S_\gamma^+)$ with the observed proportion of significant funds in the right tail, $\hat{S}_\gamma^+$, we can easily estimate the $E(F_\gamma^+)$ and $E(T_\gamma^+)$ that correspond to any chosen significance level, γ. The same approach can be used in the left tail by proxying $E(S_\gamma^-)$ in equation (4) with the observed proportion of significant funds in the left tail, $\hat{S}_\gamma^-$. This implies the following estimates of the proportions of unlucky and lucky funds:

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

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

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

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

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

where γ^* is a sufficiently high significance level—similar to the choice of λ^*, we select γ^* with a bootstrap procedure that minimizes the estimated MSE of $\hat{\pi}_A^-$ and $\hat{\pi}_A^+$ (see the Internet Appendix). Although this method is entirely data driven, there is some flexibility in the choice of γ^*, as long as it is sufficiently high. In the Internet Appendix, we find that simply setting γ^* to pre-specified values (such as 0.35 or 0.45) produces estimates similar to the MSE approach.

B. Comparison of Our Approach with Existing Methods

The previous literature has followed two alternative approaches when estimating the proportions of unskilled and skilled funds. The "full luck" approach

[9] This estimation procedure cannot be used in a one-sided multiple test because 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 that are not uniformly distributed over [0, 1].

[10] The MSE is the expected squared difference between $\hat{\pi}_0(\lambda)$ and the true value, π_0 : $MSE(\hat{\pi}_0(\lambda)) = E(\hat{\pi}_0(\lambda) - \pi_0)^2$. Because π_0 is unknown, it is proxied with $\min_\lambda \hat{\pi}_0(\lambda)$ to compute the estimated MSE (see Storey (2002)).

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 $\gamma/2$.[11] 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 ($\pi_0 = 0$).

What are the errors introduced by assuming, a priori, that the proportion of zero-alpha funds, π_0, equals zero or one, 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 different possible 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^+. For each π_0, the ratio π_A^-/π_A^+ is held fixed to 11.5 (0.23/0.02), as in Figure 1, to ensure that the proportion of skilled funds remains low compared to the unskilled funds. Second, we use these sampled t-statistics to estimate the proportion of unlucky ($\alpha = 0$, significant with $\hat{\alpha} < 0$), lucky ($\alpha = 0$, significant with $\hat{\alpha} > 0$), unskilled ($\alpha < 0$, significant with $\hat{\alpha} < 0$), and skilled ($\alpha > 0$, significant with $\hat{\alpha} > 0$) 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, then compare the average value of each estimator with its true population value.

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 shown by equation (2). Although the average value of the FDR estimator closely tracks $E(F_\gamma^-)$, this is not the case for the other two approaches. By assuming that $\pi_0 = 0$, the no luck approach consistently underestimates $E(F_\gamma^-)$ when the true proportion of zero-alpha funds is higher ($\pi_0 > 0$). Conversely, the full luck approach, which assumes that $\pi_0 = 1$, overestimates $E(F_\gamma^-)$ when $\pi_0 < 1$. To illustrate the extent of the bias, consider the case where $\pi_0 = 75\%$. Although 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, $E(F_\gamma^+)$, in Panel B are exactly the same because $E(F_\gamma^+) = E(F_\gamma^-)$.

Estimates of the expected proportions of unskilled and skilled funds, $E(T_\gamma^-)$ and $E(T_\gamma^+)$, provided by the three approaches are shown in Panels C and D, respectively. As we move to higher true proportions 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

[11] Jensen (1968 p. 910) summarizes the full luck approach in his study of 115 mutual funds as follows: "... if all 115 of these funds had a true alpha equal to zero, we would expect (merely because of random chance) to find 5% of them or about 5 or 6 funds yielding t-values 'significant' at the 5% level."

[12] We choose $\gamma = 0.20$ to examine a large portion of the tails of the cross-sectional t-distribution. As shown in the Internet Appendix, the results using $\gamma = 0.10$ are similar.

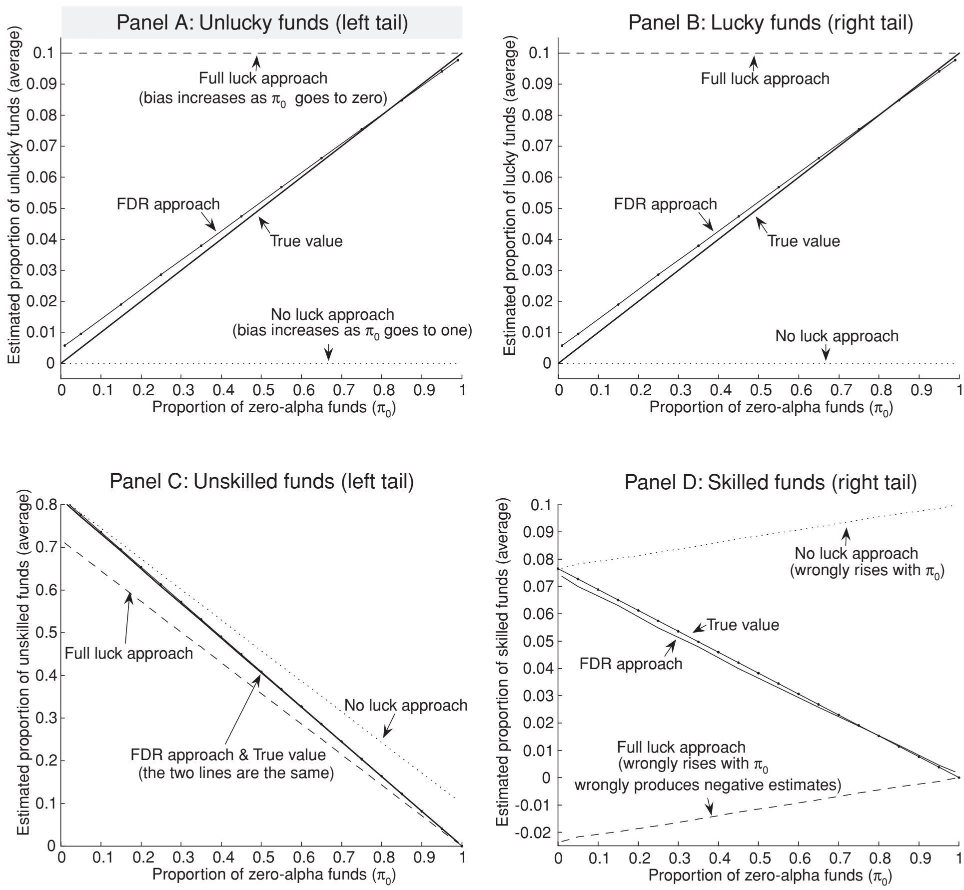

Figure 3. Measuring luck: comparison with existing approaches. This figure examines the bias of different estimators produced by the three approaches (no luck, full luck, and FDR approach) as a function of the true proportion of zero-alpha funds, π_0. We examine the estimators of the proportions of unlucky, 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 the FDR 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 ($i = 1, \ldots, 2{,}076$) 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.

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 $E(T_\gamma^-) + E(T_\gamma^+)$ (17.3% instead of 22.3%).

Panel D reveals that the no luck and full luck approaches also exhibit a nonsensical positive relation between π_0 and $E(T_\gamma^+)$. This result is a consequence of the low proportion of skilled funds in the population. As π_0 rises, the additional lucky funds drive the proportion of significant funds up, making the no luck and full luck approaches wrongly indicate that more skilled funds are present. Further, the excessive luck adjustment of the full luck approach produces estimates of $E(T_\gamma^+)$ below zero.

In addition to the bias properties exhibited by 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 $\hat{\pi}_0$ (the same arguments apply to the other estimators). Because $\hat{\pi}_0$ is a proportion estimator that depends on the proportion of p-values higher than λ^*, the Law of Large Numbers drives it close to its true value with our large sample size. For instance, taking $\lambda^* = 0.6$ and $\pi_0 = 75\%$, the standard deviation of $\hat{\pi}_0$, $\sigma_{\hat{\pi}_0}$, is as low as 2.5% with independent p-values (1/30th the magnitude of π_0).[13] In the Internet Appendix, we provide further evidence of the remarkable accuracy of our estimators using Monte Carlo simulations.

C. 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 because 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 show that a complicated bootstrap is necessary to test the significance of fund performance of a fund located at a particular alpha rank because this test depends on the joint distribution of all fund estimated alphas, that is, 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 produce zero alphas ($\pi_0 = 100\%$), 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 true p-value distribution, as shown in Figure 2. Clearly, we would make serious errors no matter where we set λ^*.

[13] Specifically, $\hat{\pi}_0 = (1-\lambda^*)^{-1} \cdot 1/M \sum_{i=1}^{M} x_i$, where x_i follows a binomial distribution with probability of success $p_{\lambda^*} = \text{prob}(P_i > \lambda^*) = 0.075 \cdot 4 = 0.30$, where P_i denotes the fund p-value (p_{λ^*} equals the rectangle area delimited by the horizontal black line and the vertical line at $\lambda^* = 0.6$ in Figure 2). Therefore, from the standard deviation of a binomial random variable, $\sigma_x = (p_{\lambda^*}(1-p_{\lambda^*}))^{\frac{1}{2}} = 0.46$ and $\sigma_{\hat{\pi}_0} = (1-\lambda^*)^{-1} \cdot \sigma_x/\sqrt{M} = 2.5\%$.

In our sample, we are not overly concerned with dependencies because we find that the average correlation between four-factor model residuals of pairs of funds is only 0.08. 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. Therefore, we believe that cross-sectional dependencies are sufficiently low to allow consistent estimators.[14]

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 data set, 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. These results, as well as further details on the simulation experiment, are discussed in the Internet Appendix.

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 risk-free rate (proxied by the monthly 30-day T-bill beginning-of-month yield); $r_{m,t}$ is the month t excess return on the CRSP NYSE/Amex/NASDAQ 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)),

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

where z_{t-1} denotes the $J \times 1$ vector of predictive variables measured at the end of month t (minus their mean values over 1975 to 2006), and B is the

[14] It is well known that the sample average, $\overline{x} = 1/M \sum x_i$, is a consistent estimator under many forms of dependence (i.e., $\overline{x}$ converges to the true mean value when M is large; see Hamilton (1994), p. 47). Because our FDR estimators can be written as sample averages (see footnote 13), it is not surprising that they are also consistent under cross-sectional dependence among funds (for further discussion, see Storey, Taylor, and Siegmund (2004)).

$J \times 1$ vector of coefficients. The four predictive variables are the 1-month T-bill yield; the dividend yield of the Center for Research in Security Prices (CRSP) value-weighted NYSE/Amex stock index; the term spread, proxied by the difference between yields on 10-year treasuries and 3-month T-bills; and the default spread, proxied by the yield difference between Moody's Baa-rated and Aaa-rated corporate bonds. We also compute fund alphas using the CAPM and the Fama and French (1993) models. These results are summarized in Section III.D.2.

To compute each fund t-statistic, we use the Newey and West (1987) heteroskedasticity and autocorrelation consistent estimator of the standard deviation, $\hat{\sigma}_{\hat{\alpha}_i}$. Further, KTWW find that the finite-sample distribution of the t-statistic is nonnormal for approximately half of the funds. Therefore, we use a bootstrap procedure (instead of asymptotic theory) to compute fund p-values for the two-sided tests with equal tail significance level, $\gamma/2$ (see the Internet Appendix). In order to estimate the distribution of the t-statistic 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 $\{\hat{\varepsilon}_{i,t}\}$.[15] For each fund, we implement 1,000 bootstrap replications. The reader is referred to KTWW for details on this bootstrap procedure.

B. Mutual Fund Data

We use monthly mutual fund return data provided by the 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 share classes 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 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 because CRSP fills this with the cumulated return since the last nonmissing return. In results presented in the Internet Appendix, we find that reducing the minimum fund return requirement to 36 months has no material impact on our main results, and 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

[15] To determine whether assuming homoskedasticity and temporal independence in individual fund residuals is appropriate, we have checked for heteroskedasticity (White test), autocorrelation (Ljung-Box test), and Arch effects (Engle test). We find 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.

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 for growth, aggressive growth, and growth and income funds. The regressions are based on monthly data between January 1975 and December 2006. Each panel contains the estimated annualized alpha ($\hat{\alpha}$), the estimated exposures to the market ($\hat{b}_m$), size ($\hat{b}_{smb}$), book-to-market ($\hat{b}_{hml}$), and momentum factors ($\hat{b}_{mom}$), 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 and autocorrelation consistent estimates of p-values under the null hypothesis that the regression parameters are equal to zero.

	$\hat{\alpha}$	$\hat{b}_m$	$\hat{b}_{smb}$	$\hat{b}_{hml}$	$\hat{b}_{mom}$	R^2
			Panel A: Unconditional Four-Factor Model			
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	−0.53%	1.04	0.43	−0.17	0.09	95.8%
Growth (388)	(0.22)	(0.00)	(0.00)	(0.00)	(0.00)	
Growth &	−0.47%	0.87	−0.04	0.17	−0.03	98.2%
Income (384)	(0.09)	(0.00)	(0.02)	(0.00)	(0.01)	
			Panel B: Conditional Four-Factor Model			
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	−0.49%	1.05	0.43	−0.19	0.08	96.2%
Growth (388)	(0.24)	(0.00)	(0.00)	(0.00)	(0.00)	
Growth &	−0.58%	0.87	−0.04	0.16	−0.03	98.3%
Income (384)	(0.05)	(0.00)	(0.02)	(0.00)	(0.02)	

investment categories: growth (1,304 funds), aggressive growth (388 funds), and growth and income (384 funds). If an investment objective is missing, the prior nonmissing 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.

Similar to results previously documented in the literature, we find that unconditional estimated alphas for each category are negative, ranging from −0.45% to −0.60% per annum. Aggressive growth funds tilt toward small capitalization, low book-to-market, and momentum stocks, while the opposite holds for growth and income funds. Introducing time-varying market betas provides similar results (Panel B). In further tests shown in the Internet Appendix, we find that using the unconditional or conditional version of the four-factor model has no material impact on our main results. For brevity, in the next section, we present only results from the unconditional four-factor model.

III. Empirical Results

A. The Impact of Luck on Long-Term Performance

We begin our empirical analysis by measuring the impact of luck on long-term mutual fund performance, measured as the lifetime performance of each fund (over the period 1975 to 2006) using the monthly four-factor model of equation (9). Panel A of Table II shows estimated proportions of zero-alpha, unskilled, and skilled funds in the population ($\hat{\pi}_0$, $\hat{\pi}_A^-$, and $\hat{\pi}_A^+$), 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 of Genovese and Wasserman (2004), which is described in the Internet Appendix.

Among the 2,076 funds, we estimate that the majority—75.4%—are zero-alpha funds. Managers of these funds exhibit stock-picking 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 prediction 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 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 prior research.

To be specific, Panel B shows the proportion of significant alpha funds in the left and right tails ($\hat{S}_\gamma^-$ and $\hat{S}_\gamma^+$, 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—$\hat{S}_\gamma^+$ peaks at 8.2% of the total population (170 funds) when $\gamma = 0.20$ (i.e., these 170 funds have a positive estimated alpha with a two-sided p-value below 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 proportions of lucky zero-alpha funds and skilled funds ($\hat{F}_\gamma^+$ and $\hat{T}_\gamma^+$, respectively) using the technique described in Section I.A.3. Clearly, we cannot reject that all of the right tail funds are merely lucky outcomes among the large number (1,565) of zero-alpha funds, and that none have truly skilled managers (i.e., $\hat{T}_\gamma^+$ is not significantly different from zero for any significance level γ).

It is interesting (Panel A) that 24% of the population (499 funds) are truly unskilled fund managers, unable to pick stocks well enough to recover their trading costs and other expenses.[16] Left tail funds, which are overwhelmingly

[16] 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 Pástor and Stambaugh (2002a)).

Table II

Impact of Luck on Long-Term Performance

Long-term performance is measured with the unconditional four-factor model over the entire period 1975 to 2006. Panel A displays the estimated proportions of zero-alpha, unskilled, and skilled funds ($\hat{\pi}_0,\ \hat{\pi}_A^-$, and $\hat{\pi}_A^+$) in the entire fund population (2,076 funds). Panel B counts the proportions of significant funds in the left and right tails of the cross-sectional t-statistic distribution ($\hat{S}_\gamma^-,\ \hat{S}_\gamma^+$) at four significance levels ($\gamma = 0.05$, 0.10, 0.15, 0.20). In the leftmost columns, the significant group in the left tail, $\hat{S}_\gamma^-$, is decomposed into unlucky and unskilled funds ($\hat{F}_\gamma^-,\ \hat{T}_\gamma^-$). In the rightmost columns, the significant group in the right tail, $\hat{S}_\gamma^+$, is decomposed into lucky and skilled funds ($\hat{F}_\gamma^+,\ \hat{T}_\gamma^+$). The bottom of Panel B also presents the characteristics of each significant group ($\hat{S}_\gamma^-,\ \hat{S}_\gamma^+$): the average estimated alpha (% per year), expense ratio (% per year), and turnover (% per year). Figures in parentheses denote the standard deviation of the different estimators.

Panel A: Proportion of Unskilled and Skilled Funds

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	75.4 (2.5)	24.0 (2.3)	0.6 (0.8)
Number	1,565	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. $\hat{S}_\gamma^-$ (%)	11.7	17.4	21.8	25.8	8.2	6.0	4.2	2.2	Signif. $\hat{S}_\gamma^+$ (%)
	(0.7)	(0.8)	(0.9)	(0.9)	(0.6)	(0.5)	(0.4)	(0.3)	
Unlucky $\hat{F}_\gamma^-$ (%)	1.9	3.8	5.6	7.6	7.6	5.6	3.8	1.9	Lucky $\hat{F}_\gamma^+$ (%)
	(0.0)	(0.1)	(0.2)	(0.3)	(0.3)	(0.2)	(0.1)	(0.0)	
Unskilled $\hat{T}_\gamma^-$ (%)	9.8	13.6	16.1	18.2	0.6	0.4	0.4	0.3	Skilled $\hat{T}_\gamma^+$ (%)
	(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)
	(0.2)	(0.2)	(0.1)	(0.1)	(0.3)	(0.4)	(0.5)	(0.7)	
Exp. (%/year)	1.4	1.4	1.4	1.4	1.3	1.2	1.2	1.2	Exp. (%/year)
Turn. (%/year)	100	97	95	95	94	95	95	104	Turn. (%/year)

comprised of unskilled (and not merely unlucky) funds, have a relatively long fund life—12.7 years, on average. Further, these funds generally perform poorly over their entire lives, making their survival puzzling. Perhaps, as discussed by Elton, Gruber, and Busse (2004), 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)).

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), while turnover does not vary systematically among the various tail segments.

In the Internet Appendix, we repeat the long-term performance test described above for investment objective subgroups—growth, aggressive growth, and growth and income. The overall results are as follows. Growth funds show similar results to the overall universe of funds: 76.5% have zero alphas, 23.5% are unskilled, while none are skilled. Performance is somewhat better for aggressive growth funds, as 3.9% of them show true skills. Finally, growth and income funds consist of the largest proportion of unskilled funds (30.7%), but have no skilled funds. The long-term survival of these actively managed funds, which includes "value funds" and "core funds" is remarkable in light of these poor results.

As noted by Wermers (2000), the universe of U.S. domestic equity mutual funds has expanded substantially since 1990. Accordingly, the proportions of unskilled and skilled funds estimated over the entire period 1975 to 2006 may not accurately describe the performance generated by the industry prior to this rapid expansion. To address this issue, we next examine the evolution of the long-term proportions of unskilled and skilled funds over time. At the end of each year from 1989 to 2006, we estimate the proportions of unskilled and skilled funds ($\hat{\pi}_A^-$ and $\hat{\pi}_A^+$, respectively) using the entire return history for each fund up to that point in time. As we move forward in time, we add new mutual funds once they exhibit a 60-month record. To illustrate, our initial estimates, on December 31, 1989, cover the first 15 years of the sample, 1975 to 1989 (427 funds), while our final estimates, on December 31, 2006, are based on the entire 32 years, 1975 to 2006 (2,076 funds; these are the estimates shown in Panel A of Table II).[17] The results in Panel A of Figure 4 show that the proportion of funds with non-zero alphas (equal to the sum of the proportions of skilled and unskilled funds) remains fairly constant over time. However, there are dramatic changes in the relative proportions of unskilled and skilled

[17] The dynamic proportion estimators, $\hat{\pi}_0$, $\hat{\pi}_A^-$, and $\hat{\pi}_A^+$, measured at the end of each year treat the universe of existing funds as a new fund population (to be included, a fund must have at least 60 return observations, ending with that year). For these estimators to be accurate (in terms of bias and variability), it is necessary that the cross-sectional fund dependence at each point in time remains sufficiently low (see Section I.C).

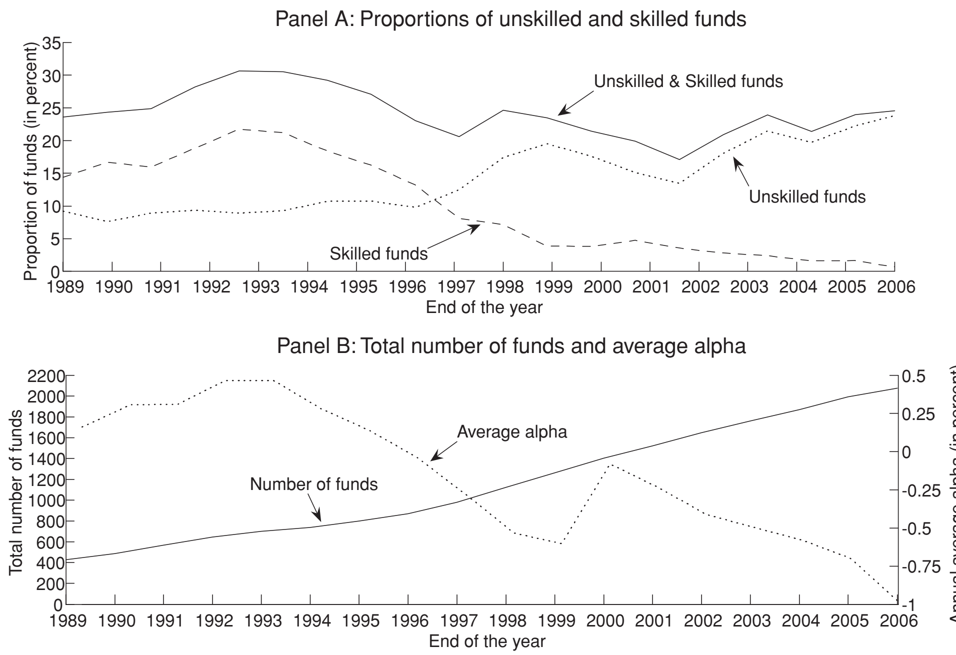

Figure 4. Evolution of mutual fund performance over time. Panel A plots the evolution of the estimated proportions of unskilled and skilled funds ($\hat{\pi}_A^-$ and $\hat{\pi}_A^+$) between 1989 and 2006. At the end of each year, we measure $\hat{\pi}_A^-$ and $\hat{\pi}_A^+$ using the entire fund return history up to that point. The initial estimates at the end of 1989 cover the period 1975 to 1989, while the last ones in 2006 use the period 1975 to 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 $\hat{\pi}_A^-$ and $\hat{\pi}_A^+$ over time), as well as its average alpha (in % per year).

funds from 1989 to 2006. Specifically, the proportion of skilled funds declines from 14.4% to 0.6%, while the proportion of unskilled funds 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. (Note that this is averaged across funds, while Table 1 computes the alpha of a monthly equal-weighted portfolio of funds.)

Further, Panel B shows the yearly count of funds included in the estimated proportions of Panel A. From 1996 to 2005, there are more than 100 additional actively managed domestic equity mutual funds (having a 60-month history) per year. Interestingly, this coincides with the time-variation in the proportions of unskilled and skilled funds shown in Panel A, which can be attributed to two distinct sources. First, new funds created during the 1990s generate very poor performance, as we find that 24% of them are unskilled, while none are skilled (i.e, $\hat{\pi}_A^- = 24.0\%$ and $\hat{\pi}_A^+ = 0\%$). Because these 1,328 new funds account for more than 60% of the total population (2,076), they greatly contribute to the performance decline shown in Panel A. Second, our results suggest that the growth in the industry has also affected the alpha of the older funds created before January 1990. Although many of these 748 funds exhibit truly positive performance up to December 1996 ($\hat{\pi}_A^+ = 14.4$, see Panel A), the decline is breathtaking afterwards. Specifically, we estimate that, during 1997 to 2006, 34.8% of these older funds are truly unskilled, while none produce truly positive alphas (i.e., $\hat{\pi}_A^- = 34.8\%$, $\hat{\pi}_A^+ = 0\%$).[18] Either the growth of the fund industry has coincided with greater levels of stock market efficiency, making stock-picking 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 (as shown by Kostovetsky (2007)), and indexed investing increased substantially.

Although increased competition may have decreased the average level of alpha, it is also possible that funds do not achieve superior performance in the long run because flows compete away any alpha surplus. However, we might find evidence of funds with superior short-term alphas before investors become fully aware of such outperformers due to search costs. Because our long-term performance estimates average alphas over time, they are not able to detect such dynamics. To address this issue, in the next section, we investigate whether funds exhibit superior alphas over the short run.[19]

B. The Impact of Luck on Short-Term Performance

To test for short run mutual fund performance, we partition our data into six non-overlapping subperiods of 5 years, beginning with 1977 to 1981 and ending with 2002 to 2006. For each subperiod, we include all funds that have 60 monthly return observations and then compute their respective alpha *p*-values—in other words, we treat each fund during each 5-year period as a separate "fund." We pool these 5-year records together across all time periods to represent the average experience of an investor in a randomly chosen fund during a randomly chosen 5-year period. After pooling, we obtain a total of 3,311 *p*-values from which we compute our different estimators. The results are shown in Table III.

First, Panel A of Table III shows that a small fraction of funds (2.4% of the population) exhibit skill over the short run (with a standard deviation of 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% (i.e., funds having a positive estimated alpha with a two-sided *p*-value below 10%), we capture almost all skilled funds, as $\hat{T}_\gamma^+$ reaches

[18] Under a structural change, the long-term alpha is a time-weighted average of the two subperiod alphas. A zero or negative performance after 1996 progressively drives the long-term alphas of the skilled funds towards zero. This explains why our estimate of the proportion of skilled funds at the end of 2006 is close to zero ($\hat{\pi}_A^+ = 0.6\%$). We have verified this pattern using the Monte Carlo setting described in the Internet Appendix. Assuming that all skilled funds become zero-alpha (unskilled) after 1996, we find that the average value of $\hat{\pi}_A^+$ (1,000 iterations) over the entire period equals 2.9% (0.3%).

[19] Time-varying betas may also affect the inference on the estimated alpha. As mentioned earlier, we have measured performance using the conditional version of the four-factor model (equation (10)), and find that the results remained qualitatively unchanged (see the Internet Appendix).

Table III

Impact of Luck on Short-Term Performance

Short-term performance is measured with the unconditional four-factor model over non-overlapping 5-year periods between 1977 and 2006. The different estimates shown in the table are computed from the pooled alpha p-values across all 5-year periods. Panel A displays the estimated proportions of zero-alpha, unskilled, and skilled funds ($\hat{\pi}_0$, $\hat{\pi}_A^-$, and $\hat{\pi}_A^+$) in the population (3,311 funds). Panel B counts the proportions of significant funds in the left and right tails of the cross-sectional t-statistic distribution ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) at four significance levels (γ=0.05, 0.10, 0.15, 0.20). In the leftmost columns, the significant group in the left tail, $\hat{S}_\gamma^-$, is decomposed into unlucky and unskilled funds ($\hat{F}_\gamma^-$, $\hat{T}_\gamma^-$). In the rightmost columns, the significant group in the right tail, $\hat{S}_\gamma^+$, is decomposed into lucky and skilled funds ($\hat{F}_\gamma^+$, $\hat{T}_\gamma^+$). The bottom of Panel B also presents the characteristics of each significant group ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$): the average estimated alpha (% per year), expense ratio (% per year), and turnover (% per year). Figures in parentheses denote the standard deviation of the different estimators.

Panel A: Proportion of Unskilled and Skilled Funds

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	72.2 (2.0)	25.4 (1.7)	2.4 (0.7)
Number	2,390	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. $\hat{S}_\gamma^-$ (%)	11.2	16.8	21.4	24.9	9.6	7.8	5.9	3.5	Signif. $\hat{S}_\gamma^+$ (%)
	(0.5)	(0.6)	(0.7)	(0.8)	(0.5)	(0.5)	(0.4)	(0.3)	
Unlucky $\hat{F}_\gamma^-$ (%)	1.8	3.6	5.4	7.2	7.2	5.4	3.6	1.8	Lucky $\hat{F}_\gamma^+$ (%)
	(0.0)	(0.0)	(0.1)	(0.2)	(0.2)	(0.1)	(0.0)	(0.0)	
Unskilled $\hat{T}_\gamma^-$ (%)	9.4	13.2	16.0	17.7	2.4	2.4	2.3	1.7	Skilled $\hat{T}_\gamma^+$ (%)
	(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)
	(0.2)	(0.2)	(0.1)	(0.1)	(0.3)	(0.4)	(0.4)	(0.6)	
Exp. (%/year)	1.4	1.3	1.3	1.3	1.2	1.2	1.2	1.2	Exp. (%/year)
Turn. (%/year)	98	95	94	93	80	80	81	78	Turn. (%/year)

2.3% (close to its maximum value of 2.4%). Proceeding toward the center of the distribution (by increasing γ to 0.10 and 0.20) produces almost no additional skilled funds, and almost entirely additional zero-alpha funds that are lucky ($\hat{F}_{\gamma}^{+}$). Thus, skilled fund managers, while rare, may be somewhat easy to find because 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.

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, $\hat{T}_{\gamma}^{-}$, is roughly five times the proportion of unlucky funds, $\hat{F}_{\gamma}^{-}$ (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 short-term skills seem to be inversely related to turnover, as indicated by the substantially higher levels of turnover of left tail funds (which are mainly unskilled funds). Unskilled managers apparently trade frequently, in the short run, to appear skilled, which ultimately hurts their performance. Perhaps poor governance of some funds (Ding and Wermers (2009)) explains why they end up in the left tail (net of expenses)—they overexpend on both trading costs (through high turnover) and other expenses relative to their skills.

In the Internet Appendix, we repeat the short-term performance test for investment objective subgroups (growth, aggressive growth, and growth and income funds). We find that the proportions of unskilled funds within the three categories are similar to that of the entire universe (from Table III), with some notable differences. Although aggressive growth funds exhibit somewhat higher skills ($\hat{\pi}_{A}^{+} = 4.2\%$) than growth funds ($\hat{\pi}_{A}^{+} = 2.6\%$), no growth and income funds are able to produce positive short-term alphas.

Because we find evidence of short-term fund manager skills that disappear in the long term, it is interesting to further examine the mechanism through which skills disappear. The model of BG provides guidance for how this process may unfold. Specifically, if competing fund investors chase winning funds (which have higher proportions of truly skilled funds), then superior fund management companies (which are in scarce supply) may capture the majority of the rents they produce. We examine this conjecture in Table IV. Specifically, at the beginning of each (non-overlapping) 5-year period from 1977 to 2006 (similar to Table III), we rank funds into quintiles based on their (1) size (total net assets under management), (2) age (since first offered to the public), and (3) prior-year flows, as a percentage of total net assets. Then, we measure the proportions of zero-alpha, unskilled, and skilled funds ($\hat{\pi}_0$, $\hat{\pi}_{A}^{-}$, and $\hat{\pi}_{A}^{+}$, respectively) within each fund size quintile (Panel A), fund age quintile (Panel B), and fund flow quintile (Panels C and D).

The BG model implies that larger and older funds should exhibit lower alphas because they have presumably grown (or survived) to the point where they provide no superior alphas, net of fees—partly due to flows that followed past superior performance. Smaller and newer funds, on the other hand, may exhibit some skills before investors learn about their superior abilities. Consistent with

Table IV

Fund Characteristics and Performance Dynamics

We examine the relation between short-term performance and fund size (Panel A), age (Panel B), and annual flows (Panels C and D). At the beginning of each non-overlapping 5-year period between 1977 and 2006, funds are ranked according to each characteristic and grouped into quintiles (Low, 2, 3, 4, High). Short-term performance is measured with the unconditional four-factor model over the next 5 years, except for Panel C (Annual Flow-Past Performance), where we use the previous 5 years. For each quintile, we pool the fund alpha *p*-values, characteristic levels, and estimated alphas across all 5-year periods to compute the estimated proportions of zero-alpha, unskilled, and skilled funds ($\hat{\pi}_0$, $\hat{\pi}_A^-$, and $\hat{\pi}_A^+$), average characteristic levels, and average estimated alphas ($\hat{\alpha}$). Median Size denotes the median quintile total net asset under management (million USD), while Avg. Age and Flow denote the average quintile age (years), and annual flow (%). Figures in parentheses denote the standard deviation of the different estimators.

Quintile	Low	2	3	4	High	High–Low
			Panel A: Size (TNA)			
Zero-alpha ($\hat{\pi}_0$)	81.0 (3.5)	72.2 (4.0)	77.7 (3.8)	64.2 (4.2)	62.1 (4.2)	−18.9
Unskilled ($\hat{\pi}_A^-$)	16.4 (3.1)	23.1 (3.7)	22.3 (3.5)	33.5 (3.9)	34.3 (3.9)	+17.9
Skilled ($\hat{\pi}_A^+$)	2.6 (1.6)	4.7 (1.7)	0.0 (1.5)	2.3 (1.5)	3.6 (1.6)	+1.0
Median Size (million $)	9.8	52.9	166.0	453.1	1,651.7	+1,641.9
Avg. $\hat{\alpha}$ (%/year)	−0.5 (0.1)	−0.6 (0.1)	−1.1 (0.1)	−1.1 (0.1)	−0.9 (0.1)	−0.4
			Panel B: Age			
Zero-alpha ($\hat{\pi}_0$)	79.6 (3.5)	65.0 (4.2)	72.5 (3.7)	70.2 (4.0)	70.1 (4.2)	−9.5
Unskilled ($\hat{\pi}_A^-$)	16.5 (3.0)	29.8 (3.9)	25.5 (3.4)	26.7 (3.6)	29.9 (4.0)	+13.4
Skilled ($\hat{\pi}_A^+$)	3.9 (1.7)	5.2 (1.6)	2.0 (1.5)	3.1 (1.5)	0.0 (1.3)	−3.9
Avg. Age (years)	2.1	5.2	8.6	15.5	37.8	+35.7
Avg. $\hat{\alpha}$ (%/year)	−0.3 (0.1)	−0.8 (0.1)	−0.9 (0.1)	−0.7 (0.1)	−1.4 (0.1)	−1.1
			Panel C: Annual Flow—Past Performance			
Zero-alpha ($\hat{\pi}_0$)	52.9 (4.0)	73.5 (3.8)	84.0 (2.7)	71.0 (3.8)	78.6 (3.5)	+25.7
Unskilled ($\hat{\pi}_A^-$)	47.1 (3.8)	26.5 (3.5)	16.0 (2.4)	22.5 (3.5)	3.4 (1.6)	−43.7
Skilled ($\hat{\pi}_A^+$)	0.0 (1.2)	0.0 (1.2)	0.0 (1.3)	6.5 (1.8)	18.0 (3.0)	+18.0
Avg. Flow (%/year)	−26.8	−11.0	−3.2	7.5	67.5	+94.3
Avg. $\hat{\alpha}$ (%/year)	−2.8 (0.1)	−1.7 (0.1)	−0.9 (0.1)	0.1 (0.1)	1.2 (0.1)	+4.0
			Panel D: Annual Flow—Future Performance			
Zero-alpha ($\hat{\pi}_0$)	69.9 (4.6)	59.7 (4.4)	70.6 (3.6)	73.8 (4.3)	80.6 (2.9)	+10.7
Unskilled ($\hat{\pi}_A^-$)	27.0 (4.2)	37.6 (4.0)	26.8 (3.3)	25.7 (3.5)	17.0 (2.5)	−10.0
Skilled ($\hat{\pi}_A^+$)	3.1 (1.7)	2.7 (1.6)	2.6 (1.6)	0.5 (1.5)	2.4 (1.7)	−0.7
Avg. Flow (%/year)	−23.2	−7.1	3.0	24.0	205.3	+228.5
Avg. $\hat{\alpha}$ (%/year)	−0.9 (0.1)	−1.4 (0.1)	−1.0 (0.1)	−1.0 (0.1)	−0.7 (0.1)	+0.2

this conjecture, Panels A and B show that larger and older funds are populated with far more unskilled funds than smaller and newer funds.

Perhaps more directly, the BG model also implies that flows should disproportionately move to truly skilled funds, and that these funds should exhibit the largest reduction in future skills. Panel C shows, for each past-year flow quintile, the proportions of each fund type during the 5 years ending with the flow measurement year, while Panel D shows similar statistics for these quintiles during the following 5 years. Here, the results are strongly supportive of the BG model. Specifically, the highest flow quintile exhibits the highest proportion of skilled funds (18%) during the 5 years prior to the flow year, and the largest reduction in skilled funds during the 5 years subsequent to the flow year (from 18% to 2.4%). Conversely, funds in the lowest flow quintile exhibit high proportions of unskilled funds prior to the flow year, but appear to improve their skills during the following years (perhaps due to a change in strategy or portfolio manager in response to the outflows). However, consistent with prior research (e.g., Sirri and Tufano (1998)), it appears that investors should have withdrawn even more money from these funds, as they continue to exhibit poor skills (27% are unskilled, compared to 17% for high inflow funds). Although the BG model does not capture the behavior of these apparently irrational investors, our results are generally consistent with the predictions of their model.

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, because we only know in which segment of the tails they lie, and not their identities, such an approach is not feasible.

Nonetheless, 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 stand a greater chance of capturing the superior alphas of the truly skilled funds. For instance, Panel B of Table III shows that when the significance level γ is low ($\gamma = 0.05$), the proportion of skilled funds among all significant funds, $\hat{T}_\gamma^+/\hat{S}_\gamma^+$, is about 50%, which is much higher than the proportion of skilled funds in the entire universe, 2.4%.

In order to choose the significance level, γ, that determines the significant funds, S_γ^+, included in the portfolio, we explicitly account for the location of the skilled funds by using the False Discovery Rate in the right tail, FDR^+. The FDR_γ^+ is defined as the expected proportion of lucky funds included in the portfolio at the significance level γ:

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

The FDR^+ makes possible a simple portfolio formation rule.[20] When we set a low FDR^+ target, we allow only a small proportion of lucky funds (false discoveries) in the chosen portfolio. Specifically, we set a sufficiently low significance

[20] 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)) because the latter does not distinguish between bad and good luck. The traditional measure is $FDR_\gamma = E(F_\gamma/S_\gamma)$, where $F_\gamma = F_\gamma^+ + F_\gamma^-$, $S_\gamma = S_\gamma^+ + S_\gamma^-$.

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 opposing effects on a portfolio: It decreases the portfolio expected future performance because the proportion of lucky funds in the portfolio is higher, and it increases the portfolio diversification because more funds are selected—reducing the volatility of the out-of-sample performance. Accordingly, we examine five FDR^+ target levels, z^+, in our persistence test, namely, $z^+ = 10\%$, 30%, 50%, 70%, and 90%.[21]

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 5-year period. Using these p-values, we estimate the FDR^+_γ over a range of chosen significance levels ($\gamma = 0.01, 0.02, \ldots, 0.60$). Following Storey (2002), we implement the following straightforward estimator of the FDR^+_γ:

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

where $\hat{\pi}_0$ is the estimator of the proportion of zero-alpha funds described in Section I.A.3. For each FDR^+ target level z^+, we determine the significance level, $\gamma(z^+)$, that provides an $\widehat{FDR}^+_{\gamma(z^+)}$ as close as possible to this target. Then, only funds with p-values smaller than $\gamma(z^+)$ are included in an equally weighted portfolio. This portfolio is held for 1 year, after which the selection procedure is repeated. If a selected fund does not survive after a given month during the holding period, 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 5 years of returns have been observed), while the last is December 31, 2005.

In Panel A of Table V, we show the FDR level ($\widehat{FDR}^+_{\gamma(z^+)}$) of the five portfolios, as well as the proportion of funds in the population that they include ($\hat{S}^+_{\gamma(z^+)}$) during the 5-year formation period, averaged over the 27 formation periods (ending from 1979 to 2005), and their respective distributions. First, we observe (as expected) that the achieved FDR increases with the FDR target assigned to a portfolio. However, the average $\widehat{FDR}^+_{\gamma(z^+)}$ does not always match its target. For instance, $FDR10\%$ achieves an average of 41.5%, instead of the targeted 10%—during several formation periods, the proportion of skilled funds in the population is too low to achieve a 10% FDR target.[22] Of course, a higher FDR

[21] Besides its financial interpretation, the FDR also has a natural statistical meaning, as it is the extension of the Type I error (i.e., rejecting the null, H_0, when it is correct) from single to multiple hypothesis testing. In the single case, the Type I error is controlled by using the significance level γ (i.e., the size of the test). In the multiple case, we replace γ with the FDR, which is a *compound* Type I error measure. In both cases, we face a similar trade-off: In order to increase power, we have to increase γ or the FDR, respectively (see the survey of Romano, Shaikh, and Wolf (2008)).

[22] 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 $\widehat{FDR}^+_{\gamma(z^+)}$ distribution for the portfolio $FDR10\%$ in Panel A, we observe that in 6 years out of 27, the $\widehat{FDR}^+_{\gamma(z^+)}$ is higher than 70%.

Table V

Performance Persistence Based on the FDR

For each of the five *FDR* targets (z^+ =10%, 30%, 50%, 70%, and 90%), Panel A contains descriptive statistics on the *FDR* level ($\widehat{FDR}^+_{\gamma(z^+)}$) achieved by the chosen portfolio, respectively, as well as the proportion of funds in the population that it includes ($\hat{S}_{\gamma(z^+)}$). The panel shows the average values of $\widehat{FDR}^+_{\gamma(z^+)}$ and $\hat{S}^+_{\gamma(z^+)}$ over the 27 annual formation dates (from December 1979 to 2005), as well as their respective distributions. Panel B displays the performance of each portfolio over the period 1980 to 2006. We estimate the annual four-factor alpha ($\hat{\alpha}$) with its bootstrap p-value, its annual residual standard deviation ($\hat{\sigma}_\varepsilon$), its annual information ratio ($IR = \hat{\alpha}/\hat{\sigma}_\varepsilon$), its loadings on the market ($\hat{b}_m$), size ($\hat{b}_{smb}$), book-to-market ($\hat{b}_{hml}$), and momentum factors ($\hat{b}_{mom}$), 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.

	Achieved FDR ($F\hat{D}R^+_{\gamma(z^+)}$)					Included Proportion of Funds ($\hat{S}^+_{\gamma(z^+)}$)				
Target (z^+)	Mean	10–30	30–50	50–70	>70%	Mean	0–6	6–12	12–24	>24%
Panel A: Portfolio Statistics										
*FDR*10%	41.5%	14	6	1	6	3.0%	25	2	0	0
*FDR*30%	47.5%	8	12	1	6	8.2%	15	7	3	2
*FDR*50%	60.4%	0	14	7	6	20.9%	5	7	4	11
*FDR*70%	71.3%	0	4	12	11	29.7%	1	5	5	16
*FDR*90%	75.0%	0	4	9	14	33.7%	0	3	4	20

Panel B: Performance Analysis									
Target (z^+)	$\hat{\alpha}$(p-value)	$\hat{\sigma}_\varepsilon$	IR	$\hat{b}_m$	$\hat{b}_{smb}$	$\hat{b}_{hml}$	$\hat{b}_{mom}$	Mean	Std dev
*FDR*10%	1.45%(0.04)	4.0%	0.36	0.93	0.16	−0.04	−0.02	8.3%	15.4%
*FDR*30%	1.15%(0.05)	3.3%	0.35	0.94	0.17	−0.02	−0.03	8.1%	15.4%
*FDR*50%	0.95%(0.10)	2.9%	0.33	0.96	0.20	−0.06	−0.01	8.1%	16.1%
*FDR*70%	0.68%(0.15)	2.7%	0.25	0.97	0.19	−0.06	−0.01	7.9%	16.1%
*FDR*90%	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 Remaining in the Portfolio (%)				
Target (z^+)	After 1 Year	After 2 Years	After 3 Years	After 4 Years	After 5 Years
*FDR*10%	36.7	12.8	3.4	0.8	0.0
*FDR*30%	40.0	14.7	5.1	1.7	1.3
*FDR*50%	48.8	23.5	12.3	4.7	2.6
*FDR*70%	52.2	29.0	17.4	9.5	6.3
*FDR*90%	55.9	33.8	20.4	13.0	8.5

target means an increase in the proportion of funds included in a portfolio, as shown in the rightmost columns of Panel A because our selection rule becomes less restrictive.

In Panel B, we present the average out-of-sample performance (during 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, $\hat{\alpha}$, along with its bootstrapped p-value; annualized residual standard deviation, $\hat{\sigma}_{\varepsilon}$; information ratio, IR$= \hat{\alpha}/\hat{\sigma}_{\varepsilon}$; 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 *FDR*10% 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 ($\hat{\sigma}_{\varepsilon}$ 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 *FDR*90% portfolio.

Panel C examines portfolio turnover. We determine the proportion of funds that are still selected using a given false discovery rule 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 funds remain in portfolios *FDR*10% and 30%, while after 3 years, these percentages drop below 6%.

Finally, in Figure 5, we examine how the estimated alpha of the portfolio *FDR*10% evolves over time using expanding windows. The initial value, on December 31, 1989, is the yearly out-of-sample alpha measured over the period 1980 to 1989, while the final value, on December 31, 2006, is the yearly out-of-sample alpha measured over the entire 1980 to 2006 period (i.e., this is the estimated alpha shown in Panel B of Table V). Again, these are the entire history (back to 1980) of persistence results that would be observed by a researcher at the end of each year. The similarity with Figure 4 is striking. Although the alpha accruing to the *FDR*10% 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 managers. However, this change is not sufficient to prevent the performance of *FDR*10% from dropping substantially.

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), and 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 (1 to 3 years). The proportionate size of fractile portfolios (e.g., deciles) are held fixed, with no regard to the changing estimated proportion of

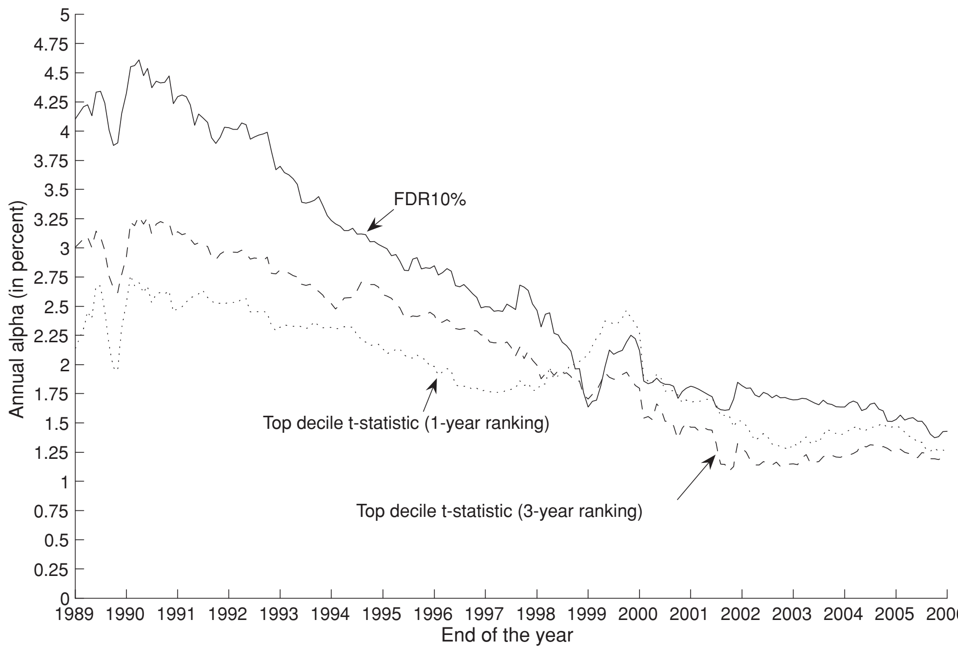

Figure 5. Performance of the portfolio FDR10% over time. The graph plots the evolution of the estimated annual four-factor alpha of the portfolio $FDR10\%$. To construct this portfolio, we estimate the (alpha) p-values of each existing fund at the end of each year using the previous 5-year period. After determining the significance level, $\gamma(z^+)$, such that the estimated FDR, $\widehat{FDR}^+_{\gamma(z^+)}$, is closest to 10%, we include all funds in the right tail of the cross-sectional t-statistic distribution with p-values lower than $\gamma(z^+)$ in an equally weighted portfolio. At the end of each year from 1989 to 2006, the portfolio alpha is estimated using the portfolio return history up to that point. The initial estimates cover the period 1980 to 1989 (the first 5 years are used for the initial portfolio formation on December 31, 1979), while the last ones use the entire portfolio history from 1980 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 1 and 3 years, respectively.

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 top decile portfolios that are formed based on ranking funds by their alpha t-statistic, estimated over the previous 1 and 3 years, respectively. 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.

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 stock-picking skills 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 recently reported annual expense ratio prior to that month) to net returns for each fund, and then revisit the long-term performance of the mutual fund industry.[23]

Panel A of Table VI contains the estimated proportions of zero-alpha, unskilled, and skilled funds in the population ($\hat{\pi}_0$, $\hat{\pi}_A^-$, and $\hat{\pi}_A^+$), 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 24.0% to 4.5%. This result indicates that only a small fraction of fund managers have stock-picking 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 slightly lower when performance is measured prior to expenses (1.3% versus 1.4% per year), indicating that high fees (potentially charged to unsophisticated investors) are one reason why funds end up in the extreme left tail, net of expenses. In addition, there is a negative relation between turnover and pre-expense performance, indicating that some unskilled managers trade too much relative to their abilities, although it is also possible that some skilled managers trade too little.

In the right tail, we find that 9.6% of fund managers have stock-picking skills sufficient to more than compensate for trading costs (Panel A). Because 75.4% of funds produce zero net-of-expense alphas, it seems surprising that we do not find more pre-expense skilled funds. However, this is due to the relatively small impact of expense ratios on the performance of funds located 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 it difficult to detect the presence of skill.[24]

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

[24] The average expense ratio across funds with $|\hat{\alpha}_i| < 1\%$ 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 $\sigma_\varepsilon = 0.021$). This implies that the null and alternative t-statistic distributions are extremely difficult to distinguish. To illustrate, for a hypothetical fund with a (pre-expense) t-statistic mean of 0.9, the probability of observing a negative (pre-expense) t-statistic equals 18%.

Table VI

Impact of Luck on Long-Term Pre-expense Performance

We add monthly expenses to the net return of each fund, and measure long-term performance with the unconditional four-factor model over the entire period 1975 to 2006. Panel A displays the estimated proportions of zero-alpha, unskilled, and skilled funds ($\hat{\pi}_0$, $\hat{\pi}_A^-$, and $\hat{\pi}_A^+$) in the entire fund population on a pre-expense basis (1,836 funds). Panel B counts the proportions of significant funds in the left and right tails of the cross-sectional t-statistic distribution ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) at four significance levels ($\gamma = 0.05, 0.10, 0.15, 0.20$). In the leftmost columns, the significant group in the left tail, $\hat{S}_\gamma^-$, is decomposed into unlucky and unskilled funds ($\hat{F}_\gamma^-$, $\hat{T}_\gamma^-$). In the rightmost columns, the significant group in the right tail, $\hat{S}_\gamma^+$, is decomposed into lucky and skilled funds ($\hat{F}_\gamma^+$, $\hat{T}_\gamma^+$). The bottom of Panel B also presents the characteristics of each significant group ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$): the average estimated alpha prior to expenses (% per year), expense ratio (in % per year), and turnover (% per year). Figures in parentheses denote the standard deviation of the different estimators.

Panel A: Proportion of Unskilled and Skilled Funds

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	85.9 (2.7)	4.5 (1.0)	9.6 (1.5)
Number	1,577	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. $\hat{S}_\gamma^-$ (%)	4.3	7.5	10.2	12.8	17.3	13.0	9.3	5.7	Signif. $\hat{S}_\gamma^+$ (%)
	(0.5)	(0.6)	(0.7)	(0.8)	(0.9)	(0.8)	(0.7)	(0.5)	
Unlucky $\hat{F}_\gamma^-$ (%)	2.1	4.3	6.4	8.6	8.6	6.4	4.3	2.1	Lucky $\hat{F}_\gamma^+$ (%)
	(0.0)	(0.1)	(0.1)	(0.2)	(0.2)	(0.1)	(0.1)	(0.0)	
Unskilled $\hat{T}_\gamma^-$ (%)	2.2	3.2	3.8	4.2	8.7	6.6	5.0	3.6	Skilled $\hat{T}_\gamma^+$ (%)
	(0.5)	(0.6)	(0.8)	(0.9)	(1.0)	(0.9)	(0.7)	(0.5)	
Pre-expense Alpha (%/year)	−5.9	−5.2	−4.8	−4.5	4.4	4.8	5.0	5.3	Pre-expense Alpha (%/year)
	(0.5)	(0.3)	(0.2)	(0.2)	(0.2)	(0.2)	(0.3)	(0.4)	
Exp. (%/year)	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.2	Exp. (%/year)
Turn. (%/year)	105	107	108	108	90	89	91	84	Turn. (%/year)

Finally, in results included in the Internet Appendix, we find that the proportion of pre-expense skilled funds in the population decreases from 27.5% at 1996 to 9.6% at 2006. This implies that the decline in net-expense skills noted in Figure 4 is driven mostly by a reduction in stockpicking skills over time (as opposed to an increase in expenses for pre-expense skilled funds). In contrast, the proportion of pre-expense unskilled funds remains equal to zero until the end of 2003. Thus, poor stock-picking skills alone (net of trading costs) cannot explain the large increase in the proportion of unskilled funds (net of both trading costs and expenses) from 1996 onwards. This increase is likely to be due to rising expenses charged by funds with weak stock selection abilities, or the introduction of new funds with high expense ratios and marginal stock-picking skills.

D.2. Performance Measured with Other Asset Pricing Models

Our estimation of the proportions of unskilled and skilled funds, $\hat{\pi}_A^-$ and $\hat{\pi}_A^+$, obviously depends on the choice of the asset pricing model. To examine the sensitivity of our results, we repeat the long-term (net of expense) performance analysis using the (unconditional) CAPM and Fama–French models. Based on the CAPM, we find that $\hat{\pi}_A^-$ and $\hat{\pi}_A^+$ are equal to 14.3% and 8.6%, respectively, which is much more supportive of active management skills, compared to Section III.A. 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).

Turning to the Fama–French (1993) model, we find that $\hat{\pi}_A^-$ and $\hat{\pi}_A^+$ amount to 25.0% and 1.7%, respectively. These proportions are very close to those obtained with the four-factor model because only one factor is omitted. As expected, the 1.1% difference in 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.[25] To see this, we denote by G_i a random variable that 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

[25] 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 $F_\gamma = F_\gamma^+ + F_\gamma^-$, $S_\gamma = S_\gamma^+ + S_\gamma^-$.

Table VII

Loadings on Omitted Factors

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

	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)
				Panel A: Unconditional CAPM					
Size ($\hat{b}_{smb}$)	0.06	0.07	0.09	0.09	0.27	0.28	0.28	0.36	Size ($\hat{b}_{smb}$)
Book ($\hat{b}_{hml}$)	−0.14	−0.14	−0.13	−0.14	0.34	0.35	0.36	0.37	Book ($\hat{b}_{hml}$)
Mom. ($\hat{b}_{mom}$)	0.00	0.00	0.00	0.01	−0.01	−0.01	−0.02	−0.01	Mom. ($\hat{b}_{mom}$)
				Panel B: Unconditional Fama–French model					
Mom. ($\hat{b}_{mom}$)	−0.02	−0.03	−0.02	−0.03	0.09	0.10	0.11	0.12	Mom. ($\hat{b}_{mom}$)

possible values (-1, 0, +1) are given by the proportion of each skill group 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 fund i has a zero alpha given that its t-statistic, denoted by T_i, is positive and significant: $fdr_\gamma^+ = \text{prob}(G_i = 0 \mid T_i \in \Gamma^+(\gamma))$, where $\Gamma^+(\gamma) = (t_\gamma^+, +\infty)$. Using Bayes's theorem, we have

$$fdr_\gamma^+ = \frac{\text{prob}(T_i \in \Gamma^+(\gamma) \mid G_i = 0) \cdot \text{prob}(G_i = 0)}{\text{prob}(T_i \in \Gamma^+(\gamma))} = \frac{\gamma/2 \cdot \pi_0}{E(S_\gamma^+)}. \tag{13}$$

Stated differently, the fdr_γ^+ indicates how the investor changes his prior probability that fund i has a zero alpha ($G_i = 0$) after observing that its t-statistic is significant. In light of equation (13), our estimator $\widehat{FDR}_\gamma^+ = (\gamma/2 \cdot \hat{\pi}_0)/\hat{S}_\gamma^+$ 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.[26]

In the recent Bayesian literature on mutual fund performance (e.g., Baks et al. (2001) and Pástor and Stambaugh (2002a)), attention is given to the posterior distribution of the fund alpha, α_i, as opposed to the posterior distribution of G_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 skill groups (zero-alpha, unskilled, and skilled). In light of our estimates, an empirically based alpha prior 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. Because $\hat{\pi}_A^-$ is higher than $\hat{\pi}_A^+$, 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 et al. (2001). Further, we estimate that 9.6% of the funds have an alpha greater than zero, before expenses. Although Baks et al. (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.

Finally, we can also interpret the mutual fund selection (Section III.C) from a Bayesian perspective. In her attempt to determine whether to include fund i ($i = 1, \ldots, M$) in her portfolio, the Bayesian investor is subject to two sorts of misclassification. First, she may wrongly include a zero-alpha fund in the portfolio (i.e., falsely rejecting H_0). Second, she may fail to include a skilled fund in the portfolio (i.e., falsely accepting H_0). Following Storey (2003), the investor's loss function, *BE*, can be written as a weighted average of each misclassification type:

[26] A full Bayesian estimation of fdr_γ^+ requires that one posits prior distributions for the proportions π_0, π_A^-, and π_A^+, and for the distribution parameters of T_i for each skill group. This method, based on additional assumptions (including independent p-values) as well as intensive numerical methods, is applied by Tang, Ghosal, and Roy (2007) to estimate the traditional *FDR* in a genomics study.

$$BE(\Gamma^+) = (1-\psi)\text{prob}(T_i \in \Gamma^+) \cdot fdr_\gamma^+(\Gamma^+) + \psi \cdot \text{prob}(T_i \notin \Gamma^+) \cdot fnr_\gamma^+(\Gamma^+), \quad (14)$$

where $fnr^+(\Gamma^+) = \text{prob}(G_i = +1 \mid T_i \notin \Gamma^+)$ is the "false nondiscovery rate" (i.e., the probability of failing to detect skilled funds), and ψ is a cost parameter that can be interpreted as the investor's regret after failing to detect skilled funds.[27] The decision problem consists of choosing the significance threshold, $t^+(\psi)$, such that $\Gamma^+(\psi) = (t^+(\psi), +\infty)$ minimizes equation (14) (equivalently, we could work with p-values and determine the optimal significance level, $\gamma(\psi)$). Contrary to the frequentist approach used in the paper, the Bayesian analysis requires an extensive parameterization, which includes, among other things, the exact specification of the null and alternative distributions of T_i, as well as the cost parameter ψ (see Efron et al. (2001) for an application in genomics).

If we decide to make this additional parameterization, we can determine the optimal Bayesian decision implied by the FDR^+ targets used in our persistence tests (z^+ =10%, 30%, 50%, 70%, and 90%). One way to do this is to consider our simple example shown in Figure 1, where the null and alternative distributions of T_i are assumed to be normal. We find that a high FDR^+ target z^+ (such as 90%) is consistent with the behavior of a Bayesian investor with a high cost of regret ($\psi(90\%)$ =0.997). Therefore, she chooses a very high significance level ($\gamma(90\%) = 0.477$), in order to include the vast majority of the skilled funds in the portfolio. In contrast, a low FDR^+ target z^+ (such as 10%) implies a lower regret ($\psi(10\%)$ =0.318), and a lower significance level ($\gamma(10\%) = 0.003$) (further details can be found in the Internet Appendix).

IV. Conclusion

In this paper, we apply a new method for measuring the skills of fund managers in a group setting. Specifically, the FDR approach provides a simple and straightforward method to estimate the proportion of skilled funds (those with a positive alpha, net of trading costs and expenses), zero-alpha funds, and unskilled funds (those with a negative alpha) in the entire population. Further, we use these estimates to provide accurate counts of skilled funds within various intervals in the right tail of the cross-sectional alpha distribution, as well as unskilled funds within segments of the left tail.

We apply the FDR technique to show that the proportion of skilled fund managers has diminished rapidly over the past 20 years, while the proportion of unskilled fund managers has increased substantially. Our paper also shows that the long-standing puzzle of actively managed mutual fund underperformance is due to the long-term survival of a 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

[27] See Bell (1982) and Loomes and Sugden (1982) for a presentation of Regret Theory, which includes in the investor's utility function the cost of regret about forgone investment alternatives.

minority of funds that produce negative alphas, when an increasing array of passively managed funds have become available (such as ETFs).

Although our paper focuses on mutual fund performance, our approach has potentially wide applications in finance. It can be used to control for luck in any setting in which a multiple hypothesis test is run and a large sample is available. This is the case, for instance, when we assess the performance of the myriad of trading rules used in technical trading (e.g., Sullivan, Timmermann, and White (1999), Bajgrowicz and Scaillet (2009)), or when we determine how many individual stocks have a commonality in liquidity (e.g., Chordia, Roll, and Subrahmanyam (2000)). With our approach, controlling for luck in multiple testing is trivial: The only input required is a vector of p-values, one for each individual test.

REFERENCES

Avramov, Doron, and Russ Wermers, 2006, Investing in mutual funds when returns are predictable, *Journal of Financial Economics* 81, 339–377.

Bajgrowicz, Pierre, and Olivier Scaillet, 2009, Technical trading revisited: False discoveries, persistence tests, and transaction costs, Working paper, University of Geneva.

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.

Bell, David E., 1982, Regret theory in decision making under uncertainty, *Operations Research* 30, 961–981.

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.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in liquidity, *Journal of Financial Economics* 56, 3–28.

Christoffersen, Susan E. K., and David K. Musto, 2002, Demand curves and the pricing of money management, *Review of Financial Studies* 15, 1495–1524.

Ding, Bill, and Russ Wermers, 2009, Mutual fund performance and governance structure: The role of portfolio managers and boards of directors, Working paper, University of Maryland.

Efron, Bradley, and Robert Tibshirani, 2002, Empirical Bayes methods and false discovery rates for microarrays, *Genetic Epidemiology* 23, 70–86.

Efron, Bradley, Robert Tibshirani, John D. Storey, and Virginia Tusher, 2001, Empirical Bayes analysis of a microarray experiment, *Journal of the American Statistical Association* 96, 1151–1160.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133–157.

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.

Elton, Edwin J., Martin J. Gruber, and Jeffrey Busse, 2004, Are investors rational? Choices among index funds, *Journal of Finance* 59, 261–288.

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.

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.

Ferson, Wayne E., and Meijun Qian, 2004, *Conditional Performance Evaluation, Revisited*, in (The Research Foundation of CFA Institute, Charlottesville, Virginia).

Ferson, Wayne E., and Rudi W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–461.

Genovese, Christopher, and Larry Wasserman, 2004, A stochastic process approach to false discovery control, *Annals of Statistics* 32, 1035–1061.

Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* 62, 393–416.

Hall, Peter, Joel L. Horowitz, and Bing-Yi Jing, 1995, On blocking rules for the bootstrap with dependent data, *Biometrika* 82, 561–574.

Hamilton, James D., 1994, *Times-Series Analysis* (Princeton University Press, Princeton).

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.

Jensen, Michael C., 1968, The performance of mutual funds in the period 1945–1964, *Journal of Finance* 23, 389–416.

Jones, Christopher S., and Jay Shanken, 2005, Mutual fund performance with learning across funds, *Journal of Financial Economics* 78, 507–552.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Halbert White, 2006, Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis, *Journal of Finance* 61, 2551–2595.

Kostovetsky, Leonard, 2007, Brain drain: Are mutual funds losing their best minds? Working paper, Princeton University.

Loomes, Graham, and Robert Sugden, 1982, Regret theory: An alternative theory of rational choice under uncertainty, *Economic Journal* 92, 805–824.

Lynch, Anthony W., and David Musto, 2003, How investors interpret past fund returns, *Journal of Finance* 58, 2033–2058.

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

Pástor, Ľuboš, and Robert F. Stambaugh, 2002a, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63, 315–349.

Pástor, Ľuboš, and Robert F. Stambaugh, 2002b, Investing in equity mutual funds, *Journal of Financial Economics* 63, 351–380.

Rea, John D., and Brian K. Reid, 1998, Trends in the ownership cost of equity mutual funds, *Investment Company Institute Perspective*, November.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf, 2008, Formalized data snooping on generalized error rates, *Econometric Theory* 24, 404–447.

Sirri, E., and P. Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Storey, John D., 2002, A direct approach to false discovery rates, *Journal of the Royal Statistical Society* 64, 479–498.

Storey, John D., 2003, The positive false discovery rate: A Bayesian interpretation and the q-value, *Annals of Statistics* 31, 2013–2035.

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.

Sullivan, Ryan, Allan Timmermann, and Halbert White, 1999, Data-snooping, technical trading rule performance and the bootstrap, *Journal of Finance* 54, 1647–1691.

Tang, Yongqiang, Subhashis Ghosal, and Anindya Roy, 2007, Nonparametric Bayesian estimation of positive false discovery rates, *Biometrics* 63, 1126–1134.

Wermers, Russ, 1999, Mutual fund herding and the impact on stock prices, *Journal of Finance* 55, 581–622.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses, *Journal of Finance* 55, 1655–1695.