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

Laurent Barras†, Olivier Scaillet‡ and Russ Wermers§

First version, September 2005
This version, March 2008

JEL Classification: G11, G23, C12
Keywords: Mutual Fund Performance, Multiple-Hypothesis Test, Luck, False Discovery Rate

*We are grateful to S. Brown, B. Dumas, A. Metrick, L. Pedersen, E. Ronchetti, R. Stulz, M.-P. Victoria-Feser, M. Wolf, as well as seminar participants at Banque Cantonale de Genève, BNP Paribas, Bilgi University, CREST, Grequam, INSEAD, London School of Economics, Maastricht University, MIT, Princeton University, Queen Mary, Solvay Business School, Stern School, Universita della Svizzeria Italiana, University of Geneva, University of Notre-Dame, University of Pennsylvania, the Imperial College Risk Management Workshop (2005), the Swiss Doctoral Workshop (2005), the Research and Knowledge Transfer Conference (2006), the Zeuthen Financial Econometrics Workshop (2006), the annual meetings of EC² (2005), ESEM (2006), EURO XXI (2006), ICA (2006), AFFI (2006), and SGF (2006), WHU Campus for Finance (2007) for their helpful comments. 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).

†Swiss Finance Institute at HEC-University of Geneva, Boulevard du Pont d’Arve 40, 1211 Geneva 4, Switzerland. Tel: +41223798141. E-mail: barras@hec.unige.ch

‡Swiss Finance Institute at HEC-University of Geneva, Boulevard du Pont d’Arve 40, 1211 Geneva 4, Switzerland. Tel: +41223798816. E-mail: scaillet@hec.unige.ch

§University of Maryland - Robert H. Smith School of Business, Department of Finance, College Park, MD 20742-1815, Tel: +13014500572. E-mail: rwermers@rhsmith.umd.edu
ABSTRACT

Prior approaches to identifying skilled funds in a population examine the performance of each fund in isolation, without regard to the role of luck in this multiple fund setting. Our paper develops a new, simple technique to properly account for “false discoveries,” or funds which exhibit significant alphas by luck alone. As such, our approach precisely identifies the proportion of funds with truly positive or negative performance in any segment of the cross-sectional alpha distribution, even with cross-fund dependencies in estimated alphas. We find that 26.6% of U.S. domestic-equity funds exhibit truly negative four-factor alphas (net of expenses and trading costs), while only 0.6% exhibit truly positive alphas over the 1975 to 2006 period. We find much higher proportions of skilled fund managers when we examine Aggressive-Growth funds, or when we examine funds prior to 1990. We also find a much higher proportion of skilled fund managers (9.6%) when we examine alphas before expenses (but after trading costs). Our findings carry some important implications. First, the large growth in the number of actively managed funds has resulted in a much lower proportion of truly skilled funds, and, second, the large number of actively managed funds has not resulted in a competitive level of fund fees and expenses.
Investors and academic researchers have long searched for outperforming mutual fund managers. Although several researchers document negative average fund alphas, net of expenses and trading costs (e.g., Jensen (1968), Lehman and Modest (1987), Elton et al. (1993), and Carhart (1997)), recent papers show that some fund managers have stock-selection skills. For instance, Kosowski, Timmermann, Wermers, and White (KTWW (2006)) use a bootstrap technique to document outperformance by some funds, while Baks, Metrick, and Wachter (2001), Pastor and Stambaugh (2002b), and Avramov and Wermers (2007) illustrate the benefits of investing in actively-managed funds from a Bayesian perspective. While these papers are useful in uncovering whether, on the margin, outperforming mutual funds exist, they are not particularly informative regarding their prevalence in the entire fund population. For instance, it is natural to wonder how many fund managers possess true stockpicking skills (according to a chosen asset pricing model), and where these funds are located in the cross-sectional estimated alpha distribution. From an investment perspective, precisely locating skilled funds maximizes our chances of achieving persistent outperformance.¹

Of course, we cannot observe the true alphas of each fund in the population. Therefore, a seemingly reasonable way to estimate the prevalence of skilled fund managers is to simply count the number of funds with sufficiently high estimated alphas, $\tilde{\alpha}$. In implementing such a procedure, we are actually conducting a multiple (hypothesis) test, since we simultaneously examine the performance of several funds in the population (instead of just one fund).² However, a simple count of significant alphas does not properly adjust for luck in such a multiple test setting—many of the funds have significant estimated alphas purely by luck (i.e., their true alphas are zero). To illustrate, consider a population of funds with no skills. With the usual chosen significance level of 5%, we know that, on average, 5% of the funds will have significant estimated alphas—some of them will be unlucky ($\tilde{\alpha} < 0$) while others are lucky ($\tilde{\alpha} > 0$), but all will be “false discoveries”—funds with significant estimated alphas, but zero true alphas.

This paper provides a new approach to controlling for false discoveries in such a multiple fund setting. Our approach much more accurately identifies (1) the proportions of unskilled and skilled funds in the population (those with truly negative and positive alphas, respectively), and (2) their respective locations in the left and right tails of the

¹From an investor’s perspective, “skill” is manager talent in generating abnormal performance (through stock selection efforts) that is sufficient to generate a positive alpha, net of trading costs and fund expenses.

²This multiple test should not be confused with the joint hypothesis test that all fund alphas are equal to zero in a sample. This test, which is employed by several past research papers (e.g., Grinblatt and Titman (1989, 1993)), addresses only whether at least one fund has a non-zero alpha among several funds, but is silent on the prevalence of these non-zero alpha funds.
cross-sectional *estimated* alpha (or estimated *t*-statistic) distribution. A main virtue of our approach is its simplicity—to determine the proportions of unlucky and lucky funds, the only parameter needed is the proportion of zero-alpha funds in the population, \( \pi_0 \). Rather than arbitrarily imposing a prior assumption on \( \pi_0 \), we estimate it with a straightforward manipulation of the *p*-values of the individual fund estimated alphas—no further econometric tests are necessary. A second advantage of our approach is its accuracy. Using a simple Monte-Carlo experiment, we confirm that our approach provides much more accurate estimates of the proportion of unskilled and skilled funds in the population than previous approaches that impose an a priori assumption about the proportion of zero-alpha funds (under the null hypothesis).\(^3\)

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

We apply our new approach to the monthly returns of 2,076 U.S. open-end, domestic equity mutual funds, and revisit several important themes examined in the previous literature. We start with an examination of the long-term performance (net of trading costs and expenses) of the mutual fund industry during 1975 to 2006. Our decomposition of the population reveals that 72.8% are zero-alpha funds—funds for which managers do have some stockpicking abilities, but that extract all of the rents of these abilities. Among remaining funds, only 0.6% are skilled (true \( \alpha > 0 \)), while 26.6% are unskilled (true \( \alpha < 0 \)). While our finding of a majority of zero-alpha funds—those that capture the entire surplus created by their managers (through fees)—is supportive of the long-run equilibrium of Berk and Green (2004), it is surprising that we find so many truly negative-alpha funds—those that overcharge relative to the skills of their managers. Indeed, we find that such unskilled funds are not particularly young, indicating that investors have had some time to evaluate and identify them as underperformers. Apparently, unskilled funds continue to exist by attracting unsophisticated or inatten-

\(^3\)The reader should note the difference between our approach and that of KTWW (2006). Our approach simultaneously estimates the prevalence and location of outperforming funds in a group, while KTWW test for the skills of a single fund that is chosen from the universe of funds. As such, our approach examines fund performance from a different perspective, with a richer set of information about active fund manager skills.
itive investors (Capon, Fitzsimmons, and Prince (1996) and Elton, Gruber, and Busse (2003)).

We also find some notable time trends in our study. Examining the evolution of each skill group between 1990 and 2006, we observe that the proportion of skilled funds decreases substantially from 14.4% to 0.6%, while the proportion of unskilled funds increases sharply from 9.2% to 26.6%. These findings are consistent with an exodus of skilled fund managers to hedge funds over this period, as well as a secular increase in the incidence of “closet indexing,” by our active fund managers, as documented by Cremers and Petajisto (2007). Thus, although the number of actively managed funds has substantially increased, skilled managers (those capable of picking stocks well enough to overcome their trading costs and expenses) have become increasingly rare.

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

Our analysis shows that the proportion of short-term skilled funds in the population is low (2.4%), but that they are concentrated in the extreme right tail of the estimated alpha distribution. We exploit this fact by implementing a persistence test that forms a portfolio of funds located at the extreme right tail. The exact portion of the tail that is included in this portfolio is determined by the frequency of “false discoveries”–during years when our tests indicate higher numbers of lucky, zero-alpha funds in the right tail, we move further to the extreme tail. Forming such a false-discovery controlled portfolio at the beginning of each year from January 1980 to 2006, we find an equal-weighted four-factor alpha of 1.45% per year, which is statistically significant. This strategy outperforms prior persistence strategies used by Carhart (1997) and others, where constant top-decile portfolios of funds are chosen. However, the performance advantage of our strategy decreases over time, as truly skilled managers become more scarce.

Our final tests examine the performance of fund managers before expenses (but after trading costs) are subtracted. Specifically, while fund managers may be able to pick stocks well enough to cover their trading costs, they do not usually exert direct control
over the level of fund expenses and fees—management companies set these expenses, with the approval of fund directors. We find evidence that indicates a very large impact of fund fees—on a pre-expense basis, we find a large increase in skilled funds that deliver positive alphas—9.6%, compared to our prior finding of 0.6% after expenses. However, this proportion of skilled managers (before fees) declines substantially over time, again indicating that portfolio managers with skills are increasingly rare.

We also observe a large reduction in the proportion of unskilled funds that deliver negative alphas (from 26.6% to 4.5%), indicating that funds produce negative net alphas because they charge excessive fees, relative to the stockpicking skills of their managers. As such, our study highlights that the expense ratios of a large minority of domestic equity mutual funds do not seem to be competed away.

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

I The Impact of Luck on Mutual Fund Performance

A Overview of the Setting

A.1 Luck in a Multiple Performance Setting

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

- **Unskilled funds**: funds having managers with stockpicking skills insufficient to recover their trading costs and expenses, such that there is an “alpha shortfall” \((\alpha < 0)\)
- **Zero-alpha funds**: funds having managers with stockpicking skills sufficient to just recover trading costs and expenses \((\alpha = 0)\)
- **Skilled funds**: funds having managers with stockpicking skills sufficient to provide an “alpha surplus,” beyond simply recovering their trading costs and expenses
(\(\alpha > 0\)).

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

Of course, we cannot observe the true alphas of each fund in the population. Therefore, how do we best infer the prevalence of each of the above skill groups from simple regression output—the estimated alphas, \(\hat{\alpha}_i\), of the individual funds \((i = 1, \ldots, M)\)? First, we use the \(t\)-statistic, \(t_i = \hat{\alpha}_i / \hat{\sigma}_{\hat{\alpha}_i}\), where \(\hat{\sigma}_{\hat{\alpha}_i}\) is the estimated standard deviation of \(\hat{\alpha}_i\). KTTW (2006) show that performance measured with the \(t\)-statistic has superior properties over performance measured with a fund’s alpha—since alphas are estimated with differential precision across funds with differing lives and portfolio volatilities. Second, after setting an appropriate significance level, \(\gamma\) (e.g., 10%), we execute, for each fund \(i\), a two-sided test of the null hypothesis that its true alpha equals zero: \(H_{0,i} : \alpha_i = 0\). If \(t_i\) (in absolute value) lies above the thresholds implied by \(\gamma\) (denoted by \(t^-_\gamma\) and \(t^+_\gamma\)), fund \(i\) is labeled “significant,” as there is strong statistical evidence that it is an outlier (either good or bad). It is important to note that this procedure corresponds to a multiple-hypothesis test, since we simultaneously test the performance of all funds in the population:

\[
\begin{align*}
H_{0,1} & : \alpha_1 = 0, & H_{A,1} : \alpha_1 \neq 0, \\
\vdots & : \vdots \\
H_{0,M} & : \alpha_M = 0, & H_{A,M} : \alpha_M \neq 0.
\end{align*}
\]

(1)

To illustrate the setup and likely outcome of this multiple test, we show, in Figure 1, a hypothetical distribution of fund \(t\)-statistics that is based on our empirical findings to be presented later.\(^5\) Specifically, the mean \(t\) is fixed at -2.5 for unskilled and 3.0 for skilled funds, which corresponds to an annual four-factor alpha (see, for example, Carhart (1997) of -3.2\% and 3.8\%, respectively (the relation of these values to our sample results are explained in the appendix). Then, the \(t\) distribution shown in Panel B is the

\(^4\)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, such as record-keeping) costs that the manager usually has little control over. In a later section of this paper, we redefine stockpicking skill in an absolute sense (net of trading costs only) and revisit some of our basic tests to be described.

\(^5\)For simplicity, this figure plots the (hypothetical) distributions of the fund \(t\)-statistics in Panel A as being normal, even though this distribution is non-normal for U.S. domestic equity funds (KTTW (2006)). In our empirical section to follow, we use a bootstrap approach to determine the distribution of \(t\)-statistics for each fund (and its associated p-value), as well as their cross-sectional distribution.
cross-section that (hypothetically) would be observed by a researcher—it is a mixture of the three distributions in Panel A, where the weight on each distribution is equal to the proportion of zero-alpha, unskilled, and skilled funds in our sample (specifically, \( \pi_0 = 75\% \), \( \pi_A^- = 23\% \), and \( \pi_A^+ = 2\% \), respectively).

Please insert Figure 1 here

To illustrate further, suppose that we choose a significance level, \( \gamma \) of 10\% \( (t^-_\gamma = -1.65, t^+_\gamma = 1.65) \). By conducting the test shown in Equation (1), the researcher expects to find 5.4\% of funds with a positive and significant \( t \)-statistic.\(^6\) This proportion, denoted by \( E(S^+_\gamma) \), is represented by the dark dashed area 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 darker shaded area in Panel A, zero-alpha funds are the recipient of good luck when their observed \( \hat{t} \)'s are positive and significant. By the same token, the proportion of funds with a negative and significant \( t \)-statistic (the lighter shaded region in Panel B), \( E(S^-_\gamma) \), overestimates the proportion of unskilled funds, because it includes some unlucky zero-alpha funds (the grey area in Panel A).\(^7\)

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

### A.2 Measuring Luck

We next describe our approach to estimating the proportion of funds that lie in the tails of the cross-sectional \( t \) distribution purely by (good or bad) luck. At a given significance level \( \gamma \), 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 \( \pi_0 \), the expected proportion of “lucky funds”

\(^6\)From Panel A, the probability that the observed \( t \)-statistic is greater than \( t^+_\gamma = 1.65 \) equals 5\% for a zero-alpha fund and 84\% for a skilled fund. Multiplying these two probabilities by the proportions of these two categories, \( \pi_0 \) and \( \pi_A^+ \), gives 5.4\%.

\(^7\)Note that we have not considered the possibility that skilled funds could be very unlucky, and exhibit a negative and significant \( t \)-stat. In our hypothetical distribution, the probability that the \( t \)-statistic of a skilled fund being lower than \( t^-_\gamma = -1.65 \) is less than 0.001\%. This probability is negligible. The same comment applies to unskilled funds who are very lucky.
(zero-alpha funds with positive and significant $t$-statistics) is equal to

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

Now, to determine the expected proportion of skilled funds, $E(T^+_{\gamma})$, we simply adjust $E(S^+_{\gamma})$ for the presence of these lucky funds:

$$E(T^+_{\gamma}) = E(S^+_{\gamma}) - E(F^+_{\gamma}) = E(S^+_{\gamma}) - \pi_0 \cdot \gamma/2. \quad (3)$$

The expected proportion of unlucky funds in the left tail, denoted by $E(F^-_{\gamma})$, is obtained using a similar approach. Since the probability of being unlucky is also equal to $\gamma/2$ (i.e., the grey and black areas in Panel A of Figure 1 are identical), $E(F^-_{\gamma})$ is equal to $E(F^+_{\gamma})$. As a result, the expected proportion of unskilled funds, $E(T^-_{\gamma})$, is similarly given by

$$E(T^-_{\gamma}) = E(S^-_{\gamma}) - E(F^-_{\gamma}) = E(S^-_{\gamma}) - \pi_0 \cdot \gamma/2. \quad (4)$$

What is the role played by the significance level, $\gamma$, chosen by the researcher? By defining the significance thresholds $t^-_{\gamma}$ and $t^+_{\gamma}$, $\gamma$ determines the portion of the left and right tails which are examined for lucky vs. skilled funds, as described by Equations (3) and (4). By varying $\gamma$, we can measure the proportion of skilled funds in any segment of the cross-section of funds (ranked by their $t$-statistics).

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

### A.3 Estimation Procedure

The key to our approach to measuring luck in a group setting, as shown in Equation (2), is the estimator of the proportion, $\pi_0$, of zero-alpha funds in the population. Here, we turn to a novel estimation approach developed by Storey (2002)—called the “False
Discovery Rate” (FDR) approach. The FDR approach is very straightforward, as its sole inputs are the \( p \)-values (two-sided) associated with the (alpha) \( t \)-statistics for each of the \( M \) funds. By definition, zero-alpha funds satisfy the null hypothesis, \( H_{0,i} : \alpha_i = 0 \), and, therefore, have \( p \)-values that are uniformly distributed over the interval \([0, 1]\).\(^8\) On the contrary, \( p \)-values of unskilled and skilled funds tend to be very small because their \( t \)'s tend to be far from zero (see Panel A of Figure 1). We can exploit this information to estimate \( \pi_0 \) without knowing the exact distribution of the \( p \)-values of the unskilled and skilled funds.

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

Please insert Figure 2 here

The dark grey zone near zero represents the proportion of \( p \)-values in the population that correspond to the unskilled and skilled funds (\( \pi_{-A} + \pi_{+A} = 25\% \)). The area of the rectangle between the horizontal black line and the abscissa is equal to the proportion \( \pi_0 \) that we want to estimate (75%), since the zero-alpha funds have uniformly distributed \( p \)-values. To approximate this area using the histogram of observed \( p \)-values, we proceed as follows. If we take a sufficiently high threshold \( \lambda^* \) (e.g., \( \lambda^* = 0.6 \)), we know that the vast majority of \( p \)-values higher than \( \lambda^* \) come from zero-alpha funds. Thus, we first estimate the area covered by the four light grey bars on the right of \( \lambda^* \), \( \hat{W}(\lambda^*)/M \) (where \( \hat{W}(\lambda^*) \) denotes the number of funds having \( p \)-values exceeding \( \lambda^* \)), then simply extrapolate this area over the entire interval \([0, 1]\) by multiplying it by \( 1/(1 - \lambda^*) \) (if

\[^8\]To see this, let us denote by \( t \) and \( p \) the \( t \)-statistic and \( p \)-value of the zero-alpha fund. We have \( p = 1 - F(|t|) \), where \( F(t) = \text{prob}(|\hat{t}_{i}| < |t| | \alpha_i = 0) \). The \( p \)-value is uniformly distributed over \([0, 1]\) since its cdf, \( G(p) = \text{prob}(\hat{p}_i < p) = \text{prob}(1 - F(|\hat{t}_{i}|) < p) = \text{prob}(|t_{\hat{p}_i}| > F^{-1}(1 - p)) = 1 - F(F^{-1}(1 - p)) = p \).
\( \lambda^* = 0.6 \), the area is multiplied by 2.5:

\[
\hat{\pi}_0(\lambda^*) = \frac{\hat{W}(\lambda^*)}{M} \cdot \frac{1}{1 - \lambda^*}.
\]  

(5)

To select \( \lambda^* \), we use a data-driven approach suggested by Storey (2002) and explained in detail in the appendix.

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

\[
\hat{F}_\gamma^- = \hat{F}_\gamma^+ = \hat{\pi}_0 \cdot \gamma / 2.
\]  

(6)

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

\[
\hat{T}_\gamma^- = \hat{S}_\gamma^- - \hat{F}_\gamma^- = \hat{S}_\gamma^- - \hat{\pi}_0 \cdot \gamma / 2,
\]

\[
\hat{T}_\gamma^+ = \hat{S}_\gamma^+ - \hat{F}_\gamma^+ = \hat{S}_\gamma^+ - \hat{\pi}_0 \cdot \gamma / 2.
\]  

(7)

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

\[
\hat{\pi}_A^- = \hat{T}_\gamma^- \gamma^*, \quad \hat{\pi}_A^+ = \hat{T}_\gamma^+ \gamma^*,
\]  

(8)

where \( \gamma^* \) is a sufficiently high significance level—we choose \( \gamma^* \) with a simple data-driven method explained in the appendix.10

**B Comparison of Our Approach with Existing Methods**

The previous literature has followed two alternative approaches when estimating the proportion of unskilled and skilled funds. The “full luck” approach proposed by Jensen (1968) and Ferson and Qian (2004) assume that all funds in the population have zero

---

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

10 We assume that the density of the \( t \)-statistic of the unskilled and skilled funds decreases monotonically as \( \gamma \) rises. This feature is shared by most test statistics when the sample size grows to infinity. Standard test statistics are asymptotically distributed as a normal (chi-square) variable under the null and as a non-central normal (chi-square) variable under the alternative.
alphas: \( \pi_0 = 1 \). Thus, for a given significance level \( \gamma \), this approach implies an estimate of the proportions of unlucky and lucky funds equal to \( \gamma/2 \).\(^{11}\) Instead of making a luck adjustment that can be excessive, the “no luck” approach reports the observed number of significant funds (for instance, Ferson and Schadt (1996) show how this number changes when performance is measured with unconditional vs. conditional models). This approach correctly estimates the proportions of unskilled and skilled funds if there are no zero-alpha funds in the population: \( \pi_0 = 0 \).

What are the errors committed by setting \( \pi_0 = 0 \) or \( 1 \) when it does not hold in the population? To address this question, we compare the bias produced by these two approaches to ours across the possible values for \( \pi_0 \) \((\pi_0 \in [0, 1])\) using our simple framework of Figure 1. Our procedure consists of four steps. First, for a chosen value for \( \pi_0 \), we create a simulated sample of 2,076 fund \( t \)-statistics (corresponding to our fund sample size) using the distributions in Panel A of Figure 1 according to the proportions \( \pi_0 \), \( \pi^-_A \), and \( \pi^+_A \). For each \( \pi_0 \), the values for \( \pi^-_A \) and \( \pi^+_A \) are determined such that their ratio remains unchanged \((\pi^-_A/\pi^+_A = 0.23/0.02)\). Second, we use these sampled \( t \)-statistics to estimate the proportion of unlucky \((\alpha = 0, \alpha < 0)\), lucky \((\alpha = 0, \alpha > 0)\), unskilled \((\alpha < 0, \alpha < 0)\), and skilled \((\alpha > 0, \alpha > 0)\) funds using each of the three approaches ("no luck", "full luck", and our FDR approach).\(^{12}\) Third, we repeat these two steps 1,000 times in order to compare the average value of each estimator with its true population value. The results are displayed in Figure 3. In each Panel, the true population value is given by the solid line, the “no luck” estimator by the dotted line, the “full luck” estimator by the dashed line, while our estimator is marked with a solid line with dots.

Please insert Figure 3 here

Panel A compares the estimators of the expected proportion of unlucky funds. The true population value, \( E(F^-) \), is an increasing function of \( \pi_0 \), as indicated in Equation (2). While the average value of our estimator precisely tracks \( E(F^-) \), this is not the case for the two other approaches. Note that, by assuming that \( \pi_0 = 0 \), the “no luck” approach consistently underestimates \( E(F^-) \) for \( \pi_0 > 0 \). Conversely, the “full luck” approach, which assumes that \( \pi_0 = 1 \), overestimates \( E(F^-) \) for \( \pi_0 < 1 \). To illustrate the extent of the bias, consider the case where \( \pi_0 = 75\% \). While the “no luck” approach underestimates \( E(F^-) \) by 100\% (0\% instead of 7\%), the “full luck” approach overestimates

\(^{11}\)Jensen (1968) summarizes the “full luck” approach as follows: “...if all the funds had a true \( \alpha \) equal to zero, we would expect (merely because of random chance) to find 5\% of them yielding \( t \) values ‘significant’ at the 5\% level.” KTWW (2006) also use this approach in one of their graphs (Figure 3)).

\(^{12}\)We set \( \gamma = 0.20 \) so as to examine a large portion of the tails of the cross-sectional \( t \)-statistic distribution (other values for \( \gamma \) provide the same insights).
The analysis for the lucky funds $E(F^+)$ shown in Panel B is exactly the same since $E(F^+ - \gamma) = E(F^- - \gamma)$.

The estimators of the expected proportions of unskilled and skilled funds ($E(T^+ - \gamma)$ and $E(T^+)$) are shown in Panel C and D, respectively. As the proportion of zero-alpha funds rises, there are fewer unskilled and skilled funds in the population ($1 - \pi_0 = \pi_A - \pi_A^+$). For this reason, we observe that the true population values, $E(T^+ - \gamma)$ and $E(T^+)$, are negatively related to $\pi_0$. In both Panels, our FDR estimator is able to capture these negative relations. For the other approaches, the bad measurement of luck leads to a poor assessment of the prevalence of both unskilled and skilled funds. For instance when $\pi_0 = 75\%$, using the “no luck” approach leads to an upward bias of the total proportion of unskilled and skilled funds equal to 52% (a total of 204 funds when $M = 2,076$). At the other extreme, the “full luck” underestimates this number by 22% (86 funds). In addition, Panel D reveals that these two approaches produce an inconsistent relation between $\pi_0$ and $E(T^+)$.

In addition to the bias properties of our FDR estimators, their variability is also low because of the large cross-section of funds ($M = 2,076$). To understand this, consider our main estimator $\pi_0$ (the same arguments apply to the other estimators). Since $\pi_0$ is a proportion estimator, it can be written in the form of a simple average (Davidson and MacKinnon (2004), p. 147): $\pi_0 = 1/M \sum_{i=1}^{M} x_i$, where $x_i$ is the indicator function taking the value $(1 - \lambda^*)^{-1}$ if $\hat{p}_i > \lambda^*$, and 0 otherwise. With independent p-values, its standard deviation, $\sigma_{\pi_0}$, is simply equal to $\sigma_x/\sqrt{M}$, implying that large $M$ drives $\sigma_{\pi_0}$ to zero.\footnote{The precision of $\pi_0$ as a sample average contrasts with the inaccuracy of equity return sample average (e.g., Jorion and Goetzman (1999)). The reason is that our estimator builds its strength on cross-sectional instead of time-series observations.}

For instance, with $\lambda^* = 0.6$ and $\pi_0 = 75\%$, we find that $\sigma_{\pi_0}$ is as low as 2.5% (30 times lower than $\pi_0$).\footnote{We have $P_{\lambda^*} = \text{prob}(\hat{p}_i > \lambda^*) = 0.30$ (i.e., the rectangle area delimited by the horizontal black line and the vertical line at $\lambda^* = 0.6$ in Figure 2). Since $(1 - \lambda^*) x_i$ follows a binomial distribution with probability $P_{\lambda^*}$ of success, we have $\sigma_x = (1 - \lambda^*)^{-1} (P_{\lambda^*} (1 - P_{\lambda^*}))^{1/2} = 1.14$, and $\sigma_{\pi_0} = \sigma_x/\sqrt{M} = 2.5$.}

\section{C Estimation under Cross-Sectional Dependence among Funds}

Funds can have correlated residuals if they hold concentrated portfolios with similar security or industry bets, or load on similar non-priced factor (e.g., Wermers (1999)).
In general, cross-sectional dependence cannot be ignored, and greatly complicates performance measurement. Any inference test in the presence of potential dependencies becomes quickly intractable as $M$ rises, since it 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 dependencies are present in the prior beliefs about fund alphas. Further, KTWW (2006) show that a complicated bootstrap is necessary to test the significance of performance of a fund located at a particular alpha rank, since this test depends on the joint distribution of all fund estimated alphas—and, cross-correlated fund residuals must be bootstrapped simultaneously.

An important advantage of our approach is that we can 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 a simple example, suppose that $\pi_0 = 100\%$, and that the correlation across all $p$-values is equal to one (perfect herding). In this case, all $p$-values take the same value, and there no uniform distribution as in Figure 2: every time the value happens to be lower than $\lambda^*$, we wrongly estimate $\pi_0 = 0$.

We are not overly concerned with a large level of dependencies in our sample, since the average (monthly) pairwise residual correlation equals only 0.08, based on the standard four-factor model (with market, size, book-to-market, and momentum factors). Further, many of our funds do not have highly overlapping return data, and, therefore, cannot have highly correlated residuals by construction. Specifically, we find that 15% of the fund pairs do not have a single monthly return observations in common; on average, only 55% of the return observations of fund pairs is overlapping. As a result, it is likely that the cross-sectional dependence is sufficiently low to produce consistent estimators (i.e., mutual fund residuals satisfy the ergodicity conditions discussed in Storey, Talyor, and Siegmund (2004)).

In order to explicitly verify the properties of our estimators, we also run a Monte-Carlo experiment with cross-sectional correlation. In our baseline model, we consider a universe of 1,500 funds, in which about 900 of them are all correlated with one another. In order to closely reproduce the complex relations across these funds, we estimate the residual covariance matrix directly from the data. We also consider other dependence cases, such residual block correlation and residual factor dependence, as in Jones and Shanken (2005). We find that not only the average values of our estimators are equal

15 The reason is that the hypothesis test for each fund $i$, $H_{0,i} : \alpha_i = 0$, depends only on the distribution of the fund $i$ estimated alpha (and not on the joint distribution of estimated alphas)
to their respective true values, but also that their confidence intervals are comparable to those computed under independence. In order to save space, these results, as well as with further details on the simulation experiment, are shown in the appendix.

II Performance Measurement and Data Description

A Asset Pricing Models

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

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

(9)

where $r_{i,t}$ is the month $t$ excess return of fund $i$ over the riskfree rate (proxied by the monthly T-bill rate); $r_{m,t}$ is the month $t$ excess return on the value-weighted market portfolio; $r_{smb,t}$, $r_{hml,t}$, 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; and $\varepsilon_{i,t}$ stands for the residual term. Adding momentum to the three-factor Fama-French model (1996) allows us to control for the momentum strategies followed by many funds, especially growth and aggressive growth funds (Grinblatt, Titman, and Wermers (1995)).

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

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

(10)

where $z_{t-1}$ denotes the $J \times 1$ vector of centered predictive variables, and $B$ is the $J \times 1$ vector of coefficients. Four predictive variables are considered. The first is the one-month T-bill rate, while the second is the dividend yield of the CRSP value-weighted NYSE and AMEX stock index. The third is the term spread, proxied by the difference between the 10-year T-bond yield and the three-month T-bill rate, while the fourth is the default spread, proxied by the yield difference between Baa-rated and Aaa-rated corporate bonds. We have also computed fund alphas using the CAPM and the Fama-French models. These results are summarized in Section 3.5.

To compute each fund $t$-statistic, we use the Newey-West (1987) heteroscedasticity and autorcorrelation consistent standard deviation, $\hat{\sigma}_{\hat{\alpha}_i}$. KTWW (2006) find that the finite-sample distribution of the fund $t$-stat, $\hat{t}_i$, is non-normal for approximately half of
the funds. Therefore, we use a bootstrap procedure (instead of asymptotic theory) to compute the fund $p$-values. In order to approximate the distribution of $\hat{\epsilon}_i$ under the null hypothesis $\alpha_i = 0$, we use a semi-parametric bootstrap procedure, which draws with replacement from the regression estimated residuals $\{\hat{\epsilon}_{i,t}\}$.\textsuperscript{16} For each fund, we implement 1,000 bootstrap iterations. Since our procedure is similar to the one implemented by KTWW (2006), the reader is referred there for further details.

B Mutual Fund Data

We use monthly mutual fund return data provided by the Center for Research in Security Prices (CRSP) between January 1975 and December 2006 to estimate fund alphas. Each fund return is computed by weighting the net return of each shareclass by its total net asset value at the beginning of each month. The CRSP database is matched with the Thomson/CDA database using the MFLINKs product of Wharton Research Data Services (WRDS) in order to use the Thomson fund investment-objective information. Wermers (2000) gives a precise description of these two databases. Our original sample is free of survivorship bias, but we further select only funds having at least 60 monthly return observations in order to get precise four-factor alpha estimates. These monthly returns need not be contiguous. When we observe a missing return, we delete the following-month return, since CRSP fills this with the cumulated return since the last non-missing return. In unreported results, we find that reducing the minimum length to 36 observations leaves our results unchanged, thus, we believe that our results are not substantially impacted by survival bias considerations.

Our final universe of funds is composed of 2,076 open-end, domestic equity mutual funds that exist for at least 60 months between 1975 and 2006. Funds are then classified into three investment categories: growth funds (1,304 funds), aggressive growth funds (388 funds), and growth and income funds (384 funds). A fund is included in a given investment category if its investment objective corresponds to the investment category for at least 60 months. Table I shows the estimated annualized alpha and factor loadings of an equally-weighted portfolio of all funds across the different investment categories. 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

\textsuperscript{16}To know whether assuming homoscedasticity and temporal independence in the fund residuals is appropriate, we have checked for heteroscedasticity (White test), autocorrelation (Ljung-Box test), and Arch effects (Engle test). We have found that only a few funds present such regularities. We have also implemented a block bootstrap methodology with a block length equal to $T^{1/2}$ (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.
Similar to results previously documented in the literature, we find that the unconditional estimated alpha for all categories is negative, ranging from -0.45% to -0.60% (annualized). Aggressive growth funds tilt towards 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 tests available upon request from the authors, we find that all results to be discussed in the next section are qualitatively similar between the unconditional and conditional four-factor models. Therefore, we present results in the next section only for its unconditional version.

III Empirical Results

A Impact of Luck on Long-Term Performance

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

Among the 2,076 funds, we estimate that a large majority—72.8%—are zero-alpha funds. Managers of these funds exhibit stockpicking skills just sufficient to cover their trading costs and other expenses (including fees). These funds, therefore, capture all of the economic rents that they generate—consistent with the long-run predictions of Berk and Green (2004).

Further, it is quite surprising that the estimated proportion of skilled funds is statistically indistinguishable from zero (see “Skilled” column). This result may seem surprising in light of some prior studies, such as Ferson and Schadt (1996), which find that a small group of top mutual fund managers appear to outperform their benchmarks, net of costs. However, a closer examination—in Panel B—shows that our adjustment for luck is key in understanding the difference between our study and those of prior researchers.
To be specific, Panel B shows the proportion of significant alpha funds in the left and right tails ($\tilde{S}_γ^-$ and $\tilde{S}_γ^+$, 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—$\tilde{S}_γ^+$ peaks at 7.8% (162 funds) of the total population at $\gamma = 0.20$. However, of course, “significant alpha” does not always mean “skilled fund manager.” Illustrating this point, the right side of Panel B decomposes these significant funds into the proportions of lucky zero-alpha funds and skilled funds ($\tilde{F}_γ^+$ and $\tilde{T}_γ^+$, respectively). Clearly, we cannot reject that all of the right tail funds are merely lucky zero-alpha funds.

The bottom of Panel B presents characteristics of the average fund at 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 1,512 zero-alpha funds in the population. Results for the three investment-objective subgroups (Aggressive Growth, Growth, and Growth & Income) are similar—these results are available upon request from the authors.

Please insert Table II here

It is also interesting that Panel A shows that 26.6% of the population (552 funds) are unskilled fund managers—unable to pick stocks well enough to recover their trading costs and other expenses.\textsuperscript{17} While the majority of funds in our sample are zero-alpha, consistent with the long-run predicted equilibrium of Berk and Green (2004), it is puzzling that a substantial number of negative alpha funds can survive. In fact, in untabulated results, we find that left-tail funds, which are overwhelmingly comprised of unskilled (and not merely unlucky) funds, have a relatively long fund life—12.6 years, on average. Perhaps, as discussed by Elton, Gruber, and Busse (2003), such funds exist if they are able to attract a sufficient number of unsophisticated investors; Christoffersen and Musto (2002) further explain that this category of investors may be charged higher fees, which is confirmed, since funds in the left tail exhibit higher average expense ratios than those in the right tail (Panel B). However, fee differences do not appear to completely explain the underperformance. In a later section of this paper, we will shed further light on this issue by exploring the role of trading costs—by examining pre-expense alphas of the funds.

As mentioned earlier, the universe of U.S. domestic equity mutual funds has increased substantially since 1990. Accordingly, we next examine the evolution of the proportion

\textsuperscript{17}This minority of funds is the driving force explaining the negative average estimated alpha that is widely documented in the literature (e.g., Jensen (1968), Carhart (1997), Elton et al. (1993), and Pastor and Stambaugh (2002a)).
of unskilled and skilled funds over time. To accomplish this, at the end of each year from 1989 to 2006, we estimate the proportion of unskilled and skilled funds using the entire return history for each fund up to that point in time. For instance, our initial estimates, on December 31, 1989, cover the first ten years of the sample (1975-89), while our final estimates, on December 31, 2006, are based on the entire period 1975-2006 (i.e., these are the estimates shown in Panel A of Table II).\textsuperscript{18} The results in Panel A of Figure 4 show that the proportion of funds with non-zero alphas (equal to the sum of the proportion of skilled plus unskilled) remains fairly constant over time. However, there are dramatic changes in the relative proportions of unskilled and skilled funds: from 1989 to 2006, the proportion of skilled funds declines from 14.4\% to 0.6\%, while the proportion of unskilled rises from 9.2\% to 26.6\% of the entire universe of funds. These changes are also reflected in the population average estimated alpha, which drops from 0.16\% to -0.97\% per year over the same period.

The dark line in Panel B displays the yearly increase in numbers of funds that we analyze. From 1996 to 2005, the yearly increase in funds exceeds 100.\textsuperscript{19} Interestingly, this coincides with the time trend in unskilled and skilled funds shown in Panel A. Thus, the huge increase in numbers of mutual funds has resulted in a much larger proportion of unskilled funds, at the expense of skilled funds. Either the growth of the fund industry has coincided with greater levels of stock market efficiency, making stockpicking a more difficult and costly endeavor, or the large number of new managers simply have inadequate skills. It is also interesting that during our period of analysis, many fund managers with good track records left the sample to manage hedge funds, and that indexed investing increased substantially.

\textbf{B Impact of Luck on Short-Term Performance}

Our above results indicate that funds do not appear to provide superior long-term performance, but that they may provide outperformance over the short-term. This prompts us to more formally examine mutual fund performance over short time horizons. Specifically, we partition our data into six non-overlapping subperiods of five years, beginning with 1977-1981 and ending with 2002-2006. For each subperiod, we select all funds having 60 monthly return observations, then compute their respective alpha $p$-values—

\textsuperscript{18} To be included in a given year, a fund must have at least 60 monthly return observations, although these observations need not be contiguous.

\textsuperscript{19} Since we require 60 monthly observations to measure the fund performance, this rise reflects the massive entry of new funds over the period 1993-2001.
other words, we treat each fund during each five-year period as a separate fund. After pooling all of these subperiods together, we obtain a total of 3,311 \( p \)-values from which we compute our different estimators. Results for the entire population (All Funds), as well as for Growth, Aggressive Growth, and Growth & Income funds are displayed in Panels A, B, C, and D of Table III, respectively. The rightmost columns examine the right tail of the cross-sectional \( t \)-distribution. For each significance level \( (\gamma = 0.05, 0.10, 0.15, 0.20) \), we show the proportion of significant funds \( \left( \hat{S}_{\gamma}^+ \right) \), and its decomposition into lucky and skilled funds \( (\hat{F}_{\gamma}^+, \hat{T}_{\gamma}^+) \) (in parentheses are the standard deviations of the point estimates). The leftmost columns repeat this analysis for the left tail.

Please insert Table III here

The performance analysis in the right tail for all funds (Panel A) provides two interesting results. First, an examination of \( \hat{T}_{\gamma}^+ \) shows that 2.4% of the population is skilled (of a total of 9.6% right-tail funds). This finding implies that 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—with a \( \gamma \) of only 10%, we capture all skilled funds, as \( \hat{T}_{\gamma}^+ \) reaches its maximum value. Thus, proceeding toward the center of the distribution (by increasing \( \gamma \) to 0.15 and 0.20) produces only additional zero-alpha funds that are lucky. This finding indicates that skilled fund managers, while rare, may be somewhat easy to find, since they have extremely high \( t \)-statistics (extremely low \( p \)-values). It is notable that we find evidence of short-term outperformance of some funds, but no evidence of long-term outperformance. This finding is consistent with the model of Berk and Green (2004), where outperforming funds may exist only until investors are successfully able to locate them.

In the left tail, we observe that the great majority of funds are unskilled, and not merely unlucky zero-alpha funds. For instance, in the extreme left tail (at \( \gamma = 0.05 \)), the proportion of unskilled funds, \( \hat{T}_{\gamma}^- \), is roughly five times larger than the proportion of unlucky funds, \( \hat{F}_{\gamma}^- \) (9.4% versus 1.8%). Here, the short-term results are similar to the long-term results discussed previously—the great majority of left-tail funds are truly unskilled.

Among the different investment categories, the proportion of skilled Growth funds in the right-tail (Panel B) is similar to that of the entire universe (Panel A). However, Aggressive-Growth funds contain the highest proportion of skilled funds (Panel C). For 20Note that reducing the number of observations comes at a cost: it greatly increases the standard deviation of the estimated alphas, making the \( p \)-values of non zero-alpha funds harder to distinguish from those of zero-alpha funds.
instance, at $\gamma = 0.05$, fewer than 40% of significant Aggressive-Growth funds are lucky (1.8/4.4), indicating that most extreme right-tail funds are skilled. On the contrary, Panel D reveals that no Growth & Income funds are truly skilled, but that a substantial proportion of them are unskilled. The long-term existence of this category of actively-managed funds, which includes “value funds” and “core funds” is remarkable in light of these poor results.

Please insert Table IV here

C  Performance Persistence

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

To circumvent this difficulty, we build on one of our main empirical findings in this paper: skilled funds are located in the extreme right tail. By forming portfolios containing all funds in this extreme tail, we have a greater chance of capturing the superior alphas of the truly skilled funds. For instance, Panel A of Table III shows that, at $\gamma = 0.05$, 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 among the entire universe, 2.4%.

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

$$FDR^+_{\gamma} = E\left(\frac{F^+_{\gamma}}{S^+_{\gamma}}\right).$$  \hspace{1cm} (11)

The $FDR^+_{\gamma}$ provides a simple portfolio formation rule. When we set a low $FDR^+$ target, we only allow a small proportion of lucky funds ("false discoveries") in the chosen portfolio. Specifically, we set a sufficiently low significance level, $\gamma$, so as to include skilled funds along with a small number of zero-alpha funds that are extremely lucky. Conversely, increasing the $FDR^+$ target has two opposite effects on a portfolio. First, it decreases the portfolio’s expected future performance, since the proportion of lucky

$^{21}$Our new measure, $FDR^+_{\gamma}$, is an extension of the traditional $FDR$ introduced in the statistical literature (e.g., Benjamini and Hochberg (1995), Storey (2002)), since the latter does not distinguish between bad and good luck: $FDR_{\gamma} = E\left(F_{\gamma}/S_{\gamma}\right)$, where $F_{\gamma} = F^+_{\gamma} + F^-_{\gamma}$, $S_{\gamma} = S^+_{\gamma} + S^-_{\gamma}$.
funds in the portfolio is higher. However, it also increases its diversification, since more funds are selected. Accordingly, we examine five $FDR^+$ target levels in our persistence test: 10%, 30%, 50%, 70%, and 90%.

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

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

(12)

where $\hat{\pi}_0$ is the estimator of the proportion of zero-alpha funds described in Section I.A.3. For each $FDR^+_\gamma$ target level, we determine the significance level, $\gamma^P$, that provides an $FDR^+_{\gamma^P}$ as close as possible to this target. Then, only funds with $p$-values smaller than $\gamma^P$ are included in an equally-weighted portfolio. This portfolio is held for one year, after which the selection procedure is repeated. If a selected fund does not survive after a given month, its weight is reallocated to the remaining funds during the rest of the year to mitigate survival bias. The first portfolio formation date is December 31, 1979 (after the first five years), while the last is December 31, 2005.

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

In Panel B, we present the average annual performance of these five portfolios between January 1, 1980 and December 31, 2006. We compute the estimated annual alpha, $\hat{\alpha}$, along with its bootstrapped $p$-value; annual residual standard deviation, $\hat{\sigma}_\varepsilon$; information.
The information ratio $\text{IR} = \frac{\hat{\alpha}}{\hat{\sigma}_\varepsilon}$; four-factor model loadings; and annual mean (minus T-bills) and standard deviation. 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 portfolio.

Finally, Panel C examines portfolio turnover. For each portfolio, we determine the proportion of funds which are still selected 1, 2, 3, 4, and 5 years after their initial inclusion. The results sharply illustrate the short-term nature of truly outperforming funds. After 1 year, fewer than $50\%$ of the funds are still included in the portfolios $FDR_{10\%}$ and $30\%$, while after 3 years, these percentages further fall to $3.4\%$ and $5.1\%$, respectively.

Finally, we examine, in Figure 5, how the estimated alpha of the portfolio $FDR_{10\%}$ evolves during the period 1990-2006. The similarity with Figure 4 is striking. While the alpha is extremely high at the beginning of the period, it consistently declines during the 1990s. As the proportion, $\pi^*_A$, of skilled funds falls, the $FDR$ approach moves much further to the extreme right tail of the cross-sectional $t$-distribution (from $10.7\%$ of all funds in 1990 to $1.7\%$ in 2000) in search of skilled funds. However, this change is not sufficient to prevent the performance 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), Carhart (1997)). These prior papers generally classify funds into fractile portfolios based on their past performance (past returns, estimated alpha, or $t$-statistic of alpha) over a previous ranking period (one to three years). Here, funds are ranked on their estimated performance without testing whether this performance is due to luck alone. As a result, the signal used to form portfolios is likely to be noisier, than using our $FDR$ approach. To compare these approaches, Figure 5 displays the performance evolution of two top decile portfolios which are formed based on ranking funds by their alpha $t$-statistic, estimated over the previous one and three years, respectively.\footnote{We use the $t$-statistic to be consistent with the rest of our paper, but the results are qualitatively similar when we rank on the estimated alpha.} 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, we define a fund as skilled if it generates a positive alpha net of trading costs, fees, and other expenses. Alternatively, skill could be defined in an absolute sense as the manager’s ability to produce a positive alpha before expenses are deducted. Measuring performance on a pre-expense basis allows one to disentangle the manager’s stockpicking skills, net of trading costs, 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 to net returns for each fund, then revisit the long-term performance of the mutual fund industry.\footnote{We discard funds which do not have at least 60 pre-expense return observations over the period 1975-2006. This leads to a small reduction in our sample from 2,076 to 1,836 funds.}

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 26.6% to 4.5%. This result indicates that only a small fraction of fund managers have stockpicking skills that are not sufficient to at least compensate for their trading costs. Instead, mutual funds produce negative net-of-expense alphas chiefly because they charge excessive fees, in relation to the selection abilities of their managers. In Panel B, we further find that the average expense ratio across funds in the left tail is lower when performance is measured prior to expenses (1.4% versus 1.6% per year), indicating that high fees (potentially charged to unsophisticated investors) are a chief reason why funds end up in the extreme left tail.

Please insert Table VI here

In the right tail, we find that 9.6% of fund managers have stockpicking skills sufficient to more than compensate for trading costs (Panel A). Consistent with Berk and Green (2004), the rents stemming from their skills are extracted through fees and expenses, driving the proportion of net-expense skilled funds to zero.
Since 72.8% of funds produce zero net-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 funds in the center of the cross-sectional t-distribution. Adding back these expenses leads only to a marginal increase in the alpha t-statistic, making the power of the tests rather low.\(^{25}\)

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

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

### D.2 Performance Measured with Other Asset Pricing Models

Our estimation of the proportions of unskilled and skilled funds, \((\hat{\pi}_A^- \text{ and } \hat{\pi}_A^+)\) obviously depends on the choice of the asset pricing model. To examine the sensitivity of our results, we repeat the long-term 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.1. However, this result may be due to the omission of the size, book-to-market, and momentum factors. This conjecture is confirmed in Panel A of Table VII: the funds located in the right tail (according to the CAPM) have substantial loadings on the size and the book-to-market factors, which carry positive risk premia over our sample period (3.7% and 5.4% per year, respectively).

Turning to the Fama-French 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-

\(^{25}\)The average expense ratio across funds with \(|\tilde{\alpha}_t| < 1\%\) is approximately 10 bp per month. Adding back these expenses to a fund with zero net-expense alpha only increases its t-statistic from 0 to 0.9 (based on \(T \hat{\alpha}_A / \hat{\sigma}_\varepsilon\), with \(T = 384\), and \(\hat{\sigma}_\varepsilon = 0.021\)). It implies that the null and alternative t-stat distributions are extremely difficult to distinguish (i.e., under the alternative, the probability of observing a negative pre-expense t-statistic is equal to 18%!).

23
factor model, since 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,
also has a natural Bayesian interpretation.\(^{26}\) To see this, we denote, by \( H_i \),
a random variable which takes the value of -1 if fund \( i \) is unskilled, 0 if it has zero alpha,
and +1 if it is skilled. The prior probabilities for the three possible values (-1, 0, +1)
are given by the proportion of each category in the population, \( \pi^{-A} \), \( \pi_{0} \), and \( \pi^{+A} \). The
Bayesian version of our \( FDR^+ \) measure, denoted by \( fdr^+_{\gamma} \), is defined as the posterior
probability that the fund \( i \) has a zero alpha given that its \( t \)-statistic, \( \hat{t}_i \), is positive and
significant: \( fdr^+_{\gamma} = \text{prob} \left( H_i = 0 \mid \hat{t}_i \in \Gamma^+_A (\gamma) \right) \), where \( \Gamma^+_A (\gamma) = (t^+_\gamma, +\infty) \). Using Bayes
theorem, we have:

\[
fdr^+_{\gamma} = \frac{\text{prob} \left( \hat{t}_i \in \Gamma^+_A (\gamma) \mid H_i = 0 \right) \cdot \text{prob}(H_i = 0)}{\text{prob} \left( \hat{t}_i \in \Gamma^+_A (\gamma) \right)} = \frac{\gamma/2 \cdot \pi_0}{E(S^+_\gamma)}. \tag{13}
\]

Stated differently, the \( fdr^+_{\gamma} \) indicates how the investor changes his prior probability that
fund \( i \) has a zero alpha (\( H_i = 0 \)) after observing that its \( t \)-statistic is significant. In light
of Equation (13), our estimator \( \hat{FDR}^+_{\gamma} = (\gamma/2 \cdot \pi_0)/\hat{S}^+_\gamma \) can therefore be interpreted as
an empirical Bayes estimator of \( fdr^+_{\gamma} \), where \( \pi_0 \) and \( E(S^+_\gamma) \) are directly estimated from
the data.\(^{27}\)

In the recent Bayesian literature on mutual fund performance (e.g., Baks, Metrick,
and Wachter (2001) and Pastor and Stambaugh (2002a)), attention is given to the pos-
terior distribution of the fund alpha, \( \alpha_i \), as opposed to the posterior distribution of \( H_i \).
Interestingly, our approach also provides some relevant information for modelling the
fund alpha prior distribution in an empirical Bayes setting. The parameters of the prior
can be specified based on the relative frequency of the three fund categories (zero-alpha,
unskilled, and skilled funds). In light of our estimates, an empirically-based alpha prior distribution is characterized by a point mass at $\alpha = 0$, reflecting the fact that 72.8% of the funds yield zero alphas. Since $\hat{\pi}_A^+ > \hat{\pi}_A^-$, the probability of observing a negative rather than a positive alpha is higher. These empirical constraints yield an asymmetric prior distribution. A tractable way to model the left and right parts of this distribution is to exploit two truncated normal distributions in the same spirit as in Baks, Metrick, and Wachter (2001). Further, our estimates imply that 73.4\% of the funds have an alpha greater than or equal to zero. While Baks, Metrick, and Wachter (2001) set this probability to 1\% in order to examine the portfolio decision made by a skeptical investor, our analysis reveals that this level represents an overly skeptical belief.

IV Conclusion

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

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

It is puzzling why investors seem to increasingly tolerate the existence of a large number of funds that provide negative alphas, when a large number of passively managed funds have become available (such as ETFs). Perhaps a class of unsophisticated or inattentive investors remain shareholders in funds after they have clearly demonstrated (over time) their inferior returns. Or, as Elton, Gruber, and Blake (2007; EGB) document, maybe investors are forced to make constrained rational decisions—since EGB document that many 401(k) plans offer inefficient choices of mutual funds.
V Appendix

A Estimation Procedure

A.1 Determining the Value for $\lambda^*$ from the Data

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

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

We choose $\lambda^*$ such that $\lambda^* = \arg\min_{\lambda} \widehat{MSE}(\lambda)$. In unreported results (available upon request), we find that fixing $\lambda$ to 0.5 or 0.6 yields similar results to those obtained with the bootstrap procedure (see also Storey (2002)). Still, the main advantage of the bootstrap approach is that it is entirely data-driven.

A.2 Determining the Value for $\gamma^*$ from the Data

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

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

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

26
A.3 Determining the Standard Deviation of the Estimators

We rely on the large-sample theory proposed by Genovese and Wasserman (2004). The essential idea is to recognize that the estimators $\pi_0(\lambda^*)$, $\hat{S}_\gamma^+$, $\hat{F}_\gamma^+$, $\hat{T}_\gamma^+$, $\hat{S}_\gamma^-$, $\hat{F}_\gamma^-$, and $\hat{T}_\gamma^-$ are all stochastic processes indexed by $\lambda^*$ or $\gamma$ which converge to a Gaussian process when the number of funds, $M$, goes to infinity. Proposition 3.2 of Genovese and Wasserman (2004) shows that $\pi_0(\lambda^*)$ is asymptotically normally distributed when $M \to \infty$, with standard deviation $\hat{\sigma}_{\pi_0}(\lambda^*) = \left(\frac{\hat{W}(\lambda^*)(M-\hat{W}(\lambda^*))}{M^2(1-\lambda^*)^2}\right)^{1/2}$. Similarly, we have $\hat{\sigma}_{\hat{F}_\gamma^+} = (\gamma/2)\hat{\sigma}_{\pi_0(\lambda^*)}$, $\hat{\sigma}_{\hat{S}_\gamma^+} = \left(\frac{\hat{S}_\gamma^+(1-\hat{S}_\gamma^+)}{M}\right)^{1/2}$, and $\hat{\sigma}_{\hat{F}_\gamma^-} = \left(\frac{\hat{S}_\gamma^-+\hat{\sigma}_{\pi_0(\lambda^*)}^2+2(\gamma/2)\hat{W}(\lambda^*)/M}{\hat{S}_\gamma^-+\hat{F}_\gamma^-+\hat{T}_\gamma^-}\right)^{1/2}$ (using the equality $\hat{S}_\gamma^+ - \hat{F}_\gamma^- - \hat{T}_\gamma^- = \hat{S}_\gamma^+$). Standard deviation for the estimators in the left tail $(\hat{S}_\gamma^-, \hat{F}_\gamma^-, \hat{T}_\gamma^-)$ are obtained by simply replacing $\hat{S}_\gamma^+$ with $\hat{S}_\gamma^-$ in the above formulas.

Finally, if $\gamma^* = \gamma^+$, the standard deviation of $\pi_A^+$ and $\pi_A^*$ are respectively given by $\hat{\sigma}_{\pi_A^+} = \hat{\sigma}_{\pi_A^*} = \left(\frac{\hat{\sigma}_{\pi_0(\lambda^*)}^2+\hat{\sigma}_{\pi_0(\lambda^*)}^2}{1-\lambda^*}\right)^{1/2}$ (using the equality $\hat{\pi}_A^* = 1 - \hat{\pi}_A^0 - \hat{\pi}_A^+$). Otherwise if $\gamma^* = \gamma^-$, we just reverse the signs $+/-$ in the two formulas above.

B Monte-Carlo Analysis

B.1 Under Cross-Sectional Independence

We use Monte-Carlo simulations to examine the performance of all estimators used in the paper: $\pi_0$, $\pi_A^*$, $\pi_A^+$, $\hat{S}_\gamma^-$, $\hat{F}_\gamma^-$, $\hat{T}_\gamma^-$, and $\hat{S}_\gamma^+$, $\hat{F}_\gamma^+$, $\hat{T}_\gamma^+$. We generate the $M \times 1$ vector of fund monthly excess returns, $r_t$, according to the four-factor model (market, size, book-to-market, and momentum factors):

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

(16)

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

Our estimators are compared with their respective true population values defined as follows. After setting values for the parameters $\pi_0$, $\pi_A^*$, $\pi_A^+$, and for the $t$-statistic means
(t mean) of the unskilled and skilled funds, we can compute $E(F^-_t) = E(F^+_t) = \pi_0 \cdot \gamma/2$.

To determine the expected proportion of unskilled and skilled funds, $E(T^-_t)$ and $E(T^+_t)$, we use the fact that, under the alternative hypothesis $\alpha_i \neq 0$, the fund $t$-statistic follows a non-central student distribution with $T - 5$ degrees of freedom and a noncentrality parameter equal to $T^{1/2} \alpha_A / \sigma_\varepsilon$ (Davidson and MacKinnon (2004), p. 169):

$$
E(T^-_t) = \pi^-_A \cdot \text{prob}\left(t < t_{T-5,\gamma/2} | H_A, \alpha_A < 0\right),
$$

$$
E(T^+_t) = \pi^+_A \cdot \text{prob}\left(t > t_{T-5,1-\gamma/2} | H_A, \alpha_A > 0\right),
$$

(17)

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

The total number of funds, $M$, used in the simulation is equal to 1,500.28 The input for $\beta$ is equal to the empirical loadings of a random draw of 1,500 funds (among the total population of 2,076 funds). The vector of fund alphas, $\alpha$, is built by randomly choosing the identity of the unskilled and skilled funds among the 1,500 funds. Consistent with our database, we set $T = 384$, $\sigma_\varepsilon = 0.021$ (equal to the empirical average across the 1,500 funds), and proxy $\Sigma_F$ by its empirical counterpart.

Realistic values for the parameters $\pi_0$, $\pi^-_A$, $\pi^+_A$, and the $t$ means are determined as follows. To set values for the proportions of the three skill groups, we estimate $\pi_0$, $\pi^-_A$, and $\pi^+_A$ at the end of each of the final 5 years of our sample (2002-2006) using the entire return history for each fund up to that point in time. These estimates are then averaged to produce values that reflect the recent trend observed in Figure 4: $\pi_0 = 75\%$, $\pi^-_A = 23\%$, and $\pi^+_A = 2\%$. To determine the $t$ means of the unskilled and skilled funds, we use a simple calibration method. We start by computing the average $\bar{T}_t^-$ and $\bar{T}_t^+$ (at $\gamma = 0.20$) over the final 5 years of our sample (2002-2006). Inserting these values along with $\pi^-_A = 23\%$ and $\pi^+_A = 2\%$ in Equation (17), we can determine what are the $t$-means which satisfy both equalities. The values obtained are equal to -2.5 and 3, and correspond to an annual four-factor alpha of -3.2% and 3.8%, respectively (using the equality $t_A = T^{1/2} \alpha_A / \sigma_\varepsilon$).

After drawing randomly $F_t$ and $\varepsilon_t$ ($t = 1, \ldots, 384$), we construct the fund return time-series according to Equation (14), and compute their $t$-statistic by regressing the fund returns on the four-factor model. To determine the $p$-values, we use the fact that the fund $t$-statistic follows a Student distribution with $T - 5$ degrees of freedom under

---

28 We use this sample size to allow for comparison with the dependence case (described hereafter), which uses a sample of 1,500 correlated fund returns. Since our original sample of funds is higher than 1,500 ($M = 2,076$), our assessment of the precision of the estimators is conservative.
the null hypothesis $\alpha_i = 0$. Then, we compute $\hat{\pi}_0$, $\hat{\pi}_A$, and $\hat{\pi}_A^+$ using Equations (5) and (8). $\hat{S}_-^i$ and $\hat{S}_+^i$ correspond to the observed number of significant funds with negative and positive alphas, respectively. $\hat{F}_-^\gamma$ and $\hat{F}_+^\gamma$ are computed with Equation (6). $\hat{T}_-^\gamma$ and $\hat{T}_+^\gamma$, are given in Equation (7). We repeat this procedure 1,000 times.

In Table VIII, we compare the average value of each estimator (over the 1,000 replications) with the true values. The figures in parentheses denote the lower and upper bounds of the estimator 90%-confidence interval. We set $\gamma$ equal to 0.05 and 0.20. In all cases, the simulation results reveal that the average values of our estimators closely match the true values, and that their 90%-confidence intervals are narrow. This result is not surprising in light of the large cross-section of funds available in our sample.

Please insert Table VII here

B.2 Under Cross-Sectional Dependence

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

$$\varepsilon_t \sim N(0, \Sigma),$$

where $\Sigma$ denotes the $M \times M$ residual covariance matrix. The main constraint imposed to $\Sigma$ is that it must be semi-positive definite. To achieve this goal, we proceed as follows. We select all funds with 60 valid return observations over the final 5 years (2002-2006), so as to have the largest possible cross-section of funds. This procedure gives a number of 898 funds, whose covariance matrix, $\Sigma_1$, is directly estimated from the data. To assess the precision of our estimators, we also need to account for the non-overlapping return observations observed in the data. We address this issue by introducing 602 uncorrelated funds. This yields the following covariance matrix:

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \sigma_\varepsilon^2 I \end{pmatrix}. $$

---

29 The 25%, 50%, and 75% pairwise correlation quantiles amount to -0.09, 0.05, and 0.19, respectively. Therefore, our simulations account for the diversity observed in the data, and for the fact that the correlations may either be positive or negative.

30 The first 200 funds proxy for the 15% of fund pairs which do not have a single common return observation. The remaining 400 funds capture the partial overlapping across the fund pairs (55% on average).
An an input for $\beta$, we use the empirical factor loadings of the 898 along with the loadings of a random draw of the 602 remaining funds. The vector of fund alphas, $\alpha$, is built by randomly choosing the identity of the unskilled and skilled funds among the 1,500 funds. The results in Table IX indicate that all estimators remain nearly unbiased ($\widehat{\pi}_0$, $\widehat{\pi}_A$, and $\widehat{\pi}_A^+$ only present a minor bias). Looking at the 90% confidence intervals, we logically observe that the dispersion of the estimators widens under cross-sectional dependence. However, the performance of the estimators is still very good.

Please insert Table IX here

Apart from this baseline dependence scenario, we also examine two other cases. First, we introduce correlation by block among each skill group (zero-alpha, unskilled, and skilled funds) to account for their possible similar bets. Inside each block (representing 10% of each skill group), we set the pairwise correlation equal to 0.15 or 0.30. In the second dependence case, we use the residual factor specification proposed by Jones and Shanken (2005) in order to capture the role of non-priced factors. We assume that all fund residuals depend on a common residual factor, and that the unskilled and skilled funds are affected by specific residual factors. The results, available upon request, show that the precision of the estimators remain very close to those obtained under the independence case.
References

[1] Avramov, Doron, and


Table I
Performance of the Equally-Weighted Portfolio of Funds

Results for the unconditional and conditional four-factor models are shown in Panels A and B for the entire fund population (All funds), as well as Growth, Aggressive Growth, and Growth 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-consistent $p$-values, under the null hypothesis that the regression parameters are equal to zero.

### Panel A Unconditional Four-Factor Model

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

### Panel B Conditional Four-Factor Model

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\alpha}$</th>
<th>$\hat{b}_m$</th>
<th>$\hat{b}_{smb}$</th>
<th>$\hat{b}_{hml}$</th>
<th>$\hat{b}_{mom}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (2,076)</td>
<td>-0.60%</td>
<td>0.96</td>
<td>0.17</td>
<td>-0.02</td>
<td>0.02</td>
<td>98.2%</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.23)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>Growth (1,304)</td>
<td>-0.59%</td>
<td>0.96</td>
<td>0.16</td>
<td>-0.03</td>
<td>0.03</td>
<td>98.2%</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.08)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>Aggressive</td>
<td>-0.49%</td>
<td>1.05</td>
<td>0.43</td>
<td>-0.19</td>
<td>0.08</td>
<td>96.2%</td>
</tr>
<tr>
<td>Growth (388)</td>
<td>(0.24)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Growth &amp;</td>
<td>-0.58%</td>
<td>0.87</td>
<td>-0.04</td>
<td>0.16</td>
<td>-0.03</td>
<td>98.3%</td>
</tr>
<tr>
<td>Income (384)</td>
<td>(0.05)</td>
<td>(0.00)</td>
<td>(0.02)</td>
<td>(0.00)</td>
<td>(0.02)</td>
<td></td>
</tr>
</tbody>
</table>
Table II
Impact of Luck on Long-Term Performance

Panel A displays the estimated proportions of zero-alpha, unskilled, and skilled funds in the entire fund population (2,076 funds). We measure fund performance with the unconditional four-factor model over the entire period 1975-2006. Panel B counts the proportions of significant funds in the left and right tails of the cross-sectional t-statistic distribution \( (\hat{S}_{-\gamma}, \hat{S}_{+\gamma}) \) 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 \( (\hat{F}_{+\gamma}, \hat{T}_{+\gamma}) \). Figures in parentheses denote the standard deviation of the different estimators. The bottom of Panel B also presents the characteristics of each significant group \( (\hat{S}_{-\gamma}, \hat{S}_{+\gamma}) \): the average estimated alpha (in % per year), expense ratio (in % per year), turnover (in % per year), and median size measured by the total net asset under management (in millions of dollars).

| Panel A Proportion of Unskilled and Skilled Funds | Zero alpha(\( \hat{\pi}_0 \)) | Non-zero alpha | Unskilled(\( \hat{\pi}_A \)) | Skilled(\( \hat{\pi}_A^+ \)) |
| Proportion | 72.8 (2.7) | 27.2 | 26.6 (4.1) | 0.6 (0.8) |
| Number | 1,512 | 564 | 552 | 12 |

| Panel B Impact of Luck in the Left and Right Tails |
| Signif. level(\( \gamma \)) | Left Tail | Right Tail | Signif. level(\( \gamma \)) |
| | 0.05 | 0.10 | 0.15 | 0.20 | 0.20 | 0.15 | 0.10 | 0.05 | Signif. level(\( \gamma \)) |
| Signif. \( \hat{S}_{-\gamma} \) (%) | 11.6 | 17.2 | 21.5 | 25.4 | 7.8 | 5.8 | 4.0 | 2.1 | Signif. \( \hat{S}_{+\gamma} \) (%) |
| (0.7) | (0.9) | (1.0) | (1.1) | (0.7) | (0.6) | (0.5) | (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) | (0.1) | (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) | (0.8) | (0.9) | (0.9) | (0.6) | (0.5) | (0.4) | (0.3) | |
| Alpha(\% year) | -5.5 | -5.0 | -4.7 | -4.6 | 4.8 | 5.2 | 5.5 | 6.5 | Alpha(\% year) |
| Expense(\% year) | 1.6 | 1.6 | 1.5 | 1.5 | 1.3 | 1.2 | 1.2 | 1.2 | Expense(\% year) |
| Turnover(\% year) | | | | | | | | | Turnover(\% year) |
| Size(million $) | | | | | | | | | Size(million $) |

35
**Table III**  
**Impact of Luck on Short-Term Performance**

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

### Panel A Proportion of Unskilled and Skilled Funds

<table>
<thead>
<tr>
<th>Proportion</th>
<th>Zero alpha($\widehat{\pi}_0$)</th>
<th>Non-zero alpha</th>
<th>Unskilled($\widehat{\pi}_A^-$)</th>
<th>Skilled($\widehat{\pi}_A^+$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion</td>
<td>72.8 (2.7)</td>
<td>27.2</td>
<td>26.6 (4.1)</td>
<td>0.6 (0.8)</td>
</tr>
<tr>
<td>Number</td>
<td>1,512</td>
<td>564</td>
<td>552</td>
<td>12</td>
</tr>
</tbody>
</table>

### Panel B Impact of Luck in the Left and Right Tails

<table>
<thead>
<tr>
<th>Signif. level($\gamma$)</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.20</th>
<th>0.10</th>
<th>0.10</th>
<th>0.05</th>
<th>Signif. level($\gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signif. $\widehat{S}_\gamma^-$ (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Signif. $\widehat{S}_\gamma^+$ (%)</td>
</tr>
<tr>
<td>(0.5)</td>
<td>(0.6)</td>
<td>(0.7)</td>
<td>(0.8)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.4)</td>
<td>(0.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unlucky $\widehat{F}_\gamma^-$ (%)</td>
<td>1.8</td>
<td>3.6</td>
<td>5.4</td>
<td>7.2</td>
<td>7.2</td>
<td>5.4</td>
<td>3.6</td>
<td>1.8</td>
<td>Lucky $\widehat{F}_\gamma^+$ (%)</td>
</tr>
<tr>
<td>(0.0)</td>
<td>(0.0)</td>
<td>(0.1)</td>
<td>(0.2)</td>
<td>(0.2)</td>
<td>(0.1)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unskilled $\widehat{T}_\gamma^-$ (%)</td>
<td>9.4</td>
<td>13.2</td>
<td>16.0</td>
<td>17.7</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>1.7</td>
<td>Skilled $\widehat{T}_\gamma^+$ (%)</td>
</tr>
<tr>
<td>(0.6)</td>
<td>(0.7)</td>
<td>(0.8)</td>
<td>(0.8)</td>
<td>(0.6)</td>
<td>(0.5)</td>
<td>(0.4)</td>
<td>(0.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha(% year)</td>
<td>-5.5</td>
<td>-5.0</td>
<td>-4.7</td>
<td>-4.6</td>
<td>4.8</td>
<td>5.2</td>
<td>5.5</td>
<td>6.5</td>
<td>Alpha(% year)</td>
</tr>
<tr>
<td>Expense(% year)</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>1.5</td>
<td>1.3</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>Expense(% year)</td>
</tr>
<tr>
<td>Turnover(% year)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Turnover(% year)</td>
</tr>
<tr>
<td>Size(million $)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Size(million $)</td>
</tr>
</tbody>
</table>
Table IV

Short-Term Performance across Investment Categories

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

Panel A Growth funds

<table>
<thead>
<tr>
<th>Signif. level($\gamma$)</th>
<th>Left Tail</th>
<th>Right Tail</th>
<th>Signif. level($\gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05 0.10 0.15 0.20</td>
<td>0.20 0.15 0.10 0.05</td>
<td></td>
</tr>
<tr>
<td>Signif. $S_-^\gamma$(%)</td>
<td>11.4 16.8 21.5 24.8</td>
<td>9.9 8.1 6.2 3.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.7) (0.8) (0.9) (1.0)</td>
<td>(0.7) (0.6) (0.5) (0.4)</td>
<td></td>
</tr>
<tr>
<td>Unlucky $\tilde{F}_-^\gamma$(%)</td>
<td>1.8 3.6 5.5 7.3</td>
<td>7.3 5.5 3.6 1.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0) (0.1) (0.2) (0.2)</td>
<td>(0.2) (0.2) (0.1) (0.0)</td>
<td></td>
</tr>
<tr>
<td>Unskilled $\tilde{T}_-^\gamma$(%)</td>
<td>9.6 13.1 16.0 17.6</td>
<td>2.6 2.6 2.6 1.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.7) (0.9) (1.0) (1.1)</td>
<td>(0.8) (0.7) (0.6) (0.4)</td>
<td></td>
</tr>
<tr>
<td>Alpha(% year)</td>
<td>-5.5 -5.0 -4.7 -4.6</td>
<td>4.8 5.2 5.5 6.5</td>
<td></td>
</tr>
<tr>
<td>Expense(% year)</td>
<td>1.6 1.6 1.5 1.5</td>
<td>1.3 1.2 1.2 1.2</td>
<td></td>
</tr>
<tr>
<td>Turnover(% year)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size(million $)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

37
Table IV
Short-Term Performance across Investment Categories (Continued)

Panel B Aggressive Growth funds

<table>
<thead>
<tr>
<th>Signif. level($\gamma$)</th>
<th>Left Tail</th>
<th>Right Tail</th>
<th>Signif. level($\gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05 0.10 0.15 0.20</td>
<td>0.20 0.15 0.10 0.05</td>
<td></td>
</tr>
<tr>
<td>Signif. $S^-_\gamma$ (%)</td>
<td>11.2 16.3 19.6 22.4</td>
<td>11.6 9.8 7.5 4.4</td>
<td>Signif. $S^+_\gamma$ (%)</td>
</tr>
<tr>
<td></td>
<td>(1.3) (1.5) (1.6) (1.7)</td>
<td>(1.3) (1.2) (1.0) (0.8)</td>
<td></td>
</tr>
<tr>
<td>Unlucky $\tilde{F}^-_\gamma$ (%)</td>
<td>1.8 3.5 5.3 7.1</td>
<td>7.1 5.3 3.5 1.8</td>
<td>Lucky $\tilde{F}^+_\gamma$ (%)</td>
</tr>
<tr>
<td></td>
<td>(0.1) (0.2) (0.3) (0.4)</td>
<td>(0.4) (0.3) (0.2) (0.1)</td>
<td></td>
</tr>
<tr>
<td>Unskilled $\tilde{T}^-_\gamma$ (%)</td>
<td>9.4 12.9 14.3 15.3</td>
<td>4.5 4.5 4.0 2.6</td>
<td>Skilled $\tilde{T}^+_\gamma$ (%)</td>
</tr>
<tr>
<td></td>
<td>(1.3) (1.6) (1.7) (1.9)</td>
<td>(1.4) (1.3) (1.1) (0.8)</td>
<td></td>
</tr>
<tr>
<td>Alpha(% year)</td>
<td>-5.5 -5.0 -4.7 -4.6</td>
<td>4.8 5.2 5.5 6.5</td>
<td>Alpha(% year)</td>
</tr>
<tr>
<td>Expense(% year)</td>
<td>1.6 1.6 1.5 1.5</td>
<td>1.3 1.2 1.2 1.2</td>
<td>Expense(% year)</td>
</tr>
<tr>
<td>Turnover(% year)</td>
<td>Size(million $)</td>
<td>Size(million $)</td>
<td>Size(million $)</td>
</tr>
</tbody>
</table>
| Size(million $) | 38
Table V
Performance Persistence Based on the False Discovery Rate

For each of the five FDR targets (10%, 30%, 50%, 70%, and 90%), Panel A contains descriptive statistics on the FDR level (\( \hat{FDR}_{p} \)) achieved by each portfolio, as well as the proportion of funds in the population that they include (\( \hat{S}_{p}^{+} \)). The panel shows the average values of \( \hat{FDR}_{p} \) and \( \hat{S}_{p}^{+} \) over the 27 annual formation dates (from December 1979 to 2005), as well as respective distributions. Panel B displays the performance of each portfolio over the period 1980-2006. We estimate the annual four-factor alpha (\( \hat{\alpha} \)) with its bootstrap p-value, its annual residual standard deviation (\( \hat{\sigma}_{e} \)), its annual information ratio (IR=\( \hat{\alpha}/\hat{\sigma}_{e} \)), 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.

Panel A Portfolio Statistics

<table>
<thead>
<tr>
<th>Achieved False Discovery Rate (( \hat{FDR}_{p} ))</th>
<th>Included proportion of funds (( \hat{S}_{p}^{+} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 10-30 30-50 50-70 &gt;70%</td>
<td>Mean 0-6 6-12 12-24 &gt;24%</td>
</tr>
<tr>
<td><strong>FDR10%</strong></td>
<td>41.5%</td>
</tr>
<tr>
<td><strong>FDR30%</strong></td>
<td>47.5%</td>
</tr>
<tr>
<td><strong>FDR50%</strong></td>
<td>60.4%</td>
</tr>
<tr>
<td><strong>FDR70%</strong></td>
<td>71.3%</td>
</tr>
<tr>
<td><strong>FDR90%</strong></td>
<td>75.0%</td>
</tr>
</tbody>
</table>

Panel B Performance Analysis

<table>
<thead>
<tr>
<th>( \hat{\alpha} ) (p-value)</th>
<th>( \hat{\sigma}_{e} )</th>
<th>IR</th>
<th>( \hat{b}_{m} )</th>
<th>( \hat{b}_{smb} )</th>
<th>( \hat{b}_{hml} )</th>
<th>( \hat{b}_{mom} )</th>
<th>Mean</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FDR10%</strong></td>
<td>1.45%(0.04)</td>
<td>4.0%</td>
<td>0.36</td>
<td>0.93</td>
<td>0.16</td>
<td>-0.04</td>
<td>-0.02</td>
<td>8.3%</td>
</tr>
<tr>
<td><strong>FDR30%</strong></td>
<td>1.15%(0.05)</td>
<td>3.3%</td>
<td>0.35</td>
<td>0.94</td>
<td>0.17</td>
<td>-0.02</td>
<td>-0.03</td>
<td>8.1%</td>
</tr>
<tr>
<td><strong>FDR50%</strong></td>
<td>0.95%(0.10)</td>
<td>2.9%</td>
<td>0.33</td>
<td>0.96</td>
<td>0.20</td>
<td>-0.06</td>
<td>-0.01</td>
<td>8.1%</td>
</tr>
<tr>
<td><strong>FDR70%</strong></td>
<td>0.68%(0.15)</td>
<td>2.7%</td>
<td>0.25</td>
<td>0.97</td>
<td>0.19</td>
<td>-0.06</td>
<td>-0.01</td>
<td>7.9%</td>
</tr>
<tr>
<td><strong>FDR90%</strong></td>
<td>0.39%(0.30)</td>
<td>2.7%</td>
<td>0.14</td>
<td>0.97</td>
<td>0.19</td>
<td>-0.05</td>
<td>-0.00</td>
<td>7.8%</td>
</tr>
</tbody>
</table>

Panel C Portfolio Turnover

<table>
<thead>
<tr>
<th>Proportion of funds</th>
</tr>
</thead>
<tbody>
<tr>
<td>After 1 year</td>
</tr>
<tr>
<td><strong>FDR10%</strong></td>
</tr>
<tr>
<td><strong>FDR30%</strong></td>
</tr>
<tr>
<td><strong>FDR50%</strong></td>
</tr>
<tr>
<td><strong>FDR70%</strong></td>
</tr>
<tr>
<td><strong>FDR90%</strong></td>
</tr>
</tbody>
</table>

39
Table VI

Impact of Luck on Long-Term Pre-Expense Performance

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

Panel A Proportion of Unskilled and Skilled Funds

<table>
<thead>
<tr>
<th>Proportion</th>
<th>Zero alpha ($\hat{\pi}_0$)</th>
<th>Non-zero alpha</th>
<th>Unskilled ($\hat{\pi}^-_A$)</th>
<th>Skilled ($\hat{\pi}^+_A$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>1,577</td>
<td>259</td>
<td>176</td>
<td>83</td>
</tr>
</tbody>
</table>

Panel B Impact of Luck in the Left and Right Tails

<table>
<thead>
<tr>
<th>Signif. level ($γ$)</th>
<th>Left Tail</th>
<th>Right Tail</th>
<th>Signif. level ($γ$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>Signif. $S^-_γ$ (%)</td>
<td>4.3</td>
<td>7.5</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(0.6)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>Unlucky $F^-_γ$ (%)</td>
<td>2.1</td>
<td>4.3</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>(0.0)</td>
<td>(0.1)</td>
<td>(0.1)</td>
</tr>
<tr>
<td>Unskilled $T^-_γ$ (%)</td>
<td>2.2</td>
<td>3.2</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(0.6)</td>
<td>(0.8)</td>
</tr>
<tr>
<td>Pre Expense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha(%) year</td>
<td>-5.8</td>
<td>-5.2</td>
<td>-4.8</td>
</tr>
<tr>
<td>Exp.(%) year</td>
<td>1.4</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Size(million $)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We determine the proportions of significant funds in the left and right tails ($\hat{S}_-, \hat{S}_+$) at four significance levels ($\gamma=0.05, 0.10, 0.15, 0.20$) according to each asset-pricing model over the period 1975-2006. For each of these significant groups, we compute their average loadings on the omitted factors from the four-factor model: size ($\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.

**Table VII**

**Loadings on Omitted Factors**

Panel A Unconditional CAPM

<table>
<thead>
<tr>
<th>Signif. level ($\gamma$)</th>
<th>Left Tail</th>
<th>Right Tail</th>
<th>Signif. level ($\gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>Size ($\hat{b}_{smb}$)</td>
<td>0.06</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>BTM ($\hat{b}_{hml}$)</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.13</td>
</tr>
<tr>
<td>Mom. ($\hat{b}_{mom}$)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Panel B Unconditional Fama-French model

<table>
<thead>
<tr>
<th>Signif. level ($\gamma$)</th>
<th>Left Tail</th>
<th>Right Tail</th>
<th>Signif. level ($\gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>Mom. ($\hat{b}_{mom}$)</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.02</td>
</tr>
</tbody>
</table>
Table VIII
Monte-Carlo Analysis under Cross-Sectional Independence

We examine the average value and the 90%-confidence interval (in parentheses) of the different estimators based on 1,000 replications. For each replication, we generate monthly fund returns for 1,500 funds and 384 periods using the four-factor model (market, size, book-to-market, and momentum factors). Fund residuals are independent from one another. The true parameter values for the proportions of zero-alpha, unskilled, and skilled funds ($\pi_0, \pi_{-A},$ and $\pi_{+A}$) are set to 75%, 23%, and 2%. The mean of the t-statistic of the unskilled and skilled funds are set to -2.5 and 3 (corresponding to an annual four-factor alpha of -3.2% and 3.8%, respectively). In each tail (left and right), we assess the precision of the different estimators at two significance levels ($\gamma=0.05$ and 0.20).

<table>
<thead>
<tr>
<th>Fund Proportion</th>
<th>True</th>
<th>Estimator (90% interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-alpha funds ($\pi_0$)</td>
<td>75.0</td>
<td>75.1 (71.8,78.5)</td>
</tr>
<tr>
<td>Unskilled funds ($\pi_{-A}$)</td>
<td>23.0</td>
<td>22.9 (19.9,26.2)</td>
</tr>
<tr>
<td>Skilled funds ($\pi_{+A}$)</td>
<td>2.0</td>
<td>2.0 (0.3,3.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Significance level $\gamma = 0.05$</th>
<th>Significance level $\gamma = 0.20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Tail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significant funds $E\left(S_{-\gamma}\right)$</td>
<td>18.1</td>
<td>18.1 (16.4,19.6)</td>
</tr>
<tr>
<td>Unlucky funds $E\left(F_{-\gamma}^{-}\right)$</td>
<td>1.8</td>
<td>1.8 (1.8,1.9)</td>
</tr>
<tr>
<td>Unskilled funds $E\left(T_{-\gamma}\right)$</td>
<td>16.2</td>
<td>16.2 (14.6,17.7)</td>
</tr>
<tr>
<td>Right Tail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significant funds $E\left(S_{+\gamma}\right)$</td>
<td>3.6</td>
<td>3.6 (2.8,4.4)</td>
</tr>
<tr>
<td>Lucky funds $E\left(F_{+\gamma}^{+}\right)$</td>
<td>1.8</td>
<td>1.8 (1.8,1.9)</td>
</tr>
<tr>
<td>Skilled funds $E\left(T_{+\gamma}\right)$</td>
<td>1.7</td>
<td>1.7 (0.9,2.5)</td>
</tr>
</tbody>
</table>
Table IX
Monte-Carlo Analysis under Cross-Sectional Dependence

We examine the average value and the 90%-confidence interval (in parentheses) of the different estimators based on 1,000 replications. For each replication, we generate monthly fund returns for 1,500 funds and 384 periods using the four-factor model (market, size, book-to-market, and momentum factors). We assume that funds are cross-sectionally correlated. In order to closely reproduce the relations across all funds, we use the empirical covariance matrix of the fund residuals as the true covariance matrix. The true parameter values for the proportions of zero-alpha, unskilled, and skilled funds ($\pi_0$, $\pi_A^-$, and $\pi_A^+$) are set to 75%, 23%, and 2%. The mean of the $t$-statistic of the unskilled and skilled funds are set to -2.5 and 3 (corresponding to an annual four-factor alpha of -3.2% and 3.8%, respectively). In each tail (left and right), we assess the precision of the different estimators at two significance levels ($\gamma=0.05$ and 0.20).

<table>
<thead>
<tr>
<th>Fund Proportion</th>
<th>True</th>
<th>Estimator (90% interval)</th>
<th>Significance level $\gamma = 0.05$</th>
<th>Significance level $\gamma = 0.20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-alpha funds ($\pi_0$)</td>
<td>75.0</td>
<td>75.2 (69.9,80.4)</td>
<td>18.1 (15.5,20.4)</td>
<td>27.9 (24.5,32.0)</td>
</tr>
<tr>
<td>Unskilled funds ($\pi_A^-$)</td>
<td>23.0</td>
<td>22.8 (17.3,28.6)</td>
<td>1.8 (1.6,2.1)</td>
<td>7.5 (6.6,8.5)</td>
</tr>
<tr>
<td>Skilled funds ($\pi_A^+$)</td>
<td>2.0</td>
<td>1.9 (0.0,5.8)</td>
<td>16.2 (13.5,19.0)</td>
<td>20.4 (16.5,24.2)</td>
</tr>
</tbody>
</table>

**Left Tail**

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>Estimator (90% interval)</th>
<th>True</th>
<th>Estimator (90% interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significant funds $E(S^-_\gamma)$</td>
<td>1.8</td>
<td>1.8 (1.6,2.1)</td>
<td>7.5</td>
<td>7.6 (6.6,8.5)</td>
</tr>
<tr>
<td>Unlucky funds $E(F^-_\gamma)$</td>
<td>16.2</td>
<td>16.2 (13.5,19.0)</td>
<td>20.4</td>
<td>20.4 (16.5,24.2)</td>
</tr>
</tbody>
</table>

**Right Tail**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>True</th>
<th>Estimator (90% interval)</th>
<th>True</th>
<th>Estimator (90% interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significant funds $E(S^+_\gamma)$</td>
<td>3.5</td>
<td>3.6 (2.4,5.2)</td>
<td>9.4</td>
<td>9.4 (6.8,12.4)</td>
<td></td>
</tr>
<tr>
<td>Lucky funds $E(F^+_\gamma)$</td>
<td>1.8</td>
<td>1.8 (1.6,2.1)</td>
<td>7.5</td>
<td>7.6 (6.6,8.5)</td>
<td></td>
</tr>
<tr>
<td>Skilled funds $E(T^+_\gamma)$</td>
<td>1.7</td>
<td>1.7 (0.5,3.8)</td>
<td>1.9</td>
<td>1.9 (0.2,5.5)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1
Outcome of the Multiple Performance Test

Panel A shows the distribution of the fund t-stat across the three fund categories (zero-alpha, unskilled, and skilled funds). We set the t-stat mean equal to -2.5 for the unskilled funds and +3 for the skilled ones. Panel B displays the cross-sectional t-stat distribution obtained after plotting the t-stat for all funds in the population. 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 ($\pi_0$, $\pi_{-A}$, and $\pi_{+A}$). In this example, we set $\pi_0 = 75\%$, $\pi_{-A} = 23\%$, and $\pi_{+A} = 2\%$. 

- The proportion of significant funds is equal to 20.3%
- But are all these funds truly unskilled?

- The proportion of significant funds is equal to 5.4%
- But are all these funds truly skilled?

Panel A: T-Stat Distribution

Panel B: Cross-Sectional T-Stat Distribution
Figure 2

Histogram of Fund p-values

This figure represents the p-value histogram of 2,076 funds (as in our database). For each fund, we draw its t-stat from one of the distributions in Figure 1 (Panel A) in proportion to the relative importance of zero-alpha, unskilled, and skilled funds in the population ($\pi_0$, $\pi_A^-$, and $\pi_A^+$). In this example, we set $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$. Then, we compute the two-sided p-values of each fund from its respective t-stat, and plot all of them in the histogram.
This figure examines the bias of different estimators produced by the three approaches as a function of the proportion of zero-alpha funds, $\pi_0$. We examine the estimators of the proportions of unlucky, lucky, unskilled, and skilled funds in Panel A, B, C, and D, respectively. The "no luck" approach assumes that $\pi_0=0$, the "full luck" approach assumes that $\pi_0=1$, while "our approach" estimates $\pi_0$ 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 $t$-stat for each fund $i$ ($i=1,...,2,076$) from one of the distributions in Figure 1 (Panel A) according to the weights $\pi_0$, $\pi_{A}^{-}$, and $\pi_{A}^{+}$, and compute the different estimators at the significance level $\gamma = 0.20$. For each $\pi_0$, the ratio $\pi_{A}^{-}$ over $\pi_{A}^{+}$ is held fixed to $11.5$ ($0.23/0.02$) as in Figure 1.
Figure 4

Evolution of Mutual Fund Performance over Time

Panel A plots the evolution of the estimated proportions of unskilled and skilled funds ($\hat{\pi}^-$ and $\hat{\pi}^+$) between 1989 and 2006. At the end of each year, we measure $\hat{\pi}^-$ and $\hat{\pi}^+$ using the entire fund return history up to that point. The initial estimates at the end of 1989 cover the period 1975-1989, while the last ones in 2006 use the period 1975-2006. The performance of each fund is measured with the unconditional four-factor model. Panel B displays the increase over the previous year in the number of funds used to compute $\hat{\pi}^-$ and $\hat{\pi}^+$ over time.

PANEL A: PROPORTIONS OF UNSKILLED AND SKILLED FUNDS

PANEL B: GROWTH IN THE FUND POPULATION
Performance of the Portfolios FDR 10% over Time

The graph plots the evolution of the estimated annual four-factor alpha of the portfolios FDR 10%. At the end of each year from 1989 to 2006, we estimate the portfolio alpha using the entire fund return history up to that point. The initial estimates cover the period 1980-1989 (the first five years are used for the initial portfolio formation), while the last ones use the period 1980-2006. For comparison purposes, we also show the performance of top decile portfolios formed according to the t-stat estimated over the prior one and three years, respectively.