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 Svizzera 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 EC^2 (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 correctly 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. While the unskilled funds reside throughout the left tail, skilled ones are located only in the extreme right tail. 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 much 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.

Introduction

Investors and academic researchers have long searched for outperforming portfolio managers. Several researchers document that the average fund alpha (net of expenses and trading costs) is negative (e.g., Jensen (1968), Lehman and Modest (1987), Elton et al (1993), and Carhart (1997)). More recent papers show that at least some funds generate positive alphas. For instance, Kosowski, Timmermann, Wermers, and White (KTWW (2006)) document outperformance of individual funds with extreme positive alphas using a bootstrap technique, while Baks, Metrick, and Wachter (2001) and Pastor and Stambaugh (2002b) illustrate the benefits of including actively-managed funds in an investor's portfolio 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 funds possess truly positive or negative, and where these funds are located in the cross-sectional alpha distribution. From an investment perspective, precisely locating skilled funds maximizes our chances of achieving persistent outperformance through our selection of managers.

A seemingly reasonable way to estimate the prevalence of skilled fund managers is to simply count the number of funds having significant estimated alphas, according to a chosen asset pricing model. However, it is important to note that this procedure is actually a multiple (hypothesis) test, because we simultaneously examine the performance of all funds in the population (instead of a just one fund).[1] As such, 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). Without properly accounting for luck in this multiple fund setting, it is impossible to properly measure the number of truly skilled funds.

What makes modeling luck so different in this multiple test, compared with the usual single (hypothesis) test of whether the alpha of a given fund equals zero? Consider such a single test for the alpha of a given fund, α_j, where we set the significance level γ (i.e., the size of the test) equal to 5%.[2] The probability that fund j is lucky (i.e. that

[1] This multiple test should not be confused with the joint test $\alpha_1 = ... = \alpha_M = 0$ employed by some past research (e.g., Grinblatt and Titman (1989, 1993)). A joint test addresses only whether at least one fund has an non-zero alpha among M funds, but cannot determine the number of funds with truly negative and positive alphas.

[2] This fund can be a specific fund, such as the Magellan fund, or the fund with the highest estimated alpha, as in KTWW (2006). Alternatively, it can also be a portfolio of funds, such as the equally-weighted

its estimated alpha is significant, while its true alpha, α_j, equals zero), equals 5% by definition. If we use 1,000 different samples to compute 1,000 estimated alphas for this fund, we expect it to exhibit (good or bad) luck 50 times (1,000 times 5%). What if, instead of testing the same fund 1,000 times, we test 1,000 different funds? Would we also expect to find 50 lucky funds? If all 1,000 funds truly have zero alphas ($\alpha_i = 0$), the answer is yes. But suppose that only half of them are zero-alpha funds. In this case, we would expect to find 25 lucky funds (500 times 5%). Therefore, setting the expected number of lucky funds to 50 is valid only with the strong prior assumption that all 1,000 funds have $\alpha_i = 0$. The problem, of course, is that it is impossible to know, a priori, whether the assumption that $\alpha_i = 0$ for all funds in the population is a reasonable one.

This paper provides the first approach to properly measuring the prevalence of lucky funds, or "false discoveries," in such a multiple fund setting.[3] By doing so, we conduct a new performance analysis, which correctly identifies (1) the number (proportion) of truly unskilled and skilled funds in the fund population, (2) their respective locations in the left and right tails of the cross-sectional alpha distribution.[4] A main virtue of our approach is its simplicity–after estimating individual fund alphas, the only remaining parameter that we must estimate is the proportion, π_0, of zero-alpha funds in the population.[5] Rather than imposing any assumption on π_0, we estimate this parameter through a straightforward manipulation of the *p*-values of the individual fund estimated alphas, without any further econometric tests. Using the estimate $\widehat{\pi}_0$, we further develop a new methodology to determine the number of lucky funds in the left and right tails of the alpha distribution, respectively. A second advantage of our approach is its accuracy. Using a simple Monte-Carlo experiment, we show our approach precisely estimates the number of unskilled and skilled funds at any chosen significance level, γ. On the contrary, the previous literature which assumes, a priori, that the proportion of zero-alpha funds is either equal to zero or to one ($\pi_0 = 0$ or 1) produces large errors in the much more likely situation where a fraction of fund managers are skilled.

portfolio of all funds which might be used to test the average alpha in the mutual fund industry.

[3] When someone finds a fund with a significant estimated alpha, he thinks he has made a discovery. However, if this fund has, in reality, an alpha equal to zero, so it turns out to be a false discovery.

[4] Since we compute alphas net of trading costs and fund expenses, our notion of "skill" is manager talent in generating abnormal performance through stock selection sufficient to generate a positive alpha, net of these costs. "Unskilled funds" are those with stock selection skills insufficient to compensate for trading costs and expenses (i.e., negative net-of-cost alphas).

[5] "Zero-alpha funds" are those having managers with stockpicking skills *just* sufficient to recover trading costs and expenses.

A further important advantage to our approach to multiple testing, as shown in extensive 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 existing at any time between 1975 and 2006. When we examine funds with high alpha t-statistics (which we use due to its superior statistical properties, compared to alpha), we find that a large proportion of these right-tail funds are substantial. Specifically, in the extreme right tail (all funds having p-values lower than $\gamma = 0.05$), about 86% of funds are lucky, while only 14% are truly skilled. Investigating a larger portion of the tail (by increasing γ to 0.20), we find that the proportion of lucky funds increases to about 92%. This evidence indicates that the few skilled mutual funds in the U.S. domestic equity universe are located in the extreme (very high t-statistic) right tail. Further, our results highlight the importance of properly measuring the proportion of zero-alpha funds, π_0. For instance, using the strong assumption that $\pi_0 = 0$ would lead to a much higher estimate of the proportion of truly skilled funds. Equivalently, assuming, a priori, that $\pi_0 = 1$ gives a result where the estimated proportion of lucky funds is higher than the proportion of significant funds, which clearly cannot be the case. When we examine the left tail of the alpha t-statistic distribution, we find quite different results–performance in this left tail is only marginally impacted by luck. In the extreme left tail, 84% of funds with significantly negative t-statistics are truly unskilled (negative net-of-cost alpha) funds. This proportion remains fairly constant as we move toward the center of the t-statistic distribution, implying that unskilled funds are spread throughout the left tail, and are not concentrated in one specific underperformance region.

When we analyze subperiods, we expose an interesting dynamic in the U.S. fund industry. Specifically, we find a much higher proportion of truly skilled funds prior to 1990, compared to the full sample–accompanied by a much larger proportion of truly unskilled (negative net-of-cost alpha) funds after 1990. This finding indicates that the large growth in the number of actively managed mutual funds has been detrimental to the average investor, and is consistent with Cremers and Petajisto (2006), who show that "closet indexing" increased significantly during the 1990s. We repeat our analysis for investment-objective subgroups, and find that the proportion of unskilled funds, as well as their location in the left tail are similar across the Growth, Aggressive Growth, and Growth & Income subgroups. However, while skilled funds are similarly located in the extreme right tail within all three groups, as they are for the entire sample, their relative proportions vary substantially across the subgroups. Most notably, a larger proportion of Aggressive Growth fund managers are truly skilled (about 5%), while no skilled managers exist among Growth & Income funds. This result has a practical implication: investors looking for truly positive alphas should concentrate on Aggressive Growth funds, and avoid Growth & Income funds–all significant alpha Growth & Income funds are merely lucky.

A further application of our approach allows for a more informative performance analysis of the U.S. domestic-equity mutual fund industry. We find that the proportion of zero-alpha funds, $\widehat{\pi}_0$, equals 72.8%; the managers of these funds possess skills just sufficient to recover costs (including expenses and trading costs). This high proportion is consistent with the equilibrium model of Berk and Green (2004), in which managers are able to extract all rents generated by their skills. Further, the estimated proportion of unskilled funds, $\widehat{\pi}_A^-$, equals 26.6%, while the proportion of skilled funds, $\widehat{\pi}_A^+$, is only 0.6%. Thus, our findings indicate that the negative average alpha documented in previous studies does not reflect the performance of the whole industry, but is generated by about 1/4 of the population–the truly unskilled funds. Our above-mentioned performance analysis shows that the proportion of skilled funds in the population is low (0.6%), suggesting that finding such funds is extremely difficult. However, since skilled funds are located in the extreme right tail, they can be, to a large extent, separated from the zero-alpha funds. We illustrate this point by implementing a persistence analysis that forms portfolios of funds while explicitly controlling the proportion of lucky funds. These portfolios are formed at the beginning of each year from 1980 to January 2006, using only available data at each point in time. Our results show that selecting funds in the extreme right tail produces a persistent performance. Specifically, the annual equal-weighted net return alphas range from 1.15 to 1.45% per year, and are, in all cases, highly statistically significant. A key to this positive performance is the flexibility of our selection procedure, which automatically changes the proportion of funds included in the "winners" portfolio in response to changes in the distribution of skilled funds over time. For instance, during years where our model indicates that true manager skills are relatively rare, we focus on the few funds in the extreme right tail of the alpha distribution.[6]

Our final tests examine the skills of fund managers before expenses (but after trading costs) are subtracted. Specifically, fund managers may be able to pick stocks well enough to cover their trading costs, but they usually do not control the level of expenses and fees–perhaps fund management companies systematically overcharge for their services. We find evidence supportive of this–on a pre-expense basis, we find a very large increase in skilled funds–9.6%, compared to our prior finding of 0.6% after expenses. This finding suggests that there is further room for competition to reduce expenses in the U.S. active mutual fund industry.

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 pro-

[6] By contrast, the top alpha portfolios commonly used in the persistence literature (e.g., Carhart (1997)) use a constant proportion of funds over time and, thus, generate lower alphas.

vide 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. 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 generating high administrative costs)–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 only trading costs) and revisit some of our basic tests to be described.

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, $\widehat{\alpha}_i$, of the individual funds ($i = 1, ..., M$)? First, we use the t-statistic, $\widehat{t}_i = \widehat{\alpha}_i / \widehat{\sigma}_{\widehat{\alpha}_i}$, where $\widehat{\sigma}_{\widehat{\alpha}_i}$ is the estimated standard deviation of $\widehat{\alpha}_i$.[7] Second, after setting an appropriate significance level, γ (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 $\widehat{t}_i$ (in absolute value) lies above the thresholds implied by γ (denoted by t_γ^- and t_γ^+), 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:

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

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. 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).[8] Then, the t distribution shown in Panel

[7] Kosowski, Timmermann, Wermers, and White (2006) show that performance measured with the t–stat 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.

[8] For simplicity of illustration, Figure 1 plots the distribution of (hypothetical) t-statistics as being normal. In our empirical section to follow, we allow account for potential departures from asymptotic

B is the cross-section that (hypothetically) would be observed by a researcher–it is a mixture of the three distributions in Panel A, where the probability of observing each distribution is equal to the proportions 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, γ 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.[9] 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 $\widehat{t}_i$'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).[10]

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 non-zero estimated alphas (i.e., their true alphas are zero).

A.2 Measuring Luck

We next describe our approach to estimating the proportion of funds in the tails of the alpha t distribution purely by (good or bad) luck. At a given significance level, γ, the probability that a zero-alpha fund (as defined in the last section) exhibits good luck (or bad luck) equals $\gamma/2$, as shown in the shaded regions of Panel A of Figure 1. If

normality (in the distribution of t-statistics of our actual sample of funds) through the use of a bootstrap approach.

[9] 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, π_0 and π_A^+, gives 5.4%.

[10] Note that we have not considered the possibility that skilled funds could be *very* unlucky, and exhibit a negative and significant t-stat. Our 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.

the proportion of zero-alpha funds in the population is π_0, the expected proportion of "lucky funds" (zero-alpha funds with positive and significant t-statistics) is computed as

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

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

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

To measure the expected proportion of unlucky funds in the left tail, denoted by $E(F_\gamma^-)$, we use 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

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

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

This flexibility in choosing γ provides us with important opportunities for insights into the merits of active fund management. First, by choosing a large γ, we can estimate the proportions of unskilled and skilled funds, as defined in the last section, in a larger portion of the left and right tails of the t-distribution, respectively–thus, giving us an appreciation of the prevalence of skilled funds in the entire population. That is, as we increase γ, $E(T_\gamma^-)$ and $E(T_\gamma^+)$ converge to π_A^- and π_A^+. Alternatively, by reducing γ, we can determine the precise location of unskilled or skilled funds in the tails of the t-distribution. For instance, choosing a very low γ (i.e., very large t_γ^- and t_γ^+, in absolute value) allows us to determine whether extreme tail funds are simply lucky (or unlucky), or truly 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, π_0, of zero-alpha funds in the population. Here, we turn to a novel estimation approach developed by Storey (2002). This 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 distribution under the null hypothesis, $H_{0,i} : \alpha_i = 0$, which implies that their p-values are uniformly distributed over the interval $[0,1]$.[11] On the contrary, the 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 π_0 without knowing the exact distribution of the p-values of the unskilled and skilled funds.

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 our population of funds, $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$. From these t's, we compute two-sided p-values, $\widehat{p}_i$, for each of the 2,076 funds, then plot them in Figure 2. The dark grey area near zero represents the proportion of p-values in the population that correspond to the unskilled and skilled funds ($\pi_A^- + \pi_A^+ =25\%$). It is important to note that the area of the rectangle between the horizontal black line and the abscissa provides us with a precise estimate of the proportion of zero-alpha funds, $\widehat{\pi}_0 =75\%$, since zero-alpha funds have uniformly distributed p-values and since unskilled and skilled funds tend to have very small p-values. To explain further, taking a sufficiently high threshold, λ^* (e.g., $\lambda^* = 0.6$), implies that virtually all of the p-values that exceed λ^* come from zero-alpha funds. Thus, to estimate π_0, we first estimate the area covered by the four light grey bars on the right of λ^*, $\widehat{W}(\lambda^*)/M$ (where $\widehat{W}(\lambda^*)$ denotes the number of funds having p-values exceeding λ^*), then simply extrapolate this area over the entire interval $[0,1]$ by multiplying it by $1/(1-\lambda^*) > 0$ (if $\lambda^* = 0.6$, the area is muliplied by $1/(1-0.6) = 2.5$):[12]

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

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

[12] 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 $[0,1]$.

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

Please insert Figure 2 here

Substituting the estimate $\widehat{\pi}_0$ in Equations (2), (3), and replacing $E(S_\gamma^+)$ with the observed proportion of significant funds in the right tail, $\widehat{S}_\gamma^+$, we can easily estimate $E(F_\gamma^+)$, and $E(T_\gamma^+)$ corresponding to any chosen critical value, γ. The same approach can be used in the left tail by replacing $E(S_\gamma^-)$ with the observed proportion of significant funds in the left tail, $\widehat{S}_\gamma^-$. This implies the following estimates of the proportions of unlucky and lucky funds:

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

Using Equation (6), the estimated proportion of unskilled and skilled funds (at the chosen critical value, γ), $\widehat{T}_\gamma^+$ and $\widehat{T}_\gamma^-$ are respectively equal to

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

Finally, we estimate the proportions of unskilled and skilled funds in the population, π_A^- and π_A^+, with the following quantities:

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

where γ^* is a sufficiently high critical value–we choose γ^* with a simple data-driven method explained in the appendix.[13]

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 "no luck" approach estimates these proportions with the observed proportion of significant funds (Ferson and Schadt (1996), Ferson and Qian (2004)). Since there is no adjustment for luck, this approach implicitly assumes that there are no zero-alpha funds in the population: $\pi_0 = 0$. The "full luck" approach initially proposed by Jensen (1968) infers, from the number of funds in each

[13]We assume that the density of the t-statistic of the unskilled and skilled funds decreases monotonically as γ 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.

tail, an estimate of the proportion of false discoveries equal to $\gamma/2$.[14] In light of Equation (2), this approach assumes there are only zero-alpha funds: $\pi_0 = 1$.

What are the errors committed by assuming that π_0 equals 0 or 1 when it is not true? Using our simple example in Figure 1, we examine the estimator bias produced by these two approaches as well as ours across the possible values for π_0 ($\pi_0 \in [0,1]$). Our procedure consists of four steps. First, we draw for each fund i ($i = 1, ..., 2,076$) its t-stat from one of the distributions in Panel A of Figure 1 according to the weights π_0, π_A^-, and π_A^+. For each π_0, the values for π_A^- and π_A^+ are determined such that their ratio remains unchanged: $\pi_A^-/\pi_A^+ = 0.23/0.02 = 11.5$. Second, we estimate the proportion of unlucky, lucky, unskilled, and skilled funds for the three approaches ("no luck", "full luck", and our approach). We set $\gamma = 0.20$ so as to examine a large portion of the tails of the cross-sectional t-stat distribution (other values for γ provide the same insights). 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 has a solid line with dot markers.

Please insert Figure 3 here

Panel A compares the estimators of the expected proportion of unlucky funds. The true population value, $E(F_\gamma^-)$, is an increasing function of π_0, as indicated in Equation (2). While the average value of our estimator precisely tracks $E(F_\gamma^-)$, this is not the case for the two other approaches. By assuming that $\pi_0 = 0$, the "no luck" approach always underestimates $E(F_\gamma^-)$ when $\pi_0 > 0$. On the contrary, the "full luck" approach, which assumes that $\pi_0 = 1$, overestimates $E(F_\gamma^-)$ as long as $\pi_0 < 1$. To illustrate the extent of the bias, consider the case where $\pi_0 = 75\%$. While the "no luck" approach underestimates $E(F_\gamma^-)$ by 100% (0% instead of 7%), the "full luck" approach overestimates $E(F_\gamma^-)$ by 35% (10% instead of 7.4%). The analysis for the lucky funds $E(F_\gamma^+)$ 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 are shown in Panel C and D, respectively. Since an increase in π_0 reduces both π_A^- and π_A^+, we

[14] Jensen (1968) summarizes the "full luck" approach as follows: "...if all the funds had a true α 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)).

observe that the true population values, $E(T_\gamma^-)$ and $E(T_\gamma^+)$, are negatively related to π_0. In both Panels, we observe that our 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 inconsistent patterns. First, the "no luck" approach wrongly infers that as π_0 rises, the proportion of skilled funds increases. Second, while $E(T_\gamma^+)$ cannot be inferior to zero, the values obtained by the "full luck" approach turn out to be negative.

In addition to the bias properties of our estimators, their variability is also low because we benefit from a large cross-section of funds ($M = 2,076$). To understand this fact, consider our main estimator $\widehat{\pi}_0$ (the same arguments apply to the other estimators). Since $\widehat{\pi}_0$ is a proportion estimator, it can be written in the form of a simple average (Davidson and MacKinnon (2004), p. 147): $\widehat{\pi}_0 = 1/M \sum_{i=1}^{M} x_i$, where x_i is the indicator function taking the value $(1 - \lambda^*)^{-1}$ if $\widehat{p}_i > \lambda^*$, and 0 otherwise. With independent p-values, its standard deviation, σ_{π_0}, is simply equal to $\sigma_x/\sqrt{M}$, implying that large M drives σ_{π_0} to zero.[15] For instance, with $\lambda^* = 0.6$ and $\pi_0 = 75\%$, we find that σ_{π_0} is as low as 2.5% (30 times lower than π_0).[16] In the appendix, we provide further evidence of the remarkable accuracy of our estimators based on Monte-Carlo simulations.

C Estimation under Cross-Sectional Dependence among Funds

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. For instance, the joint test on fund alphas ($\alpha_1 = ... = \alpha_M = 0$) proposed by Grinblatt and Titman (1989, 1993) becomes quickly intractable as M rises, since it requires the estimation and inversion of the $M \times M$ residual covariance matrix. In a bayesian framework, Jones and Shanken (2005) show that performance measure-

[15] The precision of π_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.

[16] We have $P_{\lambda^*} = prob(\widehat{p}_i > \lambda^*) = 0.30$ (i.e., the rectangle area delimited by the horizontal black line and the vertical line at $\lambda^* = 0.6$ in Figure 2). Since $(1 - \lambda^*)\, x_i$ follows a binomial distribution with probability P_{λ^*} of success, we have $\sigma_x = (1 - \lambda^*)^{-1} \left(P_{\lambda^*}\,(1 - P_{\lambda^*})\right)^{\frac{1}{2}} = 1.14$, and $\sigma_{\pi_0} = \sigma_x/\sqrt{M} = 2.5$.

ment requires intensive numerical methods when dependences among the fund prior alphas are properly accounted for. KTWW (2006) propose a bootstrap approach to test the performance of a fund located at a particular quantile of the ranking based on the estimated alphas of all funds (e.g. the top fund, the fund at the 10 percentile point...). Since this test depends on the joint distribution of all fund estimated alphas, cross-correlated fund residuals must be bootstrapped simultaneously.[17] This approach is difficult because many funds have non-synchronous data.

An important advantage of our approach is that we can estimate the p-value of each fund without any information regarding the dependence structure across funds. Therefore, we avoid the problems caused by non-synchronous data. 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). However, high cross-sectional dependence could 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 λ^*, we wrongly estimate $\widehat{\pi}_0 = 0$.

Do we expect high levels of dependence in our fund population? Surely not, for two reasons. First, the dependence across p-values is only a function of the dependence across fund residuals[18] (rather than total returns). But if the asset-pricing model used to compute performance correctly captures the common sources of risk across funds, these residual cross-correlations are low. This is exactly what we find in our sample, since the average monthly pairwise residual correlation only equals 0.08 based on the four-factor model (market, size, value, and momentum factors). Second, funds with non-synchronous data cannot have highly correlated p-values by construction. We find that not only 15% of the fund pairs do not have any single observations in common, but also that, on average, only 55% of the return obervations on the fund pairs is overlapping. As a result, it is likely that the cross-sectional dependence is sufficiently low in order to produce consistent estimators (i.e., mutual fund residuals satisfy the ergodicity conditions discussed in Storey, Talyor, and Siegmund (2004)).

[17] Consider the alpha of the top fund denoted by α_i^{top}. Since the null and the alternative hypotheses are defined as: $H_0 : \alpha_i^{top} = \max_{i=1,...,M} \{\alpha_i\} \leqslant 0$ and $H_A : \alpha_i^{top} = \max_{i=1,...,M} \{\alpha_i\} > 0$, the distribution of the test statistic depends on the joint distribution of the fund estimated alphas.

[18] Two funds with higher than average residual values both face a simultaneous increase in their estimated alphas, and a simultaneous decrease in their p-values.

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 900 of them are all correlated with one another. In order to closely reproduce the complex relations across these 900 funds, we estimate the 900×900 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 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):

$$r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + \varepsilon_{i,t}, \quad (8)$$

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 ε_{it} 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)). This conditional model is similar to the one proposed by KTWW (2006), and is written as:

$$r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + B' \left(z_{t-1} \cdot r_{m,t} \right) + \varepsilon_{i,t}, \quad (9)$$

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 one is the one-month T-bill interest rate. The second one is the dividend yield of the CRSP value-weighted NYSE and AMEX stock index. The third one is the term spread proxied by the difference between the yield of a 10-year T-bond and the three-month T-bill interest rate. The fourth one 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, as well as their conditional versions. These results are summarized in Section 3.5.

KTWW (2006) find that the finite-sample distribution of the fund t-stat, $\widehat{t}_i$, 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 $\widehat{t}_i$ under the null hypothesis $\alpha_i = 0$, we use a semi-parametric bootstrap procedure, which draws with replacement from the regression estimated residuals $\{\widehat{\varepsilon}_{i,t}\}$.[19] 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 net return data provided by the Center for Research in Security Prices (CRSP) between January 1975 and December 2006 to estimate fund alphas. If a fund consists of different shareclasses, the fund net 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 only select funds with at least 60 monthly return observations in order to get precise estimates. These monthly returns need not be contiguous. In unreported results, we find that reducing the minimum length to 36 observations leaves our results unchanged, thus, our results

[19] 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 $T^{\frac{1}{5}}$ (proposed by Hall, Horowitz, and Jing (1995)), where T denotes the length of the fund return time-series. All of our results to be presented remain unchanged.

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 shown in Panel A and B, respectively.

Please insert Table I here

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 start our empirical analysis by measuring the impact of luck on long-term performance. To this end, we estimate the performance of each fund using his entire return history over the period 1975-2006. Panel A of Table II contains the estimates of the proportions of zero-alpha, unskilled, and skilled funds in the population ($\widehat{\pi}_0$, $\widehat{\pi}_A^-$, and $\widehat{\pi}_A^+$). To compute the estimator standard deviation in parentheses, we use the method proposed by Genovese and Wasserman (2004) described in the appendix. Since the results for the three investment categories (growth, aggressive-growth, growth and income) are fairly similar, they are not shown for sake of brevity. Among the 2,076 funds, 72.8% of them produce zero alphas. Measuring π_0 offers a unique way to test the conjecture of

Berk and Green (2004). They argue that, since managerial abilities are in scarce supply, fund managers are able to capture the economic rents stemming from their abilities. Therefore, the alphas of open-end funds in rational, competitive markets must be equal to zero. The empirical evidence suggests that the vast majority of the fund managers (72.8%) do have stockpicking abilities, but extract rents up to the point where the alpha is equal to zero. Given the simplifying assumptions of their rational model, this high proportion provides a strong empirical support to its main prediction.

Please insert Table II here

Further, we find that the estimated proportion of skilled funds is too low to be statistically distinguished from zero. This result may seem surprising in light of the sizable proportion of significant funds in the right tail documented by Ferson and Qian (2004). In Panel B, we count the proportion of significant funds in the left and right tails, $\left(\widehat{S}_{\gamma}^{-},\ \widehat{S}_{\gamma}^{+}\right)$, at four different significance levels ($\gamma = 0.05$, 0.10, 0.15, 0.20). As in Ferson and Qian (2004), there are many significant funds in the right tail. We find that $\widehat{S}_{\gamma}^{+}$ logically increases with γ, and reaches 7.8% of the total population at $\gamma = 0.20$ (162 funds). In addition, the average estimated annual alphas in these significant groups of funds seems rather high (between 4.8% and 6.5%). However, "significant" does not mean "skilled". To see this, we decompose these significant funds into lucky and skilled funds, $\left(\widehat{F}_{\gamma}^{+},\ \widehat{T}_{\gamma}^{+}\right)$. The results in Panel C reveal that the vast majority of the significant funds in the right tail are simply lucky funds producing zero alphas. This result is not surprising in light of the huge number of zero-alpha funds in the population (more than 1,500). We clearly expect to find a sizeable number of them located in the right tail of the cross-sectional t-stat distribution.

Although the model proposed by Berk and Green (2004) can describe the core performance of the mutual industry, Panel A shows that about 26.6% of the population (552 funds) truly produce negative alphas. This minority of funds is likely to be the driving force explaining the negative average estimated alpha documented in the literature (e.g., Jensen (1968), Elton et al. (1993), and Pastor and Stambaugh (2002a)). In Panel B, we observe that the poor-performing funds in the left tail do not have a short return history (12.6 years), so we may wonder why they have been operating for so long. As noted by Elton, Gruber, and Busse (2003), the informed investor cannot make use of arbitrage to eliminate unskilled funds. As a result, the latter can continue to exist (and even grow) if they attract a sufficient number of unsophisticated investors. Christoffersen and Musto (2002) go further by explaining that this category of investors should be charged higher fees. This is confirmed in our sample, since we find that the funds in the left tail exhibit higher average expense ratios than the ones in the right tail (Panel B). While many papers provide convincing evidence on the existence of unsophisticated investors[20], our results suggest that their importance has been underestimated.

Next, we examine the evolution of the estimated proportions of unskilled and skilled funds over time. At the end of each year from 1989 to 2006, we estimate $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$ using the entire return history up to that point. For instance, our initial estimates on December, 31 1989 cover the first ten years of the sample (1975-89), while the last ones on December, 31 2006 are based on the entire period 1975-2006 (i.e., these are the final estimates shown in Panel A of Table II). To be selected at the end of a given year, the fund must have at least 60 monthly return observations. The results in Panel A of Figure 4 show that the proportion of funds with non-zero alphas (equal to $\widehat{\pi}_A^- + \widehat{\pi}_A^+$) remains fairly constant over time. The main difference comes from the change in the relative importance of unskilled and skilled funds. Between 1989 and 2006, the proportion of skilled funds declines from 14.4% to 0.6%, while it rises from 9.2% to 26.6% for unskilled funds. This change is 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 increase over the previous year in the number of funds used to measure $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$. During the period 1996-2005, this number spikes up above 100 funds per year.[21] Interestingly, this coincides with the strong variations observed in $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$. The rapid expansion of the mutual fund industry may have improved market efficiency, or created capacity constraints, as many funds looks for the same bets (Naik, Ramadorai, and Stromqvist (2007) advance a similar argument for hedge funds). But there are other possible explanations to this performance decline. Skilled funds may have disappeared because talented managers move to the hedge fund industry to benefit from higher wages and lower reglementary constraints. Another interpretation is "closet indexing". Using their active share measure, Cremers and Petajisto (2006) show that the fraction of passive funds which claim to be active increases

[20] Capon, Fitzsimmons, and Prince (1996), find that individual investors have little knowledge of the mutual funds they buy. Elton, Gruber, and Busse (2003) show that although the returns of index funds are easily predictable, many investors continue to hold poorly performing funds. Elton, Gruber, and Blake (2007) document that many 401(k) plan offer inefficient choices of mutual funds.

[21] 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.

significantly since the early 1990's.

Please insert Figure 4 here

B Impact of Luck on Short-Term Performance

It is well documented that some funds generate positive short-term alphas (e.g., Grinblatt and Titman (1992), Elton, Gruber, and Blake (1996), KTWW (2006)). This prompts us to examine mutual fund performance over a shorter time horizon. To address this issue, we partition the entire period into 6 non-overlapping subperiods of five years.[22] For each of these subperiods, we select all funds with valid 60 monthly return observations and compute their respective p-values.[23] Pooling all 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 and 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-stat distribution. For each significance level ($\gamma = 0.05, 0.10, 0.15, 0.20$), we show the proportion of significant funds $\left(\widehat{S}_{\gamma}^{+}\right)$, and its decomposition into lucky and skilled funds ($\widehat{F}_{\gamma}^{+}$, $\widehat{T}_{\gamma}^{+}$). Each figure in parentheses denotes the estimator standard deviation. The leftmost columns repeat this analysis for the left tail.

The performance analysis in the right tail for all funds (Panel A) provides two interesting results. First, examination of $\widehat{T}_{\gamma}^{+}$ reveals that 2.4% of the population is skilled. It implies that a scarce number of funds are able to produce a short-term performance which does not persist over time. Second, these skilled funds are located in the extreme right tail. By setting a low γ of 0.10, we capture all skilled funds, as $\widehat{T}_{\gamma}^{+}$ reaches its maximum value. Therefore, as we proceed further towards the center of the distribution (by increasing γ to 0.15 and 0.20), all new significant funds turn out to be lucky rather than skilled. The main implication of this result is that low p-values may provide a useful information signal to detect the skilled funds.[24] Turning to the left tail, we ob-

[22] The full period is split as follows: 1977-1981, 1982-1986, 1987-1991, 1992-1996, 1997-2001, 2002-2006. Using overlapping subperiods may generate highly correlated p-values, as the p-values of the same fund estimated over overlapping subperiods are likley to be close to one another.

[23] Reducing the number of observations greatly increases the standard deviation of the estimated alphas (e.g., it rises by 33% with a three-year window), and makes the p-values of the non zero-alpha funds much larger. As a result, the latter are not distinguishable from the zero-alpha funds.

[24] Note that our analysis provides much more information about the proportion and the location of unskilled or skilled funds than KTWW (2006), since that paper merely assesses the performance of a single fund located at a given estimated alpha rank. Our tests provide information about the prevalence

serve that bad luck only marginally explains the massive presence of significant funds. For instance, in the extreme left tail (at $\gamma = 0.05$), the proportion of unskilled funds, $\widehat{T}_{\gamma}^{-}$, is approximately five time bigger than the proportion of unlucky funds, $\widehat{F}_{\gamma}^{-}$ (9.6% versus 1.8%). As a result, the short-term performance analysis coincides to its long-term counterpart as it detect the important presence of funds with truly negative alphas.

Please insert Table III here

Among the different investment categories, aggressive-growth funds contain the highest proportion of skilled funds (Panel C). At $\gamma = 0.05$, less than 40% of the significant funds are lucky (1.8/4.4), indicating that the skilled funds are mostly located in the extreme right tail. On the contrary, Panel D reveals that no growth and income fund produce positive alphas, and that a substantial proportion of them are unskilled. These results suggest that investors willing to generate short-term alphas should ignore this category.

C Performance Persistence

Our previous analysis reveals that only 2.4% of the funds produce positive alphas. Can we detect these skilled funds over time to form portfolios with persistent performance? In other terms, what is the economic significance of this short-term performance? Ideally, we would like to form a portfolio containing only these skilled funds. But since we do not know their identity, this approach cannot be implemented. To circumvent this difficulty, we build on one of our main empirical findings: the skilled funds are located in the extreme right tail. Therefore, by forming portfolios containing only the significant funds in the right tail, we can separate more accurately the skilled and zero-alpha funds. For instance, Panel A of Table III shows that at $\gamma = 0.05$, the proportion of skilled funds among the significant funds, $\widehat{T}_{\gamma}^{+}/\widehat{S}_{\gamma}^{+}$, is around 50%, which is much higher than the 2.4% found in the population.

To determine the proportion of funds included in these portfolios, we use a new measure called the False Discovery Rate (FDR). The FDR_{γ}^{+} at the significance level γ is defined

and the location of all such funds, simultaneously.

as the expected proportion of lucky funds among the significant funds in the right tail:[25]

$$FDR_{\gamma}^{+} = E\left(\frac{F_{\gamma}^{+}}{S_{\gamma}^{+}}\right). \tag{10}$$

The FDR_{γ}^{+} provides a simple portfolio formation rule.[26] When we set a low FDR^{+} target, we only allow for a small proportion of lucky funds in the portfolio. To achieve this objective, we set a low signifiance level γ so as to select only the few highly significant funds located in the extreme right tail (i.e., very few zero-alpha funds are lucky enough to be so highly significant). Conversely, by choosing a higher FDR^{+} target, we allow for a greater proportion of lucky funds in the portfolio. Since the selection becomes less restrictive, we can increase γ in order to include a larger number of funds. Increasing the FDR^{+} target has two impact on the portfolio. On one hand, it decreases its performance since the proportion of lucky funds is higher. On the other hand, it increases its diversification because more funds are selected. In our persistence test, we examine five FDR^{+} target levels: 10%, 30%, 50%, 70%, and 90%.

The construction of the portfolios proceeds as follows. At the end of each year, we estimate the p-values of all existing funds using the previous five-year subperiod. Using these p-values, we estimate the FDR_{γ}^{+} over a grid of significance levels (γ =0.01, 0.02,..., 0.60). Following Storey (2002) and Storey and Tibshirani (2003), we propose the following straightforward estimator of the FDR_{γ}^{+} :

$$\widehat{FDR}_{\gamma}^{+} = \frac{\widehat{F}_{\gamma}^{+}}{\widehat{S}_{\gamma}^{+}} = \frac{\widehat{\pi}_0 \cdot \gamma/2}{\widehat{S}_{\gamma}^{+}}. \tag{11}$$

For each FDR^{+} target level, we determine the significance level, γ^{P}, such that $\widehat{FDR}_{\gamma^{P}}^{+}$ is closest to this target. Then, all funds with p-values smaller than γ^{P} are included in an equally-weighted portfolio. This portfolio is held for one year until 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. Therefore, the portfolio is not subject to survivorship bias. The first portfolio formation date is December 31,

[25] Strictly speaking, the FDR^{+} is defined as $E\left(F_{\gamma}^{+}/S_{\gamma}^{+} \middle| S_{\gamma}^{+} > 0\right) \cdot prob(S_{\gamma}^{+} > 0)$. In light of our large fund database, this distinction becomes irrelevant as the probability $prob(S_{\gamma}^{+} > 0)$ tends to one. To illustrate it, suppose that all 2,076 funds have zero-alphas. If their p-values are independent, we have: $prob(S_{\gamma}^{+} = 0) = (1-\gamma/2)^{M} = 1 \cdot 10^{-23}$ at $\gamma = 0.05$.

[26] Our new measure, FDR_{γ}^{+}, is an extension of the traditional FDR_{γ} 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 $F_{\gamma} = F_{\gamma}^{+} + F_{\gamma}^{-}$, $S_{\gamma} = S_{\gamma}^{+} + S_{\gamma}^{-}$.

1979 (after the first five years), and the last is December 31, 2005.

Our empirical analysis starts with some descriptive statistics on the FDR level $(\widehat{FDR}^+_{\gamma^P})$ achieved by the five portfolios, and the proportion of funds in the population $(\widehat{S}^+_{\gamma^P})$ they include. In Panel A of Table IV, we show the average values of $\widehat{FDR}^+_{\gamma^P}$ and $\widehat{S}^+_{\gamma^P}$ over the 27 formation dates (from 1979 to 2005), as well as their respective distribution across the different intervals. First, we observe as expected that the achieved FDR increases with the FDR target assigned to the portfolios. However, the average $\widehat{FDR}^+_{\gamma^P}$ does not always closely match its target. For example, the first portfolio achieves an average of 41.5%, instead of the targeted 10%. The reason for this mismatch is simple. During some period, the proportion of skilled fund in the population is too low to achieve the desired FDR target.[27] Consistent with the intuition, we also observe a positive relation between the FDR target and the proportion of funds included in the portfolio. For instance the average proportion of funds included in the portfolio FDR 10% is equal to 3.0% and rises up to 33.7% for the portfolio FDR 90%.

In Panel B, we examine the performance of these five portfolios between January 1, 1980 and December 31, 2006. We compute the estimated annual alpha, $\widehat{\alpha}$, along with its bootstrap p-value, the annual residual standard deviation, $\widehat{\sigma}_\varepsilon$, the information ratio IR$= \widehat{\alpha}/\widehat{\sigma}_\varepsilon$, the factor loadings, as well as the annual excess mean and standard deviation. The results reveal to our FDR portfolios are able to detect funds with short-term skills. The portfolios FDR 10% and 30% produce an estimated alphas (net of expenses) of 1.45% and 1.15% per year, respectively. Their small p-values indicate that this performance is significant at the 5% level. As the FDR target rises up to 90%, the number of funds in the portfolio increase, which improves diversification ($\widehat{\sigma}_\varepsilon$ falls from 4.0% to 2.7%). However, we observe a sharp decrease in the alpha (from 1.45% to 0.39%), reflecting the important proportion of lucky funds contained in the porfolio.

Please insert Table IV here

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 confirm the short-term nature of this positive performance. After 1 year,

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

less than 50% of the funds are still included in the portfolios FDR 10% and 30%, and after 3 years, these percentages fall to 3.4% and 5.1%, respectively.

Next, 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 progressively decline during the 90's. As the proportion π_A^+ of skilled funds falls, the FDR approach greatly reduces $\widehat{S}^+_{\gamma^P}$ (from 10.7% in 1990 to 1.7% in 2000) so as to select only the highly significant funds in the extreme right tail. However, this change is not sufficient to prevent the performance from falling.

The previous literature proposes an alternative approach to examine persistence (e.g., Hendricks, Patel, and Zeckhauser (1993), Elton, Gruber, and Blake (1996), Carhart (1997)). It consists in classifying funds into decile or octile portfolios based on their past performance (past returns, estimated alpha or t-stat) over a previous ranking period (one or three years). This approach differs from the FDR method in two aspects. First, it ranks funds based on their estimated performance without testing whether the latter is significantly different from zero. As a result, the performance signal used to form portfolios is likely to be noisier. Second, it uses a shorter ranking period, which can make it more robust to short term trends in the performance of the mutual fund industry. To compare these approaches, Figure 5 displays the performance evolution of two top decile portfolios based on the t-stat measured over the previous one and three years, respectively.[28] In the early part of the period, the FDR approach performs much better, consistent with the idea that it detect more precisely the presence of skilled funds. However, the performance of the decile portfolios are less affected by the strong reduction in π_A^+ observed in Figure 4. In addition, the 1 year-decile portfolio is able to take advantage of the rally observed in the late 90's.[29] Therefore, we find that the superior performance of the FDR portfolio is tightly linked to the prevalence of skilled funds in the population.

[28] We use the t-stat to be consistent with the rest of the paper, but the results obtained with the estimated alpha are qualitatively similar.

[29] To examine whether this performance is due to the exposure to the high tech industry, we regress the 1-year decile portfolio on an index including the telecom, computer, and electric equipment industries (downloaded from Ken French's website). We find that its beta in 1998-1999 is 0.21 higher than the FDR portfolio beta (0.73-0.52). Since the four-factor alpha of the high tech index is huge during this period (23% per year!), small differences in loadings have a large impact on the portfolio performance.

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 deduced. Measuring performance on a pre-expense basis allows to disentangle the manager's stock picking skills from the fund's expense policy. To address this issue, we add the monthly expenses of each fund to its return time-series, and revisit the long-term performance of the mutual fund industry.[30]

Panel A of Table V contains the estimated proportions of zero-alpha, unskilled, and skilled funds in the population ($\widehat{\pi}_0$, $\widehat{\pi}_A^-$, and $\widehat{\pi}_A^+$) on a pre-expense basis. Comparing these estimates with those shown in Table II, we observe a striking reduction in the proportion of unskilled funds from 26.6% to 4.5%. This result indicates that only a small fraction of funds have managers with negative stock picking skills. Indeed, funds produce negative net alphas because they charge excessive fees in relation to the selection abilities of their managers. In Panel B, we also find that the average expense ratio across funds in the left tail is lower when performance is measured before expenses (1.4% versus 1.6% per year). This is consistent with the interpretation that that poor net-expense performance is mainly due to high expenses charged to unsophisticated investors.

Please insert Table VI here

Turning to the analysis of the pre-expense skilled funds, we find that 10% of the fund managers do have stock picking skills (Panel A). Consistent with Berk and Green (2004), the rents stemming from their skills are extracted through the fees and expenses, driving the proportion of net-expense skilled funds to zero. Since 72.8% of funds produce zero net-expense alpha, we would expect to find more pre-expense skilled funds. The reason for this result is the small expense ratios observed for funds in the center of the cross-sectional net-expense t-stat distribution. Adding these expenses leads to a marginal increase in the t-stat mean, making the power of the tests rather low.[31] As a result, our

[30] We discard funds which do not have a minimum of 60 pre-expense return observations over the period 1975-2008. This leads to a small reduction in our sample from 2,076 to 1,836 funds.

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

estimate $\widehat{\pi}_A^+ = 10\%$ is likely to be a conservative estimate of the proportion of managers with stock picking skills.

Finally, we examine the evolution of the pre-expense proportions over time. We find that $\widehat{\pi}_A^+$ decreases by 17.5% (27.5%-10%) between 1996 and 2006. This implies that the reduction in stock-picking skills is the main driver of the decline in net-expense skills (see Figure 4), as opposed to an increase in expenses. On the contrary, $\widehat{\pi}_A^-$ remains equal to zero until the end of 2003, and cannot expain the important increase in the proportion of net-expense unskilled funds from 1996 on. Therefore, this positive trend is likely to be due to excessive 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, ($\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$) obviously depends on the choice of the asset pricing model. To examine the sensitivity of the results, we repeat the long-term performance analysis using the (unconditional) CAPM and Fama-French models. Based on the CAPM, we find that $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$ are equal to 14.3% and 8.6%, and lead to a remarkable performance improvement over the four-factor model. This result is due to the omission of the size, value, and momentum factors. The average loadings (among the 2,076 funds) on these factors amount to 0.20, 0.05, and 0.02, respectively. Multiplying them with the positive premia of these factors (3.7%, 5.4%, and 9.4% per year over 1975-2006), the average annual CAPM alpha is boosted by 1.2% per year. For the same reason, we expect that these CAPM-skilled funds have high loadings on the omitted factors. This is confirmed in Panel A of Table VI: the funds located in the right tail (according to the CAPM) have substantial loadings on the size and the value factors. By forming portfolios including these CAPM-skilled funds, we could generate a positive CAPM alpha, and thus increase the Sharpe ratio of the market porfolio. But this higher performance would only accrue to the size and value characteristics of the funds.

Please insert Table VI here

Turning to the analysis of the Fama-French model, we find that $\widehat{\pi}_A^-$ and $\widehat{\pi}_A^+$ amount to 25.0% and 1.7%, respectively. These proportions are very close to those obtained with the four-factor model, since only one factor is omitted. But if our previous arguments are

observing a negative pre-expense t-stat is equal to 18%!).

correct, the 1.1% difference in the proportion of skilled funds between the two models (1.7%-0.6%) should come from the momentum factor. This is confirmed in Panel B, since the average momentum loading of funds located in the right tail (according to the Fama-French model) is five times higher than average.

D.3 Bayesian Interpretation

Although we operate in a classical frequentist framework, our new FDR measure, FDR^+, has also a natural Bayesian interpretation.[32] To see this, we denote by H_i the random variable which takes the value of 0 if the fund i has a zero alpha, -1 if it is unskilled, and +1 if it is skilled. The prior probabilities for the three possible values (0, -1, +1) are given by the proportion of each category in the population, π_0, π_A^-, and π_A^+. The Bayesian version of our FDR^+ measure, denoted by fdr_γ^+, is defined as the posterior probability that the fund i has a zero alpha given that its t-stat, $\widehat{t}_i$, is positive and significant: $prob\left(H_i = 0 | \, \widehat{t}_i \in \Gamma_A^+(\gamma)\right)$, where $\Gamma_A^+(\gamma) = \left(t_\gamma^+, +\infty\right)$. Using Bayes theorem, we have:

$$fdr_\gamma^+ = \frac{prob\left(\widehat{t}_i \in \Gamma_A^+(\gamma) \middle| \, H_i = 0\right) \cdot prob(H_i = 0)}{prob\left(\widehat{t}_i \in \Gamma_A^+(\gamma)\right)} = \frac{\gamma/2 \cdot \pi_0}{E(S_\gamma^+)}. \tag{12}$$

Stated differently, the fdr_γ^+ indicates how the investor changes his prior probability that the fund i has a zero alpha ($H_i = 0$) after observing that its t-stat is significant. In light of Equation (12), our estimator $\widehat{FDR}_\gamma^+ = (\gamma/2 \cdot \widehat{\pi}_0)/\widehat{S}_\gamma^+$ can therefore be interpreted as an empirical bayes estimator of fdr_γ^+, where π_0 and $E(S_\gamma^+)$ are directly estimated from the data.[33]

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

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

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

distribution is characterized by a point mass at $\alpha = 0$, reflecting the fact that 72.8% of the funds yield zero alphas. Since $\widehat{\pi}_A^-$ is much higher than $\widehat{\pi}_A^+$, 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% ($\widehat{\pi}_0 + \widehat{\pi}_A^+$) 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 Appendix

A Estimation Procedure

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

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

$$\widehat{MSE}(\lambda) = \frac{1}{1,000} \sum_{b=1}^{1,000} \left[\widehat{\pi}_0^b(\lambda) - \min_{\lambda} \widehat{\pi}_0(\lambda) \right]^2 . \tag{13}$$

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

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

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

$$\widehat{MSE}^-(\gamma) = \frac{1}{1,000} \sum_{b=1}^{1,000} \left[\widehat{\pi}_A^{b-}(\gamma) - \max_{\gamma} \widehat{\pi}_A^-(\gamma) \right]^2 . \tag{14}$$

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

A.3 Determining the Standard Deviation of the Estimators

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

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

B Monte-Carlo Analysis

B.1 Under Cross-Sectional Independence

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

$$\begin{aligned} r_t &= \alpha + \beta F_t + \varepsilon_t, & t &= 1, ..., T, \\ F_t &\sim N(0, \Sigma_F), & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2 I), \end{aligned} \tag{15}$$

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

Our estimators are compared with their respective true population values defined hereafter. First, we set values for the parameters π_0, π_A^-, π_A^+, and for the t-stat mean of the unskilled and skilled funds. Using these values, we can then compute $E(F_\gamma^-) = E(F_\gamma^+) = \pi_0 \cdot \gamma/2$. To determine the expected number of unskilled and skilled funds, $E(T_\gamma^-)$ and $E(T_\gamma^+)$, we use the fact that under the alternative hypothesis $\alpha_i \neq 0$, the fund t-stat follows a non-central student distribution with $T-5$ degrees of freedom and a noncentrality parameter equal to $T^{\frac{1}{2}}\alpha_A/\sigma_\varepsilon$ (Davidson and MacKinnon (2004), p. 169):

$$\begin{aligned} E(T_\gamma^-) &= \pi_A^- \cdot prob\left(t < t_{T-5,\gamma/2}\,|H_A, \alpha_A < 0\right), \\ E(T_\gamma^+) &= \pi_A^+ \cdot prob\left(t > t_{T-5,1-\gamma/2}\,|H_A, \alpha_A > 0\right), \end{aligned} \quad (16)$$

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. Finally, we have $E(S_\gamma^-) = E(F_\gamma^-) + E(T_\gamma^-)$ and $E(S_\gamma^+) = E(F_\gamma^+) + E(T_\gamma^+)$.

Realistic values for the different parameters are determined as follows. The values for the proportions of zero-alpha, unskilled, and skilled funds in the population correspond to the average estimated values over the final 5 years of our sample (2002-2006, see Figure 4): $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$. To determine the t-stat means of the unskilled and skilled funds, we calibrate Equation (16) so that we match the average $\widehat{T}_\gamma^-$ and $\widehat{T}_\gamma^+$ at $\gamma = 0.20$ over the period 2002-2006. The values obtained are equal to -2.5 and 3, corresponding to an annual alpha of -3.2% and 3.8%, respectively (using the equality $t_A = T^{\frac{1}{2}}\alpha_A/\sigma_\varepsilon$). Our simulations under cross-sectional dependence (described hereafter) is based on a sample of 1,500 correlated fund returns. To allow for comparisons between the two cases, we set $M = 1,500$.[34] Consistent with our database, we set $T = 384$, and $\sigma_\varepsilon = 0.021$ (equal to the cross-sectional average). The vector of fund alphas, α, is built by randomly choosing the identity of the unskilled and skilled funds among the 1,500 funds in the population. The input for β is the empirical loading matrix of the first 1,500 funds, while Σ_F is proxied by its empirical counterpart.

After drawing randomly F_t and ε_t ($t = 1, ..., 384$), we construct the fund return time-series according to Equation (13), and compute their t-stat by regressing the fund returns on the four-factor model. To determine the p-values, we use the fact that the fund t-stat follows a Student distribution with $T-5$ degrees of freedom under the null hypothesis $\alpha_i = 0$. Then, we compute $\widehat{\pi}_0$, $\widehat{\pi}_A^-$, and $\widehat{\pi}_A^+$ using Equations (5) and (**??**). $\widehat{S}_\gamma^-$ and $\widehat{S}_\gamma^+$ correspond to the observed number of significant funds with negative and positive al-

[34]Since our original sample of funds is higher than 1,500 ($M = 2,076$), our assessment of the precision of the estimators is conservative.

phas, respectively. $\widehat{F}_{\gamma}^{-}$ and $\widehat{F}_{\gamma}^{+}$ are computed with Equation (6). $\widehat{T}_{\gamma}^{-}$ and $\widehat{T}_{\gamma}^{+}$, are given in Equation (7). We repeat this procedure 1,000 times.

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

Please insert Table VI here

B.2 Under Cross-Sectional Dependence

The return-generating process is the same as the one shown in Equation (15), except that the fund residuals are cross-correlated:

$$\varepsilon_t \sim N(0, \Sigma), \tag{17}$$

where Σ denotes the $M \times M$ residual covariance matrix. In order to construct a non-negative definite covariance matrix, we proceed as follows. First, 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. Then, we compute the 898×898 empirical covariance matrix denoted by Σ_1. This matrix reflects the complex relations linking all these funds together.[35] Finally, to account for the fact that many funds have non-synchronous data, we introduce 602 uncorrelated funds.[36] This yields the following covariance matrix:

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \sigma_{\varepsilon}^2 I \end{pmatrix}. \tag{18}$$

The results in Table VIII 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

[35] 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.

[36] 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).

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 VII here

Apart from this baseline dependence scenario, we also examine two other cases. First, we introduce correlation by block among zero-alpha, unskilled, and skilled funds to account for their possible similar bets. Inside each block, 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] Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, Should Investors Avoid all Actively Managed Mutual Funds? A study in Bayesian Performance Evaluation, Journal of Finance 56, 45-85.

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

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

[4] Capon, Noel, Gavan J. Fitzsimmons, and Russ A. Prince, 1996, An Individual Level Analysis of the Mutual Fund Investment Decision, Journal of Financial Services Research 10, 59-82.

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

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

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

[8] Cremers, Martijn, and Antti Petajisto, 2006, How Active is your Fund Manager? A New Measure that Predicts Performance, Working Paper 06-21, Yale University.

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

[10] Davison, Anthony C., and David V. Hinkley, 1997, Bootstrap Methods and Their Application, Cambridge University Press, London.

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

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

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

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

[15] Fama, Eugene F., and Kenneth R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, Journal of Finance 51, 55-84.

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

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

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

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

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

[21] Grinblatt, Mark, Sheridan Titman, and Russ Wermers, 1995, Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior, American Economic Review 85, 1088-1105.

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

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

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

[25] Horowitz, Joel L., 2001, The Bootstrap, in J.J. Heckman and E.E. Leamer, ed.: Handbook of Econometrics vol. 5, North-Holland, Netherlands, 3159-3228.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table I

Performance of the Equally-Weighted Portfolio of Funds

Results for the unconditional and conditional Carhart 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 (384 observations). Each Panel contains the estimated annualized alpha ($\widehat{\alpha}$), the estimated exposures to the market ($\widehat{b}_m$), size ($\widehat{b}_{smb}$), value ($\widehat{b}_{hml}$), and momentum factors ($\widehat{b}_{mom}$), as well as the adjusted R^2 of an equally-weighted portfolio including all funds existing at the beginning of a given month. Figures in parentheses denote the heteroskedasticity-consistent p-values under the null hypothesis that the regression parameters are equal to zero.

Panel A Unconditional Four-Factor Model

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

Panel B Conditional Four-Factor Model

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

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 proportion of significant funds in the left and right tails of the cross-sectional t-stat distribution ($\widehat{S}_{\gamma}^{-}$, $\widehat{S}_{\gamma}^{+}$) at four significance levels (γ=0.05, 0.10, 0.15, 0.20). For each these significant groups, we also compute the average estimated alpha (in % per year), expense ratio (in % per year), and fund age (proxied by the number of yearly return observations). Panel C measures the impact of luck on the performance of each significant group. In the left tail, $\widehat{S}_{\gamma}^{-}$ is decomposed into unlucky and unskilled funds ($\widehat{F}_{\gamma}^{-}$, $\widehat{T}_{\gamma}^{-}$) , while 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.

Panel A Proportion of Unskilled and Skilled Funds

	Zero alpha($\widehat{\pi}_0$)	Non-zero alpha	Unskilled($\widehat{\pi}_A^{-}$)	Skilled($\widehat{\pi}_A^{+}$)
Proportion	72.8 (2.7)	27.2	26.6 (4.1)	0.6 (0.8)
Number	1,512	564	552	12

Panel B Significant Funds in the Left and Right Tails

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_{\gamma}^{-}$(%)	11.6	17.2	21.5	25.4	7.8	5.8	4.0	2.1	Signif. $\widehat{S}_{\gamma}^{+}$(%)
	(0.7)	(0.9)	(1.0)	(1.1)	(0.7)	(0.6)	(0.5)	(0.3)	
Alpha(% year)	-5.5	-5.0	-4.7	-4.6	4.8	5.2	5.5	6.5	Alpha(% year)
Exp.(% year)	1.6	1.6	1.5	1.5	1.3	1.2	1.2	1.2	Exp.(% year)
Age(year)	12.6	12.6	12.6	12.6	15.2	15.4	16.0	15.4	Age(year)

Panel C Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_{\gamma}^{-}$(%)	11.6	17.2	21.5	25.4	7.8	5.8	4.0	2.1	Signif. $\widehat{S}_{\gamma}^{+}$(%)
	(0.7)	(0.9)	(1.0)	(1.1)	(0.7)	(0.6)	(0.5)	(0.3)	
Unlucky $\widehat{F}_{\gamma}^{-}$(%)	1.8	3.6	5.4	7.2	7.2	5.4	3.6	1.8	Lucky $\widehat{F}_{\gamma}^{+}$(%)
	(0.0)	(0.1)	(0.2)	(0.3)	(0.3)	(0.2)	(0.1)	(0.0)	
Unskilled $\widehat{T}_{\gamma}^{-}$(%)	9.8	13.6	16.1	18.2	0.6	0.4	0.4	0.3	Skilled $\widehat{T}_{\gamma}^{+}$(%)
	(0.7)	(0.8)	(0.9)	(0.9)	(0.6)	(0.5)	(0.4)	(0.3)	

Table III

Impact of Luck on Short-Term Performance

The impact of luck in the left and right tails for the entire fund population (All funds), as well as Growth, Aggressive Growth, and Growth and Income funds is presented in Panels A, B, C, and D. We measure fund performance with the unconditional four-factor model over non-overlapping 5-year subperiods during the period 1975-2006. We count the proportion of significant funds in the left and right tails of the cross-sectional t-stat distribution ($\widehat{S}_\gamma^-$, $\widehat{S}_\gamma^+$) at four significance levels (γ=0.05, 0.10, 0.15, 0.20). For each significance level γ, we decompose $\widehat{S}_\gamma^-$ into unlucky and unskilled funds ($\widehat{F}_\gamma^-$, $\widehat{T}_\gamma^-$), and $\widehat{S}_\gamma^+$ into lucky and skilled funds ($\widehat{F}_\gamma^+$, $\widehat{T}_\gamma^+$). Figures in parentheses denote the standard deviation of the different estimators.

Panel A All funds

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$(%)	11.2	16.8	21.4	24.9	9.6	7.8	6.0	3.5	Signif. $\widehat{S}_\gamma^+$(%)
	(0.5)	(0.6)	(0.7)	(0.8)	(0.5)	(0.5)	(0.4)	(0.3)	
Unlucky $\widehat{F}_\gamma^-$(%)	1.8	3.6	5.4	7.2	7.2	5.4	3.6	1.8	Lucky $\widehat{F}_\gamma^+$(%)
	(0.0)	(0.0)	(0.1)	(0.2)	(0.2)	(0.1)	(0.0)	(0.0)	
Unskilled $\widehat{T}_\gamma^-$(%)	9.4	13.2	16.0	17.7	2.4	2.4	2.4	1.7	Skilled $\widehat{T}_\gamma^+$(%)
	(0.6)	(0.7)	(0.8)	(0.8)	(0.6)	(0.5)	(0.4)	(0.3)	

Panel B Growth funds

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$(%)	11.4	16.8	21.5	24.8	9.9	8.1	6.2	3.5	Signif. $\widehat{S}_\gamma^+$(%)
	(0.7)	(0.8)	(0.9)	(1.0)	(0.7)	(0.6)	(0.5)	(0.4)	
Unlucky $\widehat{F}_\gamma^-$(%)	1.8	3.6	5.5	7.3	7.3	5.5	3.6	1.8	Lucky $\widehat{F}_\gamma^+$ (%)
	(0.0)	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.0)	
Unskilled $\widehat{T}_\gamma^-$(%)	9.6	13.1	16.0	17.6	2.6	2.6	2.6	1.7	Skilled $\widehat{T}_\gamma^+$ (%)
	(0.7)	(0.9)	(1.0)	(1.1)	(0.8)	(0.7)	(0.6)	(0.4)	

Table III

Impact of Luck on Short-Term Performance (Continued)

Panel C Aggressive Growth funds

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_{\gamma}^{-}$(%)	11.2	16.3	19.6	22.4	11.6	9.8	7.5	4.4	Signif. $\widehat{S}_{\gamma}^{+}$(%)
	(1.3)	(1.5)	(1.6)	(1.7)	(1.3)	(1.2)	(1.0)	(0.8)	
Unlucky $\widehat{F}_{\gamma}^{-}$(%)	1.8	3.5	5.3	7.1	7.1	5.3	3.5	1.8	Lucky $\widehat{F}_{\gamma}^{+}$(%)
	(0.1)	(0.2)	(0.3)	(0.4)	(0.4)	(0.3)	(0.2)	(0.1)	
Unskilled $\widehat{T}_{\gamma}^{-}$(%)	9.4	12.9	14.3	15.3	4.5	4.5	4.0	2.6	Skilled $\widehat{T}_{\gamma}^{+}$(%)
	(1.3)	(1.6)	(1.7)	(1.9)	(1.4)	(1.3)	(1.1)	(0.8)	

Panel D Growth and Income funds

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_{\gamma}^{-}$(%)	11.0	16.5	22.3	26.4	7.4	5.6	3.7	1.9	Signif. $\widehat{S}_{\gamma}^{+}$(%)
	(1.1)	(1.4)	(1.5)	(1.6)	(1.0)	(0.8)	(0.7)	(0.5)	
Unlucky $\widehat{F}_{\gamma}^{-}$ (%)	1.9	3.7	5.6	7.4	7.4	5.6	3.7	1.9	Lucky $\widehat{F}_{\gamma}^{+}$(%)
	(0.1)	(0.2)	(0.3)	(0.4)	(0.4)	(0.3)	(0.2)	(0.1)	
Unskilled $\widehat{T}_{\gamma}^{-}$ (%)	9.1	12.8	16.7	19.0	0.0	0.0	0.0	0.0	Skilled $\widehat{T}_{\gamma}^{+}$(%)
	(2.1)	(2.5)	(2.6)	(2.6)	(1.1)	(1.0)	(0.8)	(0.5)	

Table IV

Performance Persistence Based on the False Discovery Rate

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

Panel A Portfolio Statistics

	Achieved False Discovery Rate $\left(\widehat{FDR}^+_{\gamma^P}\right)$					Included proportion of funds($\widehat{S}^+_{\gamma^P}$)				
	Mean	10-30	30-50	50-70	>70%	Mean	0-6	6-12	12-24	>24%
FDR 10%	41.5%	13	6	1	6	3.0%	27	0	0	0
FDR 30%	47.5%	8	12	1	6	8.2%	19	5	2	1
FDR 50%	60.4%	0	14	7	6	20.9%	11	2	7	7
FDR 70%	71.3%	0	4	11	11	29.7%	3	8	0	16
FDR 90%	75.0%	0	4	9	14	33.7%	0	6	2	19

Panel B Performance Analysis

	$\widehat{\alpha}$(p-value)	$\widehat{\sigma}_\varepsilon$	IR	$\widehat{b}_m$	$\widehat{b}_{smb}$	$\widehat{b}_{hml}$	$\widehat{b}_{mom}$	Mean	Std dev
FDR 10%	1.45%(0.04)	4.0%	0.36	0.93	0.16	-0.04	-0.02	8.3%	15.4%
FDR 30%	1.15%(0.05)	3.3%	0.35	0.94	0.17	-0.02	-0.03	8.1%	15.4%
FDR 50%	0.95%(0.10)	2.9%	0.33	0.96	0.20	-0.06	-0.01	8.1%	16.1%
FDR 70%	0.68%(0.15)	2.7%	0.25	0.97	0.19	-0.06	-0.01	7.9%	16.1%
FDR 90%	0.39%(0.30)	2.7%	0.14	0.97	0.19	-0.05	-0.00	7.8%	16.0%

Panel C Portfolio Turnover

	Proportion of funds staying in the portfolio or leaving it...									
	After 1 year		After 2 years		After 3 years		After 4 years		After 5 years	
	In	Out	In	Out	In	Out	In	Out	In	Out
FDR 10%	36.7	63.3	12.8	87.2	3.4	96.6	0.8	99.2	0.0	100.0
FDR 30%	40.0	60.0	14.7	85.3	5.1	94.9	1.7	98.3	1.3	98.7
FDR 50%	48.8	51.2	23.5	76.5	12.3	87.7	4.7	95.3	2.6	97.4
FDR 70%	52.2	47.8	29.0	71.0	17.4	82.6	9.5	90.5	6.3	93.7
FDR 90%	55.9	44.1	33.8	66.2	20.4	79.6	13.0	87.0	8.5	91.5

Table V

Impact of Luck on Long-Term Pre-Expense Performance

Panel A Proportion of Unskilled and Skilled Funds

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

Panel B Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif. level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif. level(γ)
Signif. $\widehat{S}_\gamma^-$(%)	4.3	7.5	10.2	12.8	17.3	13.1	9.3	5.8	Signif. $\widehat{S}_\gamma^+$(%)
	(0.5)	(0.6)	(0.7)	(0.8)	(0.9)	(0.8)	(0.7)	(0.5)	
Unlucky $\widehat{F}_\gamma^-$(%)	2.1	4.3	6.4	8.6	8.6	6.4	4.3	2.1	Lucky $\widehat{F}_\gamma^+$(%)
	(0.0)	(0.1)	(0.1)	(0.2)	(0.2)	(0.1)	(0.1)	(0.0)	
Unskilled $\widehat{T}_\gamma^-$(%)	2.2	3.2	3.8	4.2	8.7	6.6	5.0	3.6	Skilled $\widehat{T}_\gamma^+$(%)
	(0.5)	(0.6)	(0.8)	(0.9)	(1.0)	(0.9)	(0.7)	(0.5)	
Pre Expense									Pre Expense
Alpha(% year)	-5.8	-5.2	-4.8	-4.5	4.4	4.8	5.0	5.3	Alpha(% year)
Exp.(% year)	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	Exp.(% year)
Age(year)	10.8	10.4	10.6	11.0	14.6	14.9	15.4	16.5	Age(year)

Table VI

Loadings on Omitted Factors

In Panel A, we determine the significant funds in the left and right tails at four significance levels (γ=0.05, 0.10, 0.15, 0.20) based on the unconditional CAPM. For each of these significant groups, we compute their average loadings on the four-factor omitted factors: size ($\widehat{b}_{smb}$), value ($\widehat{b}_{hml}$), and momentum ($\widehat{b}_{mom}$). Panel B repeats the same procedure with the unconditional Fama-French model. In both cases, performance is measured during the period 1975-2006.

Panel A Unconditional CAPM

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

Panel B Unconditional Fama-French model

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

Table VII

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, value, and momentum factors). Fund residuals are independent from one another. The true parameter values for the proportions of zero-alpha, unskilled, and skilled funds (π_0, π_A^-, and π_A^+) are set to 75%, 23%, and 2%. The *t*-stat means of the unskilled and skilled funds are set to -2.5 and 3 (corresponding to an annual alpha of -3.2% and 3.8%, respectively). In each tail (left and right), we assess the precision of the different estimators at two significance levels (γ=0.05 and 0.20).

Fund Proportion	True	Estimator (90% interval)
Zero-alpha funds (π_0)	75.0	75.1 (71.8,78.5)
Unskilled funds (π_A^-)	23.0	22.9 (19.9,26.2)
Skilled funds (π_A^+)	2.0	2.0 (0.3,3.8)

	Significance level $\gamma = 0.05$		Significance level $\gamma = 0.20$	
Left Tail	True	Estimator (90% interval)	True	Estimator (90% interval)
Significant funds $E\left(S_\gamma^-\right)$	18.1	18.1 (16.4,19.6)	27.9	27.9 (26.1,30.0)
Unlucky funds $E\left(F_\gamma^-\right)$	1.8	1.8 (1.8,1.9)	7.5	7.5 (7.1,7.9)
Unskilled funds $E\left(T_\gamma^-\right)$	16.2	16.2 (14.6,17.7)	20.4	20.4 (18.3,22.6)

	Significance level $\gamma = 0.05$		Significance level $\gamma = 0.20$	
Right Tail	True	Estimator (90% interval)	True	Estimator (90% interval)
Significant funds $E\left(S_\gamma^+\right)$	3.6	3.6 (2.8,4.4)	9.4	9.4 (8.2,10.8)
Lucky funds $E\left(F_\gamma^+\right)$	1.8	1.8 (1.8,1.9)	7.5	7.5 (7.1,7.9)
Skilled funds $E\left(T_\gamma^+\right)$	1.7	1.7 (0.9,2.5)	1.9	1.9 (0.5,3.4)

Table VIII

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, value, and momentum factors). We assume that funds are cross-sectionally correlated. In order to closely reproduce the relations across all funds, we use the emprical covariance matrix of the fund residuals as the true covariance matrix. The true parameter values for the proportions of zero-alpha, unskilled, and skilled funds (π_0, π_A^-, and π_A^+) are set to 75%, 23%, and 2%. The *t*-stat means of the unskilled and skilled funds are set to -2.5 and 3 (corresponding to an annual alpha of -3.2% and 3.8%, respectively). In each tail (left and right), we assess the precision of the different estimators at two significance levels (γ=0.05 and 0.20).

Fund Proportion	True	Estimator (90% interval)
Zero-alpha funds (π_0)	75.0	75.2 (69.9,80.4)
Unskilled funds (π_A^-)	23.0	22.8 (17.3,28.6)
Skilled funds (π_A^+)	2.0	1.9 (0.0,5.8)

	Significance level $\gamma = 0.05$		Significance level $\gamma = 0.20$	
Left Tail	True	Estimator (90% interval)	True	Estimator (90% interval)
Significant funds $E\left(S_\gamma^-\right)$	18.1	18.1 (15.5,20.4)	27.9	27.9 (24.5,32.0)
Unlucky funds $E\left(F_\gamma^-\right)$	1.8	1.8 (1.6,2.1)	7.5	7.6 (6.6,8.5)
Unskilled funds $E\left(T_\gamma^-\right)$	16.2	16.2 (13.5,19.0)	20.4	20.4 (16.5,24.2)

	Significance level $\gamma = 0.05$		Significance level $\gamma = 0.20$	
Right Tail	True	Estimator (90% interval)	True	Estimator (90% interval)
Significant funds $E\left(S_\gamma^+\right)$	3.5	3.6 (2.4,5.2)	9.4	9.4 (6.8,12.4)
Lucky funds $E\left(F_\gamma^+\right)$	1.8	1.8 (1.6,2.1)	7.5	7.6 (6.6,8.5)
Skilled funds $E\left(T_\gamma^+\right)$	1.7	1.7 (0.5,3.8)	1.9	1.9 (0.2,5.5)

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 (π_0, π_A^-, and π_A^+). In this example, we set $\pi_0 = 75\%$, $\pi_A^- = 23\%$, and $\pi_A^+ = 2\%$.

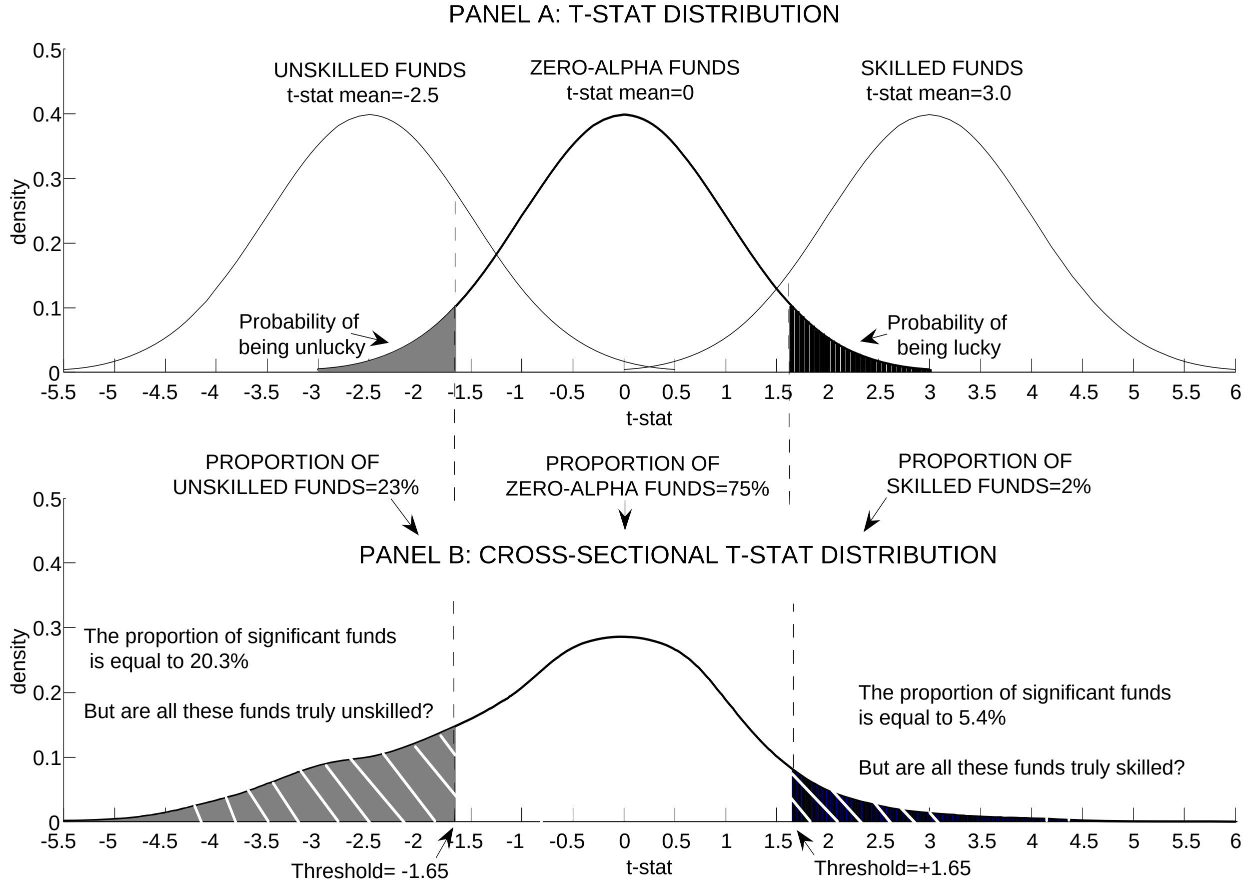


Figure 2
Histogram of Fund p-values

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 (π_0, π_A^-, and π_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.

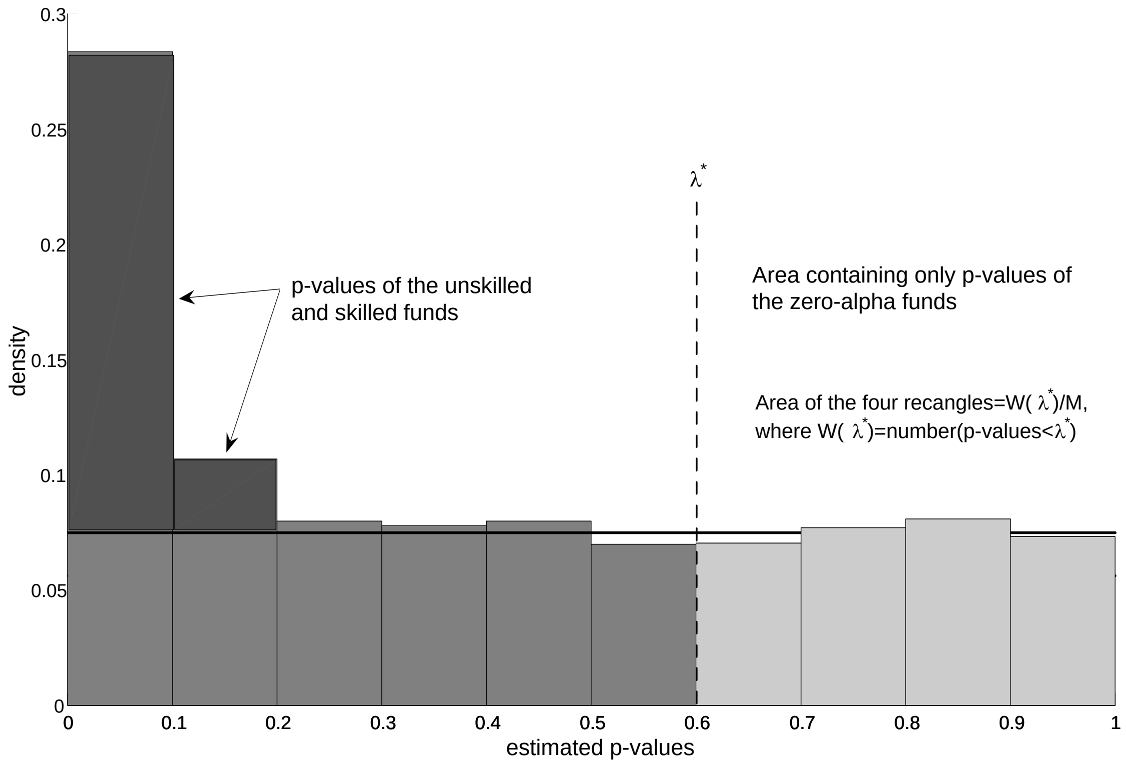
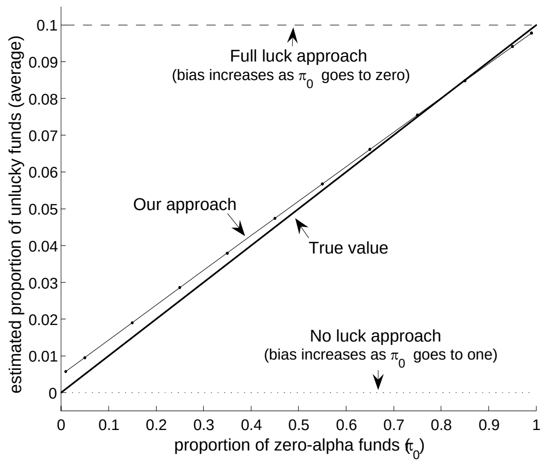


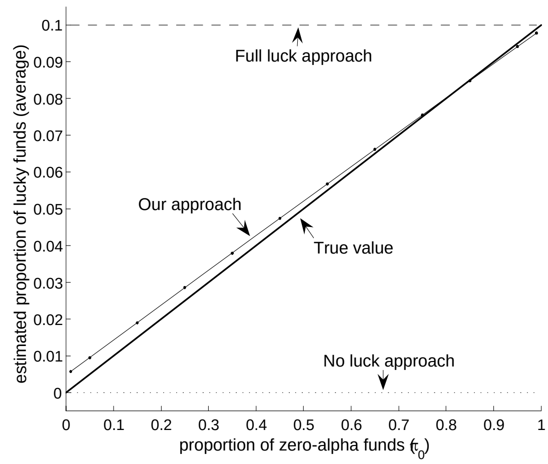
Figure 3

Measuring Luck: Comparison with Existing Approaches

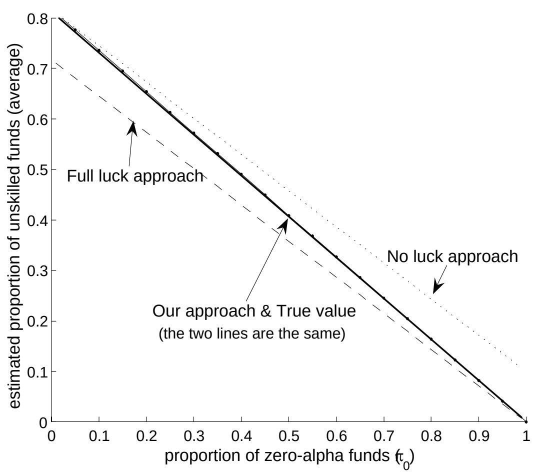
This figure examines the bias of different estimators produced by the three approaches as a function of the proportion of zero-alpha funds, π_0. We examine the estimators of the proportions of unlucky, lucky, unskilled, and skilled funds in Panel A, B, C, and D, respectively. The "no luck" approach assumes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



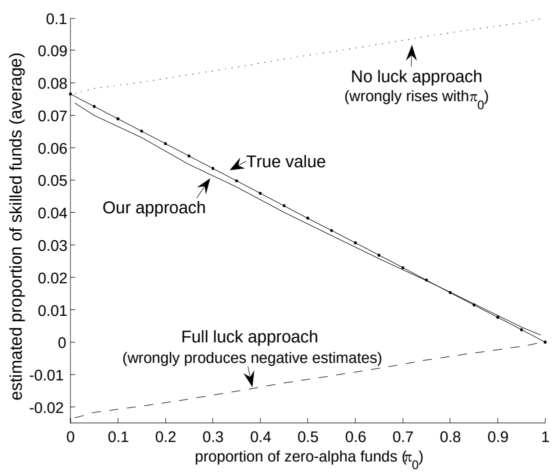
(a) Unlucky funds (left tail)



(b) Lucky funds (right tail)



(c) Unskilled funds (left tail)



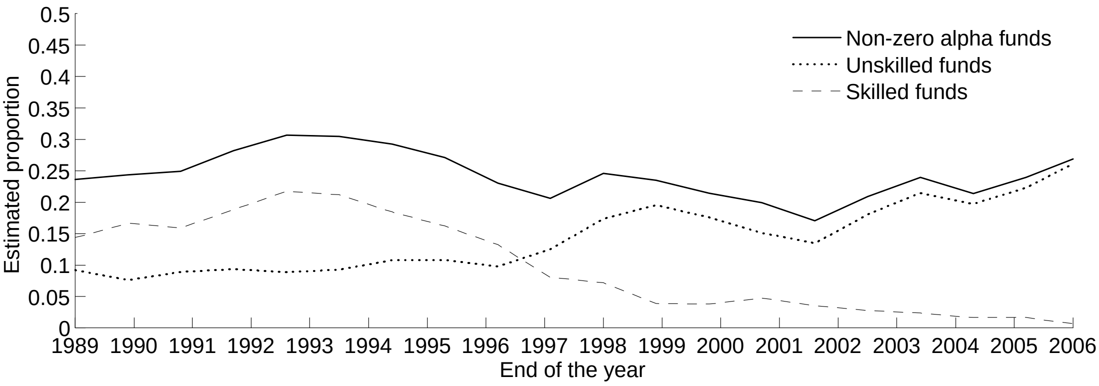
(d) Skilled funds (right tail)

Figure 4

Evolution of Mutual Fund Performance over Time

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

PANEL A: PROPORTIONS OF UNSKILLED AND SKILLED FUNDS



PANEL B: GROWTH IN THE FUND POPULATION

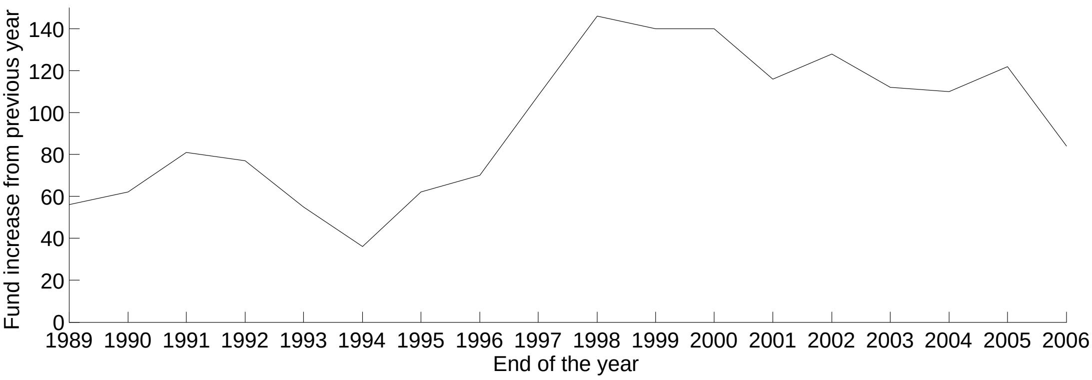


Figure 5

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.