False Discoveries in Mutual Fund Performance: Measuring the Role of Lucky Alphas*

L. Barras[†], O. Scaillet[‡] and R. Wermers[§]

First version, October 2005

Abstract

The standard tests designed to detect funds with positive and negative alphas are subject to luck. Lucky funds have significant estimated alphas whereas their true alphas are equal to zero. This paper measures the relative importance of these lucky funds among the significant funds with the False Discovery Rate (FDR). Using US funds in different investment categories, we show that luck has a substantial impact on the performance documented in previous studies. Our results have implications for mutual fund performance analysis and portfolio management. We find that most of the few funds with non-zero alphas yield negative performance. However, the small fraction of funds with positive performance is sufficient to form portfolios with positive alphas.

*We are grateful to R. Stulz and as well as seminar participants at the Bilgi University, Grequam, and Swiss Doctoral Workshop for their helpful comments. The first and second authors aknowledge financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK).

[†]HEC-University of Geneva and FAME, Boulevard du Pont d'Arve 40, 1211 Geneva 4, Switzerland. Tel: +41223798141. E-mail: barras@hec.unige.ch

[‡]HEC-University of Geneva and FAME, 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

1 Introduction

In light of the large amount of money invested in the mutual fund industry, the issue of whether actively-managed mutual funds generate superior performance has been largely studied by academics. Most studies examine performance at the macrolevel to know whether the industry as a whole delivers positive risk-adjusted returns. They generally find that the average alpha among the fund population is negative (Jensen (1968), Lehman and Modest (1987), Elton et al. (1993), Pastor and Stambaugh (2002a))[1].

Since mutual fund holdings account for a significant portion of the total market value, it is unlikely that the industry as a whole is able to beat the market by a large margin (Chen, Jegadeesh and Wermers (2000)). However, some funds with positive or negative alphas are likely to exist, but they cannot be detected with average performance measures. For instance, suppose that 10% of the funds in the population have an annual alpha equal to -5%, 10% of them an annual alpha of +5%, and the rest of the industry a zero alpha. The average fund alpha turns out to be equal to zero, despite the presence of funds with non-zero alphas. Therefore, to have a finer representation of mutual fund performance, it is important to analyse mutual fund performance at the microlevel. This is done by focusing on the performance of funds located at both tails of the cross-sectional alpha distribution.

To detect these funds with positive and negative performance, the standard approach developed in the literature (Jensen (1968), Ferson and Schadt (1996), Ferson and Qian (2004)) can be described as follows. The presence of differential performance (positive or negative alphas) is tested for each of the M funds in the population. Then, a conventional significance level γ is set (usually 0.05 or 0.10) and all funds with p-values inferior to γ have significant estimated alphas. Finally, these significant funds are counted in order to provide an estimator of the number of funds with differential performance.

Note that the standard approach is based on a multiple hypothesis test because the null hypothesis of no performance is not tested once but M times. Like in every hypothesis test, the presence of random sampling makes the inference on the fund alphas subject to luck. We define a fund as lucky if its estimated alpha (irrespective of its sign) is significant whereas its true alpha is equal to zero. When a single test is run

[1] Using fund portfolio holdings, Grinblatt and Titman (1993) and Daniel et al. (1997) find significant stock-picking ability. However, this ability is not sufficient to cover the fund expenses (Wermers (2000)).

on the estimated alpha of a given fund (or a given portfolio of funds)[2], luck is easily controlled by setting the significance level γ (or alternatively the $Size$ of the test). For instance, if γ is set to 0.05, the probability of finding one lucky fund amounts to 0.05 by construction. However, accounting for luck in a multiple test is more complex because luck cannot be measured by γ. For instance, if γ is set to 0.05 for each individual test, the probability of finding at least one lucky fund among the M funds is much higher than 5%. Even if all funds have zero alphas, we still expect that some of the M funds will be significant by pure luck.

The major drawback of the standard approach stems from its inability to assess the impact of luck on performance. Its estimation of the number of funds with differential performance is flawed because it does not account for the presence of lucky funds among the significant funds[3]. Moreover, to determine the source of differential performance, the standard approach partitions the number of significant funds according to the sign of their estimated alphas. The number of funds with positive estimated alphas (called hereafter the best funds) is used as an estimator of the number of funds with positive performance. Similarly, the funds yielding a negative performance are estimated by the number of funds with negative estimated alphas (called hereafter the worst funds). Unfortunately, these two estimators are also inaccurate because they cannot distinguish between lucky funds and funds with positive or negative alphas.

For this reason, the use of the standard approach raises important issues. First, we cannot properly measure the real fund performance. Let us suppose that 20 out of 200 funds have positive estimated alphas at a given significance level γ. The real performance of these 20 funds depends on the proportion of lucky funds. For instance, if the latter is equal to 100%, all 20 funds produce in reality zero alphas. Second, the standard approach assumes that the observed increase in the number of significant funds as γ rises is only due to the detection of new funds with differential performance. However, part of this increase is due to the inclusion of new lucky funds. Therefore, the standard approach cannot provide any information about the location of funds with differential performance in the tails of the cross-sectional alpha distribution. For instance, suppose

[2] Performance tests at the macrolevel fall in this category. They boil down to running a single test on the alpha of the equally-weighted portfolio of all funds.

[3] This issue is clearly stated in Grinblatt and Titman (1995): "While some funds achieved positive abnormal returns, it is difficult to ascertain the implications of this for the efficient market hypothesis because of multiple comparison being made. That is, even if no superior management ability existed, we would expect some funds to achieve superior risk-adjusted returns by chance."

that the number of funds with negative estimated alphas increases by 50 as γ passes from 0.05 to 0.15. If all these 50 funds are lucky, we would conclude that the few funds with negative performance are located in the extreme left tail. On the contrary, if none of them is lucky, we would say that the funds with negative performance are largely spread in the left tail. Finally, because the standard approach cannot measure the real performance, it cannot compare the different investment categories (Growth, Growth and Income funds...).

The main contribution of this paper is to address all these issues by examining the impact of luck on mutual fund performance. To this end, we use the False Discovery Rate (FDR) introduced by Benjamini and Hochberg (1995) in the statistical literature. The FDR measures the proportion of lucky funds among the significant funds. The FDR is very easy to compute from the fund estimated p-values and is therefore a simple extension of the standard approach. Besides, we develop a new staightforward methodology which allows us to separately compute the FDR among the best and worst funds. First, we use the FDR to assess the real performance of the best and worst funds at different significance levels γ. We measure the proportion of funds in the population which truly generate positive and negative alphas. Second, we examine how the FDR varies as γ rises. This indicates whether funds with differential performance are concentrated or spread in the tails of the cross-sectional alpha distribution. Third, we compare performance across the different investment categories.

Other methods have dealt with multiple testing in mutual fund performance. Grinblatt and Titman (1989, 1993) jointly test the restriction that the alphas of all funds are equal to zero (i.e. $\alpha_1 = ... = \alpha_M = 0$). However, this method can only tell us whether there is at least one fund with non-zero alpha. The second approach consists in detecting a number of significant funds such that the FamilyWise Error Rate ($FWER$) is controlled at a given level (usually 0.01, 0.05 or 0.10). The $FWER$ is defined as the probability of yielding at least one lucky fund among the M tested funds (Romano and Wolf (2005)). A famous illustration of this approach is the conservative Bonferroni method (Ferson and Schadt (1996)). A related approach is the Reality Check (White (2000)) which tests whether the performance of the best fund is significant[4]. These two methods explictly account for the presence of luck to determine whether a given fund (or the best one) is significant. But contrary to the FDR, they are not designed to

[4] Sullivan, Timmermann and White (1999) propose an application of the Reality Check to the performance analysis of technical trading strategies.

measure the proportion of lucky funds among the significant funds.

Our empirical results are based on monthly returns of 1'472 U.S. open-end equity funds from CRSP between 1975 and 2002. We investigate the performance of four investment categories, namely All (*All*), Growth (*G*), Aggressive Growth (*AG*) and Growth and Income (*GI*) funds. Following Kosowski et al. (2005), we use bootstrap techniques to obtain precise inferences on the fund alphas, namely accurate p-values. First, our results show that the impact of luck on performance is substantial. Our estimators of the number funds with differential, positive and negative performance are much lower than those obtained with the standard approach in all cases. Besides, while the standard approach concludes that a significant proportion of *All*, *G* and *GI* funds achieve positive performance, we find that is not the case once luck is accounted for.

Second, we find that luck has a stronger impact on the performance of the best rather than the worst funds. Across the four investment categories, the FDR among the worst funds is always inferior to 50%, meaning that the majority of worst funds truly yield negative alphas. Moreover, the FDR increases slowly as γ rises, because part of the new significant funds truly deliver negative alphas. As a result, the funds with negative performance are largely spread in the left tail of the cross-sectional alpha distribution. Although the FDR among the best funds varies across the different investment categories, it is generally much higher than the FDR among the worst funds. For *All* and *G* funds, the FDR is always higher than 50% and increases quickly as γ rises. It indicates that there are only a few funds with positive performance, all of them being concentrated in the extreme right tail. On the contrary, the low level of the FDR among the best *AG* funds reveals that many funds produce a positive performance. This is not the case for the *GI* funds, since none of them is able to yield a positive alpha.

The FDR allows us to determine the relative importance as well as the location of fund with differential performance. This information has important implications for mutual fund performance analysis. From an overall perspective, we observe more frequently funds with negative rather than positive performance. While the percentage of funds with negative alphas is approximately equal to 20% across all categories, the proportion of funds with positive performance is much lower (around 1.5% for the *All* and *G* funds and 8.4% for *AG* funds). However, the performance of the industry as a whole is not so bad because about 80% of the funds produce a neutral performance. In fact, the negative average performance documented in past studies is not due to the majority of funds but is caused by 20% of the funds.

Our analysis also has implications for mutual fund portfolio management. In the recent years, the quest for positive alphas has led to the creation of funds of mutual funds. The ability of these portfolios of funds to yield positive alphas crucially depends on the proportion of lucky funds they contain. Using the FDR, we show that even though the evidence of positive performance among the *All* and G funds is very low, it is strong enough to form portfolios with positive alphas. Since the funds with positive performance are located in the extreme right tail, we can separate them from the lucky funds, by adequatly choosing a sufficiently low significance level γ.

The remainder of the paper is as follows. The next section defines the standard approach and the notion of luck. Then, we define the FDR and explain our new methodology which allows us to compute the FDR among the best and worst funds separately. Section 3 presents the performance measures, the bootstrap technique to compute the p-values as well as the mutual fund data. Section 4 contains the empirical analysis of the impact of luck on performance across the four investment categories. Section 5 concludes. An appendix gathers proofs and results of a Monte-Carlo study.

2 Measuring the Impact of Luck On Mutual Fund Performance

2.1 The Standard Approach to Performance Testing

2.1.1 Testing the Performance of Individual Funds

Let us assume that the mutual fund universe is composed of M individual funds. The performance of each fund i ($i = 1, ..., M$) is measured by its alpha computed with a given asset pricing model. The null hypothesis H_0 that the fund i achieve no performance ($\alpha_i = 0$) and the alternative H_A that it delivers differential performance ($\alpha_i > 0$ or $\alpha_i < 0$) are defined as:

$$H_0 : \alpha_i = 0, \quad H_A : \alpha_i > 0 \text{ or } \alpha_i < 0. \tag{1}$$

To detect the funds with positive or negative alphas, the standard approach developed in the literature (Jensen (1968), Ferson and Schadt (1996), Ferson and Qian (2004)) consists in testing the null hypothesis H_0 for each fund i. To this end, a significance level γ is set (usually 0.05 or 0.10). All funds with estimated p-values smaller than γ have significant estimated alphas (i.e. H_0 is rejected). Then, the number of significant funds are counted in order to provide an estimator of the number of funds with differential performance

2.1.2 The Definition of Luck

The standard approach boils down to running a multiple hypothesis test because the null hypothesis H_0 of no performance is not tested once but M times. Like in every hypothesis test, the presence of random sampling makes the inference on the fund alphas subject to luck. We define a fund as lucky if its estimated alpha is significant whereas its true alpha is equal to zero. In our definition, the sign of the fund estimated alpha is not relevant. All that matters is that this fund is significant while its true alpha is equal to zero.

In the case of a single performance test of a given fund (or a given portfolio of funds), luck is easily controlled by setting the significance level γ. For instance, if γ is set to 0.10, the probability of finding one lucky fund amounts to 10% by construction. However, accounting for luck in multiple performance testing is more complex. In this case, luck cannot be measured by γ. For instance, if γ is set to 0.10 for each individual test, the probability of finding at least one lucky fund among the M funds (also called the compound type I error) is much higher than 0.10. Even if all funds have zero alphas, we still expect that some of the M funds will be significant only by pure luck.

2.1.3 The Drawback of the Standard Approach

To understand how luck spuriously affects the results obtained by the standard approach, Table 1 classifies the four possible outcomes of the multiple test. $F(\gamma)$ denotes the number of lucky funds. $T(\gamma)$ stands for the number of significant funds which truly yield differential performance. Adding $F(\gamma)$ and $T(\gamma)$, the total number of significant funds amounts to $R(\gamma)$. All these quantities depend on the chosen significance level γ.

Please Insert Table 1 here

The major drawback of the standard approach is that it cannot assess the impact of luck on performance because it cannot distinguish between luck and differential performance. Indeed, the standard approach measures differential performance by the $R(\gamma)$ significant funds. However, $F(\gamma)$ among these $R(\gamma)$ funds are simply lucky. Therefore, a correct measurement of the funds with differential performance is given by $T(\gamma) = R(\gamma) - F(\gamma)$. Obviously, the standard approach tends to overestimate the number of funds with differential performance. Besides, as γ gets higher, the test of differential performance becomes less stringent, thus increasing both the number of significant funds $R(\gamma)$, and the number of lucky funds $F(\gamma)$. However, the standard approach implicitly assumes that the observed increase in $R(\gamma)$ is only due to the detection of new funds with differential performance. Therefore, it cannot capture the proportion of the rise in $R(\gamma)$ due to the inclusion of lucky funds. To address all these issues, we propose a simple extension of the standard approach aimed at measuring the importance of these lucky funds. This additional step is made by using the False Discovery Rate.

2.2 The False Discovery Rate

2.2.1 The FDR among the Significant Funds

The FDR is defined as the expected proportion of lucky funds[5] among the significant funds at the significance level γ. It is written as[6]:

$$FDR(\gamma) = E\left(\left.\frac{F(\gamma)}{R(\gamma)}\right| R(\gamma) > 0\right). \tag{2}$$

The FDR is a direct measure of luck since it increases as the number of lucky funds rises. Stated differently, the FDR takes into account the compound type I error stemming from the fact that the null hypothesis H_0 is not tested once but M times. To identify the factors which determine the FDR, we can write the latter as (see Storey (2003)):

$$\begin{aligned} FDR(\gamma) &= \frac{\pi_0 \cdot prob(p_i < \gamma \,|H_0)}{\pi_0 \cdot prob(p_i < \gamma \,|H_0) + \pi_A \cdot prob(p_i < \gamma \,|H_A)} \\ &= \frac{\pi_0 \cdot Size(\gamma)}{\pi_0 \cdot Size(\gamma) + \pi_A \cdot Power(\gamma)} = \frac{\pi_0 \cdot \gamma}{\pi_0 \cdot \gamma + (1-\pi_0) \cdot Power(\gamma)}, \end{aligned} \tag{3}$$

where π_0 and $\pi_A = 1 - \pi_0$ represent respectively the proportion of funds with no performance ($\alpha_i = 0$) and differential performance ($\alpha_i > 0$ or $\alpha_i < 0$). The $Size$ stands for the probability of picking up a lucky fund. In statistical terms, the $Size$ corresponds to the probability of committing a type I error. The $Power$ gives the probability of finding

[5] The term false discovery is the statistical analogue of lucky fund. When someone finds a fund with a significant estimated alpha, he thinks he has made a discovery, namely a fund with differential performance. However, if this fund has in reality an alpha equal to zero (i.e. a lucky fund), it turns out be a false discovery.

[6] Strictly speaking, our definition corresponds to the positive False Discovery Rate ($pFDR$). The FDR is defined as $E\left(\left.\frac{F(\gamma)}{R(\gamma)}\right| R(\gamma) > 0\right) \cdot prob(R(\gamma) > 0)$. As the number of funds M in our database is large, the distinction between FDR and the $pFDR$ becomes irrelevant as $prob(R(\gamma) > 0)$ tends to one (see Storey (2002) for a discussion).

a fund with differential performance. It is equal to one minus the probability of making a type II error.

Equation (3) states that the FDR is a function of π_0 and the significance level γ. The FDR is positively related to π_0. If π_0 is high, there are only few funds with differential performance in the population. It implies that most significant funds are in fact lucky funds. The relation between γ and the FDR depends on the ratio $\frac{Size(\gamma)}{Power(\gamma)}$. A higher γ increases the $Size$ and thus the probability of picking up lucky funds. However, a higher γ also increases the $Power$ and thus the probability of finding funds with differential performance. Since both the $Size$ on the $Power$ are driven up as γ rises, the effect on the FDR depends on the distribution of the estimated alpha under H_0 and H_A.

2.2.2 The FDR among the Best and Worst Funds

Funds with differential performance are either characterized by positive or negative alphas. To determine the source of differential performance, the standard approach partitions the $R(\gamma)$ significant funds according to the sign of their estimated alphas. The first group contains the $R^+(\gamma)$ funds with positive estimated alphas. We refer to them as the best funds. Similarly, the second group is formed with the $R^-(\gamma)$ funds with negative estimated alphas. We call them the worst funds. At a second step, $R^+(\gamma)$ and $R^-(\gamma)$ are used as estimators of the number of funds with positive alphas and negative performance, respectively.

Unfortunately, these estimators are flawed like the estimator $R(\gamma)$ because they do not account for the presence of luck. Among the $R^+(\gamma)$ best funds, $F^+(\gamma)$ of them do not have a positive alpha, but are simply lucky. Similarly, $F^-(\gamma)$ among the $R^-(\gamma)$ worst funds do not yield a negative performance, but are lucky. As a result, the impact of luck on the performance of the best and worst funds can be very different according to the proportion of lucky funds among these two groups.

To measure the relative importance of $F^+(\gamma)$ and $F^-(\gamma)$, we use a new methodology which allows us to compute the FDR among the best and worst funds. Suppose that at a given significance level γ, $F(\gamma)$ among the $R(\gamma)$ significant ones are lucky funds. Since the test of the null hypothesis H_0 of no performance is a two-sided, symmetric test, we expect that half of these lucky funds have positive estimated alphas and half of them negative estimated alphas[7]. Because lucky funds are by definition drawn from H_0, this result is independent of the proportion of funds with positive and negative alphas in the population. We can therefore divide $F(\gamma)$ into two equal components, $F^+(\gamma)$ and $F^-(\gamma)$, which respectively denote the number of lucky funds among the best and worst funds. By analogy with the definition of the FDR, the FDR among the best and worst funds (denoted by $FDR^+(\gamma)$ and $FDR^-(\gamma)$) can be written as:

$$FDR^+(\gamma) = E\left(\frac{F^+(\gamma)}{R^+(\gamma)}\middle| R^+(\gamma) > 0\right) = E\left(\frac{\frac{1}{2}\cdot F(\gamma)}{R^+(\gamma)}\middle| R^+(\gamma) > 0\right), \tag{4}$$

$$FDR^-(\gamma) = E\left(\frac{F^-(\gamma)}{R^-(\gamma)}\middle| R^-(\gamma) > 0\right) = E\left(\frac{\frac{1}{2}\cdot F(\gamma)}{R^-(\gamma)}\middle| R^-(\gamma) > 0\right). \tag{5}$$

These new measures are used to separately estimate the proportion of lucky funds in the two tails of the cross-sectional alpha distribution.

2.3 Estimation Procedure

To estimate the empirical counterpart of the $FDR(\gamma)$ defined in Equation (2), we use the following estimator proposed by Storey (2002) and Storey and Tibshirani (2003):

$$\widehat{FDR}_\lambda(\gamma) = \frac{M\cdot\widehat{\pi}_0(\lambda)\cdot\gamma}{\#\{\widehat{p}_i < \gamma\}} = \frac{\widehat{F}(\gamma)}{\widehat{R}(\gamma)}, \tag{6}$$

where $\widehat{F}(\gamma)$ denotes the estimated number of lucky funds. It is computed as $M\cdot\widehat{\pi}_0(\lambda)\cdot\gamma$, where $\widehat{\pi}_0(\lambda)$ is the estimated proportion of funds with zero alphas in the total population of M funds. It depends on the parameter λ defined below. $\widehat{R}(\gamma)$ stands for the observed number of significant funds at the significance level γ, and is equal to funds with a p-value $\widehat{p}_i$ inferior to γ.

The $\widehat{FDR}_\lambda(\gamma)$ is easy to compute from the estimated p-values of the M funds. All that is needed is an estimator of π_0. The intuition behind the computation of $\widehat{\pi}_0$ is the following. Under H_0, the p-values are known to be uniformly distributed over the interval $[0,1]$. On the contrary, the p-values under H_A are extremely small because they are associated with extreme positive or negative estimated alphas. We can exploit this information to estimate $\widehat{\pi}_0$ without specifying the exact distribution of the p-values un-

[7] Technically speaking, the p-values associated with the $F(\gamma)$ funds with zero alphas are uniformly distributed on $[0,\gamma]$. Therefore, we expect half of them to end up in the right tail of the cross-sectional alpha distribution and half of them in the left tail.

der H_A. As an illustration, Figure 1 represents an histogram of the estimated $p-$values from a set of Monte-Carlo simulated data (the details of the design are given in the Appendix). Consistently with the size of our database, we set $M = 1'472$. We assume that 80% of the funds have an alpha equal to zero. The remaining funds yield an annual alpha of +5% or -5% with equal probabilities.

Please insert Figure 1 here

The high concentration of p-values near zero is due to the existence of 20% of the funds with differential performance. On the contrary, the histogram density is fairly flat between 0.3 and 1. In this region, the p-values are mostly drawn from the uniform distribution under H_0. Therefore, by taking a sufficiently high threshold λ (for instance 0.5), we can exploit the histogram density beyond λ to obtain a conservative estimate of the proportion π_0 of non-performing funds:

$$\widehat{\pi}_0(\lambda) = \frac{\#\{\widehat{p}_i > \lambda\}}{(1-\lambda) \cdot M} = \frac{\widehat{W}(\lambda)}{(1-\lambda) \cdot M}, \tag{7}$$

where $\widehat{W}(\lambda)$ denotes the number of estimated p-values superior to λ. The simplest way to define the parameter λ consists in eye-balling the histogram of p-values illustrated in Figure 1. In this paper, we use a more rigorous bootstrap procedure proposed by Storey (2002). The latter chooses λ such that the mean-squared error of $\widehat{\pi}_0(\lambda)$ is minimized.

An important property of $\widehat{FDR}_\lambda(\gamma)$ is that it yields a conservative estimate of $FDR(\gamma)$, meaning that $\lim_{M\to\infty} \widehat{FDR}_\lambda(\gamma) - FDR(\gamma) \geq 0$ with probability one for all γ. This result is robust to the presence of many forms of dependence in the estimated p-values such as dependence in finite blocks or ergodic dependence (Storey, Taylor and Siegmund (2004)).

Using a similar approach, we can estimate the empirical counterparts of $FDR^+(\gamma)$ and $FDR^-(\gamma)$ defined in Equations (4) and (5):

$$\widehat{FDR}^+_\lambda(\gamma) = \frac{\frac{1}{2} \cdot M \cdot \widehat{\pi}_0(\lambda) \cdot \gamma}{\#\{\widehat{p}^+_i < \gamma\}} = \frac{\widehat{F}^+(\gamma)}{\widehat{R}^+(\gamma)}, \tag{8}$$

$$\widehat{FDR}^-_\lambda(\gamma) = \frac{\frac{1}{2} \cdot M \cdot \widehat{\pi}_0(\lambda) \cdot \gamma}{\#\{\widehat{p}^-_i < \gamma\}} = \frac{\widehat{F}^-(\gamma)}{\widehat{R}^-(\gamma)}, \tag{9}$$

where $\widehat{p_i^+}$ and $\widehat{p_i^-}$ correspond to the p-values of the best and worst funds, $\widehat{F}^+(\gamma)$ and $\widehat{F}^-(\gamma)$ denote the estimated number of false discoveries among the best and worst funds and $\widehat{R}^+(\gamma)$ and $\widehat{R}^-(\gamma)$ stand for the observed number of best and worst funds. By combining Equations (6), (8) and (9), we have:

$$\widehat{FDR}_\lambda(\gamma) = w \cdot \widehat{FDR}_\lambda^+(\gamma) + (1-w) \cdot \widehat{FDR}_\lambda^-(\gamma), \tag{10}$$

where $w = \widehat{R}^+(\gamma)/\widehat{R}(\gamma)$. Therefore, the estimated FDR among the significant funds is a weighted average of the estimated FDR among the best and worst funds, where the weights are equal to the relative importance of the best and worst funds among the significant funds.

To examine the performance of the estimators $\widehat{\pi}_0(\lambda)$, $\widehat{FDR}_\lambda(\gamma)$, $\widehat{FDR}_\lambda^+(\gamma)$ and $\widehat{FDR}_\lambda^-(\gamma)$ we have run Monte-Carlo simulations which match our performance analysis setting. The results are presented in the Appendix. They show that all of these estimators are very close to the true values independently of the choice of the true parameters and the significance level γ.

3 Performance Measurement and Data Description

3.1 Asset Pricing Models

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

$$r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + \varepsilon_{i,t}, \tag{11}$$

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, whereas $r_{smb,t}$, $r_{hml,t}$, and $r_{mom,t}$ are the month t returns on zero-investment factor-mimicking portfolios for size, book-to-market and momentum. ε_{it} stands for the residual term. Adding momentum to the three-factor Fama-French model (1996) allows 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 Carhart model to account for the time-variation of factor exposures (Ferson and Schadt (1996)). This conditional model is similar to the model proposed by Kosowski et al. (2005) and is written as:

$$r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + B^{'} \left(z_{t-1} \cdot r_{m,t} \right) + \varepsilon_{i,t}, \quad (12)$$

where z_{t-1} denotes the $J \times 1$ vector of centered predictive variables and B is the $J \times 1$ vector of coefficients. Following Ferson and Schadt (1996), 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 also computed the fund alphas using the CAPM and the Fama-French model as well as conditional versions of these models. For sake of brevity, these results are summarized in the next Section.

3.2 Estimation of the p-values

Kosowski et al. (2005) show that the distribution of the fund estimated alphas in finite samples is non-normal. Therefore, we use bootstrap techniques to make inference on the fund performance. From bootstrap theory on higher order improvements, we know that the bootstrap is expected to yield better results when applied to asymptotic pivots[8]. We know that the fund estimated $t-$stat $\widehat{t}_i$ is asymptotically pivotal. It is defined as

$$\widehat{t}_i = \frac{\widehat{\alpha}_i}{\widehat{\sigma}_{\alpha_i}}, \quad (13)$$

where $\widehat{\alpha}_i$ is the fund estimated alpha and $\widehat{\sigma}_{\alpha_i}$ denotes a consistent estimator of the asymptotic standard deviation of $\widehat{\alpha}_i$ based on the Newey-West procedure (1987). For this reason, we use the t-stat instead of the alpha to compute the p-values under the null H_0 of no performance. Another advantage of the t-stat is that it reduces the presence of extreme observations due to volatile funds because the estimated alpha is divided by its standard deviation.

The bootstrap consists in approximating the distribution of $\left(\widehat{t}_i - t_i\right)$ by the distribution of $\left(\widehat{t}_i^* - \widehat{t}_i\right)$, where t_i is the fund $t-$stat and $\widehat{t}_i^*$ the bootstrapped $t-$stat. To compute the

[8] A test statistic is asymptotically pivotal if its asymptotic distribution does not depend on unknown population parameters. Pivotal test statistics have lower coverage errors and have more power than non-pivotal statistics (Davison and Hinkley (1997), Horowitz (2001), Romano and Wolf (2005)).

distribution of $\left(\widehat{t}_i^* - \widehat{t}_i\right)$ for each fund i $(i = 1, ..., M)$, we use a parametric bootstrap procedure based on residual resampling[9]. Since our procedure is similar to the one implemented by Kosowski et al. (2005), we refer to them for further details.

For each bootstrap iteration q $(q = 1, ..., Q)$, we draw with replacement from the estimated residuals $\{\widehat{\varepsilon}_{i,t}\}$. From the resampled residuals $\{\widehat{\varepsilon}_{i,t}^{*q}\}$, we create a new time-series of monthly excess return $\{r_{i,t}^{*q}\}$ by imposing that α_i is equal to zero. By regressing $r_{i,t}^{*q}$ on the factors, we compute $\widehat{\alpha}_i^{*q}$ and $\widehat{\sigma}_{\alpha_i}^{*q}$ to obtain the estimated $\widehat{t}_i^{*q}$. After repeating the same procedure Q times, the bootstrap p-value $\widehat{p}_i$ of the fund i is computed as follows:

$$\widehat{p}_i = \frac{1}{Q} \sum_{q=1}^{Q} I\{|\widehat{t}_i^{*q}| > |\widehat{t}_i|\}. \tag{14}$$

To approximate the distribution of $\left(\widehat{t}_i^* - \widehat{t}_i\right)$ with sufficient accuracy, we fix the number of iterations Q to 1'000.

3.3 Mutual Fund Data

We measure the performance of U.S. open-end, domestic equity funds on a monthly basis. We use monthly net return data provided by the Center for Research in Security Prices (CRSP) between January 1975 and December 2002[10]. The CRSP database is matched with the CDA database (from Thomson Financial) in order to obtain the fund investment objectives. We require that each fund has at least 60 monthly return observations to estimate its alpha and t-stat. Since we use the same mutual fund database as Wermers (2000) and Kosowski et al. (2005), we refer to them for a precise description of these two databases (as well as the matching technique).

Our final fund universe (denoted by *All*) is composed of 1'472 funds that exist for at least 60 months between 1975 and 2002. Funds are then classified into three investment categories: Growth funds (G), Aggressive Growth funds (AG), and Growth and Income

[9] To know whether this approach is appropriate, we checked for the presence of autocorrelation (with the Ljung-Box test), heteroscedasticity (with the White test) and Arch effects (with the Engle test) in the fund residuals. We found that only few funds presented some of these features. We 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. The results remain unchanged.

[10] If the fund proposes 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.

funds (GI)[11]. A fund is included in a given investment category if its investment objective corresponds to the investment category for at least 60 months. While there are some Balanced and Income funds among the *All* funds, we do not consider this investment category separately because there are not enough funds to accurately estimate the FDR.

Table 2 shows the average mutual fund performance across the four investment categories (*All*, G, AG, GI). For each investment category, we compute the alpha (expressed in percent per year) and factor exposures of an equally-weighted portfolio including all funds existing at a given point in time. Panel A and B show the results produced by the unconditional and conditional Carhart models, respectively.

Please insert Table 2 here

The average alpha is always negative. Similarly to Daniel et al. (1997), AG funds have significant positive momentum and negative book-to-market exposures, whereas it is the oppostive for GI funds. Introducing time-varying market betas does not greatly modify the results shown in Panel A. Since the empirical analysis of the FDR based on the two models is extremely close, the analysis presented in the next Section is based on the unconditional Carhart model.

4 Empirical Analysis

4.1 Illustrating the Drawbacks of the Standard Approach

We begin our empirical analysis by applying the standard approach to our mutual fund database. The results across the four investment categories are given in Panels A, B, C, and D of Table 3. The left part of each Panel, contains the observed number of significant, best and worst funds, respectively denoted by $\widehat{R}$, $\widehat{R}^+$, and $\widehat{R}^-$ at different significant levels γ ($\gamma = 0.05$, 0.10, 0.15 and 0.20). The right part shows the percentage of the significant, best and worst funds, respectively denoted by $\widehat{R}/M$, $\widehat{R}^+/M$, and $\widehat{R}^-/M$.

Please insert Table 3 here

Three main comments stem from the analysis of Table 3. First, by comparing $\widehat{R}^+$ and $\widehat{R}^-$ at different significance levels γ, we observe a predominance of the worst funds over the best ones across the four investment categories. This finding is also documented

[11] For more details about the equity style classifications, see Christopherson (1995), Grinblatt, Titman and Wermers (1995), and Brown and Goetzmann (1997).

by Jensen (1968) who finds a large proportion of funds with significant negative alphas. Ferson and Schadt (1996) reach the same conclusion with unconditional models[12]. Second, the percentage of significant funds varies across the various investment categories. The percentage $\widehat{R}/M$ is higher for AG and GI funds than for All and G funds independently of γ. However, the number of significant fund $\widehat{R}$ is logically always higher for All and G than for AG and GI because of the larger size of the first two categories. Third, as γ rises, $\widehat{R}$, $\widehat{R}^+$, and $\widehat{R}^-$ increase significantly.

These results suggest that some funds across the four investment categories do generate differential performance. A majority of these funds seem to produce negative alphas, but a non-negligible proportion appear to generate positive alphas. However, these statements are inaccurate because they are based on estimators, $\widehat{R}$, $\widehat{R}^+$ and $\widehat{R}^-$, which do not account for luck. Therefore, it is impossible to correctly measure the presence of differential, positive and negative performance. For instance, we find that 80 All funds have positive estimated alphas at $\gamma = 0.10$. But do all these funds generate a positive performance or are many of them simply lucky?

Second, the standard approach provides no information about the location of funds with differential performance in the tails of the cross-sectional alpha distribution. To take a concrete example, we observe that the number of the worst G funds increases by 85 as γ rises from 0.05 to 0.15. If these 85 funds are all lucky funds, we know that the few funds with negative performance have p-values inferior to 0.05. We would conclude that these funds are located at the extreme left tail of the distribution and are more likely to generate highly negative alphas. On the contrary, if none of the 85 funds are lucky, we would say that the funds with negative performance are largely spread in the left tail.

Finally, we cannot compare the performance between the different investment categories. At $\gamma = 0.20$, we observe that the percentage of best funds is similar across the G and GI funds. However, if the GI funds turn out to contain more lucky funds than the G funds, the real performance of the G funds can be much higher than the one of GI funds. In order to answer these questions, we need to determine the proportion of lucky funds with the FDR.

[12] However, they show that the percentages of worst and best funds become similar when conditional models are used. Contrary to them, we do not find striking differences between unconditional and conditional models. This can be due to the fact that our mutual fund data and asset pricing models are different and that the $p-$values are computed with bootstrap techniques instead of standard asymptotic approximations.

4.2 Estimating the Impact of Luck on Performance

4.2.1 The Proportion of Funds with Zero Alphas

The first step to compute the FDR consists in estimating the proportion π_0 of funds with zero alphas with Equation (7). The figures shown in Table 4 indicate that 76.5% of funds in the population have zero alphas. It implies that 23.5% of the funds generate either positive or negative alphas. While the percentage of G funds with zero alphas is similar to the one of *All* funds (80.2%), this proportion is lower for AG and GI funds (70.5% and 74.8% respectively). These results show that although the majority of funds are not able to beat the market, they do not yield negative risk-adjusted returns.

Please insert Table 4 here

A few papers (Jensen (1968), Kosowski et al. (2005)) have proposed a simple method to measure the impact of luck by assuming that π_0 is equal to one. It implies that the expected number and proportion of lucky funds at a given significance level γ are respectively given by $M \cdot \gamma$ and γ. However, Table 4 clearly shows that $\widehat{\pi}_0$ is never equal to one. Therefore, this method overestimates the impact of luck because it does not account for the proportion π_A of funds which truly yield non-zero alphas. This estimation of luck can become very inaccurate as γ rises. For instance, the number of lucky *All* funds are overestimated by 35 at $\gamma = 0.10$ and by 70 at $\gamma = 0.20$. Stated differently, this method assumes that the proportion of lucky funds is equal to 10% and 20% at $\gamma = 0.10$ and 0.20, while it amounts in reality to 7.5% and 15.3%. These approximations are even worse for AG funds, since $\widehat{\pi}_0^{AG}$ is only equal to 70.5%.

4.2.2 The FDR of All Funds

To measure the impact of luck on mutual fund performance, we measure the proportion of lucky funds among three sets of funds. The first one is the set of significant funds. The second and third ones correspond to the best and worst funds. We compute the FDR among these three groups at different significant levels γ ($\gamma = 0.05$, 0.10, 0.15 and 0.20). The results across the four investment categories are displayed in Panels A, B, C and D of Table 5. For the set of significant funds, the left part of each Panel displays the $\widehat{FDR}$, the number of significant funds $\widehat{R}$, the number of lucky funds $\widehat{F}$, and the number of funds with differential performance $\widehat{T}$ equal to $\widehat{R} - \widehat{F}$. The right part of each Panel shows the proportion of significant funds $\widehat{R}/M$, the proportion of lucky funds $\widehat{F}/M$ and the proportion of funds with differential performance $\widehat{T}/M$. For the set of the best or worst funds, the information provided is identical except that $\widehat{FDR}$ is respectively replaced by $\widehat{FDR}^{+}$ or $\widehat{FDR}^{-}$, $\widehat{R}$ by $\widehat{R}^{+}$ or $\widehat{R}^{-}$, $\widehat{F}$ by $\widehat{F}^{+}$ or $\widehat{F}^{-}$, and $\widehat{T}$ by $\widehat{T}^{+}$ or $\widehat{T}^{-}$.

We begin our analysis with the results of *All* funds summarized in Panel A. At the conventional level γ of 0.05, the $\widehat{FDR}$ amounts to 35.8%, indicating that 101 out of the 157 significant funds generate differential performance. As γ rises, the number of lucky funds $\widehat{F}$ increase more quickly than the number of funds with differential performance $\widehat{T}$. Therefore, the FDR is equal to 52.1% at $\gamma = 0.20$, which implies that only half of the 432 significant funds have non-zero alphas.

From Equation (10), we know that the $\widehat{FDR}$ is a weighted average of $\widehat{FDR}^{+}$ and $\widehat{FDR}^{-}$. These two components can be very different from the $\widehat{FDR}$ as long as these differences offset each other. This is exactly what we observe since the $\widehat{FDR}^{+}$ is much higher than the $\widehat{FDR}$ at all significance levels γ. At $\gamma = 0.05$, $\widehat{FDR}^{+}$ is equal to 56.3%. It means that 28 among the 50 best *All* funds have in reality alphas equal to zero. As γ rises, the number of lucky funds $\widehat{F}^{+}$ grows at a higher pace than the number of funds with positive alphas $\widehat{T}^{+}$. This increased presence of luck among the best funds leads to a sharpe increase in $\widehat{FDR}^{+}$. On the contrary, the $\widehat{FDR}^{-}$ is close to the $\widehat{FDR}$. This reflects the fact that $\widehat{FDR}$ depends more heavily on $\widehat{FDR}^{-}$ because the proportion of worst funds $\widehat{R}^{-}/\widehat{R}$ is higher than the proportion of best funds $\widehat{R}^{-}/\widehat{R}$. At $\gamma = 0.05$, $\widehat{FDR}^{-}$ only amounts to 26.3%. Stated differently, 73.7% of the worst funds truly have a negative alphas. As γ rises, the number of funds with negative alphas $\widehat{T}^{-}$ grows at a slightly lower rate than the lucky funds $\widehat{F}^{-}$. As a result, the $\widehat{FDR}^{-}$ increases only slowly.

First of all, these results clearly show that the difference between the FDR among the best and worst funds is striking. It implies that luck has a much larger impact on the performance of the best funds than the worst funds. Stated differently, the proportion of lucky funds is always higher among the best funds at any significance levels γ. Second, the results highlight the inaccuracy of the performance assessment under the standard approach. The latter concludes that 9.4% of the funds are able to achieve positive alphas at $\gamma = 0.20$. However, the $\widehat{FDR}^{+}$ shows a completely different picture. Only 1.7% of the funds generate positive alphas, while the remaining funds (7.6%) are purely lucky. Moreover, our analysis confirm that there is a larger proportion of funds with negative rather than positive performance. However, the standard approach concludes that 20% of the funds have negative alphas at $\gamma = 0.20$, while our estimation is only equal to 12.4%.

Examining, the evolution of $\widehat{T}^{+}/M$ and $\widehat{T}^{-}/M$ allows us to determine the location of the funds with differential performance in the tails of the cross-sectional alpha distribution. As γ rises, $\widehat{T}^{+}/M$ remains constant around 1.6%. It implies that the few performing funds are located at the extreme right tail since their p-values are below or equal to 0.05. On the contrary, $\widehat{T}^{-}/M$ continuously increases as γ rises. Therefore, the funds with negative performance are not located at the extreme left tail because their p-values are largely spread in the interval $[0, 0.20]$.

Please insert Table 5 here

4.2.3 The FDR of the Growth Funds

The results for G funds are summarized in Panel B of Table 5. The FDR among the significant and worst G funds are similar to those observed of the *All* funds. On the contrary, the FDR among the best funds follows a different pattern. It starts at an extremely high level equal to 73.4% at $\gamma = 0.05$. This is 17.1% higher than the $\widehat{FDR}^{+}$ of the *All* funds. It means that only 7 out of the 28 best G funds are able to deliver a positive performance. At $\gamma = 0.20$, this is even worst since almost 90% of the best funds are just lucky.

Since the $\widehat{FDR}^{+}$ is much higher than the $\widehat{FDR}^{-}$ at all significance levels γ, luck has a more pronounced impact on the best rather than the worst funds. These results also show why the standard approach leads to wrong conclusions regarding the real performance of the G funds. At $\gamma = 0.20$, the latter estimates that a non-negligible proportion of funds (8.8%) yield positive alphas. After accounting for luck, we find that only a tiny fraction of the G funds equal to 1.3% is capable of producing a positive performance. Moreover, the standard approach concludes that 18.8% of the funds yield negative alphas at $\gamma = 0.20$, while our estimate of this proportion is equal to 11.3%.

We observe that $\widehat{T}^{+}/M$ increases by 0.6% as γ rises from 0.05 to 0.15 and then remains constant at 1.3%. It implies that the funds with positive alphas are fairly concentrated in the right tail since their p-values are below 0.15. Similarly to *All* funds, $\widehat{T}^{-}/M$ increases continuously as γ rises. Therefore, the funds with negative performance are largely spread in the left tail of the distribution since their associated p-values span an interval larger than $[0, 0.20]$.

4.2.4 The FDR of the Aggressive Growth Funds

Panel C of Table 5 contains the results for the *AG* funds. This investment category has the highest proportion π_A of funds with differential performance. At a given significance threshold γ, a higher π_A reduces the number of lucky funds and increases the number of funds with differential performance. It is therefore not surprising to observe that the $\widehat{FDR}$ is lower than those of the *All* and *G* funds.

The most striking result comes from the low level of the FDR among the best funds. At $\gamma = 0.05$, the $\widehat{FDR}^+$ is only equal to 22.9%, implying that only 4 out the 18 best funds are lucky. As γ rises, the number of lucky funds $\widehat{F}^+$ increases more quickly than the number of funds with positive alphas $\widehat{T}^+$. This contributes to increase the $\widehat{FDR}^+$ by 20%. However, its level remains largely inferior to the figures documented for *All* and *G* funds. Concerning the worst funds, the $\widehat{FDR}^-$ starts at the same level as the $\widehat{FDR}^+$. However, the $\widehat{FDR}^-$ rises by only 7% as γ passes from 0.05 to 0.20.

The impact of luck on the performance of the best and worst funds is similar because the proportions of lucky funds among these two groups is approximately equal. For this reason, we reach the same qualitative conclusions as the standard approach. Indeed, we do find a significant proportion of funds with positive and negative alphas. However, the estimation proposed by the standard approach are still largely inflated. While the latter finds that 14.9% and 20.5% of the funds yield positive and negative alphas respectively, our FDR analysis leads to percentages equal to 8.8% and 14.0%.

$\widehat{T}^+/M$ becomes constant at $\gamma = 0.15$ is reached. It indicates that the funds with positive performance are fairly concentrated in the right tail since their p-values are below or equal to 0.15. Similarly to *All* and *G* funds, the increase in $\widehat{T}^-/M$ is quite strong as γ rises. Therefore, the funds with negative performance are largely spread in the left tail of the distribution.

4.2.5 The FDR of the Growth and Income Funds

The results for the *GI* funds are displayed in Panel D of Table 5. The FDR among the significant *GI* funds is similar to those observed for *All* and *G* funds. However, the patterns of the FDR among the best and worst funds are completely different from these two investment categories. First, the $\widehat{FDR}^+$ is equal to 100% independently of γ. It implies that the best funds are all lucky funds. For instance, the 21 best funds discovered at $\gamma = 0.20$ all have zero alphas. Second, the $\widehat{FDR}^-$ starts at 20.7% and increases extremely slowly as γ rises.

Our results show that the impact of luck on the performance of the best funds is extremely strong, since no single GI fund is able to produce a positive alpha. This contradicts the conclusions obtained with the standard approach, which finds that 6.7% GI funds generate positive alphas.

Finally, the constant increase in $\widehat{T}^-/M$ indicates that the funds with negative performance are largely spread in the left tail.

4.2.6 Comparative Analysis

To compare the impact of luck across the four investment categories, Figure 2 plots the FDR among the best and worst funds at different significance levels γ. The solid line represents the $\widehat{FDR}^+$ and the dashed one the $\widehat{FDR}^-$. The $\widehat{FDR}^-$ is similar across the four categories. Its initial value is low and its weak slope indicates that many funds with negative performance are discovered as γ rises. It confirms the fact that these funds are dispersed in the left tail of the distribution. Although the $\widehat{FDR}^+$ differs significantly across the four investment categories, it generally starts at higher levels than $\widehat{FDR}^-$. Moreover, it increases more steeply as γ rises because the few funds with positive performance are not largely dispersed in the right tail. The $\widehat{FDR}^+$ of the two smallest investment categories shows extreme patterns. First, the $\widehat{FDR}^+$ of the GI is always equal to one, since none of the funds is able to produce positive alphas. Second, $\widehat{FDR}^+$ of the AG funds is low indicating a non-negligible proportion of funds generate a positive performance.

Please insert Figure 2 here

4.3 Implications for Mutual Fund Performance Analysis

Table 4 show that a sizable proportion π_A of funds in the population yields non-zero alphas. An important issue is to know whether most of these funds generate positive or negative performance. In order to answer this question, we must decompose π_A into the proportion π_A^+ of funds with positive alphas and the proportion π_A^- of funds with negative alphas. By definition, the proportion π_A can be written as (see the description in Table 1):

$$\pi_A = \pi_A^+ + \pi_A^- = \frac{(T^+(\gamma) + A^+(\gamma)) + (T^-(\gamma) + A^-(\gamma))}{M}, \tag{15}$$

where $A^+(\gamma)$ $(A^-(\gamma))$ denotes the number of funds with positive (negative) alphas which do not have significant p-values (i.e. they are not rejected). Decomposing π_A is therefore not trivial since it depends on the unobservable quantities $A^+(\gamma)$ and $A^-(\gamma)$.

To tackle this issue, we use the fact that as γ increases, the test of differential performance has more power and detect more funds with differential performance. Hence, if the tails of the t-stat distribution under H_A decrease monotonically[13] both $T^+(\gamma)$ and $T^-(\gamma)$ go up while $A^+(\gamma)$ and $A^-(\gamma)$ go towards zero. By continuously increasing γ, we can find γ^* which is the minimum significance level such that either $T^+(\gamma)$ or $T^-(\gamma)$ becomes constant. For illustrative purposes. say that this level is γ^* for $T^+(\gamma)$, while $T^+(\gamma)$ is still increasing. As a result, we know that $A^+(\gamma^*) > 0$ should be close to zero for all $\gamma > \gamma^*$. This result provides a methodology to compute π_A^+ and π_A^- which can be described as follows. First, we can estimate the proportion $\widehat{\pi}_A^+$ of funds with positive performance by taking $\widehat{T}^+(\gamma^*)/M$ since $A^+(\gamma^*)$ tends to zero. Second, we can deduce the proportion of funds with negative performance by using the following equality: $\widehat{\pi}_A = \widehat{\pi}_A^+ + \widehat{\pi}_A^-$.

Our methodology can easily be applied to our database since we observe that $\widehat{T}^+(\gamma)/M$ remains constant among the four investment categories after a certain significance level γ is reached (i.e. $\gamma = 0.20$). Therefore, we can estimate $\widehat{\pi}_A^+$ by $\widehat{T}^+(\gamma = 0.20)/M$ and compute $\widehat{\pi}_A^-$ such that $\widehat{\pi}_A = \widehat{\pi}_A^+ + \widehat{\pi}_A^-$.

The decomposition presented in Table 6 indicates that the vast majority of funds with differential performance distinguish themselves by their poor performance. Except for the *AG* funds, more than 90% of the funds with differential performance generate negative risk-adjusted returns. Expressed as a percentage of the fund population, the proportion of funds with positive alphas is extremely low. The only exception comes from the

[13] Note that this feature is shared asymptotically by most test statistics. Indeed, standard tests are asymptotically distributed as a normal (or khi-square) variable under the null and as a *non-central* normal (or khi-square) variable under the alternative.

AG fund category, which contains 8.4% of funds with positive alphas[14]. Moreover, the proportion of funds with negative alphas in the fund population is approximately equal to 20% across the four investment categories. From an overall perspective, we observe more frequently funds with negative rather than positive performance. However, the performance of the mutual fund industry is not so bad since around 80% of the funds yield zero alphas. In fact, the negative average performance documented in the literature (and in Table 2) is only caused by the poor performance of 20% of the funds. Moreover, mutual funds can be close substitutes for systematics risk factors (such as those of the Carhart model) which are unavailable for investment (Pastor and Stambaugh (2000b)). For this reason, active funds can still be valuable investments even though most of them do not yield positive alphas. Finally, the negative performance generated by the funds should not be extreme because these funds are dispersed in the extreme left tail of the cross-sectional alpha distribution. If their alphas were very high, they would surely be located at the extreme left tail.

Please insert Table 6 here

4.4 Implications for Mutual Fund Portfolio Management

In the recent years, new management techniques have been developed in order to form strategies generating positive alphas (Bernstein (2003), Kung et Pohlman (2004))[15]. This quest for alpha has led to the creation of funds of mutual funds. Their objective is to build portfolios of funds with positive alphas. Our results indicate that there exists a tiny but real evidence of positive performance among *All* and G funds and, to a greater extent, among AG funds. An important issue regarding mutual fund portfolio management is to know whether this evidence is strong enough to generate portfolios of funds with positive alphas.

As it is shown in the following proposition, the FDR among the best funds forming the portfolio is a key factor determining the portfolio expected alpha. Therefore, once we know the FDR^+, we can gauge the expected portfolio alpha.

Proposition 4.1 *Let us denote by* α_P^γ *the expected alpha of an equally-weighted portfolio* P *of the best funds at the significance level* γ. *We set* $\gamma \geq \gamma^0(M)$, *where* $\gamma^0(M) =$

[14] This finding is consistent with the previous literature documenting a positive performance of AG funds (Grinblatt and Titman (1993) and Daniel et al. (1997)).

[15] One of these techniques is called portable alpha. Under this approach, the optimal porfolio is broken into a beta and an alpha portfolio. The beta portfolio return is generated by exposures to systematic sources of risk, while the alpha portfolio return is driven by selection skills.

$\inf_{\gamma}\{\gamma : prob(R^{+}(\gamma) > 0) = 1\}$. *The expected alpha of P can be written as:*

$$\alpha_P^{\gamma} = FDR^{+}(\gamma) \cdot \alpha_0 + \left(1 - FDR^{+}(\gamma)\right) \cdot \alpha_A^{+} \tag{16}$$

where $FDR^{+}(\gamma)$ is the FDR among the best funds defined in Equation (4). α_0 denotes the fund alpha under H_0. α_A^{+} stands for the fund alpha under H_A with $\alpha_A > 0$.

Proof. See the Appendix. ■

Using Equation (16), we can compute the expected alpha of an equally-weighted portfolio of the best *All*, *G* and *AG* funds. We exclude the *GI* funds since none of them produces positive alphas. We set α_0 equal to zero. To estimate α_A^{+} in a conservative way, we rank all funds in decreasing order according to their estimated alphas and select the alpha of the fund located at the 5%-quantile[16]. It respectively amounts to 5.3%, 5.8% and 7.7% for *All*, *G* and *AG* funds. The FDR among the best funds is estimated with the $\widehat{FDR}^{+}$. Table 7 displays the values taken by $\widehat{FDR}^{+}$ and α_P at different significance levels γ ($\gamma = 0.05$, 0.10, 0.15 and 0.20). $d\alpha_P/\alpha_P$ denotes the relative reduction of the portfolio alpha as γ increases by 0.05. It has the nice property of being independent of the value chosen for α_A^{+}.

We observe that the few *All* funds with positive performance are sufficient to generate a positive alpha equal to 2.62% per year at $\gamma = 0.05$. This result may be surprising in light of the small proportion of these funds. However, the possibility to form portfolios with positive alphas depends not only on the proportion of funds with positive performance, but also on their location in the right tail of the distribution. Our previous analysis shows that these few *All* funds are located at the extreme right tail. Therefore, by choosing a sufficiently low γ, we can partially separate these funds from the lucky ones. This is exactely what the $\widehat{FDR}^{+}$ tells us: although there are 1.7% of *All* funds with positive alphas, these funds represent almost half of the funds in the portfolio at $\gamma = 0.05$. As γ rises, the relative reduction of the alpha is substantial. This is not surprising because the only new funds which enter the portfolio are lucky funds, which greatly reduces the performance. The results of the portfolios of *G* funds are similar to those of *All* funds, except that the initial portfolio alpha is lower because the $\widehat{FDR}^{+}$ of *G* funds is higher at $\gamma = 0.05$. Finally, the portfolio of *AG* funds generate a substantial alpha equal to 6.16% per year at $\gamma = 0.05$. This is striking performance is due to the low level of $\widehat{FDR}^{+}$, implying that most of the funds in the portfolio generate positive

[16] Taking lower quantiles would reduce the estimated fund alpha under H_A. However, it would not overturn the main conclusion of our analysis, which depends primarily on the level of the FDR.

alphas. Moreover, as γ rises, the decline is less pronounced than for *All* and *G* funds, because the *AG* funds with positive performance are more dispersed at the right tail of the distribution.

4.5 Sensitivity Analysis

4.5.1 Alternative Asset Pricing Models

Table 8 contains the FDR among the best and worst funds at $\gamma = 0.05$ and 0.20 computed with the unconditional and conditional versions of the CAPM and Fama-French (FF) models. The results related to the four investment categories are displayed in Panels A, B, C and D, respectively. When the unconditional FF model is used, the patterns of $\widehat{FDR}^{+}$ and $\widehat{FDR}^{-}$ are similar to those found with the Carhart model. For instance, we still find a low $\widehat{FDR}^{-}$ across the four investment categories, a low $\widehat{FDR}^{+}$ for *AG* funds and a $\widehat{FDR}^{+}$ equal to 100% for *GI* funds.

On the contrary, the results obtained with the unconditional CAPM are quite different from those obtained with the Carhart model. In particular, both the $\widehat{FDR}^{+}$ and the $\widehat{FDR}^{-}$ are higher across the four investment categories. It implies that the estimated CAPM-alphas of the best funds are lower than their Carhart-alphas. Similarly, the estimated CAPM-alphas of the worst funds are lower than their Carhart-alphas. This can be easily explained by the bias introduced in the estimated alpha when relevant explanatory variables are omitted from a linear regression model (Modest and Lehman (1987)). For instance, the estimated CAPM-alphas of the best *AG* funds are negatively biased because of the negative exposures of these funds to the book-to-market factor, which has a positive premium over the period. By the same token, the estimated CAPM-alphas of the worst *GI* funds are positively biased because of the positive exposures of these funds to the size and book-to-market factors, which both have positive premia. Finally, there are no striking difference between the unconditional and conditional versions of the CAPM and FF models.

4.5.2 Subperiod Analysis

In order to see whether the results are consistent throughout the investigated period, we form two subperiods of equal lengths (168 observations). The first period starts in January 1975 and ends in December 1988. During this period, there are 276 *All* funds and only 111 *G*, 54 *AG* and 63 *GI* funds. Because of the small size of these three categories, we only compute the FDR for *All* funds. The FDR among the best funds is lower than the one observed during the entire period. It respectively amounts to 35.3% and 43.3% at $\gamma = 0.05$ and 0.20. The fact that mutual fund performance is better during this period is also documented by Daniel et al. (1997). They justify this finding by the fact that market efficiency has improved over time and that the rapid increase in the number of mutual funds has diluted performance over time.

The second subperiod begins in January 1989 and ends in December 2002. The sample contains 1417 *All* funds and 976 G, 196 AG and 277 GI funds. During this period, the levels of the FDR across the four investment categories is extremely close to those documented for the entire period.

5 Conclusion

In this paper, we examine the impact of luck on mutual fund performance. To this end, we use the False Discovery Rate (FDR) in order to measure the proportion of lucky funds among the funds with significant estimated alphas. Our approach is a straightforward extension of the standard approach developed in the literature and is very easy to compute. By using the FDR, we are able to shed light on important issues that could not be addressed with the previous methodologies. In particular, the FDR allows us to determine the relative importance as well as location of funds with differential performance in the tails of the cross-sectional alpha distribution.

Our results based on 1'472 U.S. open-end equity funds between 1975 and 2002 clearly show that the impact of luck on performance is substantial. First, our estimators of the number of funds with differential, positive and negative performance is much lower than those obtained with the standard approach. It implies that our judgement on performance across the different investment categories can substantially differ from the one implied by the standard approach. Second, we find that luck has a stronger impact on the performance of the best rather than the worst funds. Across the four investment categories, the FDR among the worst funds is always inferior to 50% and increases slowly as γ rises. It means that the majority of worst funds truly yield negative alphas and that the latter are largely spread in the left tail of the alpha distribution. The FDR among the best funds is generally much higher than the FDR among the worst funds. For *All* and G funds, the FDR is always higher than 50%, while it amounts to 100% for the GI funds. The only exception comes from the AG funds. Its low FDR reveals that a sizable proportion of AG funds produce a positive performance.

Our results has important implications for mutual fund performance analysis. From an overall perspective, we observe more frequently funds with negative rather than positive performance. However, the performance of the industry as a whole is not so bad because about 80% of the funds produce zero alphas. In fact, the negative average performance documented in the previous literature is not due to the majority of funds but is only caused by 20% of the funds. Our analysis also has implications for mutual fund portfolio management. By computing the FDR among the best funds, we show that it is possible to form portfolios of *All* and G funds with positive alphas even though the evidence of positive performance among the *All* and G funds is very low. This is due to the fact that the funds with positive performance are located at the extreme right tail of the alpha distribution. Therefore, by choosing a sufficiently low significance level γ, it is possible to separate funds with positive alphas from the lucky ones.

Because the FDR can measure the proportion of funds in a given portfolio which yield positive alphas, it is a powerful tool to measure the expected performance of this portfolio. This result suggests that there is interesting work to be done. By controlling the FDR of the portfolio, it could be possible to identify ex ante portfolios that will produce high alphas in the future. This has important implications for the persistence literature (Hendricks, Patel and Zeckhauser (1993), Elton, Gruber and Blake (1996), Carhart (1997), Kosowski et al. (2005)), where the portfolios are not formed according to the FDR, but rather by dividing funds into deciles (octiles or quintiles).

6 Appendix

6.1 Proof of Proposition 2.1

The expected alpha of the portfolio P based on a significance threshold γ can be written as:

$$\alpha_P^\gamma = E\left(\alpha_P \left| R^+(\gamma) > 0\right.\right) \cdot prob\left(R^+(\gamma) > 0\right). \tag{17}$$

Since $prob\left(R^+(\gamma) > 0\right) = 1$ by assumption, $\alpha_P^\gamma = E\left(\alpha_P \left| R^+(\gamma) > 0\right.\right)$. As each fund in the portfolio P receives a weight $\frac{1}{R^+}$, we have:

$$\alpha_P^\gamma = E\left(\frac{1}{R^+}\sum_{i=1}^{R^+}\alpha_i \middle| R^+(\gamma) > 0\right). \tag{18}$$

Each fund share the same alpha α_0 under H_0 and the same α_A^+ under H_A with $\alpha_A^+ > 0$. Thus, Equation (18) becomes:

$$\alpha_P^\gamma = E\left(\frac{F^+(\gamma)}{R^+(\gamma)} \middle| R^+(\gamma) > 0\right) \cdot \alpha_0 + E\left(\frac{T^+(\gamma)}{R^+(\gamma)} \middle| R^+(\gamma) > 0\right) \cdot \alpha_A^+. \tag{19}$$

Since $T^+ + F^+ = R^+$, we have:

$$\alpha_P^\gamma = FDR^+(\gamma) \cdot \alpha_0 + \left(1 - FDR^+(\gamma)\right) \cdot \alpha_A^+ \tag{20}$$

6.2 Monte-Carlo Simulations

In this section, we examine the precision of $\widehat{\pi}_0(\lambda)$, $\widehat{FDR}_\lambda(\gamma)$, $\widehat{FDR}_\lambda^+(\gamma)$ and $\widehat{FDR}_\lambda^-(\gamma)$ based on Monte Carlo simulations. We generate artificial monthly return data according to the market model:

$$\begin{aligned} r_{i,t} &= \alpha_i + \beta \cdot r_{m,t} + \varepsilon_{i,t}, \qquad i = 1,...,M, \quad t = 1,...,T, \\ r_{m,t} &\sim N(0, \sigma_{r_m}), \qquad \varepsilon_{i,t} \sim N(0, \sigma_\varepsilon). \end{aligned} \tag{21}$$

For each fund i $(i = 1,...,M)$, we test the null hypothesis H_0 of no performance $(\alpha_i = 0)$ against the alternative H_A of differential performance $(\alpha_i > 0$ or $\alpha_i < 0)$. Under H_0, the t-stat t_i follows the Student distribution with $T-2$ degrees of freedom. Under H_A, t_i follows a noncentral student distribution with $T-2$ degrees of freedom whose true parameter of noncentrality can be well approximated by $\frac{T^{\frac{1}{2}}\alpha_A}{\sigma_\varepsilon}$ (Davidson and MacKinnon (2004), p. 169). Consistently with the size of our database, we set $M = 1'472$ and $T = 336$. The values for β, σ_{r_m} and σ_ε are based on sample estimates from the market model. β and σ_ε correspond to the cross-sectional average accross the funds and σ_{r_m} is the standard deviation of the market return. We therefore set $\beta = 0.97$, $\sigma_\varepsilon = 0.030$ and $\sigma_{r_m} = 0.046$. Residuals are assumed to be uncorrelated across funds.

We assume that a proportion π_0 of them are drawn from H_0 and have an alpha α_0 equal to zero. A proportion π_A of funds generate differential performance. Under H_A, funds either yield a positive alpha α_A^+ with probability $\pi_A^+ = \pi_A \cdot q^-$ or a negative alpha α_A^- with probability $\pi_A^- = \pi_A \cdot (1 - q^-)$, where $q^- \in [0,1]$ is a positive scalar. We thus have:

$$\begin{aligned} H_0 &: \alpha_i \sim N(0, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with probability } \pi_0 \\ H_A &: \alpha_i \sim N(\alpha_A^+, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with probability } \pi_A^+ \\ &: \alpha_i \sim N(\alpha_A^-, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with probability } \pi_A^- \end{aligned} \tag{22}$$

The experiment is done according to different parameter values. Three sets of α_A^+ and α_A^- are considered (in percent per year): (a) 8% and -5% (b) 5% and -5% (c) 5% and -8%. These figures are close to the average estimated alphas of funds in the top and worst deciles which amount to 6.5% and 5.52% per year. Since these two deciles contain lucky funds which drive the estimated alphas near zero, our parameter values are therefore conservative estimates of the true α_A^+ and α_A^-. π_0 is set in turn to (a) 0.7 and (b)

0.9. Finally, q^- is set to (a) 0.3 and (b) 0.7. Two significance levels γ are examined: (a) 0.05 and (b) 0.10. For simplicity we set the tuning parameter λ of Equation (7) equal to 0.5. The number of Monte Carlo replications is equal to 1'000.

The simulation results displayed in Table 9 show that the performance of all estimators is extremely good. In most cases, the estimators are identical to the true values up to the third decimal. Moreover, the FDR estimators approach the true FDR by above because of its conservative property. The third and fourth columns contain the true π_0 and the average of the estimates $\hat{\pi}_0$ over the 1'000 replications, respectively. The fifth and sixth columns display the true $FDR\,(\gamma)$ and the average of the estimates $\widehat{FDR}_\lambda\,(\gamma)$ (defined by Equation (6). The true $FDR\,(\gamma)$ is computed as follows:

$$FDR\,(\gamma) = \frac{\pi_0 \cdot \gamma}{\pi_0 \cdot \gamma + \pi_A \left[prob\left(t < t_{T-2,\frac{\gamma}{2}}\, |H_A, \alpha_A < 0\right) + prob\left(t > t_{T-2,1-\frac{\gamma}{2}}\, |H_A, \alpha_A > 0\right)\right]}, \tag{23}$$

where $t_{T-2,\frac{\gamma}{2}}$ and $t_{T-2,1-\frac{\gamma}{2}}$ denotes the quantile of probability level $\frac{\gamma}{2}$ and $1-\frac{\gamma}{2}$ of the Student distribution with $T-2$ degrees of freedom.

The seventh and eighth columns display the true $FDR^+\,(\gamma)$ and the average of the estimates $\widehat{FDR}^+_\lambda\,(\gamma)$ (defined by Equation (8)). The true $FDR^+\,(\gamma)$ is computed as:

$$FDR^+\,(\gamma) = \frac{\frac{1}{2} \cdot \pi_0 \cdot \gamma}{\frac{1}{2} \cdot \pi_0 \cdot \gamma + \pi_A^+ \cdot prob\left(t > t_{T-2,1-\frac{\gamma}{2}}\, |H_A, \alpha_A > 0\right)}. \tag{24}$$

The final two columns display the true $FDR^-\,(\gamma)$ and the average of the estimates $\widehat{FDR}^-_\lambda\,(\gamma)$ (defined by Equation (9)). The true $FDR^-\,(\gamma)$ is computed as:

$$FDR^-\,(\gamma) = \frac{\frac{1}{2} \cdot \pi_0 \cdot \gamma}{\frac{1}{2} \cdot \pi_0 \cdot \gamma + \pi_A^- \cdot prob\left(t < t_{T-2,\frac{\gamma}{2}}\, |H_A, \alpha_A < 0\right)}. \tag{25}$$

References

[1] Berstein, P.L., 2003, Points of Inflection: Investment Management Tomorrow, Financial Analysts Journal, 18-23.

[2] Benjamini, Y., and Y. 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] Brown, S.J., and W.N. Goetzmann, 1997, Mutual Fund Style, Journal of Financial Economics, 373-399.

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

[5] Chen, H.-L., Jegadeesh, N., and R. Wermers, 2000, The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of Fund Managers, Journal of Financial and Quantitative Analysis 35, 343-368.

[6] Christopherson, J.A., 1995, Equity Style Classifications, Journal of Portfolio Management 21, 32-46.

[7] Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring Mutual Fund Performance with Characteristic-Based Benchmarks, Journal of Finance 52, 1-33.

[8] Davidson., R., and J.G. MacKinnon, 1993, Estimation and Inference in Econometrics, Oxford University Press, New-York.

[9] Davison, A.C., and D.V. Hinkley, 1997, Bootstrap Methods and Their Application, Cambridge University Press, London.

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

[11] Elton, E.J., M.J. Gruber, and C.R. Blake, 1996, The Persistence of Risk-Adjusted Mutual Fund Performance, Journal of Business 69, 133-158.

[12] Fama, E.F., and K.R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, Journal of Finance 51, 55-84.

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

[14] Ferson, W., and M. Qian, 2004, Conditional Performance Evaluation Revisited, in Research Foundation Monograph of the CFA Institute (formerly, AIMR), forthcoming.

[15] Grinblatt, M., and S. Titman, 1989, Mutual Fund Performance: An analysis of Quarterly Portfolio Holdings, Journal of Business 62, 393-416.

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

[17] Grinblatt, M., and S. Titman, 1995, Performance Evaluation, Handbooks in OR and MS, Vol 9, in R. Jarrow et al., ed: Elsevier Science.

[18] Grinblatt, M., S. Titman, and R. Wermers, 1995, Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior, American Economic Review 85, 1088-1105.

[19] Hall P., J.L. Horowitz and B.-Y. Jing, 1995, On Blocking Rules for the Bootstrap with Dependent Data, Biometrika 82, 561-574.

[20] Hendricks, D., J. Patel, and R. Zeckhauser, 1993, Hot Hands in Mutual Funds: The Persistence of Performance, 1974-88, Journal of Finance 48, 93-130.

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

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

[23] Kosowski, R., A. Timmermann, H. White, and R. Wermers, 2005, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, 2004, Working Paper.

[24] Kung, E., and L. Pohlman, 2004, Portable Alpha, Journal of Portfolio Management, 78-87.

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

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

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

[28] Pastor, L., and R.F. Stambaugh, 2002b, Investing in Equity Mutual Funds, Journal of Financial Economics 63, 351-380.

[29] Romano, J.P., and M. Wolf, 2004, Stepwise Multiple Testing as Formalized Data Snooping, Econometrica 73, 1237-1282.

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

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

[32] Storey, J.D., and R. Tibshirani, 2003, Statistical Significance for Genomewide Studies, PNAS 100, 9440-9445.

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

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

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

[36] White, H., 2000, A Reality Check for Data Snooping, Econometrica 68, 1097-1126.

Table 1

Outcomes of the Multiple Test of Differential Performance for the Significance Level γ

	# Accept H_0	# Reject H_0	# Total
Funds with no performance	$N(\gamma)$	$F(\gamma)$	M_0
Funds with differential performance	$A(\gamma)$	$T(\gamma)$	M_A
# Total	$W(\gamma)$	$R(\gamma)$	M

The null hypothesis H_0 of no performance ($\alpha_i = 0$) is tested against the alternative H_A of differential performance ($\alpha_i > 0$ or $\alpha_i < 0$). $N(\gamma)$ stands for the number of funds with no performance which are correctly considered as funds with zero alphas. $F(\gamma)$ denotes the number of funds with no performance which are incorrectly classified as significant funds (i.e. lucky funds). $A(\gamma)$ corresponds to the number of funds with differential performance which are incorrectly classified as funds with zero alphas. $T(\gamma)$ stands for the number of funds with differential performance which are correctly considered as significant. Among the M funds, $R(\gamma)$ funds are called significant (i.e. H_0 is rejected R times). The ratio $\pi_0 = M_0/M$ corresponds to the proportion of funds with no performance in the total population of M funds.

Table 2

Average Mutual Fund Performance

Panel A Unconditional Carhart Model

	α	β_m	β_{smb}	β_{hml}	β_{mom}	R^2
All funds	-0.44%	0.95	0.14	-0.02	0.02	97.9%
	(0.18)	(0.00)	(0.00)	(0.22)	(0.12)	
G funds	-0.43%	0.96	0.15	–0.04	0.03	97.8%
	(0.20)	(0.00)	(0.00)	(0.12)	(0.05)	
AG funds	-0.64%	1.05	-0.40	-0.26	0.08	95.8%
	(0.22)	(0.00)	(0.00)	(0.00)	(0.00)	
GI funds	-0.72%	0.88	-0.06	0.16	-0.02	97.9%
	(0.05)	(0.00)	(0.00)	(0.00)	(0.16)	

Panel B Conditional Carhart Model

	α	β_m	β_{smb}	β_{hml}	β_{mom}	R^2
All funds	-0.51%	0.96	0.15	-0.03	0.02	98.0%
	(0.16)	(0.00)	(0.00)	(0.12)	(0.11)	
G funds	-0.56%	0.96	0.15	-0.04	0.03	97.9%
	(0.16)	(0.00)	(0.00)	(0.05)	(0.03)	
AG funds	-0.70%	1.06	-0.40	-0.26	0.08	96.1%
	(0.19)	(0.00)	(0.00)	(0.00)	(0.00)	
GI funds	-0.72%	0.88	-0.06	0.15	-0.02	97.9%
	(0.05)	(0.00)	(0.00)	(0.00)	(0.09)	

This Table shows the alpha, the factor exposures and the adjusted R-square of an equally-weighted portfolio including all funds in a given investment category. Figures in parentheses denote the heteroskedasticity-consistent p-values under the null hypothesis that the regression parameters are equal to zero. Panel A and B show the coefficients of the unconditional and conditional Carhart models, respectively. The regressions are based on monthly data between January 1975 and December 2002 (336 observations). The alphas are expressed in percent per year.

Table 3

Performance Measurement with the Standard Approach

Panel A: *All* funds

γ	0.05	0.10	0.15	0.20
$\widehat{R}$	157	253	342	432
$\widehat{R}^+$	50	80	107	138
$\widehat{R}^-$	107	173	235	294

γ	0.05	0.10	0.15	0.20
$\widehat{R}/M$	10.7%	17.2%	23.2%	29.3%
$\widehat{R}^+/M$	3.4%	5.4%	7.3%	9.4%
$\widehat{R}^-/M$	7.3%	11.8%	16.0%	20.0%

Panel B: *G* funds

γ	0.05	0.10	0.15	0.20
$\widehat{R}$	97	170	226	284
$\widehat{R}^+$	28	52	72	91
$\widehat{R}^-$	69	118	154	193

γ	0.05	0.10	0.15	0.20
$\widehat{R}/M$	9.5%	16.6%	22.0%	27.1%
$\widehat{R}^+/M$	2.7%	5.0%	7.0%	8.8%
$\widehat{R}^-/M$	6.7%	11.5%	15.0%	18.8%

Panel C: *AG* funds

γ	0.05	0.10	0.15	0.20
$\widehat{R}$	35	54	72	83
$\widehat{R}^+$	18	25	32	35
$\widehat{R}^-$	17	29	40	48

γ	0.05	0.10	0.15	0.20
$\widehat{R}/M$	15.0%	23.1%	30.7%	35.5%
$\widehat{R}^+/M$	7.7%	10.7%	13.7%	14.9%
$\widehat{R}^-/M$	7.3%	12.4%	17.0%	20.5%

Panel D: *GI* funds

γ	0.05	0.10	0.15	0.20
$\widehat{R}$	35	57	82	85
$\widehat{R}^+$	6	12	16	21
$\widehat{R}^-$	29	45	66	74

γ	0.05	0.10	0.15	0.20
$\widehat{R}/M$	11.3%	18.4%	26.5%	30.7%
$\widehat{R}^+/M$	1.9%	3.8%	5.2%	6.7%
$\widehat{R}^-/M$	9.3%	14.5%	21.3%	23.8%

The results for All funds (*All*), Growth funds (*G*), Aggressive Growth funds (*AG*), and Growth and Income funds (*GI*) are presented in Panels A, B, C, and D, respectively. The left part of each Panel displays the number of significant funds $\widehat{R}$, the number of best funds $\widehat{R}^+$, and the number of worst funds $\widehat{R}^+$ at different significance levels γ. The right part of each Panel displays the proportion of significant funds $\widehat{R}/M$, the proportion of best funds $\widehat{R}^+/M$, and the proportion of worst funds $\widehat{R}^-/M$ at different significance levels γ. The best (worst) funds are defined as funds with significant positive (negative) estimated alphas. The alphas of all funds are computed with the unconditional Carhart model.

Table 4

Proportion of Funds with Zero and Non-Zero Alphas

	No performance $\widehat{\pi}_0$	Differential performance $\widehat{\pi}_A$
All funds	76.5%	23.5%
G funds	80.2%	19.8%
AG funds	70.5%	29.5%
GI funds	74.8%	25.2%

The first column contains the proportion $\widehat{\pi}_0$ of funds with no performance (zero alphas) and the second one contains the proportion $\widehat{\pi}_A$ of funds with differential performance (non-zero alpha). For each investment category, $\widehat{\pi}_0$ is estimated from the empirical cross-sectional distribution of the fund p-values. The alphas of all funds are computed with the unconditional Carhart model.

Table 5

Performance Measurement with the False Discovery Rate

Panel A: *All* funds

Significant funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}$	35.8%	44.5%	49.4%	52.1%	$\widehat{FDR}$	35.8%	44.5%	49.4%	52.1%
$\widehat{R}$	157	253	342	432	$\widehat{R}/M$	10.7%	17.2%	23.2%	29.3%
$\widehat{F}$	56	113	169	225	$\widehat{F}/M$	3.8%	7.6%	11.4%	15.3%
$\widehat{T}$	101	140	173	207	$\widehat{T}/M$	6.9%	9.6%	11.8%	14.0%

Best funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	56.3%	70.3%	78.9%	81.6%	$\widehat{FDR}^+$	56.3%	70.3%	78.9%	81.6%
$\widehat{R}^+$	50	80	107	138	$\widehat{R}^+/M$	3.4%	5.4%	7.3%	9.4%
$\widehat{F}^+$	28	56	83	113	$\widehat{F}^+/M$	1.9%	3.8%	5.7%	7.6%
$\widehat{T}^+$	22	24	24	25	$\widehat{T}^+/M$	1.5%	1.6%	1.6%	1.7%

Worst funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^-$	26.3%	32.5%	35.9%	38.3%	$\widehat{FDR}^-$	26.3%	32.5%	35.9%	38.3%
$\widehat{R}^-$	107	173	235	294	$\widehat{R}^-/M$	7.3%	11.8%	16.0%	20.0%
$\widehat{F}^-$	28	56	70	113	$\widehat{F}^-/M$	1.9%	3.8%	5.7%	7.6%
$\widehat{T}^-$	79	117	151	181	$\widehat{T}^-/M$	5.4%	8.0%	10.3%	12.4%

The results for All funds (*All*), Growth funds (*G*), Aggressive Growth funds (*AG*), and Growth and Income funds (*GI*) are presented in Panels A, B, C, and D, respectively. The impact of luck is measured among the sets of significant, best and worst funds at different significance levels γ. For the set of significant (best or worst) funds, the left part of each Panel displays the False Discovery Rate $\widehat{FDR}$ ($\widehat{FDR}^+$ or $\widehat{FDR}^-$), the number of significant (best or worst) funds $\widehat{R}$ ($\widehat{R}^+$ or $\widehat{R}^-$), the number of lucky funds $\widehat{F}$ ($\widehat{F}^+$ or $\widehat{F}^-$), and the number of funds with differential (positive or negative) performance $\widehat{T}$ ($\widehat{T}^+$ or $\widehat{T}^-$). The right part of each Panel shows the proportion of significant (best or worst) funds $\widehat{R}/M$ ($\widehat{R}^+/M$ or $\widehat{R}^-/M$), the proportion of lucky funds $\widehat{F}/M$ ($\widehat{F}^+/M$ or $\widehat{F}^+/M$), and the proportion of funds with differential (positive or negative) performance $\widehat{T}/M$ ($\widehat{T}^+/M$ or $\widehat{T}^+/M$). The best (worst) funds are defined as funds with significant positive (negative) estimated alphas. The alphas of all funds are computed with the unconditional Carhart model.

Table 5

Performance Measurement with the False Discovery Rate

Panel B: G funds

Significant funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}$	42.4%	44.5%	49.4%	52.1%	$\widehat{FDR}$	42.4%	44.5%	49.4%	52.1%
$\widehat{R}$	97	170	226	284	$\widehat{R}/M$	9.5%	16.6%	22.0%	27.1%
$\widehat{F}$	41	82	123	164	$\widehat{F}/M$	4.0%	8.0%	12.0%	16.0%
$\widehat{T}$	56	88	103	120	$\widehat{T}/M$	5.5%	8.6%	10.0%	11.1%

Best funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	73.4%	79.0%	85.6%	88.0%	$\widehat{FDR}^+$	73.4%	79.0%	85.6%	88.0%
$\widehat{R}^+$	28	52	72	91	$\widehat{R}^+/M$	2.7%	5.0%	7.0%	8.8%
$\widehat{F}^+$	21	41	61	80	$\widehat{F}^+/M$	2.0%	4.0%	5.7%	7.5%
$\widehat{T}^+$	7	10	11	11	$\widehat{T}^+/M$	0.7%	1.0%	1.3%	1.3%

Worst funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^-$	29.8%	34.8%	40.0%	42.6%	$\widehat{FDR}^-$	29.8%	34.8%	40.0%	42.6%
$\widehat{R}^-$	69	118	154	193	$\widehat{R}^-/M$	6.7%	11.5%	15.0%	18.8%
$\widehat{F}^-$	21	41	61	80	$\widehat{F}^-/M$	2.0%	4.0%	5.7%	7.5%
$\widehat{T}^-$	48	77	93	113	$\widehat{T}^-/M$	4.7%	7.5%	9.3%	11.3%

Table 5

Performance Measurement with the False Discovery Rate

Panel C: *AG* funds

Significant funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}$	23.6%	30.6%	34.3%	39.7%	$\widehat{FDR}$	23.6%	30.6%	34.3%	39.7%
$\widehat{R}$	35	54	72	83	$\widehat{R}/M$	15.0%	23.1%	30.7%	35.5%
$\widehat{F}$	8	16	25	33	$\widehat{F}/M$	3.5%	7.0%	10.6%	14.6%
$\widehat{T}$	27	38	47	50	$\widehat{T}/M$	11.5%	16.1%	20.7%	20.9%

Best funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	22.9%	33.0%	38.7%	42.8%	$\widehat{FDR}^+$	22.9%	33.0%	38.7%	42.8%
$\widehat{R}^+$	18	25	32	35	$\widehat{R}^+/M$	7.7%	10.7%	13.7%	14.9%
$\widehat{F}^+$	4	8	12	15	$\widehat{F}^+/M$	1.8%	3.5%	5.3%	6.5%
$\widehat{T}^+$	14	17	20	20	$\widehat{T}^+/M$	5.9%	7.2%	8.4%	8.4%

Worst funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^-$	24.3%	28.4%	30.9%	31.3%	$\widehat{FDR}^-$	24.3%	28.4%	30.9%	31.3%
$\widehat{R}^-$	17	29	40	48	$\widehat{R}^-/M$	7.3%	12.4%	17.0%	20.5%
$\widehat{F}^-$	4	8	12	15	$\widehat{F}^-/M$	1.8%	3.5%	5.3%	6.5%
$\widehat{T}^-$	13	21	28	33	$\widehat{T}^-/M$	5.5%	8.9%	11.7%	14.0%

Table 5

Performance Measurement with the False Discovery Rate

Panel D: *GI* funds

Significant funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}$	33.1%	40.7%	42.4%	48.8%	$\widehat{FDR}$	33.1%	40.7%	42.4%	48.8%
$\widehat{R}$	35	57	82	85	$\widehat{R}/M$	11.3%	18.4%	26.5%	30.7%
$\widehat{F}$	12	23	35	46	$\widehat{F}/M$	3.7%	7.5%	11.2%	15.0%
$\widehat{T}$	23	34	47	50	$\widehat{T}/M$	7.6%	10.9%	15.3%	15.7%

Best funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	100.0%	100.0%	100.0%	100.0%	$\widehat{FDR}^+$	100.0%	100.0%	100.0%	100.0%
$\widehat{R}^+$	6	12	16	21	$\widehat{R}^+/M$	1.9%	3.8%	5.2%	6.7%
$\widehat{F}^+$	6	12	16	21	$\widehat{F}^+/M$	1.9%	3.8%	5.2%	6.7%
$\widehat{T}^+$	0	0	0	0	$\widehat{T}^+/M$	0.0%	0.0%	0.0%	0.0%

Worst funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^-$	20.7%	28.4%	24.2%	28.3%	$\widehat{FDR}^-$	20.7%	28.4%	24.2%	28.3%
$\widehat{R}^-$	29	45	66	74	$\widehat{R}^-/M$	9.3%	14.5%	21.3%	23.8%
$\widehat{F}^-$	6	12	16	21	$\widehat{F}^-/M$	1.9%	3.8%	5.2%	6.7%
$\widehat{T}^-$	23	33	50	53	$\widehat{T}^-/M$	7.4%	10.7%	16.1%	17.1%

Table 6

Source of Differential Performance

	Differential performance $\widehat{\pi}_A$	Positive performance $\widehat{\pi}_A^+$	Negative performance $\widehat{\pi}_A^-$
All funds	23.5%	1.7%	21.8%
G funds	19.8%	1.3%	18.5%
AG funds	29.5%	8.4%	21.1%
GI funds	25.2%	0.0%	25.2%

The first column contains the proportion $\widehat{\pi}_A$ of funds with differential performance (non-zero alpha). The second column contains the proportion $\widehat{\pi}_A^+$ of funds with positive performance (positive alpha). It is estimated with $\widehat{T}^+/M$ at the significance level $\gamma = 0.20$. The second column contains the proportion $\widehat{\pi}_A^-$ of funds with negative performance (negative alpha). It is estimated with the following equation: $\widehat{\pi}_A^- = \widehat{\pi}_A - \widehat{\pi}_A^-$. The alphas of all funds are computed with the unconditional Carhart model.

Table 7

Expected Alpha of Portfolios of the Best Funds

Panel A: Best *All* funds ($\alpha_A^+ = 5.3\%$)

γ	0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	56.3%	70.3%	78.9%	81.6%
α_p	2.31%	1.57%	1.11%	0.97%
$d\alpha_p/\alpha_p$		-32.1%	-28.8%	-12.6%

Panel B: Best *G* funds ($\alpha_A^+ = 5.8\%$)

γ	0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	73.4%	79.0%	85.6%	88.0%
α_p	1.54%	1.21%	0.83%	0.70%
$d\alpha_p/\alpha_p$		-21.2%	-31.4%	-16.4%

Panel C:Best *AG* funds ($\alpha_A^+ = 7.7\%$)

γ	0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	22.9%	33.0%	38.7%	42.8%
α_p	5.93%	5.15%	4.72%	4.40%
$d\alpha_p/\alpha_p$		-13.0%	-8.5%	-6.7%

The results for All funds (*All*), Growth funds (*G*), and Aggressive Growth funds (*AG*) are presented in Panels A, B, and C, respectively. We exclude Growth and Income funds (*GI*) since none of them produce positive alphas. Each portfolio is built by equally-weighting the best funds at different significance levels γ. $\widehat{FDR}^+$ denotes the False Dicovery Rate among the funds forming the portfolio. To estimate the alpha of funds with positive performance α_A^+, we compute the estimated alpha of the fund located at the 5%-quantile of the best funds. It respectively amounts to 5.3%, 5.8% and 7.7% per year for *All*, *G*, and *AG* funds. $d\alpha_p/\alpha_p$ denotes the relative reduction of the portfolio alpha as γ rises by 0.05. It is independent of the value chosen for α_A^+. The alphas of all funds are computed with the unconditional Carhart model.

Table 8

The False Discovery Rate with Alternative Asset Pricing Models

Panel A: *All* funds

	Unconditional models					Conditional models			
	CAPM		FF			CAPM		FF	
γ	0.05	0.20	0.05	0.20	γ	0.05	0.20	0.05	0.20
$\widehat{FDR}^{+}$	100%	100%	66.7%	80.8%	$\widehat{FDR}^{+}$	100%	100%	42.9%	58.0%
$\widehat{FDR}^{-}$	45.6%	68.0%	17.7%	29.6%	$\widehat{FDR}^{-}$	49.1%	76.2%	14.4%	24.7%

Panel B: *G* funds

	Unconditional models					Conditional models			
	CAPM		FF			CAPM		FF	
γ	0.05	0.20	0.05	0.20	γ	0.05	0.20	0.05	0.20
$\widehat{FDR}^{+}$	100%	100%	78.8%	89.1%	$\widehat{FDR}^{+}$	100%	100%	46.4%	55.5%
$\widehat{FDR}^{-}$	48.5%	70.1%	22.5%	37.2%	$\widehat{FDR}^{-}$	59.2%	77.7%	18.2%	29.4%

Panel C: *AG* funds

	Unconditional models					Conditional models			
	CAPM		FF			CAPM		FF	
γ	0.05	0.20	0.05	0.20	γ	0.05	0.20	0.05	0.20
$\widehat{FDR}^{+}$	100%	100%	29.5%	36.0%	$\widehat{FDR}^{+}$	100%	100%	20.2%	34.0%
$\widehat{FDR}^{-}$	29.5%	48.8%	26.0%	55.1%	$\widehat{FDR}^{-}$	48.1%	57.6%	22.6%	40.3%

Panel D: *GI* funds

	Unconditional models					Conditional models			
	CAPM		FF			CAPM		FF	
γ	0.05	0.20	0.05	0.20	γ	0.05	0.20	0.05	0.20
$\widehat{FDR}^{+}$	100%	100%	100%	100%	$\widehat{FDR}^{+}$	100%	100%	100%	100%
$\widehat{FDR}^{-}$	64.3%	77.1%	11.5%	18.4%	$\widehat{FDR}^{-}$	64.6%	77.5%	7.8%	14.4%

The results for All funds (*All*), Growth funds (*G*), Aggressive Growth funds (*AG*), and Growth and Income funds (*GI*) are presented in Panels A, B, C, and D, respectively. The left part of each Panel contains the FDR among the best and worst funds computed with the unconditional CAPM and Fama-French (FF) models at two different significance levels γ. The right part of each Panel contains the FDR among the best and worst funds computed with the conditional CAPM and Fama-French (FF) models at two different significance levels γ.

Table 9

Precision of the FDR Estimators using Monte-Carlo Simulations

$\alpha_A^+ = 8\%, \alpha_A^- = -5\%$

q^-	γ	π_0	$\widehat{\pi}_0$	FDR	$\widehat{FDR}$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.704	0.114	0.114	0.078	0.078	0.211	0.212
		0.9	0.900	0.332	0.331	0.246	0.246	0.508	0.511
	0.10	0.7	0.700	0.198	0.200	0.143	0.144	0.321	0.324
		0.9	0.900	0.489	0.488	0.393	0.392	0.646	0.647
0.7	0.05	0.7	0.711	0.127	0.128	0.165	0.167	0.103	0.104
		0.9	0.902	0.359	0.358	0.433	0.433	0.307	0.306
	0.10	0.7	0.712	0.211	0.214	0.281	0.285	0.168	0.171
		0.9	0.902	0.508	0.508	0.602	0.603	0.439	0.440

$\alpha_A^+ = 5\%, \alpha_A^- = -5\%$

q^-	γ	π_0	$\widehat{\pi}_0$	FDR	$\widehat{FDR}$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.716	0.138	0.141	0.103	0.105	0.211	0.216
		0.9	0.904	0.382	0.382	0.307	0.308	0.508	0.508
	0.10	0.7	0.716	0.221	0.226	0.168	0.172	0.321	0.329
		0.9	0.903	0.523	0.522	0.439	0.439	0.646	0.646
0.7	0.05	0.7	0.715	0.138	0.141	0.211	0.216	0.103	0.105
		0.9	0.904	0.382	0.384	0.508	0.512	0.307	0.308
	0.10	0.7	0.713	0.221	0.226	0.321	0.329	0.168	0.172
		0.9	0.904	0.523	0.526	0.646	0.651	0.439	0.442

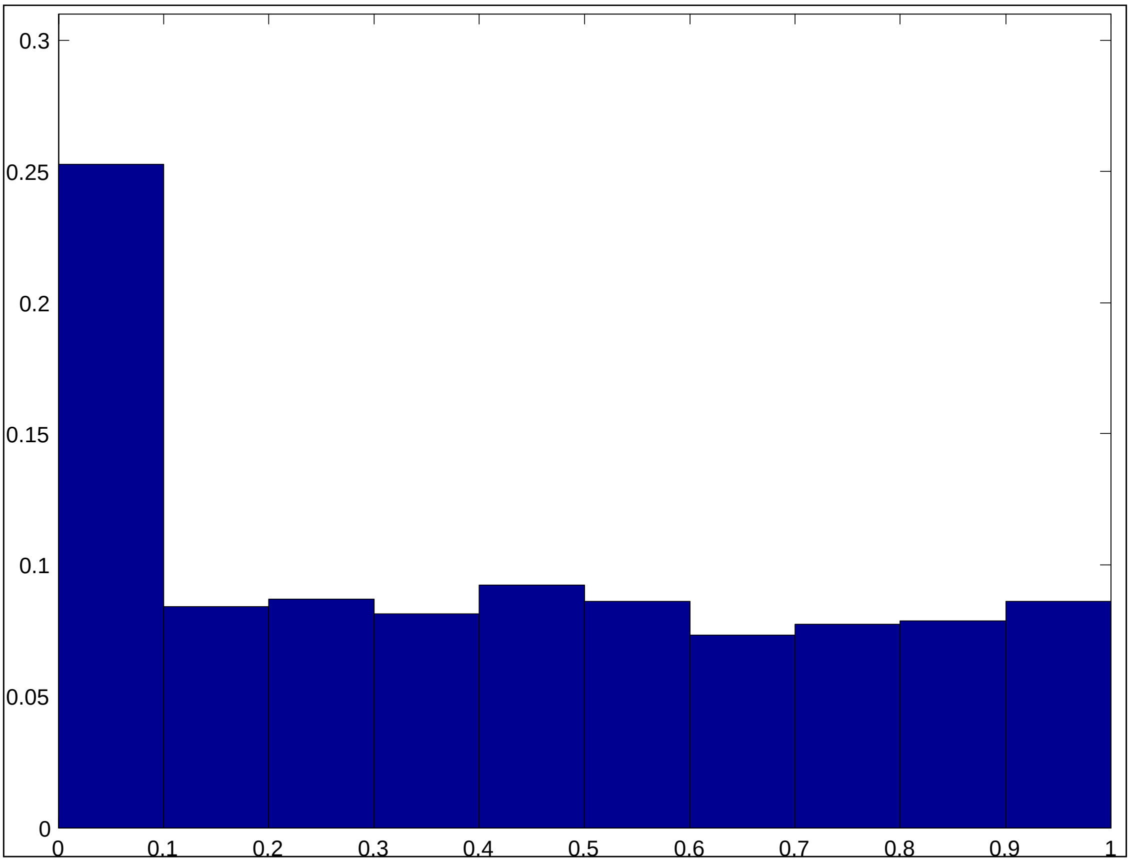
$\alpha_A^+ = 5\%, \alpha_A^- = -8\%$

q^-	γ	π_0	$\widehat{\pi}_0$	FDR	$\widehat{FDR}$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.712	0.127	0.129	0.103	0.104	0.165	0.168
		0.9	0.901	0.359	0.359	0.307	0.307	0.433	0.432
	0.10	0.7	0.710	0.211	0.214	0.168	0.171	0.281	0.285
		0.9	0.902	0.508	0.507	0.439	0.438	0.602	0.603
0.7	0.05	0.7	0.705	0.114	0.114	0.211	0.212	0.078	0.078
		0.9	0.900	0.332	0.331	0.508	0.510	0.246	0.245
	0.10	0.7	0.705	0.198	0.200	0.321	0.324	0.143	0.144
		0.9	0.901	0.489	0.489	0.646	0.647	0.393	0.393

The monthly returns are simulated according to the one-factor model for 1'472 funds and 396 periods. A proportion π_0 of funds have zero alphas. A proportion π_A of funds have differential performance. Under H_A, funds yield negative alphas with probability $\pi_A^- = \pi_A \cdot q^-$ or positive alphas with probability $\pi_A^+ = \pi_A \cdot (1 - q^-)$. FDR, FDR^+ and FDR^- correspond to the true false discovery rates. Ave $\widehat{\pi}_0$, Ave $\widehat{FDR}$, Ave $\widehat{FDR^+}$ and Ave $\widehat{FDR^-}$ stand for the average value of the estimators across 1'000 Monte-Carlo simulations.

Figure 1

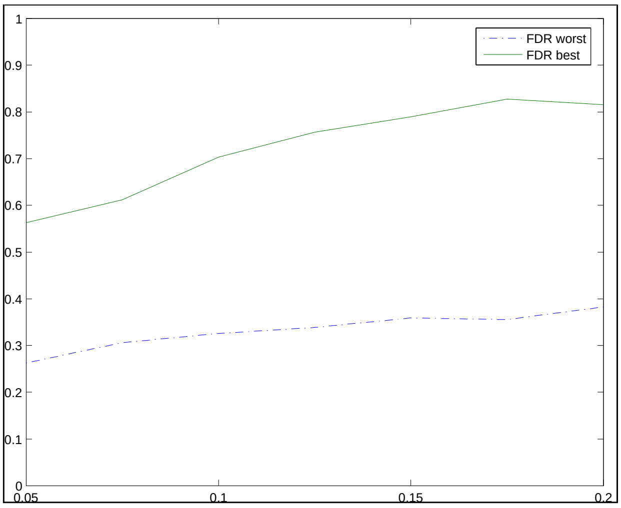
Cross-Sectional p-value Distribution



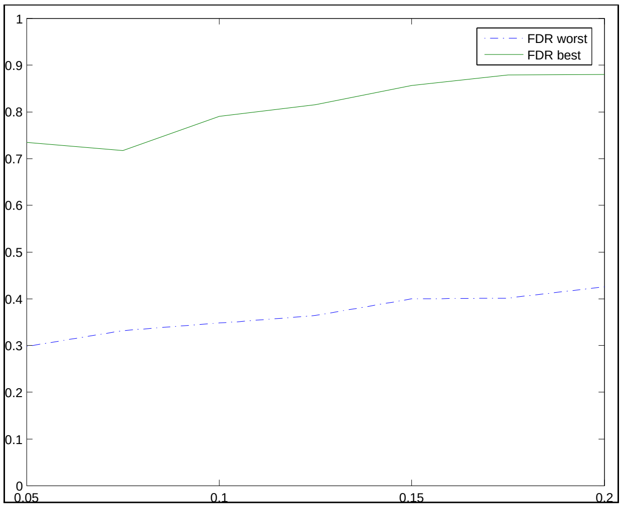
We simulate fund excess returns for 1'472 funds and 336 observations with the market model (see the Appendix for the details). From these simulated time-series, the fund alphas and p-values are estimated. The proportion π_0 of funds drawn from H_0 is equal to 80%. Under H_A, funds yield a negative alpha of -5% per year or a positive alpha of 5% per year with equal probabilities. Under H_0, the p-values are uniformly distributed over [0,1].

Figure 2

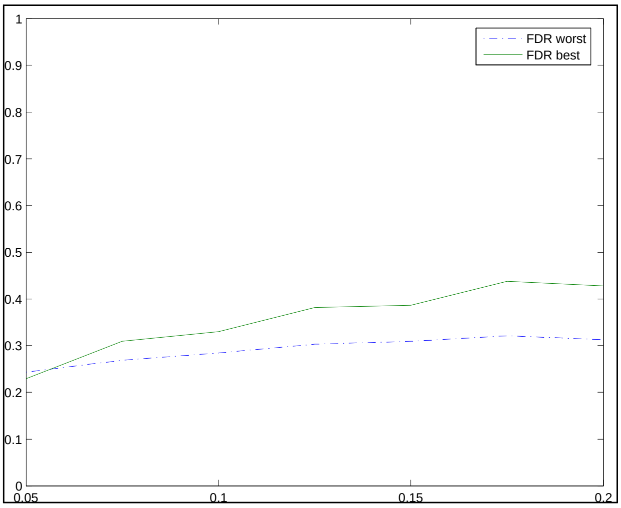
False Discovery Rates among the Best and the Worst Funds



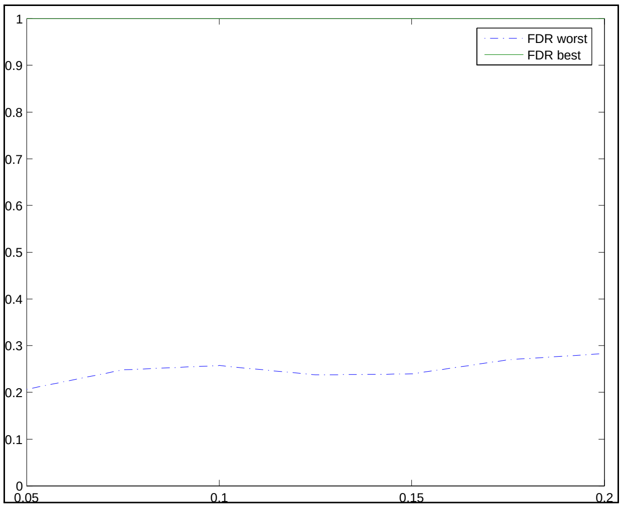
(a) *All* funds



(b) *G* funds



(c) *AG* funds



(d) *GI* funds

The figure plots the estimated FDR among the best and the worst funds at different significant levels γ. The solid line denotes the FDR among the best funds ($\widehat{FDR}^+$) and the dashed one the FDR among the worst funds ($\widehat{FDR}^-$). The best (worst) funds are defined as funds with significant positive (negative) estimated alphas. The alphas of all funds are computed with the unconditional Carhart model.