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

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

First version, September 2005
This version, November 2005

Abstract

Standard tests designed to identify mutual funds with non-zero alphas are problematic, in that they do not adequately account for the presence of lucky funds. Lucky funds have significant estimated alphas, while their true alphas are equal to zero. To address this issue, this paper quantifies the impact of luck with new measures built on the False Discovery Rate (FDR). These FDR measures provide a simple way to compute the proportion of funds with genuine positive or negative performance as well as their location in the cross-sectional alpha distribution. Using a large cross-section of U.S. domestic-equity funds, we find that about one fifth of the funds in the population truly yield negative alphas. These funds are dispersed in the left tail of the alpha distribution. We also find a small proportion of funds with truly positive performance, which are concentrated in the extreme right tail of the alpha distribution.

JEL Classification: G11, G23, C12
Keywords: Mutual Fund Performance, False Discovery Rate, Multiple Testing

---

*We are grateful to E. Ronchetti, R. Stulz and M.-P. Victoria-Feser as well as seminar participants at the Bilgi University, Grequam, Imperial College, Solvay Business School, University of Geneva, and Swiss Doctoral Workshop for their helpful comments. The first and second authors acknowledge financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK). Part of this research was done while the second author was visiting the Centre Emile Bernheim (ULB).

[†]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

Over $4 trillion is currently managed by equity mutual funds in the U.S., with roughly 90 percent invested in actively managed accounts. These mutual funds hold over 25 percent of the outstanding equity value of the average U.S. common stock. This high level of ownership makes it very unlikely that the equity fund industry as a whole is able to outperform the market by a large margin. However, several recent papers show some evidence of manager skills among subgroups of funds (see, for example, Gruber (1996)).

To detect funds with positive or negative performance, the standard approach in the academic literature (e.g., 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 (e.g., 5 percent) and all funds with p-values smaller than γ are said to have significant estimated alphas. Finally, these significant alphas are counted in order to provide an estimate of the number of funds with differential performance.

As with every hypothesis test, inference based on alpha estimates can lead to the detection of a lucky fund, namely a fund with a significant estimated alpha, while its true alpha is equal to zero. When a single performance test is run on the estimated alpha of one fund (or one portfolio of funds[1]), 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 under the hypothesis that its alpha is zero amounts to 0.05, by construction. The difficulty raised by the standard approach is that it implies a multiple hypothesis test since the null hypothesis of no performance is not tested once, but M times. Accounting for luck in a multiple testing framework is much more complex, because luck cannot be measured by γ. Specifically, 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 0.05–that is, even if all funds have true alphas of zero, we would still expect that some of the M funds will exhibit positive and significant alpha estimates purely through luck[2].

[1] For instance, testing the average performance of the mutual fund industry boils down to a single test on the alpha of the equally-weighted portfolio of all funds in the population (see, for instance, Lehman and Modest (1987) or Elton et al. (1993)).

[2] 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,

The standard approach, therefore, cannot properly measure the odds of observing a group of funds having genuine positive alphas. For example, suppose that 20 out of 200 funds have positive estimated alphas at a given significance level γ. Obviously, the true 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. Another problem with the standard approach is that it assumes that an 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 can be due to the inclusion of new lucky funds. As a consequence, the standard approach does not provide guidance on the location of funds with differential performance in the tails of the cross-sectional alpha distribution. For instance, suppose 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 more dispersed throughout the left tail.

This paper addresses all these issues by measuring the impact of luck on mutual fund performance. Specifically, 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. We extend this methodology by developing a new approach which allows us to separately compute the FDR among funds with significant positive estimated alphas (called hereafter the best funds) and funds with significant negative estimated alphas (called hereafter the worst funds). A main virtue of the FDR and these new measures is that they are very easy to compute from estimated p-values of fund alphas, and are, therefore, very simple extensions of the standard approach. The contributions brought by our approach are threefold. First, we account for luck by measuring the proportion of lucky funds at different significant levels γ. Second, we examine how the FDR varies as γ rises. This indicates whether funds with genuine differential performance are concentrated or more dispersed in the tails of the cross-sectional alpha distribution. Third, we provide explicit estimates of the proportion of funds in the population which have truly positive and negative alphas (as opposed to significant only).

Other methods have dealt with multiple testing in mutual fund performance. Grin-

we would expect some funds to achieve superior risk-adjusted returns by chance."

blatt 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 a 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)). This method explictly accounts for the presence of luck to determine how many funds should be rejected. But contrary to the FDR, it is not designed to measure the proportion of lucky funds among a given set of significant funds. Our approach can also be contrasted with the recent work by Kosowski et al. (2005). They use a bootstrap technique in order to test the significance of individual funds corresponding to various quantiles of the cross-section of estimated alphas (e.g. the top fund, the fund at the 10% quantile...). Since they do not address the issue of multiple testing, they do not quantify the relative importance of lucky funds among the significant ones.

Our empirical results are based on monthly returns of 1,472 U.S. open-end, domestic equity mutual funds existing at any time between 1975 and 2002. We investigate the performance of the entire cross-section of mutual funds, as well as the cross-section of each of three investment-objective categories, growth, aggressive-growth, and growth and income.

We first show that the impact of luck on performance is substantial. Specifically, we find that our estimators of the number of mutual funds with positive or negative performance is much lower than those obtained with the standard approach (only based on significant funds). These differences are informative, since they lead to a completely different assessment of mutual fund performance. For instance, while the standard approach concludes that 7.7% of the growth and income funds generate positive alphas (at $\gamma = 0.2$), we find that all of them are purely lucky. Finding 7.7% instead of 0% is clearly a false discovery!

Second, we find that luck has a stronger impact on the right versus the left tail of the estimated alpha distribution. Across all four categories of funds, the FDR among the worst funds is always lower than 50%, meaning that the majority of worst funds truly yield negative alphas. Moreover, the FDR among the worst funds increases slowly as γ rises, because part of the new significant funds deliver truly negative alphas. As a result, the funds with true negative performance are dispersed in the left tail of the cross-sectional estimated 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 the universe of funds and for growth funds, the FDR is always higher than 50% and increases quickly as γ rises. This indicates that there are only a few funds with truly positive performance, and that they are concentrated in the extreme right tail. On the contrary, the low level of the FDR among right-tail aggressive-growth funds reveals that many funds produce truly positive performance, while growth and income funds exhibit no truly positive alphas.

Third, we find that after accounting for luck, the percentage of funds with negative alphas in the population is approximately equal to 20% across all categories, while the proportion of funds with positive performance is much lower (it amounts to 1.9% of all funds in the population). It implies that the performance of the industry as a whole is not so bad because about 80% of the funds generate a performance sufficiently high to cover their expenses. The negative average performance documented in past studies is not due to the majority of funds but is caused by 20% of the funds.

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 estimation 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 on the accuracy of our new measures of luck.

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 achieves 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)) can be described as follows. First, the null hypothesis H_0 is tested 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 share significant estimated alphas (i.e. H_0 is rejected). Second, 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

Given the finite amount of data, it is impossible to know with certainty whether the alpha of a fund is different from zero. Therefore, the inference of the fund alpha is subject to luck. In this paper, 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 estimated alpha is not relevant. All that matters is that this fund is significant while its true alpha is equal to zero.

In the usual case where a single test is performed on the alpha of one fund (or one portfolio of funds), luck is controlled by setting the significance level γ (or equivalently the *Size* of the test). The standard approach differs from this framework because it boils down to running a multiple hypothesis test instead of a single one. The null hypothesis H_0 of no performance is tested for each of the M funds in the population. In a multiple testing framework, luck refers to the number (or the proportion) of lucky funds among the significant funds that are discovered. Accounting for luck in this situation is more complex. 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 some of the M funds to be significant only by pure luck.

To avoid any possible confusion, we stress that our definition of luck is very different from the one used by Kosowski et al. (2005). Their objective is to test the significance of the alpha of individual funds located at various quantiles of the cross-section of estimated alphas (e.g. the top fund, the fund corresponding to the 10% quantile...). Since such a fund is determined according to a pre-ranking of all fund alphas, a correct inference about its alpha must take into account the entire cross-section of the fund alphas[3]. In this context, they advocate to use a bootstrap procedure, and conclude that they account for luck in the sense that they correctly compute the p-values of the individual funds[4]. This definition of luck is not related to the issue of multiple testing investigated in this paper. In fact, we would face a similar multiple testing problem in their setting if we wanted to know how many individual funds corresponding to the various quantiles of the cross-section of estimated alphas truly yield non-zero alphas.

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

[3] Consider the alpha of the best fund denoted by α_i^{top}. Expressing the null and the alternative hypotheses as $H_0 : \alpha_i^{top} = \max_{i=1,...,M}\{\alpha_i\} \leqslant 0$ and $H_A : \alpha_i^{top} = \max_{i=1,...,M}\{\alpha_i\} > 0$ makes it clear that the distribution of the test statistic depends on the joint distribution of the alphas of all funds in the population. Note that the test statistic in our setting does not depend on the joint alpha distribution because of the form of the null and alternative hypotheses (see Equation (1)).

[4] This is summarized in the second page of their paper: "When outlier funds are selected from such an ex-post ranking of a large cross-section, the separation of luck from skill becomes extremely sensitive to the assumed distribution of the joint distribution from which the alphas are drawn".

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 to use the False Discovery Rate.

2.2 The False Discovery Rate (FDR)

2.2.1 The FDR among the Significant Funds

The FDR introduced in the statistical literature by Benjamini and Hochberg (1995) is defined in our setting as the expected proportion of lucky funds[5] among the significant funds at the significance level γ. It is written as[6]:

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

From Equation (2), it is easy to understand why the FDR is a straightforward extension of the standard approach. The FDR simply quantifies the proportion of lucky funds among the $R(\gamma)$ discovered by the standard approach in the first place. 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 the proportion of funds with no performance ($\alpha_i = 0$) and differential performance ($\alpha_i > 0$ or $\alpha_i < 0$), respectively. The $Size$ stands for the probability of picking up a lucky fund under the null hypothesis H_0 (i.e. $\alpha_i = 0$). In statistical terms, the $Size$ corresponds to the probability of committing a type I error. The $Power$ gives the probability of finding a fund with differential performance under

[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).

the alternative hypothesis H_A (i.e. $\alpha_i \neq 0$). 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 in two groups 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 true positive alpha, but are simply lucky. Similarly, $F^-(\gamma)$ among the $R^-(\gamma)$ worst funds do not yield a true 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 develop a new methodology which allows us to compute separately the FDR among the best funds and the FDR among the 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 test with equal-tailed confidence level, $\gamma/2$, 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 denote the number of lucky funds among the best and worst funds, respectively. By analogy with the definition of the FDR, we suggest to define the FDR among the best and worst funds (denoted by $FDR^+(\gamma)$ and $FDR^-(\gamma)$) as:

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

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

These measures are new, and are especially designed to deal with quantifying separately the proportion of lucky funds in the right tail of the cross-sectional alpha distribution and the proportion of lucky funds in the left tail of the cross-sectional alpha distribution.

2.3 Estimation Procedure

As we have previously mentioned, the estimation of the FDR is staightforward. All that is required is the estimated p-values $\widehat{p}_i$ of each fund i $(i = 1, ..., M)$. We use the following estimator of the FDR proposed by Storey (2002) and Storey and Tibshirani (2003):

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

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.

All that is needed to get $\widehat{FDR}_\lambda(\gamma)$ is an estimator of π_0 computed from the fund estimated p-values. The intuition is the following. Under H_0, the p-values are known to be uniformly distributed over the interval $[0,1]$[8]. On the contrary, the p-values under

[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.

[8]This feature is crucial to correctly estimate $\widehat{\pi}_0$. This method cannot be used in one-sided tests because the p-values are not necessarily uniformly distributed under H_0. In one-sided tests, the null is

H_A are extremely small because they are associated with extreme positive or negative estimated alphas. We can exploit this information to compute $\widehat{\pi}_0$ without specifying the exact distribution of the p-values under 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 divides themselves equally into funds with annual alphas of +5% and -5%.

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 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 density beyond λ to obtain an 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) and Storey, Taylor and Siegmund (2004). The latter is data-driven and chooses λ such that the mean-squared error of $\widehat{\pi}_0(\lambda)$ is minimized (see the Appendix for further details on the methodology).

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 suggest to compute the empirical counterparts of the

tested under the least favorable configuration (LFC). For instance, consider the following null hypothesis $H_0 : \alpha_i \leq 0$ against $H_A : \alpha_i > 0$. Under the LFC, $H_0 : \alpha_i \leq 0$ is replaced with $H_0 : \alpha_i = 0$. Therefore, all funds with $\alpha_i \leq 0$ have inflated p-values which are close to one.

new measures $FDR^{+}(\gamma)$ and $FDR^{-}(\gamma)$ defined in Equations (4) and (5) with:

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

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

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 respective weights are given by the ratio of the number of best and worst funds on the number of significant funds.

To examine the performance of the estimators $\widehat{\pi}_0(\lambda)$, $\widehat{FDR}_{\lambda}(\gamma)$, as well as the new measures $\widehat{FDR}_{\lambda}^{+}(\gamma)$, and $\widehat{FDR}_{\lambda}^{-}(\gamma)$ in our performance analysis setting, we have run Monte-Carlo simulations which match the empirical characteristics of our data. 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, the significance level γ, and the procedure used to choose the value of λ. Therefore, the estimation methodology can be thought as remarkably accurate.

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:

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

where z_{t-1} denotes the $J \times 1$ vector of centered predictive variables, and B is the $J \times 1$ vector of coefficients. Four predictive variables are considered. The first one is the one-month T-bill interest rate. The second one is the dividend yield of the CRSP value-weighted NYSE and AMEX stock index. The third one is the term spread proxied by the difference between the yield of a 10-year T-bond and the three-month T-bill interest rate. The fourth one is the default spread proxied by the yield difference between BAA-rated and AAA-rated corporate bonds.

We have also computed 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 last subsection of the empirical analysis.

3.2 Estimation of the p-values

Kosowski et al. (2005) finds that the distribution of fund estimated alphas in finite samples is non-normal for approximately half of the funds. This finding calls for a bootstrap procedure (instead of asymptotic theory) to compute each fund estimated p-value $\widehat{p}_i$. From bootstrap theory on higher order improvements, we know that the bootstrap is expected to yield better results when applied to asymptotic pivots[9]. We know that the fund $t-$stat $\widehat{t}_i$ is asymptotically pivotal. For this reason, we use the t-stat instead of the estimated alpha to compute the p-values under the null H_0 of no

[9] 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)).

performance. The t-stat $\widehat{t}_i$ is defined as

$$\widehat{t}_i = \frac{\widehat{\alpha}_i}{\widehat{\sigma}_{\alpha_i}}, \tag{13}$$

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). As shown in Equation (13), another advantage of the t-stat is that it reduces the presence of extreme observations due to volatile funds because the estimated alpha is scaled 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 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[10]. Since our procedure is similar to the one implemented by Kosowski et al. (2005), we refer to them for further details beyond the following brief description.

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 bootstrap t-stat $\widehat{t}_i^{*q}$. This procedure is repeated Q times, where Q is set to 1'000. Since we use a two-sided, equal-tailed test, the bootstrapped p-value $\widehat{p}_i$ of the fund i is computed as follows:

$$\widehat{p}_i = 2 \cdot \min\left(\frac{1}{Q}\sum_{q=1}^{Q} I\{\widehat{t}_i^{*q} > \widehat{t}_i\}, \frac{1}{Q}\sum_{q=1}^{Q} I\{\widehat{t}_i^{*q} < \widehat{t}_i\}\right) \tag{14}$$

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[11]. The CRSP database

[10] To know whether this approach is appropriate, we have checked for the presence of autocorrelation (with Ljung-Box test), heteroscedasticity (with White test) and Arch effects (with Engle test) in the fund residuals. We have found that only few funds presented some of these features. We have also implemented a block bootstrap methodology with a block length equal to $T^{\frac{1}{5}}$ (proposed by Hall, Horowitz and Jing (1995)), where T denotes the length of the fund return time-series. The results remain unchanged.

[11] 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.

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. We refer to Wermers (2000) 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 funds (GI). 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. The category of G funds is the biggest one with 1'025 funds, while the categories of AG and GI funds contain 234 and 310 funds, respectively.

Table 2 shows the average mutual fund performance across the four investment categories (All, G, AG, GI). For each investment category, we estimate 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 estimated 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 at different significant levels γ (γ = 0.05, 0.10, 0.15 and 0.20). 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 displays the number of significant funds $\widehat{R}$, the number of best funds $\widehat{R}^+$, and the number of worst funds $\widehat{R}^-$. The right part of each Panel shows the proportion of significant funds $\widehat{R}/M$, the proportion of best funds $\widehat{R}^+/M$, and the proportion of worst funds $\widehat{R}^-/M$.

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 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 generally higher for *AG* and *GI* funds than for *All* and *G* funds. 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 genuine differential, positive and negative performance. For instance, we find that 83 *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 87 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

[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 theory.

to generate highly negative alphas. On the contrary, if none of the 87 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.

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 78.4% of funds in the population have zero alphas. It implies that 21.6% 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 (80.4%), this proportion is lower for AG and GI funds (71.5% and 75.3% 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

Without referring to a FDR methodology, a few papers (Jensen (1968), Kosowski et al. (2005)) have in fact proposed a proxy 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 given by $M \cdot \gamma$ and γ, respectively. However, Table 4 clearly shows that $\widehat{\pi}_0$ is never equal to one. Therefore, the proxy overestimates the impact of luck because it does not account for the proportion π_A of funds which truly yield non-zero alphas. This explains the results documented in the Figure 3 of Kosowski et al. (2005): the number of lucky funds F is higher that the number of significant funds R, which cannot happen since $R = F + T$. This estimation of luck can become very inaccurate as γ rises. For instance, the number of lucky *All* funds are overestimated by 32 at $\gamma = 0.10$ and by 64 at $\gamma = 0.20$. Stated differently, the proxy 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 7.8% and 15.6%. These approximations are even worse for AG funds, since $\widehat{\pi}_0^{AG}$ is only equal to 71.5%.

4.2.2 The FDR of All Funds

To assess 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 $\widehat{FDR}$, 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, $\widehat{FDR}$ amounts to 37.0%, indicating that 98 out of the 156 significant funds generate differential performance. As γ rises, the number of lucky funds $\widehat{F}$ increases more quickly than the number of funds with differential performance $\widehat{T}$. Therefore, the FDR is equal to 54.5% at $\gamma = 0.20$, which implies that only half of the 422 significant funds have non-zero alphas.

From Equation (10), we know that $\widehat{FDR}$ is a weighted average of $\widehat{FDR}^+$ and $\widehat{FDR}^-$. These two components can be very different from $\widehat{FDR}$ as long as these differences offset each other. This is exactly what we observe since $\widehat{FDR}^+$ is much higher than $\widehat{FDR}$ at all significance levels γ. At $\gamma = 0.05$, $\widehat{FDR}^+$ is equal to 56.3%. It means that 29 among the 52 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, $\widehat{FDR}^-$ is close to $\widehat{FDR}$. This reflects 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 27.7%. Stated differently, 72.3% 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, $\widehat{FDR}^-$ increases only slowly.

First of all, these results clearly indicate 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 one of the worst funds. In other words, the proportion of lucky funds is always higher among the best funds at any significance level γ. Second, the results highlight the inaccuracy of the performance assessment under the standard approach. The latter concludes that 9.6% of the funds are able to achieve positive alphas at $\gamma = 0.20$. However, $\widehat{FDR}^{+}$ evidences a completely different picture. Only 1.7% of the funds generate positive alphas, while the remaining funds (7.9%) are purely lucky. Moreover, our analysis confirms that there is a larger proportion of funds with negative rather than positive performance. However, the standard approach concludes that 19.1% of the funds have negative alphas at $\gamma = 0.20$, while our estimation is merely 11.2%.

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 at 1.7%. 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, best and worst G funds are similar to those observed of the *All* funds. This is not surprising since approximately two thirds of the funds in the population are G funds.

Since $\widehat{FDR}^{+}$ is much higher than $\widehat{FDR}^{-}$ at all significance levels γ, luck has a more pronounced impact on the best funds rather than on the worst funds. These results also illustrate why the standard approach cannot correctly quantify the real performance of the G funds. At $\gamma = 0.20$, the latter estimates that a non-negligible proportion of funds (9.6%) yields positive alphas. After accounting for luck, we find that only a tiny fraction of the G funds equal to 1.6% is capable of producing a positive performance. Moreover, the standard approach concludes that 17.9% of the funds yields negative alphas at $\gamma = 0.20$, while our estimate of this proportion is equal to 9.9%.

We observe that $\widehat{T}^{+}/M$ remains constant at 1.6% as γ rises. 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 $\widehat{FDR}$ 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 $\gamma = 0.05$, $\widehat{FDR}^{+}$ is only equal to 23.3%, 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 $\widehat{FDR}^{+}$ by 25.9%. However, its level remains largely inferior to the figures documented for *All* and *G* funds. Concerning the worst funds, $\widehat{FDR}^{-}$ starts at the same level as $\widehat{FDR}^{+}$. However, $\widehat{FDR}^{-}$ rises by only 16.4% 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, but with a more rigorous method. 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 at $\gamma = 0.20$, 14.5% and 19.7% of the funds yield positive and negative alphas respectively, our FDR analysis leads to percentages equal to 7.2% and 14.5%.

$\widehat{T}^{+}/M$ becomes constant at $\gamma = 0.10$ 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, $\widehat{FDR}^+$ is equal to 100% independently of γ. It implies that the best funds are all lucky funds. For instance, the 24 best funds discovered at $\gamma = 0.20$ all have zero alphas. Second, $\widehat{FDR}^-$ starts at 24.3% and increases extremely slowly as γ rises.

Our results unveil that the impact of luck on the performance of the best funds is extremely strong here, since no single GI fund is able to produce a positive alpha. This is in complete contradiction with the conclusions reached by the standard approach, which wrongly infers that a sizable proportion of 7.7% GI funds generate positive alphas at $\gamma = 0.20$. This case exemplifies a clear false discovery in mutual fund performance analysis arising from using an approach which does not incorporate the presence of luck.

Finally, the constant increase in $\widehat{T}^-/M$ reveals 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 $\widehat{FDR}^+$ and the dashed one $\widehat{FDR}^-$. $\widehat{FDR}^-$ 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 that these funds are dispersed in the left tail of the distribution. Although $\widehat{FDR}^+$ 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. $\widehat{FDR}^+$ of the two smallest investment categories yield extreme patterns. First, $\widehat{FDR}^+$ 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 the Performance of the Mutual Fund Industry

Table 4 shows that a sizable proportion π_A of funds in the population yields non-zero alphas. An important financial issue is to know which proportion of funds in the population generate positive or negative performance. 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 Table 1):

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

where $A^+(\gamma)$ and $A^-(\gamma)$ respectively denote the number of funds with positive and negative alphas, which do not have significant p-values (i.e. they are not significant). 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 decreases monotonically[13] both $T^+(\gamma)$ and $T^-(\gamma)$ go up while $A^+(\gamma)$ and $A^-(\gamma)$ go towards zero. As γ increases, $T^+(\gamma)/M$ converges to π_A^+, while $T^-(\gamma)/M$ approaches π_A^-. Using this result, we suggest to take as conservative estimates of the proportion of funds with positive and negative alphas:

$$\widehat{\pi}_A^+ = \frac{\widehat{T}^+(\gamma)}{M}, \tag{16}$$

$$\widehat{\pi}_A^- = \frac{\widehat{T}^-(\gamma)}{M}. \tag{17}$$

The simplest way to choose a sufficiently high significance level γ is to find the minimum significance level such that either $T^+(\gamma)$ or $T^-(\gamma)$ becomes constant[14]. For instance, we observe in Table 5 that that $\widehat{T}^+(\gamma)/M$ remains constant among the four investment categories after a certain significance level γ is reached. In this paper, we use a bootstrap technique which is similar to the data-driven procedure used to determine $\widehat{\pi}_0(\lambda)$ in Section 2.3. γ is chosen such that the mean-squared error of $\widehat{\pi}_A^+(\gamma)$ or $\widehat{\pi}_A^-(\gamma)$

[13] Note that this feature is shared by most test statistics when the sample size grows to infinity. Indeed, standard test statistics 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.

[14] This approach is similar to the eyeballing procedure used to pick up the parameter λ to estimate the proportion of funds with zero alphas $\widehat{\pi}_0(\lambda)$.

is minimized. The Appendix contains further details on the methodology as well as the performance of these estimators based on Monte-Carlo simulations. In all cases, these estimators have a good accuracy.

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 *AG* fund category, which contains 8.1% of funds with positive alphas[15]. 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 truly yield zero alphas, meaning that they generate a sufficient performance to cover their expenses. In fact, the negative average alpha 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 left tail of the cross-sectional alpha distribution.

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))[16]. 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 reveal 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 manage-

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

[16] 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.

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

$$\alpha_P^{\gamma} = FDR^{+}(\gamma) \cdot \alpha_0 + \left(1 - FDR^{+}(\gamma)\right) \cdot \alpha_A^{+} \tag{18}$$

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 (18), 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 corresponding to the 5%-quantile[17]. These respectively amount to 5.3%, 5.8% and 7.7% per year for *All*, G and AG funds. The FDR among the best funds is estimated with $\widehat{FDR}^{+}$. 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 2.35% 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 does not only depend 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

[17] 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.

lucky ones. This is exactly what $\widehat{FDR}^+$ 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. Finally, the portfolio of AG funds generates a substantial alpha equal to 6.16% per year at $\gamma = 0.05$. This high performance is due to the low level of $\widehat{FDR}^+$, implying that most of the funds in the portfolio generate positive alphas.

Please insert Table 7 here

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 and conditional FF models are 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 and conditional versions of the CAPM are quite different from those obtained with the Carhart model. In particular, both $\widehat{FDR}^+$ and $\widehat{FDR}^-$ are higher across the four investment categories. It implies that the CAPM-alphas of the best funds are lower than their Carhart-alphas. Similarly, the CAPM-alphas of the worst funds are lower than their Carhart-alphas. This can be easily explained by the bias introduced by omitting relevant explanatory variables in a linear regression model (Lehman and Modest (1987)). For instance, the CAPM-alphas of the best AG funds are biased downwards 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 CAPM-alphas of the worst GI funds are biased upwards because of the positive exposures of these funds to the size and book-to-market factors, which both have positive premia.

Please insert Table 8 here

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. $\widehat{FDR}^+$ is lower than the one observed during the entire period. It respectively amounts to 21.3% and 38.2% 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 argue that this finding is due to the improvement of market efficiency and to the dilution of performance caused by the rapid increase in the number of mutual funds.

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 $\widehat{FDR}$ 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. To address the financial problem at hand, we further develop new measures which allows us to separately compute the proportion of lucky funds among the best and worst funds. The FDR and these new measures are very easy to compute and therefore represent straightforward extensions of the standard approach developed in the literature. By accounting for the presence of luck, we are able to shed light on important issues that could not be addressed with the previous methodologies. In particular, our approach permits to determine the relative importance as well as location of funds with genuine 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 funds 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 (the worst funds are not bad simply by luck!) 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 produces a positive performance after accounting for luck.

Our results have important implications for the performance of the mutual fund industry. 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 produces 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. The reason is 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 gauge the expected performance of this portfolio. This result suggests that there is interesting work to be done in the future. By controlling the FDR of the portfolio, it could be possible to identify ex ante portfolios that will produce high alphas ex-post. 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 arbitrarily 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:

$$\alpha_P^{\gamma} = E\left(\alpha_P \left| R^+(\gamma) > 0\right.\right) \cdot prob\left(R^+(\gamma) > 0\right). \tag{19}$$

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:

$$\alpha_P^{\gamma} = E\left(\frac{1}{R^+}\sum_{i=1}^{R^+}\alpha_i \middle| R^+(\gamma) > 0\right). \tag{20}$$

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 (20) becomes:

$$\alpha_P^{\gamma} = E\left(\frac{F^+(\gamma)}{R^+(\gamma)} \middle| R^+(\gamma) > 0\right) \cdot \alpha_0 + E\left(\frac{T^+(\gamma)}{R^+(\gamma)} \middle| R^+(\gamma) > 0\right) \cdot \alpha_A^+. \tag{21}$$

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

$$\alpha_P^{\gamma} = FDR^+(\gamma) \cdot \alpha_0 + \left(1 - FDR^+(\gamma)\right) \cdot \alpha_A^+, \tag{22}$$

which is the stated result.

6.2 Monte-Carlo Simulations

In this section, we first check the finite sample performance of the FDR methodology and its extensions based on the new FDR^+ and FDR^- measures. Then, we examine the finite sample performance of our estimators $\widehat{\pi}_A^+$ and $\widehat{\pi}_A^-$ of the proportion of funds with genuine positive and negative performance. We build on a setting matching our performance analysis problem and the mutual fund data at hand.

6.2.1 Design of the Monte-Carlo Experiment

We generate artificial monthly return data according to a one-factor model:

$$\begin{aligned} r_{i,t} &= \alpha_i + \beta \cdot r_{m,t} + \varepsilon_{i,t}, \qquad i = 1, ..., M, \quad t = 1, ..., T, \\ r_{m,t} &\sim N(0, \sigma_{r_m}), \qquad \varepsilon_{i,t} \sim N(0, \sigma_\varepsilon). \end{aligned} \tag{23}$$

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 $\widehat{t}_i$ follows the Student distribution with $T - 2$ degrees of freedom. Under H_A, $\widehat{t}_i$ 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.

A proportion π_0 of the funds comes from H_0 and has an alpha equal to zero. A proportion π_A of funds generate differential performance. Under H_A, a proportion $\pi_A^+ = \pi_A \cdot q^-$ of funds yields a positive alpha α_A^+ and a proportion $\pi_A^- = \pi_A \cdot (1 - q^-)$ of funds yields a negative alpha α_A^-. $q^- \in [0, 1]$ is a positive scalar. We thus have:

$$\begin{aligned} H_0 &: \alpha_i \sim N(0, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with proportion } \pi_0, \\ H_A &: \alpha_i \sim N(\alpha_A^+, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with proportion } \pi_A^+, \\ &: \alpha_i \sim N(\alpha_A^-, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with proportion } \pi_A^-. \end{aligned} \tag{24}$$

The experiment is realized 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. The number of Monte Carlo replications is equal to 1'000.

6.2.2 Estimators presented in Section 2.3

In this section, we successively examine the performance of the following estimators: $\widehat{\pi}_0(\lambda)$, $\widehat{FDR}_\lambda(\gamma)$, $\widehat{FDR}_\lambda^+(\gamma)$ and $\widehat{FDR}_\lambda^-(\gamma)$.

The estimator $\widehat{\pi}_0(\lambda)$ of the proportion of funds with zero alphas $\widehat{\pi}_0(\lambda)$

As shown by Equation (7), the estimator $\widehat{\pi}_0(\lambda)$ depends on the parameter λ that has to be determined. To this end, we use a bootstrap procedure which automatically chooses λ such that the mean-squared error (MSE) of $\widehat{\pi}_0(\lambda)$ is minimized. The method can be described as follows (See Storey (2002), Storey, Taylor and Sigmund (2004) for further details). First, we compute $\widehat{\pi}_0(\lambda)$ across a range of λ ($\lambda = 0.05, 0.10, ..., 0.70$). Second, we form 1'000 bootstrap versions of $\widehat{\pi}_0(\lambda)$ for each possible value of λ. These are respectively denoted by $\widehat{\pi}_0^b(\lambda)$ with $b = 1, ..., 1000$. Third, we compute the MSE for each possible value of λ :

$$\widehat{MSE}(\lambda) = \frac{1}{1'000} \sum_{b=1}^{1'000} \left[\widehat{\pi}_0^b(\lambda) - \min_\lambda \widehat{\pi}_0(\lambda) \right]^2. \tag{25}$$

we choose λ^* such that $\lambda^* = \arg\min_\lambda \widehat{MSE}(\lambda)$. Our estimate of π_0 is then equal to $\widehat{\pi}_0(\lambda^*)$.

Alternatively, we also test a more simple approach where λ is set to 0.5. This corresponds to the eyeballing procedure explained in the text, where λ is chosen at the point where the bars of the histogram shown in Figure 1 becomes flat. These two estimators are compared with the true value π_0 defined in the Monte-Carlo design.

The estimators $\widehat{FDR}_\lambda(\gamma)$, $\widehat{FDR}^+_\lambda(\gamma)$, $\widehat{FDR}^-_\lambda(\gamma)$ of the false discovery rates

The estimator $\widehat{FDR}_\lambda(\gamma)$, is defined by Equation (6). It is compared with the true $FDR(\gamma)$ computed as follows:

$$FDR(\gamma) = \frac{\pi_0 \cdot \gamma}{\pi_0 \cdot \gamma + \pi_A \left[prob\left(t < t_{T-2,\frac{\gamma}{2}} \,|H_A, \alpha_A < 0\right) + prob\left(t > t_{T-2,1-\frac{\gamma}{2}} \,|H_A, \alpha_A > 0\right)\right]}, \tag{26}$$

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 estimator $\widehat{FDR}^+_\lambda(\gamma)$ is defined by Equation (8). It is compared with the true $FDR^+(\gamma)$ computed as follows:

$$FDR^+(\gamma) = \frac{\frac{1}{2} \cdot \pi_0 \cdot \gamma}{\frac{1}{2} \cdot \pi_0 \cdot \gamma + \pi_A^+ \cdot prob\left(t > t_{T-2,1-\frac{\gamma}{2}} \,|H_A, \alpha_A > 0\right)}. \tag{27}$$

The estimator $\widehat{FDR}^-_\lambda(\gamma)$ is defined by Equation (9). It is compared with the true $FDR^-(\gamma)$ computed as follows:

$$FDR^-(\gamma) = \frac{\frac{1}{2} \cdot \pi_0 \cdot \gamma}{\frac{1}{2} \cdot \pi_0 \cdot \gamma + \pi_A^- \cdot prob\left(t < t_{T-2,\frac{\gamma}{2}} \,|H_A, \alpha_A < 0\right)}. \tag{28}$$

Table 9 shows the differences between the average values (over the 1'000 replications) of the different estimators and their theoretical counterparts. Panel A considers the case where the parameter λ is chosen with the bootstrap technique. Panel B examines the case where λ is fixed to 0.5. The simulation results 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 as expected because of its conservative property.

Please insert Table 9 here

6.2.3 Estimators presented in Section 4.3

The estimated proportions $\widehat{\pi}_A^+(\gamma)$ and $\widehat{\pi}_A^-(\gamma)$ of funds in the population with positive and negative alphas are given by Equations (16) and (17). To choose the significance level γ, we use a bootstrap technique which is similar to the one used to determine $\widehat{\pi}_0(\lambda)$.

γ is chosen such that the mean-squared error (MSE) of $\widehat{\pi}_A^+(\gamma)$ and $\widehat{\pi}_A^-(\gamma)$ is minimized. The method can be described as follows. First, we compute $\widehat{\pi}_A^+(\gamma)$ across a range of γ ($\gamma = 0.10, 0.15, ..., 0.25$). Second, we form 1'000 bootstrap versions of $\widehat{\pi}_0^+(\gamma)$ for each possible value of γ. These are respectively denoted by $\widehat{\pi}_A^{b+}(\gamma)$ with $b = 1, ..., 1000$. Third, to compute the MSE for each possible value of γ, we use $\max_\gamma \widehat{\pi}_A^+(\gamma)$ as a plug-in estimate of $\widehat{\pi}_A^+$:

$$\widehat{MSE}^+(\gamma) = \frac{1}{1'000}\sum_{b=1}^{1'000}\left[\widehat{\pi}_A^{b+}(\gamma) - \max_\gamma \widehat{\pi}_A^+(\gamma)\right]^2. \tag{29}$$

We choose γ^{+*} such that $\gamma^{+*} = \arg\min_\gamma \widehat{MSE}^+(\gamma)$. Our estimate of π_A^+ is then equal to $\widehat{\pi}_A^+(\gamma^{+*})$. We use the same procedure for $\widehat{\pi}_0^-(\gamma)$ to determine γ^{-*} such that $\gamma^{-*} = \arg\min_\gamma \widehat{MSE}^-(\gamma)$. In this case, the estimate of π_A^- is then equal to $\widehat{\pi}_A^-(\gamma^{-*})$.

Alternatively, we also test a more simple approach where γ is set to 0.2. This corresponds to an empirical eyeballing procedure where γ is chosen at the point where both $\widehat{\pi}_A^+(\gamma)$ or $\widehat{\pi}_A^-(\gamma)$ become constant. These two estimators are compared with the true values π_A^+ and π_A^- defined in the Monte-Carlo design.

Table 10 shows the differences between the average values (over the 1'000 replications) of the two estimators and their theoretical counterparts. Panel A considers the case where the significance level γ is chosen with the data-driven technique. The simulation results show that the estimators based on the bootstrap procedure have a good accuracy, up to the second decimal. Moreover, it is conservative since the average value of the estimators are most of time lower than the true parameter values. Panel B examines the case where γ is fixed to 0.2. Although the performance of the estimators are slightly worse in this case, they remain close to the true values. Unsurprisingly, we notice that this simple procedure yields better estimates when the power of the test is higher. This is the case for $\widehat{\pi}_A^+$ when $\alpha_A^+ = 8\%$ or for $\widehat{\pi}_A^-$ when $\alpha_A^- = -8\%$. Since we find empirically that funds with positive alphas are located at the extreme right tail of the alpha distribution (meaning that the power of the test of positive performance is likely to be high), fixing γ should also provide a precise approximation of the proportion of funds with positive alphas.

Please insert Table 10 here

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

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

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

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

[7] 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.

[8] 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.

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

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

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

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

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

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

[15] 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.

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

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

[18] 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.

[19] 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.

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

[21] Kosowski, R., A. Timmermann, H. White, and R. Wermers, 2005, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, 2005, Journal of Finance, Forthcoming.

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

[23] 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.

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

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

[26] Romano, J.P., and M. Wolf, 2005, Stepwise Multiple Testing as Formalized Data Snooping, Econometrica 73, 1237-1282.

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

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

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

[30] 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.

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

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
$\widehat{R}$	156	248	346	422
$\widehat{R}^{+}$	52	83	112	139
$\widehat{R}^{-}$	104	165	234	283

γ	0.05	0.10	0.15	0.20
$\widehat{R}/M$	10.6%	16.8%	23.5%	28.6%
$\widehat{R}^{+}/M$	3.5%	5.6%	7.6%	9.4%
$\widehat{R}^{-}/M$	7.1%	11.2%	15.9%	19.2%

Panel B: *G* funds

γ	0.05	0.10	0.15	0.20
$\widehat{R}$	106	171	235	282
$\widehat{R}^{+}$	36	57	78	98
$\widehat{R}^{-}$	70	114	157	184

γ	0.05	0.10	0.15	0.20
$\widehat{R}/M$	10.3%	16.7%	22.9%	27.5%
$\widehat{R}^{+}/M$	3.5%	5.6%	7.6%	9.6%
$\widehat{R}^{-}/M$	6.8%	11.1%	15.3%	17.9%

Panel C: *AG* funds

γ	0.05	0.10	0.15	0.20
$\widehat{R}$	39	54	68	80
$\widehat{R}^{+}$	18	25	30	34
$\widehat{R}^{-}$	21	29	38	46

γ	0.05	0.10	0.15	0.20
$\widehat{R}/M$	16.6%	23.1%	29.0%	34.2%
$\widehat{R}^{+}/M$	7.7%	10.7%	12.8%	14.5%
$\widehat{R}^{-}/M$	8.9%	12.4%	16.2%	19.7%

Panel D: *GI* funds

γ	0.05	0.10	0.15	0.20
$\widehat{R}$	30	52	76	91
$\widehat{R}^{+}$	6	12	16	24
$\widehat{R}^{-}$	24	40	60	69

γ	0.05	0.10	0.15	0.20
$\widehat{R}/M$	9.6%	16.7%	24.5%	29.3%
$\widehat{R}^{+}/M$	1.9%	3.8%	5.2%	7.7%
$\widehat{R}^{-}/M$	7.7%	12.9%	19.3%	22.2%

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	78.4%	21.6%
G funds	80.4%	19.6%
AG funds	71.5%	28.5%
GI funds	75.3%	24.7%

The first column contains the estimated proportion $\widehat{\pi}_0$ of funds with no performance (zero alphas), and the second one contains the estimated proportion $\widehat{\pi}_A$ of funds with differential performance (non-zero alpha). For each investment category, $\widehat{\pi}_0$ is determined with a bootstrap procedure which minimizes the mean-squared error of the estimator (details are provided in the Appendix). 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
$\widehat{FDR}$	37.0%	46.5%	50.0%	54.5%	$\widehat{FDR}$	37.0%	46.5%	50.0%	54.5%
$\widehat{R}$	156	248	346	422	$\widehat{R}/M$	10.6%	16.8%	23.5%	28.6%
$\widehat{F}$	58	115	173	231	$\widehat{F}/M$	3.9%	7.8%	11.7%	15.6%
$\widehat{T}$	98	133	173	191	$\widehat{T}/M$	6.7%	9.0%	11.7%	13.0%

Best funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	55.5%	69.5%	76.8%	82.1%	$\widehat{FDR}^+$	55.5%	69.5%	76.8%	82.1%
$\widehat{R}^+$	52	83	112	139	$\widehat{R}^+/M$	3.5%	5.6%	7.6%	9.4%
$\widehat{F}^+$	29	58	87	116	$\widehat{F}^+/M$	1.9%	3.9%	5.9%	7.9%
$\widehat{T}^+$	23	25	25	25	$\widehat{T}^+/M$	1.6%	1.7%	1.7%	1.7%

Worst funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^-$	27.7%	34.9%	37.0%	40.9%	$\widehat{FDR}^-$	27.7%	34.9%	37.0%	40.9%
$\widehat{R}^-$	104	165	234	283	$\widehat{R}^-/M$	7.1%	11.2%	15.9%	19.2%
$\widehat{F}^-$	29	58	87	116	$\widehat{F}^-/M$	1.9%	3.9%	5.9%	7.9%
$\widehat{T}^-$	75	107	147	167	$\widehat{T}^-/M$	5.1%	7.3%	10.0%	11.2%

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 estimated 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
$\widehat{FDR}$	38.8%	48.2%	52.6%	58.5%	$\widehat{FDR}$	38.8%	48.2%	52.6%	58.5%
$\widehat{R}$	106	171	235	282	$\widehat{R}/M$	10.3%	16.7%	22.9%	27.5%
$\widehat{F}$	41	82	124	165	$\widehat{F}/M$	4.0%	8.0%	12.0%	16.0%
$\widehat{T}$	65	89	111	117	$\widehat{T}/M$	6.3%	8.6%	10.8%	11.4%

Best funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	57.2%	72.3%	79.2%	84.2%	$\widehat{FDR}^+$	57.2%	72.3%	79.2%	84.2%
$\widehat{R}^+$	36	57	78	98	$\widehat{R}^+/M$	3.5%	5.6%	7.6%	9.6%
$\widehat{F}^+$	21	41	62	82	$\widehat{F}^+/M$	2.0%	4.0%	6.0%	8.0%
$\widehat{T}^+$	15	16	16	16	$\widehat{T}^+/M$	1.5%	1.6%	1.6%	1.6%

Worst funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^-$	29.4%	36.1%	39.3%	44.8%	$\widehat{FDR}^-$	29.4%	36.1%	39.3%	44.8%
$\widehat{R}^-$	70	114	157	184	$\widehat{R}^-/M$	6.8%	11.1%	15.3%	17.9%
$\widehat{F}^-$	21	41	62	82	$\widehat{F}^-/M$	2.0%	4.0%	6.0%	8.0%
$\widehat{T}^-$	49	73	95	102	$\widehat{T}^-/M$	4.8%	7.1%	9.3%	9.9%

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
$\widehat{FDR}$	21.5%	31.0%	37.5%	41.8%	$\widehat{FDR}$	21.5%	31.0%	37.5%	41.8%
$\widehat{R}$	39	54	68	80	$\widehat{R}/M$	16.6%	23.1%	29.0%	34.2%
$\widehat{F}$	8	16	25	33	$\widehat{F}/M$	3.4%	7.0%	10.6%	14.1%
$\widehat{T}$	31	38	43	47	$\widehat{T}/M$	13.2%	16.2%	18.4%	20.1%

Best funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	23.3%	33.4%	43.3%	49.2%	$\widehat{FDR}^+$	23.3%	33.4%	43.3%	49.2%
$\widehat{R}^+$	18	25	30	34	$\widehat{R}^+/M$	7.7%	10.7%	12.7%	14.5%
$\widehat{F}^+$	4	8	13	17	$\widehat{F}^+/M$	1.8%	3.5%	5.5%	6.5%
$\widehat{T}^+$	14	17	17	17	$\widehat{T}^+/M$	5.9%	7.2%	7.2%	7.2%

Worst funds

γ	0.05	0.10	0.15	0.20		0.05	0.10	0.15	0.20
$\widehat{FDR}^-$	20.0%	28.9%	33.0%	36.4%	$\widehat{FDR}^-$	20.0%	28.9%	33.0%	36.4%
$\widehat{R}^-$	21	29	38	46	$\widehat{R}^-/M$	8.9%	12.4%	16.3%	19.7%
$\widehat{F}^-$	4	8	13	17	$\widehat{F}^-/M$	1.8%	3.5%	5.5%	6.5%
$\widehat{T}^-$	17	21	25	29	$\widehat{T}^-/M$	7.1%	8.9%	10.7%	12.5%

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
$\widehat{FDR}$	38.6%	44.8%	46.0%	51.2%	$\widehat{FDR}$	38.6%	44.8%	46.0%	51.2%
$\widehat{R}$	30	52	76	91	$\widehat{R}/M$	9.6%	16.7%	24.5%	29.3%
$\widehat{F}$	12	23	35	47	$\widehat{F}/M$	3.7%	7.5%	11.2%	15.0%
$\widehat{T}$	18	29	41	44	$\widehat{T}/M$	5.8%	9.3%	13.2%	14.3%

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	24	$\widehat{R}^+/M$	1.9%	3.8%	5.2%	7.7%
$\widehat{F}^+$	6	12	16	24	$\widehat{F}^+/M$	1.9%	3.8%	5.2%	7.7%
$\widehat{T}^+$	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
$\widehat{FDR}^-$	24.3%	29.1%	29.2%	33.8%	$\widehat{FDR}^-$	24.3%	29.1%	29.2%	33.8%
$\widehat{R}^-$	24	40	60	69	$\widehat{R}^-/M$	7.7%	12.9%	19.3%	22.2%
$\widehat{F}^-$	6	12	16	24	$\widehat{F}^-/M$	1.9%	3.8%	5.2%	7.7%
$\widehat{T}^-$	18	28	44	45	$\widehat{T}^-/M$	5.8%	9.1%	14.2%	14.5%

Table 6

Source of Differential Performance

	Positive performance $\widehat{\pi}_A^+$	Negative performance $\widehat{\pi}_A^-$
All funds	1.9%	19.6%
G funds	1.5%	18.0%
AG funds	8.1%	20.3%
GI funds	0.0%	24.3%

The first column contains the estimated proportion $\widehat{\pi}_A^+$ of funds with positive performance (positive alpha). The second column contains the estimated proportion $\widehat{\pi}_A^-$ of funds with negative performance (negative alpha). For each investment category, these two proportions are determined with a bootstrap procedure which minimizes the mean-squared error of the estimator (details are provided in the Appendix). 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
$\widehat{FDR}^+$	55.5%	69.5%	76.8%	82.1%
α_p	2.35%	1.61%	1.22%	0.94%
$d\alpha_p/\alpha_p$		-31.4%	-25.5%	-22.9%

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

γ	0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	57.2%	72.3%	79.2%	84.2%
α_p	2.47%	1.60%	1.21%	0.92%
$d\alpha_p/\alpha_p$		-35.2%	-24.4%	-24.0%

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

γ	0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	23.3%	33.4%	43.3%	49.2%
α_p	5.90%	5.12%	4.36%	3.90%
$d\alpha_p/\alpha_p$		-13.2%	-14.8%	-10.5%

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 estimated 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%	55.3%	72.1%	$\widehat{FDR}^+$	100%	100%	35.5%	51.0%
$\widehat{FDR}^-$	45.9%	68.9%	17.2%	29.0%	$\widehat{FDR}^-$	49.8%	76.2%	13.4%	24.1%

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%	68.6%	80.0%	$\widehat{FDR}^+$	100%	100%	35.7%	51.2%
$\widehat{FDR}^-$	48.4%	67.6%	21.3%	35.4%	$\widehat{FDR}^-$	52.8%	76.2%	17.1%	28.2%

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%	20.2%	29.4%	$\widehat{FDR}^+$	100%	100%	18.0%	27.3%
$\widehat{FDR}^-$	32.4%	49.3%	22.6%	46.4%	$\widehat{FDR}^-$	44.0%	60.1%	22.8%	35.9%

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%
$\widehat{FDR}^-$	51.0%	76.7%	12.2%	18.6%	$\widehat{FDR}^-$	70.5%	79.5%	7.4%	12.5%

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 estimated 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 estimated 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

Performance of the FDR Estimators using Monte-Carlo Simulations

Panel A: Bootstrap Procedure

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

q^-	γ	π_0	$\widehat{\pi}_0$	FDR	$\widehat{FDR}$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.696	0.114	0.114	0.078	0.078	0.211	0.211
		0.9	0.890	0.332	0.328	0.246	0.243	0.508	0.505
	0.10	0.7	0.696	0.198	0.197	0.143	0.143	0.321	0.321
		0.9	0.892	0.489	0.482	0.393	0.387	0.646	0.640
0.7	0.05	0.7	0.704	0.127	0.128	0.165	0.166	0.103	0.104
		0.9	0.895	0.359	0.357	0.433	0.430	0.307	0.305
	0.10	0.7	0.704	0.211	0.212	0.281	0.283	0.168	0.168
		0.9	0.894	0.508	0.502	0.602	0.599	0.439	0.435

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

q^-	γ	π_0	$\widehat{\pi}_0$	FDR	$\widehat{FDR}$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.711	0.138	0.140	0.103	0.104	0.211	0.214
		0.9	0.897	0.382	0.382	0.307	0.307	0.508	0.508
	0.10	0.7	0.710	0.221	0.222	0.168	0.170	0.321	0.324
		0.9	0.901	0.523	0.524	0.439	0.440	0.646	0.647
0.7	0.05	0.7	0.710	0.138	0.140	0.211	0.214	0.103	0.104
		0.9	0.897	0.382	0.381	0.508	0.507	0.307	0.305
	0.10	0.7	0.709	0.221	0.222	0.321	0.323	0.168	0.169
		0.9	0.902	0.523	0.524	0.646	0.647	0.439	0.439

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

q^-	γ	π_0	$\widehat{\pi}_0$	FDR	$\widehat{FDR}$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.704	0.127	0.128	0.103	0.104	0.165	0.166
		0.9	0.894	0.359	0.355	0.307	0.303	0.433	0.428
	0.10	0.7	0.704	0.211	0.212	0.168	0.169	0.281	0.283
		0.9	0.903	0.508	0.510	0.439	0.440	0.602	0.603
0.7	0.05	0.7	0.697	0.114	0.114	0.211	0.211	0.078	0.078
		0.9	0.895	0.332	0.330	0.508	0.506	0.246	0.242
	0.10	0.7	0.697	0.198	0.199	0.321	0.321	0.143	0.143
		0.9	0.902	0.489	0.489	0.646	0.647	0.393	0.394

The monthly returns are simulated according to a 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, a proportion $\pi_A^- = \pi_A \cdot q^-$ of funds has a negative alpha and a proportion $\pi_A^+ = \pi_A \cdot (1 - q^-)$ of funds has a positive alpha, with $q^- \in [0, 1]$. FDR, FDR^+ and FDR^- correspond to the true false discovery rates. $\widehat{\pi}_0$, $\widehat{FDR}$, $\widehat{FDR}^+$ and $\widehat{FDR}^-$ stand for the average value of the estimators across 1'000 Monte-Carlo simulations. In Panel A, the parameter λ used to compute $\widehat{\pi}_0$ is chosen with the bootstrap procedure explained in the Appendix. In Panel B, the parameter λ is fixed to 0.5.

Table 9

Performance of the FDR Estimators using Monte-Carlo Simulations

Panel B: Fixed λ equal to 0.5

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

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

Table 10

Performance of the Estimators of the Proportion of Funds with Positive and Negative Alphas using Monte-Carlo Simulations

Panel A: Bootstrap Procedure

		$\alpha_A^+ = 8\%, \alpha_A^- = -5\%$			
q^-	π_0	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.214	0.09	0.085
	0.9	0.07	0.075	0.03	0.032
0.7	0.7	0.09	0.088	0.21	0.194
	0.9	0.03	0.035	0.07	0.068
		$\alpha_A^+ = 5\%, \alpha_A^- = -5\%$			
q^-	π_0	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.193	0.09	0.083
	0.9	0.07	0.075	0.03	0.032
0.7	0.7	0.09	0.083	0.21	0.194
	0.9	0.03	0.031	0.07	0.068
		$\alpha_A^+ = 5\%, \alpha_A^- = -8\%$			
q^-	γ	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.194	0.09	0.092
	0.9	0.07	0.068	0.03	0.033
0.7	0.7	0.09	0.084	0.21	0.210
	0.9	0.03	0.032	0.07	0.074

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, a proportion $\pi_A^- = \pi_A \cdot q^-$ of funds has a negative alpha and a proportion $\pi_A^+ = \pi_A \cdot (1 - q^-)$ of funds has a positive alpha, with $q^- \in [0, 1]$. π_A^+ and π_A^+ respectively correspond to the true proportion of funds with positive and negative alphas. $\widehat{\pi}_A^+$ and $\widehat{\pi}_A^-$ stand for the average value of the estimators across 1'000 Monte-Carlo simulations. In Panel A, the parameter γ used to compute $\widehat{\pi}_A^+$ and $\widehat{\pi}_A^-$ is chosen with the bootstrap procedure explained in the Appendix. In Panel B, the parameter γ is fixed to 0.2.

Table 10

Performance of the Estimators of the Proportion of Funds with Positive and Negative Alphas using Monte-Carlo Simulations

Panel B: Fixed γ equal to 0.2

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

q^-	π_0	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.209	0.09	0.081
	0.9	0.07	0.070	0.03	0.027
0.7	0.7	0.09	0.088	0.21	0.187
	0.9	0.03	0.030	0.07	0.064

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

q^-	π_0	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.188	0.09	0.080
	0.9	0.07	0.063	0.03	0.028
0.7	0.7	0.09	0.079	0.21	0.189
	0.9	0.03	0.027	0.07	0.063

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

q^-	π_0	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.188	0.09	0.089
	0.9	0.07	0.064	0.03	0.030
0.7	0.7	0.09	0.081	0.21	0.209
	0.9	0.03	0.028	0.07	0.070

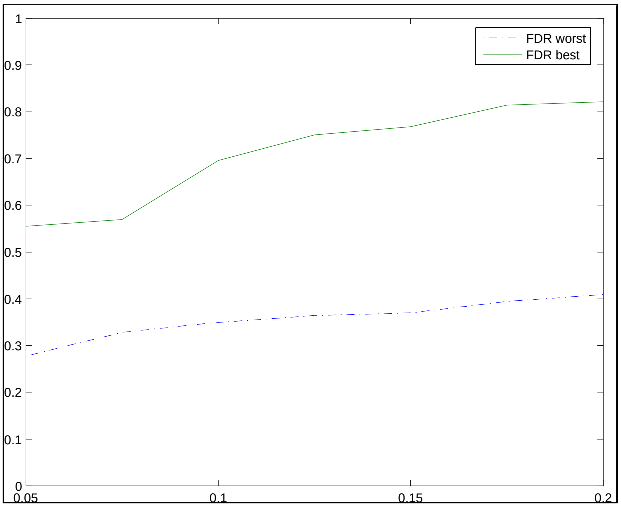
Figure 1

Histogram of the Fund Estimated p-values

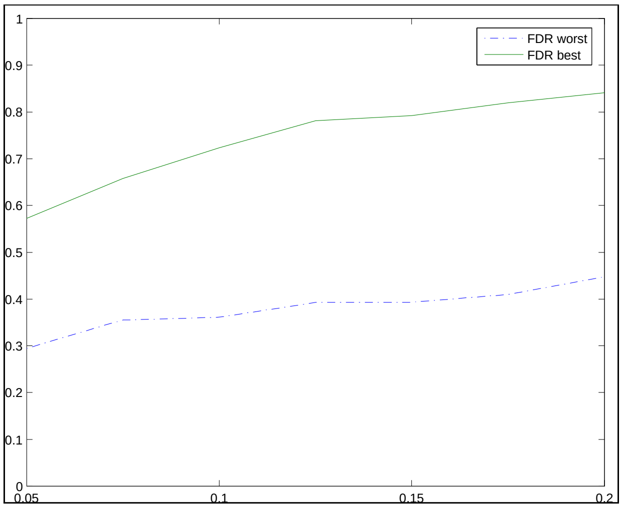
We simulate fund excess returns for 1'472 funds and 336 observations with a one-factor market model (see the Appendix for the details). From these simulated time-series, the fund alphas and p-values are estimated. The proportion π_0 of funds drawn from H_0 is equal to 80%. Under H_A, an equal proportion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

Figure 2

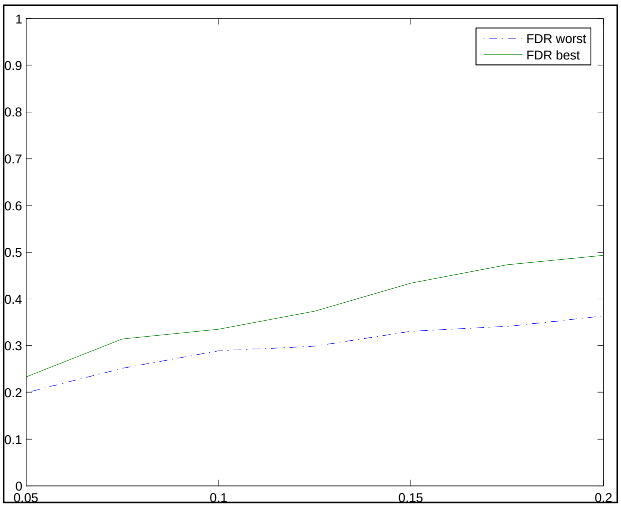
False Discovery Rates among the Best and the Worst Funds



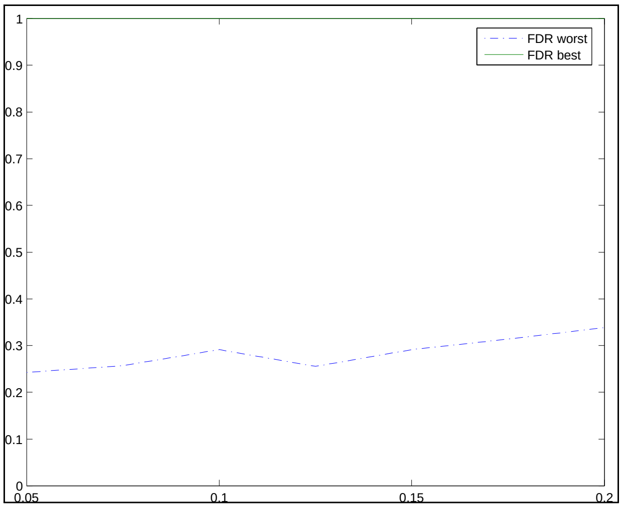
(a) *All* funds



(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 estimated FDR among the best funds ($\widehat{FDR}^{+}$) and the dashed one the estimated 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.