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

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

First version, September 2005
This version, July 2006

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

*We are grateful to B. Dumas, R. Kosowski, E. Ronchetti, R. Stulz, M.-P. Victoria-Feser, and M. Wolf, as well as seminar participants at Banque Cantonale de Genève, BNP Paribas, Bilgi University, Grequam, INSEAD, London School of Economics, Queen Mary, Solvay Business School, University of Geneva, Universita della Svizzera Italiana, the 2005 Imperial College Risk Management Workshop, the 2005 Swiss Doctoral Workshop, the 2006 Research and Knowledge Transfer Conference, the annual meetings of EC^2 2005, EURO XXI 2006, ICA 2006, AFFI 2006, and SGF 2006 for their helpful comments. The first and second authors acknowledge financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK). Part of this research was done while the second author was visiting the Centre Emile Bernheim (ULB).

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

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

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

Abstract

Standard testing approaches designed to identify funds with non-zero alphas do not account for the presence of lucky funds. Lucky funds have a significant estimated alpha, while their true alpha is equal to zero. 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 number and the proportion of funds with truly positive and negative performance in any portion of the tails of the cross-sectional alpha distribution. Using a large cross-section of U.S. domestic-equity funds, we find that 76.6% of them have zero alphas. 21.3% yield negative performance and are dispersed in the left tail of the alpha distribution. The remaining 2.1% with positive alphas are located at the extreme right tail. The same analysis is run on three investment categories (growth, aggressive growth, growth and income funds), as well as groups formed according to lagged fund characteristics (turnover, expense ratio, total net asset value).

Introduction

Measuring individual fund performance is important for investors as well as managers of fund of funds aiming at selecting performing funds. Evaluating the proportion of funds with non-zero alphas is also of central interest for academics. To detect funds with differential performance (i.e. funds with positive or negative alphas), the performance of all M funds in the population has to be examined. For each fund, the null hypothesis that its alpha is equal to zero is tested. If the fund p-value is inferior to a chosen significance level γ (e.g. 5 percent), the null hypothesis is rejected and the fund has a significant estimated alpha. This procedure corresponds to a multiple-hypothesis test, because the null hypothesis of no performance is not tested once, but M times (one time for each fund). It differs from the usual single-hypothesis test run on the estimated alpha of the equally-weighted portfolio of all funds[1]. For each of the M funds, the inference based on its estimated alpha 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. The main difficulty raised by the multiple-hypothesis test is to measure the impact of luck on individual fund performance in order to determine the number of funds with truly positive or negative performance.

The previous literature, referred to as the standard approach, proposes to measure the number of funds with differential performance by the number of significant funds (Jensen (1968), Ferson and Schadt (1996), Ferson and Qian (2004)). Stated differently, it simply counts the number of funds which are located at the tails of the cross-sectional alpha distribution. This approach is problematic, because it does not account for the presence of lucky funds among these significant funds. To illustratre this issue, 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. Even if all funds produce in reality zero alphas, we would still expect some of the 200 funds to exhibit significant positive estimated alphas simply by luck[2]. A second issue is that the standard approach cannot determine the consequences of changes in the significance level γ. As we increase γ, we go further towards the center of the cross-sectional alpha distribution, and logically find more significant funds. But without accounting for luck,

[1] This test is frequently proposed in the literature to assess the average performance of the mutual fund industry (see, for instance, Lehman and Modest (1987), 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, we would expect some funds to achieve superior risk-adjusted returns by chance."

we are unable to know whether these additional significant funds are lucky or truly yield differential performance.

This paper measures the real performance of individual funds by accounting for the impact of luck. To this end, we use a measure called the False Discovery Rate (FDR). The FDR is defined as the proportion of lucky funds among the significant funds at any significance level γ. It is a simple indicator of luck among the funds located at the tails of the alpha distribution: it takes the value of zero when all significant funds truly yield non-zero alphas, and rises to one when all significant funds are lucky. In order to evaluate the impact of luck separately in the left and right tails, we develop new FDR measures. They allow us to compute the FDR among the best funds, namely the funds with significant positive estimated alphas, and the FDR among the worst funds, namely the funds with significant negative estimated alphas. A main virtue of these FDR-based measures is that they are straightforward to compute from the fund estimated p-values, and are therefore direct extensions of the standard approach.

Our first contribution is to precisely estimate the number and the proportion of funds with truly positive and negative performance in different portion of the tails of the cross-sectional alpha distribution. This is simply done by computing the FDR at different significance levels γ (e.g. 0.05, 0.10,...). We can assess the impact of luck by comparing our estimates with those obtained with the standard approach. We also determine the location of the funds with genuine positive and negative performance in the tails of the alpha distribution. Fund location allows us to measure the difference in alphas produced by performing and unskilled funds. It is also useful for portfolio selection, by indicating whether performing funds can be easily detected by the investor. Our second contribution is to estimate the proportions of funds in the population with zero, positive, and negative alphas. This approach gives a finer representation of the performance of the mutual fund industry than the commonly-used average alpha. Moreover, our measures provide a unique test of the Berk and Green (2004) model, asserting that mutual funds must yield zero alphas in equilibrium.

Our empirical results are based on monthly returns of 1,456 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 (All), as well as the cross-section of each of three investment-objective categories, Growth (G), Aggressive-Growth (AG), and Growth and Income (GI). We find that the impact of luck on the performance of

the best funds is substantial. Except for *AG* funds, the FDR among the best funds is always superior to 50%, meaning that more than half of the best funds are simply lucky. For this reason, our estimates of the number of truly performing funds is completely different from the standard approach. For instance, while the standard approach concludes that 7.1% of the *GI* funds generate positive alphas (at $\gamma = 0.2$), we find that all of them are purely lucky. Finding 7.1% instead of 0% is clearly a false discovery[3]! The presence of luck is less pronounced among the worst funds, since the FDR is always lower than 50% across the four investment categories (*All*, *G*, *AG*, and *GI*). In this case, we reach the same qualitative conclusion as the standard approach-that is, there are many funds with negative alphas in the left tail of the alpha distribution. Yet, our measures differ. As an illustration, we estimate that, at $\gamma = 0.2$, 14.2% of *All* funds are unskilled, instead of 21.9% for the standard approach. Concerning the fund location, there are similarities across the four categories. We observe that the funds with truly negative alphas are dispersed in the left tail of the cross-sectional alpha distribution. On the contrary, the funds with genuine positive performance are concentrated at the extreme right tail. This result implies that the average alpha generated by the skilled funds is higher than the one produced by the unskilled ones. It has also implications on portfolio management. Since performing funds are located at the extreme right tail, they can be easily detected by taking a sufficiently low γ. For instance, at $\gamma = 0.10$, the FDR among the best *AG* funds only amounts to 31%, meaning that 70% of the best funds truly generate positive performance. Although there are few funds with positive performance, an investor can still form portfolios with positive alphas.

We observe that approximately 76.6% of *All* funds generate alphas equal to zero, which strongly supports the predictions made by Berk and Green (2004). From the 23.4% remaining *All* funds, 21.3% produce negative alphas. As a result, the negative average alpha documented in the previous literature does not reflect the performance of the whole industry, as it is caused by only 20% of them. Finally, a tiny fraction of 2.1% of *All* funds yield truly positive alphas. Looking at the remaining categories (*G*, *AG*, and *GI*), the proportion of unskilled funds is fairly similar. The main differences appear with the proportion of performing funds. While the proportion among *G* funds is close to the one observed for *All* funds (1.7%), a sizable proportion of 8.4% of *AG* funds produce positive alphas. On the contrary, none of the *GI* funds is skilled. Investors looking for

[3]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 to be a false discovery.

positive alphas should therefore concentrate on AG funds and discard GI funds.

Finally, we examine how the performance of All funds is related to three lagged fund characteristics, the turnover, the expense ratio, and the total net asset value (TNA). High turnover funds contain a higher proportion of funds with negative alphas than Low turnover funds (19.3% versus 14.9%). It may be the case that some unskilled funds trade frequently to convince the investors of their abilities. Since the proportion of funds with positive alphas is comparable across the High and Low turnover funds (2.6% versus 1.8%), there is no strong positive link between performance and turnover. High expense funds contain a much lower proportion of unskilled funds than Low expense funds (14.7% versus 26.3%), but absolutely no skilled ones (0% versus 4.1%). These results show that, although Low expense funds produce most performing funds in the population, more than a quarter of them yield negative alphas. Low TNA funds contain the same proportion of funds with negative alphas as High TNA funds (21.2% versus 20.7%), but absolutely no performing ones (0% versus 4.2%). Therefore, a minority of large funds seem to benefit from economies of scale. In all these High and Low groups of funds, the majority of funds yield zero alphas (the proportion ranges from 69.6% up to 83.3%). This explains why average alphas computed across the different groups cannot capture their differences (see, for instance, Grinblatt and Titman (1989) and Chen, Jegadeesh and Titman (2000)). Comparing the proportions of funds with positive and negative performance therefore allows to better understand the link between fund characteristics and performance.

The remainder of the paper is as follows. The next section defines the notion of luck and the intuition behind the FDR. In Section 2, we give the definition and the estimation procedure of the FDR among the best and worst funds, as well as the proportions of funds with positive and negative alphas. Section 3 presents the performance measures and the mutual fund data. Section 4 contains the results of the paper. An appendix contains the details of the estimation procedure, as well as a Monte-Carlo study on the accuracy of our new measures of luck.

1 Luck in Individual Fund Performance Measurement

1.1 The Definition of Luck

We consider a mutual fund universe 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. Under the null hypothesis, fund i achieves no performance ($\alpha_i = 0$), while under the alternative hypothesis, it delivers differential performance ($\alpha_i > 0$ or $\alpha_i < 0$).

To detect funds with differential performance, we must examine the performance of all funds in the population. To this end, the researcher (or the investor) runs a two-sided test on *each* fund estimated alpha. For each fund i, the researcher compares the fund estimated p-value with a significance level γ (e.g., 0.05, 0.10) previously determined. If the p-value is smaller than γ, the null hypothesis of no performance is rejected, and fund i has a significant estimated alpha at the significance level γ. This procedure differs from a standard single-hypothesis test, because we examine the performance of M different funds, instead of a single one. For this reason, this procedure is referred to as a multiple-hypothesis test, as indicated by the running subscript from 1 to M:

$$\begin{aligned} H_{0,1} &: \alpha_1 = 0, \text{ versus } H_{A,1} : \alpha_1 > 0 \text{ or } \alpha_1 < 0, \\ ... &: ... \\ H_{0,M} &: \alpha_M = 0, \text{ versus } H_{A,M} : \alpha_M > 0 \text{ or } \alpha_M < 0. \end{aligned} \tag{1}$$

Given the finite amount of data, the inference of the alpha of each fund 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 the fund estimated alpha is significant, while the true alpha is equal to zero. Since the test of no performance is run M times, we refer to luck as the number of lucky funds.

1.2 The Impact of Luck on Performance

The previous literature (Jensen (1968), Ferson and Schadt (1996), Ferson and Qian (2004)) proposes to estimate the number of funds with differential performance by the number of significant funds. These funds are, by definition, located at the tails of the cross-sectional alpha distribution. This method, which we call the standard approach, does not account for the impact of luck on the performance of these significant funds. To illustrate this issue, Table 1 classifies the outcomes of the multiple-hypothesis test displayed in Equation (1). $F(\gamma)$ denotes the number of lucky funds, while $T(\gamma)$ stands for the number of funds with differential performance. Adding $F(\gamma)$ and $T(\gamma)$ gives the total number $R(\gamma)$ of significant funds. All these quantities depend on the significance

level γ set by the researcher.

Please Insert Table 1 here

First, the standard approach measures differential performance by $R(\gamma)$, the number of 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 $T(\gamma) = R(\gamma) - F(\gamma)$. Second, the standard approach cannot measure how the significance level γ affects the assessment of individual fund performance. γ determines the portion of the tail of the cross-sectional alpha distribution being investigated. As γ is increased, we go from the extreme part of the tails towards the center of the distribution. As a result, we increase mechanically the number $R(\gamma)$ of significant funds. However, the standard approach cannot determine whether this rise is due to the inclusion of new lucky funds, or to additional funds with genuine differential performance. To address these issues, we propose a new approach, the False Discovery Rate.

1.3 The False Discovery Rate (FDR)

1.3.1 The Basic Idea

The FDR is defined as the expected proportion of the $F(\gamma)$ lucky funds among the $R(\gamma)$ significant funds at the significance level γ. Obviously, when $R(\gamma) = 0$, there are no lucky funds to detect. For this reason, the expected proportion is conditioned on positive values of $R(\gamma)$:

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

The FDR is an indicator of luck present in the tails of the cross-sectional alpha distribution. When all significant funds yield differential performance, the FDR is equal to 0. When all significant funds are simply lucky, the FDR amounts to 1. Contrary to the standard approach, it allows us to precisely estimate the number $T(\gamma)$ of funds with differential performance at any point of the tails of the alpha distribution. Another advantage of the FDR is its computational tractability: the only input required is the M fund estimated p-values delivered by any regression package.

For a single-hypothesis test on only one alpha, we commonly fix the significance level γ (or the $Size$) in order to control for luck. γ is an error measure which determines the probability of finding a lucky fund (i.e. the probability of committing a Type I error).

Similarly to γ, we use the FDR to control for luck in a multiple-hypothesis test. The FDR is a compound error measure which determines the importance of lucky funds (i.e the compound type I error) in a test of differential performance among M funds. Equation (2) shows the basic FDR formula for a two-sided multiple test proposed in the statistical literature (Benjamini and Hochberg (1995), Storey (2002))[4]. In the next section, we propose a new FDR approach, which allows us to gauge separately the number of funds with truly positive and negative alphas.

1.3.2 Comparison with Existing Methods

In a recent paper, Kosowski et al. (2006) also discuss the impact of luck on mutual fund performance. However, their objective is different; they implement a single-hypothesis test on 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...). The inference about a pre-ranked estimated alpha is more difficult, since the entire cross-section of the fund alphas must be taken into account[5]. They use the term luck to stress that they correctly compute the p-value. Our definition of luck as well as the contributions of our paper are completely different, since they are related to the multiple-hypothesis testing problem occuring when the performance of M funds is examined[6].

Other methods are closer to ours since they deal with multiple hypothesis in mutual fund performance. First, Grinblatt and Titman (1989, 1993) jointly test the restriction that the alphas of all funds are equal to zero (i.e. $\alpha_1 = ... = \alpha_M = 0$). However, this method is not informative enough, as it only indicates whether there is at least one fund with non-zero alpha. The second approach is to use another compound error measure, the FamilyWise Error Rate ($FWER$). It is defined as the probability of yielding at least one lucky fund among the M tested funds (Romano and Wolf (2005)). The procedure consists in detecting a number of significant funds such that the $FWER$ is controlled

[4] Strictly speaking, our definition corresponds to the positive False Discovery Rate ($pFDR$). The FDR is defined as $E\left(\frac{F(\gamma)}{R(\gamma)}\middle| R(\gamma) > 0\right) \cdot prob(R(\gamma) > 0)$. 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).

[5] Consider the alpha of the best fund denoted by α^{top}. Expressing the null and the alternative hypotheses as $H_0 : \alpha^{top} = \max_{i=1,...,M}\{\alpha_i\} \leqslant 0$ and $H_A : \alpha^{top} = \max_{i=1,...,M}\{\alpha_i\} > 0$ makes it clear that the distribution of the test statistic depends on the joint distribution of the alphas of all funds.

[6] In the context of Kosowski et al. (2006), we face a similar multiple testing problem if we wish to know how many individual funds above given quantiles of the cross-section of estimated alphas have truly non-zero alphas.

at a given level (usually 0.01, 0.05, or 0.10). Contrary to the FDR, the $FWER$ is not useful to address the questions raised in this paper, since it does not measure the proportion of lucky funds among the significant funds.

2 Measuring The Impact of Luck on Performance

2.1 The FDR among the Best and Worst Funds

2.1.1 Definition

Knowing the number of funds with differential performance is not so interesting per se, since these funds can either yield positive or negative alphas. To address this issue, the standard approach partitions the $R(\gamma)$ significant funds into two groups, the best and worst funds. The best funds are the $R^{+}(\gamma)$ funds with significant positive estimated alphas. They are, by construction, located in the right tail of the alpha distribution. The worst funds correspond to the $R^{-}(\gamma)$ funds with negative estimated alphas. These funds are located in the left tail. $R^{+}(\gamma)$ and $R^{-}(\gamma)$ are then used as estimators of the number of funds with truly positive and negative alphas. However, these estimators present the same drawbacks as $R(\gamma)$, because they do not account for luck. Among the $R^{+}(\gamma)$ and $R^{-}(\gamma)$ funds, some funds are just lucky and yield zero alphas.

In order to measure the impact of luck separately in the right and left tails, we develop a new methodology which extends the basic FDR formula shown in Equation (2). We call our new measures the FDR among the best funds and the FDR among the worst funds. Using these measures, we can estimate the numbers $T^{+}(\gamma)$ and $T^{-}(\gamma)$ of funds with truly positive and negative performance in any portion of the right and left tails of the alpha distribution. At a given significance level γ, we know that $F(\gamma)$ among the $R(\gamma)$ significant alphas are lucky. Since the test of the null hypothesis of no performance is a two-sided test with equal-tailed significance level, $\gamma/2$, we expect that half of these zero-alpha funds have positive estimated alphas and half of them negative estimated alphas[7]. Because lucky funds satisfy the null hypothesis by definition, this result is independent of the proportion of funds with truly positive and negative alphas in the population. We can therefore divide $F(\gamma)$ into two equal components,

[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. Our approach does not require symmetry of the distribution of the fund alpha under the null hypothesis. Estimated with a bootstrap procedure, the distribution can take any form, as long as we run an equal-tailed test (see Davidson and MacKinnon (2004), p. 187).

$F^{+}(\gamma) = F^{-}(\gamma) = F(\gamma)/2$. These are the number of lucky funds among the best and worst funds, respectively. Using this result, the FDR among the best and worst funds, denoted by $FDR^{+}(\gamma)$ and $FDR^{-}(\gamma)$, are written as:

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

$$FDR^{-}(\gamma) = E\left[\left.\frac{F^{-}(\gamma)}{R^{-}(\gamma)}\right| R^{-}(\gamma) > 0\right] = E\left[\left.\frac{\frac{1}{2}\cdot F(\gamma)}{R^{-}(\gamma)}\right| R^{-}(\gamma) > 0\right]. \tag{3}$$

2.1.2 Estimation Procedure

Storey (2002) and Storey and Tibshirani (2003) propose to estimate the basic FDR shown in Equation (2) by $\widehat{FDR}(\gamma) = \widehat{F}(\gamma)/\widehat{R}(\gamma)$, where $\widehat{F}(\gamma)$ denotes the estimated expected number of lucky funds, and $\widehat{R}(\gamma)$ the number of significant funds at the significance level γ. These authors show that the estimator $\widehat{FDR}(\gamma)$ is conservative in the sense that $E\left[\widehat{FDR}(\gamma)\right] \geq FDR(\gamma)$, for all γ. Using a similar approach, we propose the following estimators of the $FDR^{+}(\gamma)$ and $FDR^{-}(\gamma)$:

$$\widehat{FDR}^{+}(\gamma) = \frac{\widehat{F}^{+}(\gamma)}{\widehat{R}^{+}(\gamma)} = \frac{\frac{1}{2}\cdot\widehat{F}(\gamma)}{\widehat{R}^{+}(\gamma)}, \qquad \widehat{FDR}^{-}(\gamma) = \frac{\widehat{F}^{-}(\gamma)}{\widehat{R}^{-}(\gamma)} = \frac{\frac{1}{2}\cdot\widehat{F}(\gamma)}{\widehat{R}^{-}(\gamma)}, \tag{4}$$

where $\widehat{R}^{+}(\gamma)$ and $\widehat{R}^{-}(\gamma)$ stand for the observed number of significant funds with positive and negative estimated alphas, respectively. Using these FDR estimators, $\widehat{T}^{+}(\gamma)$ is given by $\left(1 - \widehat{FDR}^{+}(\gamma)\right)\cdot\widehat{R}^{+}(\gamma)$, and $\widehat{T}^{-}(\gamma)$ by $\left(1 - \widehat{FDR}^{-}(\gamma)\right)\cdot\widehat{R}^{-}(\gamma)$.

The only input required to compute Equation (4) is $\widehat{F}(\gamma)$. At the significance level γ, we know that the expected number of lucky funds is equal to $M\cdot\pi_0\cdot\gamma$, where π_0 denotes the proportion of funds in the population with zero alphas. Using this relation, we can compute $\widehat{F}(\gamma)$ as $M\cdot\widehat{\pi}_0\cdot\gamma$, where $\widehat{\pi}_0$ is an estimator of π_0. Storey (2002) proposes a simple method to compute $\widehat{\pi}_0$, which only depends on the M fund estimated p-values. This method uses the fact that, under the null hypothesis of no performance, p-values are uniformly distributed over the interval $[0,1]$ [8]. On the contrary, p-values under the

[8] This feature is crucial to correctly estimate $\widehat{\pi}_0$. This method cannot be used in a one-sided multiple test, because the p-values are not necessarily uniformly distributed under the null. In a one-sided test, the null is 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. As a result, $\widehat{\pi}_0$ is biased upward, and overestimates the true proportion π_0 of funds with $\alpha_i \leq 0$.

alternative hypothesis of differential performance are small, because they are associated with large positive or negative estimated alphas. By discarding these small p-values lying below a given threshold λ, we can use the density of the large p-values associated with zero-alpha funds to compute $\widehat{\pi}_0$. The appendix explains in detail the estimation procedure and the data-driven method used to select the threshold λ. It also contains the results of the Monte-Carlo study, which show that $\widehat{\pi}_0$, $\widehat{FDR}^{+}(\gamma)$, and $\widehat{FDR}^{-}(\gamma)$ are close to the true values, regardless of the choice of the true parameters and the significance level γ.

2.2 The Proportions of Funds with Positive and Negative Alphas

2.2.1 Definition

Our previous analysis allows us to estimate π_0 and to deduce the proportion π_A of funds with differential performance. To determine the source of this differential performance, we need to decompose π_A in two components π_A^+ and π_A^-, denoting the proportion of funds in the population with positive and negative alphas, respectively.

From Table 1, we know that, at the significance level γ, the total number of funds with truly positive alphas is equal to $T^+(\gamma) + A^+(\gamma)$, where $A^+(\gamma)$ denotes the number of funds with truly positive alphas incorrectly classified as zero-alpha funds. Similarly, the total number of funds with genuine negative performance amounts to $T^-(\gamma) + A^-(\gamma)$, where $A^-(\gamma)$ denotes the number of funds with truly negative alphas incorrectly classified as zero-alpha funds. As a result, the proportions π_A^+ and π_A^- can be written as:

$$\pi_A^+ = \frac{T^+(\gamma) + A^+(\gamma)}{M}, \qquad \pi_A^- = \frac{T^-(\gamma) + A^-(\gamma)}{M}. \tag{5}$$

2.2.2 Estimation Procedure

Estimating π_A^+ and π_A^- is 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 distribution under the alternative decreases monotonically[9], 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,

[9] 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.

we suggest to take the following estimators of the proportion of funds with positive and negative alphas:

$$\widehat{\pi}_A^+ = \frac{\widehat{T}^+(\gamma^*)}{M}, \qquad \widehat{\pi}_A^- = \frac{\widehat{T}^-(\gamma^*)}{M}. \tag{6}$$

In the appendix, we explain the method used to select γ^*, and analyse the performance of these estimators based on Monte-Carlo simulations. These estimators have a good accuracy: except for one case, the difference between the estimate and the true value is always smaller than 1%.

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):

$$r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + \varepsilon_{i,t}, \tag{7}$$

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 us to control for the momentum strategies followed by many funds, especially Growth and Aggressive Growth funds (Grinblatt, Titman and Wermers (1995)).

We also implement a conditional 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. (2006), and is written as:

$$r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + B^{'}(z_{t-1} \cdot r_{m,t}) + \varepsilon_{i,t}, \tag{8}$$

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 two 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. (2006) find that the finite-sample distribution of the fund estimated alphas is non-normal for approximately half of the funds. Therefore, to test for differential performance, we use a bootstrap procedure (instead of the asymptotic theory) to compute the fund estimated two-sided p-values. We use the t-stat $\widehat{t}_i$ instead of the alpha to compute the p-values because it is an asymptotic pivot[10]:

$$\widehat{t}_i = \frac{\widehat{\alpha}_i}{\widehat{\sigma}_{\alpha_i}}, \tag{9}$$

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 (9), 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. In order to approximate the distribution of $\widehat{t}_i$ under the null, we use a semi parametric bootstrap procedure. We draw with replacement from the regression estimated residuals $\{\widehat{\varepsilon}_{i,t}\}$[11], and impose the null hypothesis $\alpha_i = 0$. For each fund, we set the number of bootstrap iterations to 1,000. Since our procedure is similar to the one implemented by Kosowski et al. (2006), we refer to them for further details.

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. If the fund proposes different

[10] A test statistic is asymptotically pivotal if its asymptotic distribution does not depend on unknown population parameters. The bootstrap theory shows that pivotal test statistics have lower coverage errors than non-pivotal statistics (Davison and Hinkley (1997), Horowitz (2001)).

[11] To know whether this approach is appropriate, we have checked for the presence of autocorrelation (with the Ljung-Box test), heteroscedasticity (with the White test) and Arch effects (with the Engle test) in the fund residuals. We 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.

shareclasses, the fund net return is computed by weighting the net return of each share-class by its total net asset value at the beginning of each month. The CRSP database is matched with the CDA database (from Thomson Financial) in order to obtain the fund investment objectives. We refer to Wermers (2000) for a precise description of these two databases (as well as the matching technique). Although our original sample is free of survivorship bias, we require that each fund has at least 60 monthly return observations to estimate its alpha and t-stat. In unreported results, we find that reducing the minimum length to 36 observations leaves our results unchanged.

Our final fund universe (denoted by *All*) is composed of 1,456 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. These monthly returns need not be contiguous. 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 the beginning of a given month. Panel A and B show the results produced by the unconditional and conditional Carhart models, respectively.

Please insert Table 2 here

Similarly to the previous results documented in the literature, we find that the average unconditional estimated alpha for all categories is negative, ranging between -0.43% and -0.68% per year. AG funds have positive momentum, positive size and negative book-to-market exposures, whereas it is the opposite 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 Results

4.1 FDR Analysis across Investment Categories

We measure the FDR among the best and worst funds at four different significant levels γ (0.05, 0.10, 0.15, and 0.20). The objective of this approach is twofold. The first one is to measure the impact of luck on performance at different portions of the right and left tails of the cross-sectional alpha distribution. The second is to determine the location of funds with truly positive and negative performance in the right and left tails, respectively. The results for the four investment categories (*All*, *G*, *AG*, and *GI*) are displayed in Panels A, B, C, and D of Table 3. Each Panel looks like a cross-sectional alpha distribution. The left part contains the results for the worst funds, which by construction are located at the left tail of the alpha distribution. For each significance level γ, we display the estimated FDR^- $\left(\widehat{FDR}^-\right)$, and the number of worst funds, lucky funds, and funds with truly negative alphas (denoted by $\widehat{R}^-$, $\widehat{F}^-$, and $\widehat{T}^-$, respectively). We also measure the importance of these three sets of funds as a percentage of the entire fund population (denoted by $\widehat{R}^-/M$, $\widehat{F}^-/M$, and $\widehat{T}^-/M$, respectively). On the right-hand side of each Panel, we show the same information for the best funds, which are located at the right tail of the alpha distribution.

4.1.1 All Funds

We begin our analysis with the impact of luck on the performance of the worst funds. At $\gamma = 0.05$, the $\widehat{FDR}^-$ is equal to 22.8%. This low level indicates that the importance of luck is small, since 94 funds out of the 122 truly generate negative alphas. As we go further towards the center of the distribution by increasing γ up to 0.20, the number $\widehat{F}^-$ of lucky funds increases from 28 to 112. At the same time, 106 (200-94) additional funds with truly negative alphas are detected. Since the rise in the lucky funds is partly compensated by the discovery of new funds with negative performance, the $\widehat{FDR}^-$ rises moderately from 22.8% to 35.7%. The FDR among the best funds shows a different pattern. At $\gamma = 0.05$, the $\widehat{FDR}^+$ starts at 50.0%, meaning that half of the 56 best funds do not produce positive alphas, but are simply lucky. As γ rises from 0.05 to 0.20, $\widehat{F}^+$ increases to 84 (112-28), while the number $\widehat{T}^+$ of funds with genuine positive performance remains constant. This causes a jump of the $\widehat{FDR}^+$ from 50.0% to 80.0%.

First of all, our results show that luck has a strong impact on the performance of the best funds, since the $\widehat{FDR}^+$ is always higher than 50%, regardless of γ. It implies

that our performance assessment is completely different from the standard approach. While the latter concludes that 9.6% of the funds are able to achieve positive alphas at $\gamma = 0.20$, we find that only 1.9% of them can do it in reality. For the worst funds, the impact of luck is smaller, as the $\widehat{FDR}^-$ is always smaller than 40%. For this reason, we conclude, like the standard approach, that there is a non-negligible number of funds with negative alphas. However, our quantitative results are still different, since our estimate of funds with truly negative alphas amounts to 14.2%, instead of 21.9% under the standard approach. Second, previous studies propose to approximate the number of lucky funds by the product $M \cdot \gamma$ (see, for instance, Jensen (1968), and Kosowski et al. (2006)). Our results clearly show that this procedure overestimates the impact of luck, because it assumes that π_0 is equal to one. At $\gamma = 0.20$, this procedure would set $\widehat{F}^+$ equal to 146 ($1{,}456 \cdot \frac{0.20}{2}$), which is higher than $\widehat{R}^+$ equal to 140. This estimation is incorrect because, by definition, $R^+ = F^+ + T^+$. On the contrary, our estimate $\widehat{F}^+$ is equal to 112 and is inferior to $\widehat{R}^+$.

In order to determine the location of the funds with negative and positive alphas in the tails of the cross-sectional alpha distribution, we examine the evolution of $\widehat{T}^-/M$ and $\widehat{T}^+/M$. As γ rises, we observe that $\widehat{T}^-/M$ increases continuously from 6.5% to 13.7%. It indicates that we keep on detecting funds with truly negative alphas as we go from the extreme tail to the center of the distribution. Therefore, the funds with negative performance are dispersed in the left tail. On the contrary, $\widehat{T}^+/M$ remains constant at 1.9%, meaning that the few performing funds are located at the extreme right tail of the alpha distribution. By going towards the center, all new significant funds are lucky and yield zero alphas.

Determining fund location has two interesting implications which we illustrate with *All* funds. First, since funds with positive and negative performance do not have the same location, their average alphas (denoted by α_A^+ and α_A^-, respectively) needs not be identical. Because funds with positive performance are located at the extreme right tail, we expect α_A^+ to be higher (in absolute value) than α_A^-. Since these two quantities are unobservable, we assess the performance difference by computing the average alphas of funds at both tails. To account for the difference in location, we estimate α_A^+ with the average estimated alpha of the best funds at $\gamma = 0.05$, and α_A^- with the average alphas of the worst funds at $\gamma = 0.20$. We find that $\widehat{\alpha}_A^+$ amounts to 6.6% per year, while $\widehat{\alpha}_A^-$ is equal to -4.8% per year. This difference of 2% confirms that funds with positive per-

formance generate a higher alpha[12]. Second, fund location has implications for mutual fund portfolio management. Although the funds with truly positive alphas represent a tiny fraction of funds in the population ($\widehat{T}^{+}/M$ is equal to 1.9% at $\gamma = 0.20$), they are located at the extreme right tail. A manager of fund of funds can use this information to form a portfolio with positive performance. By choosing a sufficiently low γ, he is able to partially separate these funds from the lucky ones. For instance, at $\gamma = 0.05$, the funds with truly positive alphas represent 50.0% of the best funds. This percentage, which is much higher than 1.9%, reveals that an equally-weighted portfolio including these 56 best funds produces a positive performance.

Please insert Table 3 here

4.1.2 Growth Funds

The results for G funds are summarized in Panel B of Table 3. The initial level as well as the evolution of $\widehat{FDR}^{+}$ and $\widehat{FDR}^{-}$ are very close to those observed on *All* funds. This is not surprising, since approximately two thirds of the population are G funds. Moreover, the location of the funds with differential performance in the tails of the alpha distribution is also similar. First, $\widehat{T}^{-}/M$ rises from 5.4% to 11.9%, as γ increases from 0.05 to 0.20. It implies that funds with negative performance are largely spread in the left tail. Second, $\widehat{T}^{-}/M$ jumps from 1.3% to 1.6% at $\gamma = 0.10$, but remains constant afterwards. It implies that the funds with genuine positive alphas are concentrated at the far end of the right tail.

4.1.3 Aggressive Growth Funds

Panel C of Table 3 contains the results for AG funds. The $\widehat{FDR}^{-}$ starts at a low level of 20.0%. As a result, the impact of luck on performance is weak, as 17 out of the 21 worst funds truly generate negative alphas. As γ increases to 0.20, the rise of 12.8% (32.8%-20.0%) in the $\widehat{FDR}^{-}$ is small, because 30 additional funds with truly negative performance are detected. The most striking result comes from the low level of the FDR among the best funds. At $\gamma = 0.05$, the $\widehat{FDR}^{+}$ is only equal to 22.1%, meaning that only 4 out the 19 best funds are lucky. As γ rises, the number of lucky funds $\widehat{F}^{+}$ in-

[12] The estimator $\widehat{\alpha}_A^{+} - \left|\widehat{\alpha}_A^{-}\right|$ of the performance difference can potentially be biased, because of the presence of lucky funds among the best and worst funds. However, this bias is minor in your case, because the FDR level is approximately the same in the two groups (35.9% for the worst funds, 49.7% for the best ones). Moreover, our estimate of 2% represents a lower bound, since only 64.1% and 50.3% of the funds produce negative and positive alphas. If we account for it, we obtain a higher difference equal to 5.6% ($6.6\% \cdot \frac{1}{50.3\%} - \left|-4.8\% \cdot \frac{1}{64.1\%}\right|$).

creases progressively from 4 to 17, while the number of funds with positive alphas $\widehat{T}^{+}$ remains almost constant. This contributes to increase $\widehat{FDR}^{+}$ by 25.1% (47.2%-22.1%). However, this level remains largely inferior to the figures documented for *All* and *G* funds.

Our analysis shows that the impact of luck among the best and worst funds is moderate, especially at low significance levels γ. For this reason, our conclusions are in line with those obtained by the standard approach. We find a sizable proportion of funds with truly positive and negative performance. However, the quantitative assessment of these proportions by the standard approach are still largely inflated. While the latter finds that, at $\gamma = 0.20$, 21.8% and 15.4% of the funds yield positive and negative alphas, respectively, our FDR analysis produces estimates equal to 14.5% and 8.1% only.

Concerning the fund location, we find that $\widehat{T}^{-}/M$ doubles, as γ goes from 0.05 to 0.20. Therefore, by going further towards the center of the alpha distribution, we still find many funds with truly negative performance. Similarly to *All* and *G* funds, $\widehat{T}^{+}/M$ remains constant at 8.1% after $\gamma = 0.10$ is reached. This result implies that the funds with positive alphas are all located at the extreme right tail.

4.1.4 Growth and Income Funds

The results for *GI* funds displayed in Panel D of Table 3 present unique characteristics. First of all, the $\widehat{FDR}^{-}$ is extremely low. At $\gamma = 0.05$, it amounts to 17.3%, and reveals that 17 out of the 21 worst funds truly generate negative alphas. Because $\widehat{T}^{-}$ increases continuously from 26 to 61, as we go towards the center of the alpha distribution, the $\widehat{FDR}^{-}$ rises very slowly. Second, the $\widehat{FDR}^{+}$is always equal to 100%, regardless of γ. It implies that all the best funds are purely lucky and do not generate a positive performance. For instance, this is the case for the 5 best funds discovered at $\gamma = 0.05$.

Our results reveal that the impact of luck on the performance of the best funds is enormous 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 7.1% *GI* funds generate positive alphas at $\gamma = 0.20$. This case documents a clear false discovery in mutual fund performance analysis caused by an approach which does not incorporate the presence of luck.

Concerning the fund location, we observe that the funds with truly negative alphas

are largely spread in the left tail, because $\widehat{T}^{+}/M$ rises from 8.4% to 19.7%. Obviously, the location of funds with positive performance cannot be determined, since we do not find any of them.

4.1.5 Comparative Analysis

To compare the impact of luck across the four investment categories, Figure 1 plots the FDR among the worst and best funds at the different significance levels γ. The dashed line represents the $\widehat{FDR}^{-}$, and the solid one the $\widehat{FDR}^{+}$. The $\widehat{FDR}^{-}$ follows the same pattern across the four categories. Its initial value is low and its mild slope indicates that many funds with negative performance are discovered, as we go further towards the center of the alpha distribution. It confirms that these funds are dispersed in the left tail. Although the $\widehat{FDR}^{+}$ differs significantly across the four investment categories, it always starts at higher levels than the $\widehat{FDR}^{-}$. Moreover, it increases more steeply as γ rises, because the few funds with positive performance are located at the extreme right tail. The $\widehat{FDR}^{+}$ of the two smallest investment categories yield extreme patterns. First, the $\widehat{FDR}^{+}$ of the GI is always equal to one, since none of the funds is able to produce positive alphas. Second, the $\widehat{FDR}^{+}$ of the AG funds is low, indicating a non-negligible proportion of funds with positive performance.

Please insert Figure 1 here

4.2 Performance of the Mutual Fund Industry

In the previous literature, the performance of the mutual fund industry is generally measured by the average alpha computed across all funds in the population. Similarly to the results shown in Table 2, these studies find that the average alpha is negative (see, for instance, Jensen (1968), Lehman and Modest (1987), Elton et al. (1993), Pastor and Stambaugh (2002a)). However, an average measure cannot determine the proportions of funds with zero, negative, and positive alphas. Many different proportion combinations are compatible with any given average alpha. By computing the three proportions $\widehat{\pi}_0$, $\widehat{\pi}_A^{-}$, and $\widehat{\pi}_A^{+}$, our approach offers a finer representation of the true performance of the mutual fund industry. The results are shown in Table 4.

Please insert Table 4 here

We start by examining the proportion $\widehat{\pi}_0$ of funds with zero alphas. For *All* funds, this proportion amounts to 76.6%. It indicates that the vast majority of funds (1,115

out of the 1,456) produce a zero-alpha. This proportion is slightly higher for G funds (80.1%), while it is around 70% for the two smaller investment categories AG and GI. Berk and Green (2004) argue that, since managerial abilities are in scarce supply, fund managers are able to capture the economic rent stemming from these abilities. Therefore, the alphas of open-end funds in rational markets must be equal to zero. Based on a calibration of their theoretical model, the prior distribution of alphas indicates that about 80% of the funds generate a sufficient performance to cover their fees. Measuring $\widehat{\pi}_0$ offers a unique way to test empirically the conclusions reached by these authors. We find that the proportion $\widehat{\pi}_0$ for *All* funds is not only high, but also close to 80%. This empirical finding strongly supports the prediction, as well as the calibration proposed by Berk and Green (2004). As fund alphas are net of all expenses, our results indicate that about 76.6% of the fund managers do have abilities. However, they extract the economic rents up to the point where the alpha is equal to zero. The proportion $\widehat{\pi}_0$ can also be used to specify the prior distribution in an empirical bayesian setting. Baks, Metrick and Wachter (2001) examine the portfolio decision made by a skeptical investor who thinks that only 1% of the funds have an alpha equal or superior to zero. A level of 1% is much lower than our estimate of 76.6%, and therefore represents a highly skeptical belief.

Although the model proposed by Berk and Green (2004) describes the performance of the mutual industry fairly well, Table 4 shows that between 20 and 30% of the funds across the four investment categories generate differential performance. The vast majority of these funds distinguish themselves by their poor performance. We find that $\widehat{\pi}_A^-$ for *All* funds amounts to 21.3%, representing a total of 310 funds. The percentage is similar for G and AG funds, but increases to 29.1% for the GI funds. These results shed some light on the negative average alpha documented in the previous literature. Cochrane (1999) finds this negative performance surprising, as professional managers are expected to outperform the market. In fact, most of them do, as about 80% of the funds perform well enough to cover their expenses. This negative performance is only due to a minority of 20% of the funds. Several reasons may explain the survival of these funds with truly negative alphas. One obvious reason is that mutual funds cannot be sold short. Since the only way to penalize them is to remove money out of these funds, the elimination process is longer. Gruber (1996) also explains the presence of poor performing funds by the presence of unsophisticated investors, who partly base their allocation on advertising. Finally, mutual funds can be used to replicate systematic risk factors unavailable for investment (Pastor and Stambaugh (2002b)), as well as dynamic strategies based on public information (Avramov and Wermers (2006)). For this reason, active funds can

still be valuable investments even though some of them yield negative alphas.

Finally, we observe that the proportion $\widehat{\pi}_A^+$ of funds with positive alphas is generally very low. The only exception comes from the *AG* fund category, which contains 8.4% of funds with positive alphas. This finding is consistent with the previous literature (Grinblatt and Titman (1993) and Daniel et al. (1997)). On the contrary, we cannot detect any performing *GI* fund. For *All* and *G* funds, $\widehat{\pi}_A^+$ only represents about 2% of the funds. Our results therefore 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.

4.3 Performance and Fund Characteristics

Our previous analysis reveals that approximately 20% of *All* funds in the population yield negative alphas, while 2% of them generate positive alphas. We now examine whether this differential performance is related to three fund characteristics, namely turnover, expense ratio, and total net asset value (TNA). Using the CRSP database, we determine the annual turnover, annual expense ratio, and TNA of each fund three years after its first return observation. This time lag allows for a wide dispersion of fund characteristics across the different funds. Since the average levels of turnover, expense ratio, and TNA vary over time (Wermers (2000)), we cannot compare fund characteristics computed during different years. To address this issue, we divide the characteristic of each fund i by the fund characteristic average corresponding to the fund i measurement year. We use this ratio as our new measure of the fund characteristic. The fund alpha is computed with data following the measurement year in order to avoid any spurious correlation between fund characteristics and performance[13]. As we still require 60 observations to compute the alpha, our final sample of funds is equal to 1,093.

For each of the three characteristics, the funds are ranked and three groups of 364 funds denoted by Low, Medium, and High are formed[14]. Then, we use our FDR approach to compare the performance of the High versus Low-characteristic groups. The results for turnover, expense ratio, and TNA are displayed in Panels A, B, and C of

[13] If we measure the fund characteristic and performance over the same period, we expect that funds with higher performance also have higher TNA, and may also have higher expense ratios as a response to prior positive performance.

[14] It is common in the literature to form quintiles of funds, instead of thirds (see, for instance, Grinblatt and Titman (1989), Chen, Jegadeesh and Wermers (2000)). Although forming quintiles leads to similar results, we choose thirds to increase the number of funds and improve the precision of the FDR estimators. To be precise, the Low and High groups contain 364 funds, and the Medium one 365.

Table 5. For each Panel, the left part contains the results for the worst funds. For each significance level γ $(0.05, 0.10, 0.15, 0.20)$, we display the estimated FDR^- $\left(\widehat{FDR}^-\right)$, the number of worst funds, lucky funds, and funds with truly negative alphas (denoted by $\widehat{R}^-$, $\widehat{F}^-$, and $\widehat{T}^-$, respectively). We also determine the proportion of funds in the population with zero, negative, and positive performance (denoted by $\widehat{\pi}_0$, $\widehat{\pi}_A^-$, and $\widehat{\pi}_A^+$, respectively). On the right-hand side of each Panel, we show the same information for the best funds.

4.3.1 Turnover

Despite being all actively managed, the turnover levels of the High and Low groups are very different. Taking the 2002 average annual turnover equal to 95.5% as the reference level, the annual turnover is equal to 191.4% for the High group and 24.6% for the Low one. The population in these two groups is mainly composed of funds with zero alphas, since $\widehat{\pi}_0$ is equal to 78.1% and 83.3% for the High and Low groups, respectively. The main difference comes from the proportion $\widehat{\pi}_A^-$ of funds with negative alphas. It amounts to 19.3% for the High group and 14.9% for the Low one. This finding is consistent with the idea that unskilled funds simply trade on noise to convince investors that they can successfully pick stocks. The $\widehat{FDR}^-$ across the High and Low groups shows a similar pattern, since it starts at a level inferior to 30% and increases only slowly. As γ rises, the difference between the two $\widehat{FDR}^-$ falls from -3.5% to -8.7%, and indicates that the 13 additional unskilled funds in the High group are dispersed in the left tail of the alpha distribution.

In rational markets, we would expect a positive relation between management skills and the trading frequency. However, we find that the proportion $\widehat{\pi}_A^+$ of funds with positive performance in the High group is low. $\widehat{\pi}_A^+$ is only equal to 2.6%, implying that about 10 funds out of the 364 funds generate positive alphas. But even though the $\widehat{\pi}_A^+$ difference between the High and Low groups is inferior to 1%, it produces a very different FDR level, because these additional funds are located at the extreme right tail of the alpha distribution. At $\gamma = 0.05$, the $\widehat{FDR}^+$ difference between the High and Low groups is equal to -25.4%. While a portfolio formed with the 16 best High turnover funds contains 56.3% performing funds, a portfolio of the 11 best Low turnover funds has only 30.9% of them.

Chen, Jegadeesh and Wermers (2000) and Wermers (2000) find no statistically significant difference between the average performance of High and Low turnover funds. This

is not surprising, since the average alpha is mainly influenced by the large proportion π_0 of funds with zero alphas in the two groups. Our results reveal that one way to capture the small differences between High and Low turnover funds is to estimate the proportions π_A^+ and π_A^- of funds with positive and negative alphas. Using a cross-sectional regression of the fund alpha on its turnover, Carhart (1997) finds that funds with higher turnover generate lower performance. Our results show that this relation is not due to the majority of the funds in the two groups, since most of them yield zero alphas. This relation may be due to the additional 4.4% of funds with negative alphas observed in the High turnover group.

Please insert Table 5 here

4.3.2 Expense Ratio

Fund expenses include management, administration, as well as marketing fees. Taking the 2002 average annual expense ratio equal to 1.37% as the reference unit, the expense ratio for the Low group is equal to 0.83% per year, while the one of the High group amounts to 2.16% per year. The results summarized in Panel B of Table 5 indicate that the composition of the two groups is very different. We find that 85.3% of the High-group funds yield zero alphas, while the remaining 14.7% produce negative performance. This finding confirms Carhart's (1997) results, that the persistence of the unskilled funds is partly due to excessive expense ratios. Since no fund is able to achieve a positive performance (i.e. the $\widehat{FDR}^+$ is always equal to 100%), investors looking for positive alphas should discard High expense funds.

Compared to the High group, the proportion $\widehat{\pi}_A^-$ of funds with negative alpha is 11.6% higher in the Low group. Although these funds charge low fees to their clients, they are not sufficiently skilled to generate positive alphas. Contrary to what might be believed, Low Expense funds contain more unskilled funds than its High counterpart. Interestingly, the $\widehat{FDR}^-$ difference between the High and Low groups has a reversed pattern. At $\gamma = 0.05$, we detect 10 more poorly performing funds in the High versus Low group. As we increase γ, the number $\widehat{T}^-$ of funds with truly negative alphas increases at a faster pace for the Low group (from 17 to 61 funds). These additional funds detected in the Low group are therefore all spread in the left tail. The Low Expense group also contains 4.1% of funds with positive alphas. This proportion is particularly high in light of the 2% previously found for the *All* funds. The low initial level of $\widehat{FDR}^+$ indicates that these funds are located at the extreme right tail of the alpha distribution. A portfolio including the 18 best funds at $\gamma = 0.05$ would contain only 35.3% of lucky funds. As

we go further into the right tail, the $\widehat{FDR}^{+}$ rises quickly up to 65.1%, because all but 2 new significant funds are lucky.

Elton et al. (1993) and Gruber (1996) find that High expense funds yield a lower average alpha than Low expense funds[15]. Our results seem to contradict their findings, since the proportion difference between the Low and High groups is much higher for the unskilled funds (11.6%) than for the skilled ones (4.1%). However, the additional performing funds are located at the extreme right tail. As they are likely to generate higher alphas than the additional 11.6% unskilled funds, the average alpha in the Low group can perfectly be higher.

4.3.3 Total Net Asset (TNA)

Similarly to the previous characteristics, there are substantial TNA differences across funds. Using the 2002 average TNA equal to 1,293 millions as the reference unit, the TNA of the Low group amounts to 50 millions, compared to 1,667 millions for the High TNA group. The results displayed in Panel C of Table 5 reveals that the two groups present many similarities. First, the proportion $\widehat{\pi}_0$ of funds with zero alphas is high, since it amounts to 75.1% and 78.8% for the High and Low groups, respectively. Second, both groups contain the same proportion $\widehat{\pi}_A^-$ of funds with negative alphas, approximately equal to 20%. Looking at the slow increase in the $\widehat{FDR}^{-}$ in the two groups, the funds with negative alphas are dispersed in the left tail of the alpha distribution.

The main difference between the High and Low groups comes from the proportion $\widehat{\pi}_A^+$ of funds with positive alphas. High TNA funds contains 4.2% of performing funds. At $\gamma = 0.05$, most of these funds are detected, as the 12 funds with positive alphas represent 3.3%. The sharp increase in the $\widehat{FDR}^{+}$ from 35.9% to 65.1% indicates once again that the performing funds are located at the extreme right tail. It seems that a few High TNA funds can benefit from economies of scales. On the contrary, no funds in the Low group can yield positive performance. Similarly to High expense funds, an investor looking for positive alphas should avoid Low TNA funds.

Grinblatt and Titman (1989) document no difference between the average alphas com-

[15] Over the 1975-84 period, Elton et al. (1993) document an average annual alpha of -4.37% for the highest expense decile and -1.68% for the lowest one. Similarly, Gruber (1996) shows that the highest and lowest expense deciles yield respectively annual alphas of -1.84% and 0.02%, respectively, during the following one-year holding period.

puted across the various TNA quintiles. Similarly to the turnover, this finding is probably due to the large proportion π_0 of funds with zero alphas in the various quintiles, which drives the average performance towards zero. Cross-sectional regressions of fund alphas on TNA do not produce an unambiguous relation. While Chen et al. (2004) find a negative coefficient, Ferson and Qian (2004) document a positive one. Moreover, these coefficients are not significant[16]. A strong relation between TNA and performance is unlikely, because the High and Low TNA groups are indeed very similar. Based on our results, the presence of 4.2% of funds with positive alphas in the High TNA group may produce a positive relation between performance and TNA. However, this small percentage makes the evidence tenuous.

4.4 Sensitivity Analysis

4.4.1 Alternative Asset Pricing Models

Table 6 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. When the unconditional and conditional FF models are used, the patterns of the $\widehat{FDR}^{+}$ and the $\widehat{FDR}^{-}$ 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 the $\widehat{FDR}^{+}$ and the $\widehat{FDR}^{-}$ 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

[16] Both papers document coefficients for different performance measures and/or different investment categories. None of the Ferson and Qian (2004) coefficient are significant (at the 10% level). Chen et al. (2004) obtain significant negative coefficients based on the CAPM and the Fama-French models. However, the relation obtained with Carhart alphas is not significant.

factors, which both have positive premia.

Please insert Table 6 here

4.4.2 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 268 *All* funds and only 111 *G*, 54 *AG*, and 63 *GI* funds. Because of the small size of these three categories, we only compute the FDR for *All* funds. The $\widehat{FDR}^{+}$ is lower than the one observed during the entire period. It respectively amounts to 23.1% and 37.2% 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 1,404 *All* funds and 976 *G*, 196 *AG* and 277 *GI* funds. During this period, the levels of $\widehat{FDR}$ across the four investment categories are close to those documented for the entire period. These results are available upon request.

5 Conclusion

Detecting funds with differential performance requires to examine the performance of all funds in the population. The main issue raised by this multiple-hypothesis test is how to control for the presence of lucky funds, namely funds which have significant estimated alphas, while their true alphas is equal to zero. This paper uses the False Discovery Rate (FDR) to account for luck in individual fund performance measurement. To address the financial problem at hand, we further develop new FDR measures, the FDR among the best funds and the FDR among the worst funds. These measures allow us to separately measure the impact of luck on the performance of the funds located at the right and left tails of the cross-sectional alpha distribution. These new measures are very easy to compute from the fund estimated p-values. 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. Using our approach, we are able to estimate the number of funds with genuine positive and negative alphas at any portion of the tails of the alpha distribution. Moreover, we obtain a better representation of the performance of

the mutual fund industry by measuring the proportions of funds with zero, positive, and negative performance.

Our results based on 1,456 U.S. open-end equity funds between 1975 and 2002 show that the impact of luck on the performance of the best funds is substantial. We find that, at any significance level γ, more than half of the best *All*, *G*, and *GI* funds are lucky and yield in reality zero alphas. For this reason, our conclusions regarding the number of performing funds are completely different from the standard approach. A striking illustration concerns *GI* funds: while the standard approach finds that 7.1% are skilled at $\gamma = 0.2$, none of them yields a positive alphas after accounting for luck. Looking at the worst funds, we show that the impact of luck is less pronounced, as the FDR is always lower than 50% across the four investment categories (*All*, *G*, *AG*, and *GI*). Fund location reveals common patterns across the four categories. While the funds with negative alphas are spread in the left tail of the cross-sectional alpha distribution, the funds with positive alphas are located at the extreme right tail. It first implies that performing funds generate higher alphas in absolute value than the unskilled ones. Second, although there exists a few funds with positive performance, we find that an investor can still build a portfolio of funds with a positive alpha. Because of their location, these performing funds can be separated from the lucky funds simply by taking a sufficiently low significance level γ.

Our analyis of the performance of the mutual fund industry shows that approximately 76.6% of *All* funds have zero alphas. This confirms the predictions of the Berk and Green (2004) model, asserting that, in equilibrium, open-end mutual funds yield no performance. Among the remaining funds, 21.3% of them yield negative alphas. It implies that the average negative alpha documented in the previous literature does not reflect the performance of the majority of funds, but is rather driven by a minority of 20%. Finally, we find a tiny proportion of 2.1% of funds with positive alphas. We also observe that *AG* funds produce the highest proportion of performing funds (8.4%), while no *GI* funds yield a positive performance.

Using our FDR analysis, we investigate the relation between performance and three lagged fund characteristics, the turnover, the expense ratio, and the total net asset value (TNA). The main difference between High and Low turnover funds is the proportion of funds with negative alphas (19.3% versus 14.9%). Low expense funds contain more skilled funds that High expense funds (4.1% versus 0%), as well as more funds with

negative alphas (26.3% versus 14.7%). Both High and Low TNA funds have the same proportion of unskilled funds (21.2% versus 20.7%). But contrary to Low TNA funds, High TNA funds contain a positive proportion of performing funds (4.1%). All these High and Low groups of funds are mostly composed of zero-alpha funds (the proportion ranges from 69.6% up to 83.3%). In light of this high proportion, an average alpha is unlikely to capture the small differences between these groups. This reinforces the need to compute the proportions of funds with positive and negative alphas in order to understand the relation between performance and fund characteristics.

Our paper gives attention to mutual fund performance, but the FDR approach, as well as its extensions (the FDR^+ and FDR^-), have wide applications in finance. These measures can be used in any setting in which a given hypothesis is tested many times. We list here some illustrative examples. First, technical trading can be implemented with a myriad of trading rules (see Sullivan, Timmermann and White (1999)). The FDR could be used similarly to our setting to determine the impact of luck on the performance of the best and worst trading rules. Second, testing the presence of commonality in liquidity boils down to regressing an individual stock liquidity measure on the market liquidity measure (see Chordia, Roll and Subrahmanyam (2000)). Since this regression is run for a large number of stocks, we are dealing with multiple testing and the FDR msut be applied to control for luck.

6 Appendix

6.1 Estimation Procedure

In this section, we describe in detail the estimation procedure of the FDR estimators, $\widehat{FDR}^{+}(\gamma)$ and $\widehat{FDR}^{-}(\gamma)$. In particular, we explain the data-driven approach used to compute the proportion $\widehat{\pi}_0$ of funds with zero alphas. Then, we show how to compute the estimated proportions $\widehat{\pi}_A^{+}$ and $\widehat{\pi}_A^{-}$ of funds with positive and negative alphas.

6.1.1 The FDR Estimators

To estimate the FDR among the best and worst funds, we need to compute the estimated expected number of lucky funds $\widehat{F}(\gamma)$ written as $M \cdot \widehat{\pi}_0 \cdot \gamma$. The key point consists in correctly estimating the proportion π_0 of funds with zero alphas in the population. Following Storey (2002) and Storey and Tibshirani (2003), we note that, under the null of no performance, p-values are uniformly distributed over the interval $[0,1]$. On the contrary, under the alternative hypothesis of differential performance, funds with truly positive or negative alphas typically have small p-values. We can exploit this information to compute $\widehat{\pi}_0$ without specifying the exact distribution of the p-values under the alternative. To illustrate it, Figure 2 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 Monte-Carlo subsection). Consistently with the size of our database, we set $M = 1,456$ as our number of mutual funds in the simulation. In this simulation, we set the alpha of 80% of the mutual funds to zero. The remaining funds are divided into equal numbers with annual alphas of +5% and -5%.

Please insert Figure 2 here

Clearly, 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 the null hypothesis of no performance. Therefore, by taking a sufficiently high threshold λ, we can exploit the density beyond λ to estimate the proportion π_0 of non-performing funds:

$$\widehat{\pi}_0(\lambda) = \frac{\widehat{W}(\lambda)}{(1-\lambda)\cdot M}, \tag{10}$$

where $\widehat{W}(\lambda)$ denotes the number of estimated p-values larger than λ. The simplest way to determine λ^* is to eye-ball the flat portion of the histogram of p-values shown in

Figure 2. In this paper, we use a more rigorous bootstrap procedure proposed by Storey (2002) and Storey, Taylor and Siegmund (2004). We stress that, when the number of funds is high (which is the case in our study), $\widehat{\pi}_0(\lambda)$ is not sensitive to the choice of λ. Intermediate values of λ between 0.3 and 0.7 all produce similar values for $\widehat{\pi}_0$.

The data-driven approach based on bootstrapping chooses λ such that the mean-squared error (MSE) of $\widehat{\pi}_0(\lambda)$ is minimized. First, we compute $\widehat{\pi}_0(\lambda)$ across a range of λ ($\lambda = 0.05, 0.10, ..., 0.70$). Second, for each possible value of λ, we form 1,000 bootstrap versions of $\widehat{\pi}_0(\lambda)$ by drawing with replacement from the $M \times 1$ vector of fund estimated p-values. These are respectively denoted by $\widehat{\pi}_0^b(\lambda)$ with $b = 1, ..., 1,000$. Third, we compute the MSE for each possible value of λ:

$$\widehat{MSE}(\lambda) = \frac{1}{1,000} \sum_{b=1}^{1,000} \left[\widehat{\pi}_0^b(\lambda) - \min_{\lambda} \widehat{\pi}_0(\lambda) \right]^2. \tag{11}$$

We choose λ^* such that $\lambda^* = \arg\min_\lambda \widehat{MSE}(\lambda)$. The estimate of π_0 is then equal to $\widehat{\pi}_0(\lambda^*)$. Using $\widehat{\pi}_0(\lambda^*)$, the FDR estimators can be written as:

$$\widehat{FDR}_\lambda^+(\gamma) = \frac{\frac{1}{2} \cdot M \cdot \widehat{\pi}_0(\lambda^*) \cdot \gamma}{\widehat{R}^+(\gamma)}, \quad \widehat{FDR}_\lambda^-(\gamma) = \frac{\frac{1}{2} \cdot M \cdot \widehat{\pi}_0(\lambda^*) \cdot \gamma}{\widehat{R}^-(\gamma)} \tag{12}$$

An important property of the FDR estimators is strong control in the sense that $E\left(\widehat{FDR}_\lambda(\gamma)\right) \geq FDR(\gamma)$ for all γ and all π_0. 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 and Tibshirani (2001), Storey, Taylor and Siegmund (2004)).

6.1.2 The Proportion Estimators

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 the following equation:

$$\widehat{\pi}_A^+ = \frac{\widehat{T}^+(\gamma^*)}{M}, \qquad \widehat{\pi}_A^- = \frac{\widehat{T}^-(\gamma^*)}{M}. \tag{13}$$

The simplest way to choose a sufficiently high significance level γ^* is to find the minimum significance level such that either $\widehat{T}^+(\gamma)$ or $\widehat{T}^-(\gamma)$ becomes constant. Let us suppose that $\widehat{T}^+(\gamma)$ is constant after γ^* is reached. In this case, $\widehat{\pi}_A^+$ is given by $\widehat{T}^+(\gamma^*)/M$, while

$\widehat{\pi}_A^-$ is set to $\widehat{\pi}_A - \widehat{\pi}_A^+$ to preserve the equality $\pi_A = \pi_A^+ + \pi_A^-$. This approach is similar to the visual procedure used to pick up the parameter λ. In this paper, we use a bootstrap technique which minimizes the MSE of $\widehat{\pi}_A^+(\gamma)$ and $\widehat{\pi}_A^-(\gamma)$. 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 $\widehat{\pi}_A^+(\gamma)$ for each possible value of γ. These are respectively denoted by $\widehat{\pi}_A^{b+}(\gamma)$ with $b = 1, ..., 1,000$. Third, we compute the MSE for each possible value of γ:

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

We choose γ^+ such that $\gamma^+ = \arg\min_\gamma \widehat{MSE}^+(\gamma)$. Our estimate of π_A^+ is then equal to $\widehat{\pi}_A^+(\gamma^+)$. We use the same procedure for $\widehat{\pi}_A^-(\gamma)$ to determine $\gamma^- = \arg\min_\gamma \widehat{MSE}^-(\gamma)$ and $\pi_A^- = \widehat{\pi}_A^-(\gamma^-)$. If $\arg\min_\gamma \widehat{MSE}^+(\gamma) < \arg\min_\gamma \widehat{MSE}^-(\gamma)$, we set $\widehat{\pi}_A^+(\gamma^*) = \widehat{\pi}_A^+(\gamma^+)$ and $\widehat{\pi}_A^-(\gamma^*) = \widehat{\pi}_A - \widehat{\pi}_A^+(\gamma^*)$ to preserve the equality $\pi_A = \pi_A^+ + \pi_A^-$. Otherwise, we set $\widehat{\pi}_A^-(\gamma^*) = \widehat{\pi}_A^-(\gamma^-)$ and $\widehat{\pi}_A^+(\gamma^*) = \widehat{\pi}_A - \widehat{\pi}_A^-(\gamma^*)$.

6.2 Monte-Carlo Simulations

In this section, we first check the finite sample performance of the estimators of our 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:

$$\begin{aligned} r_{i,t} &= \alpha_i + \beta \cdot r_{m,t} + \varepsilon_{i,t}, \qquad i = 1, ..., M, \quad t = 1, ..., T, \\ r_{m,t} &\sim N(0, \sigma_{r_m}), \qquad \varepsilon_{i,t} \sim N(0, \sigma_\varepsilon). \end{aligned} \tag{15}$$

For each fund i ($i = 1, ..., M$), we test the null hypothesis of no performance ($\alpha_i = 0$) against the alternative hypothesis of differential performance ($\alpha_i > 0$ or $\alpha_i < 0$). Under the null, the t-stat $\widehat{t}_i$ follows the Student distribution with $T - 2$ degrees of freedom. Under the alternative, $\widehat{t}_i$ follows a noncentral student distribution with $T - 2$ degrees of freedom whose true parameter of noncentrality can be well approximated by $T^{\frac{1}{2}} \alpha_A / \sigma_\varepsilon$ (see Davidson and MacKinnon (2004), 169). Consistently with the size of our database, we set $M = 1,456$ 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 across the funds and σ_{r_m} is the standard deviation of the market return. We 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 the null and yield zero alphas. A proportion π_A of funds generate differential performance. Among these funds, a proportion $\pi_A^- = \pi_A \cdot q^-$ of funds yield a negative alpha α_A^+, and a proportion $\pi_A^+ = \pi_A \cdot (1 - q^-)$ of funds yield a positive alpha α_A^-. $q^- \in [0, 1]$ is a positive scalar. We thus have:

$$\begin{aligned} H_{0,i} &: \alpha_i \sim N(0, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with proportion } \pi_0, \\ H_{A,i} &: \alpha_i \sim N(\alpha_A^+, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with proportion } \pi_A^+, \\ &: \alpha_i \sim N(\alpha_A^-, T^{-\frac{1}{2}}\sigma_\varepsilon) \quad \text{with proportion } \pi_A^-. \end{aligned} \tag{16}$$

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 5.5% per year, respectively. 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, (b) 0.9. Finally, q^- is set to (a) 0.3, (b) 0.7. Two significance levels γ are examined: (a) 0.05, (b) 0.10. The number of Monte-Carlo replications is equal to 1,000.

The FDR among the Best and Worst Funds

The estimator $\widehat{\pi}_0(\lambda)$ is computed with the bootstrap procedure explained previously. We also test a more simple approach where λ is set to 0.5. These two estimators are compared with the true value π_0 defined in the Monte-Carlo design. The estimator $\widehat{FDR}_\lambda^+(\gamma)$ is defined in Equation (12). It is compared with the true $FDR^+(\gamma)$ computed as follows:

$$FDR^+(\gamma) = \frac{\frac{1}{2} \cdot \pi_0 \cdot \gamma}{\frac{1}{2} \cdot \pi_0 \cdot \gamma + \pi_A^+ \cdot prob\left(t > t_{T-2,1-\frac{\gamma}{2}} \,|H_A, \alpha_A > 0\right)}, \tag{17}$$

where $t_{T-2,1-\frac{\gamma}{2}}$ denotes the quantile of probability level $1 - \gamma/2$ of the non-central Student distribution with $T-2$ degrees of freedom. The estimator $\widehat{FDR}_\lambda^-(\gamma)$ is defined

by Equation (12). It is compared with the true $FDR^{-}(\gamma)$ computed as follows:

$$FDR^{-}(\gamma) = \frac{\frac{1}{2} \cdot \pi_0 \cdot \gamma}{\frac{1}{2} \cdot \pi_0 \cdot \gamma + \pi_A^{-} \cdot prob\left(t < t_{T-2,\frac{\gamma}{2}} \,|H_A, \alpha_A < 0\right)}, \tag{18}$$

where $t_{T-2,\frac{\gamma}{2}}$ denotes the quantile of probability level $\gamma/2$ of the non-central Student distribution with $T-2$ degrees of freedom. Table 7 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 (strong control).

Please insert Table 7 here

6.2.2 The Proportion Estimators

The estimators $\widehat{\pi}_A^{+}(\gamma)$ and $\widehat{\pi}_A^{+}(\gamma)$ are computed with the bootstrap procedure explained in the previous subsection. Alternatively, we also test a more simple approach where γ^{+} is set to 0.2. We set $\widehat{\pi}_A^{+} = \widehat{\pi}_A^{+}(0.2)$ and $\widehat{\pi}_A^{-} = \widehat{\pi}_A^{-}(0.2)$. These two estimators are compared with the true values π_A^{+} and π_A^{-} defined in the Monte-Carlo design. Table 8 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 bootstrap technique. The simulation results show that the estimators based on the bootstrap procedure have a good accuracy, up to the second decimal. 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 cross-sectional 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 8 here

References

[1] Avramov, D., and R. Wermers, 2006, Investing in Mutual Funds when Returns are Predictable, Journal of Financial Economics, Forthcoming.

[2] Baks, K.P., A. Metrick, and J. Wachter, 2001, Should Investors avoid all Actively Managed Mutual Funds? A study in Bayesian Performance Evaluation, Journal of Finance 56, 45-85.

[3] 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.

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

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

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

[7] Chen, J., H. Hong, H. Huang, and M. Kubrik, 2004, Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization, American Economic Review 94, 1276-1302.

[8] Chordia, T., R. Roll, and A. Subrahmanyam, 2000, Commonality in Liquidity, Journal of Financial Economics 56, 3-28.

[9] Cochrane, J., 1999, New Facts in Finance, Economic Perspectives Federal Reserve Bank of Chicago 23, 36-58.

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

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

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

[13] 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.

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

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

[16] Ferson, W., and M. Qian, 2004, Conditional Performance Evaluation Revisited, in Research Foundation Monograph of the CFA Institute.

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

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

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

[20] 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.

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

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

[23] 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.

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

[25] Kosowski, R., A. Timmermann, H. White, and R. Wermers, 2006, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, 2006, Journal of Finance, Forthcoming.

[26] 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.

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

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

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

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

[31] Storey, J.D. and R. Tibshirani, 2001, Estimating False Discovery Rates under Dependence, with Applications to DNA Microarrays, Technical Report 28-2001, Department of Statistics, Stanford University.

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

[33] Storey, J.D., and R. Tibshirani, 2003, Statistical Significance for Genomewide Studies, Proceedings of the National Academy of Science 100, 9440-9445.

[34] 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.

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

[36] 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-Hypothesis Test of No Performance at the Significance Level γ

	# Accept $H_{0,i}$	# Reject $H_{0,i}$	# 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

For each fund, the null hypothesis $H_{0,i}$ of no performance ($\alpha_i = 0$) is tested against the alternative $H_{A,i}$ 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. These are the 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, a total of $R(\gamma)$ funds are called significant (i.e. $H_{0,i}$ is rejected R times).

Table 2

Average Mutual Fund Performance

Panel A Unconditional Carhart Model

	α	β_m	β_{smb}	β_{hml}	β_{mom}	R^2
All funds	-0.45%	0.95	0.15	-0.02	0.02	97.9%
	(0.18)	(0.00)	(0.00)	(0.21)	(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	-0.68%	0.88	-0.06	0.17	-0.02	97.6%
	(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	-0.54%	0.96	0.15	-0.03	0.02	98.0%
	(0.15)	(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%
	(0.20)	(0.00)	(0.00)	(0.00)	(0.00)	
GI funds	-0.71%	0.88	-0.06	0.15	-0.02	97.7%
	(0.05)	(0.00)	(0.00)	(0.00)	(0.11)	

The results for the unconditional and conditional Carhart models are shown in Panels A and B for All funds (*All*), Growth funds (*G*), Aggressive Growth funds (*AG*), and Growth and Income funds (*GI*). Each Panel contains the annual alpha, the factor exposures, and the adjusted R-square of an equally-weighted portfolio including all funds existing at the beginning of a given month. The regressions are based on monthly data between January 1975 and December 2002 (336 observations). Figures in parentheses denote the heteroskedasticity-consistent p-values under the null hypothesis that the regression parameters are equal to zero.

Table 3

Impact of Luck on the Performance of the Best and Worst Funds

Panel A *All* funds

	Worst funds				Best funds				
γ	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	γ
$\widehat{FDR}^-$	22.8%	29.5%	32.2%	35.7%	80.0%	75.0%	66.6%	50.0%	$\widehat{FDR}^+$
$\widehat{R}^-$	122	188	260	312	140	112	84	56	$\widehat{R}^+$
$\widehat{F}^-$	28	56	84	112	112	84	56	28	$\widehat{F}^+$
$\widehat{T}^-$	94	132	176	200	28	28	28	28	$\widehat{T}^+$
$\widehat{R}^-/M$	8.4%	12.9%	17.9%	21.4%	9.6%	7.7%	5.7%	3.8%	$\widehat{R}^+/M$
$\widehat{F}^-/M$	1.9%	3.8%	5.8%	7.7%	7.7%	5.8%	3.8%	1.9%	$\widehat{F}^+/M$
$\widehat{T}^-/M$	6.5%	9.1%	12.1%	13.7%	1.9%	1.9%	1.9%	1.9%	$\widehat{T}^+/M$

Panel B G funds

	Worst funds				Best funds				
γ	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	γ
$\widehat{FDR}^-$	27.0%	32.3%	36.9%	40.3%	83.7%	79.4%	72.1%	60.4%	$\widehat{FDR}^+$
$\widehat{R}^-$	76	127	167	204	98	78	57	34	$\widehat{R}^+$
$\widehat{F}^-$	21	41	62	82	82	62	41	21	$\widehat{F}^+$
$\widehat{T}^-$	55	86	105	122	16	16	16	13	$\widehat{T}^+$
$\widehat{R}^-/M$	7.4%	12.4%	16.3%	19.9%	9.6%	7.6%	5.6%	3.3%	$\widehat{R}^+/M$
$\widehat{F}^-/M$	2.0%	4.0%	6.0%	8.0%	8.0%	6.0%	4.0%	2.0%	$\widehat{F}^+/M$
$\widehat{T}^-/M$	5.4%	8.4%	10.3%	11.9%	1.6%	1.6%	1.6%	1.3%	$\widehat{T}^+/M$

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. The left part shows the FDR analysis for the worst funds (i.e. funds with significant negative estimated alphas) at different significance levels γ. We display the estimated False Discovery Rate $\widehat{FDR}^-$, the number of worst funds, lucky funds, and funds with truly negative alphas (denoted by $\widehat{R}^-$, $\widehat{F}^-$, and $\widehat{T}^-$, respectively), as well as the proportion of worst funds, lucky funds, and funds with truly negative alphas in the population (denoted by $\widehat{R}^-/M$, $\widehat{F}^-/M$, and $\widehat{T}^-/M$, respectively). The right part contains the same information for the best funds (i.e. funds with significant positive estimated alphas). The alphas of all funds are computed with the unconditional Carhart model.

Table 3

Impact of Luck on the Performance of the Best and Worst Funds

Panel C *AG* funds

	Worst funds				Best funds				
γ	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	γ
$\widehat{FDR}^{-}$	20.0%	28.6%	31.8%	32.8%	47.2%	40.6%	31.0%	22.1%	$\widehat{FDR}^{+}$
$\widehat{R}^{-}$	21	28	41	51	36	32	27	19	$\widehat{R}^{+}$
$\widehat{F}^{-}$	4	8	13	17	17	13	8	4	$\widehat{F}^{+}$
$\widehat{T}^{-}$	17	20	28	34	19	19	19	15	$\widehat{T}^{+}$
$\widehat{R}^{-}/M$	9.0%	12.0%	17.5%	21.8%	15.4%	13.7%	11.5%	8.1%	$\widehat{R}^{+}/M$
$\widehat{F}^{-}/M$	1.7%	3.4%	5.6%	7.3%	7.3%	5.6%	3.4%	1.7%	$\widehat{F}^{+}/M$
$\widehat{T}^{-}/M$	7.3%	8.6%	11.9%	14.5%	8.1%	8.1%	8.1%	6.4%	$\widehat{T}^{+}/M$

Panel D *GI* funds

	Worst funds				Best funds				
γ	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	γ
$\widehat{FDR}^{-}$	17.3%	18.9%	22.3%	26.5%	100%	100%	100%	100%	$\widehat{FDR}^{+}$
$\widehat{R}^{-}$	31	58	74	83	22	16	11	5	$\widehat{R}^{+}$
$\widehat{F}^{-}$	5	11	16	22	22	16	11	5	$\widehat{F}^{+}$
$\widehat{T}^{-}$	26	47	58	61	0	0	0	0	$\widehat{T}^{+}$
$\widehat{R}^{-}/M$	10.0%	18.7%	23.9%	26.8%	7.1%	5.2%	3.5%	1.6%	$\widehat{R}^{+}/M$
$\widehat{F}^{-}/M$	1.6%	3.5%	5.2%	7.1%	7.1%	5.2%	3.5%	1.6%	$\widehat{F}^{+}/M$
$\widehat{T}^{-}/M$	8.4%	15.2%	18.7%	19.7%	0.0%	0.0%	0.0%	0.0%	$\widehat{T}^{+}/M$

Table 4

Source of Differential Performance

		No Performance	Differential Performance	Negative Performance	Positive Performance
All funds	Proportion	76.6%	23.4%	21.3%	2.1%
	Number	1,115	341	310	31
G funds	Proportion	80.1%	19.9%	18.3%	1.6%
	Number	822	203	187	16
AG funds	Proportion	71.6%	28.4%	20.0%	8.4%
	Number	167	67	47	20
GI funds	Proportion	70.9%	29.1%	29.1%	0.0%
	Number	220	90	90	0

For All funds (*All*), Growth funds (*G*), Aggressive Growth funds (*AG*), and Growth and Income funds (*GI*), we compute the proportion as well as the number of funds with zero alphas in the population. The proportion and number of funds with differential performance is then decomposed into two groups. The first group contains funds with truly negative performance (i.e. negative alphas), while the second one contains funds with truly positive performance (i.e. positive alphas). The alphas of all funds are computed with the unconditional Carhart model.

Table 5

Performance and Fund Characteristics

Panel A Turnover

	Worst funds					Best funds				
γ	0.05	0.10	0.15	0.20		0.20	0.15	0.10	0.05	γ
					High Turnover					
$\widehat{FDR}^-$	23.6%	25.9%	29.6%	34.1%		73.6%	67.7%	58.3%	43.7%	$\widehat{FDR}^+$
$\widehat{R}^-$	31	54	71	82		38	31	24	16	$\widehat{R}^+$
$\widehat{F}^-$	7	14	21	28		28	21	14	7	$\widehat{F}^+$
$\widehat{T}^-$	24	40	50	54		10	10	10	9	$\widehat{T}^+$
				$\widehat{\pi}_A^-$	$\widehat{\pi}_0$	$\widehat{\pi}_A^+$				
				19.3%	78.1%	2.6%				
					Low Turnover					
$\widehat{FDR}^-$	27.1%	36.2%	39.3%	42.8%		84.4%	78.6%	72.4%	69.1%	$\widehat{FDR}^+$
$\widehat{R}^-$	28	42	57	71		36	29	21	11	$\widehat{R}^+$
$\widehat{F}^-$	8	15	23	30		30	23	15	8	$\widehat{F}^+$
$\widehat{T}^-$	20	27	34	41		6	6	6	3	$\widehat{T}^+$
				$\widehat{\pi}_A^-$	$\widehat{\pi}_0$	$\widehat{\pi}_A^+$				
				14.9%	83.3%	1.8%				
					High minus Low					
$\Delta\widehat{FDR}^-$	-3.5%	-10.3%	-9.7%	-8.7%		-10.8%	-10.9%	-14.1%	-25.4%	$\Delta\widehat{FDR}^+$
$\Delta\widehat{R}^-$	3	12	14	11		2	2	3	5	$\Delta\widehat{R}^+$
$\Delta\widehat{F}^-$	-1	-1	-2	-2		-2	-2	-1	-1	$\Delta\widehat{F}^+$
$\Delta\widehat{T}^-$	4	13	16	13		4	4	4	6	$\Delta\widehat{T}^+$
				$\Delta\widehat{\pi}_A^-$	$\Delta\widehat{\pi}_0$	$\Delta\widehat{\pi}_A^+$				
				4.4%	-5.2%	0.8%				

The results for Turnover, Expense Ratio, and Total Net Asset (TNA) are presented in Panels A, B, and C. In each Panel, we examine the performance of the High and Low-characteristic groups, as well as their differences (High minus Low). The left part shows the FDR analysis for the worst funds (i.e. funds with significant negative estimated alphas) at different significance levels γ. We display the estimated False Discovery Rate $\widehat{FDR}^-$, the number of worst funds, lucky funds, and funds with truly negative alphas (denoted by $\widehat{R}^-$, $\widehat{F}^-$, and $\widehat{T}^-$, respectively). The right part contains the same information for the best funds (i.e. funds with significant positive estimated alphas). We also display the proportions of funds with negative, zero, and positive alphas (denoted by $\widehat{\pi}_A^-$, $\widehat{\pi}_0$, and $\widehat{\pi}_A^+$, respectively). The alphas of all funds are computed with the unconditional Carhart model.

Table 5

Performance and Fund Characteristics

Panel B Expense Ratio

	Worst funds					Best funds				
γ	0.05	0.10	0.15	0.20		0.20	0.15	0.10	0.05	γ
					High Expense					
$\widehat{FDR}^-$	22.0%	27.0%	35.5%	39.5%		100.0%	100.0%	100.0%	100.0%	$\widehat{FDR}^+$
$\widehat{R}^-$	35	57	65	78		31	23	15	8	$\widehat{R}^+$
$\widehat{F}^-$	8	15	23	31		31	23	15	8	$\widehat{F}^+$
$\widehat{T}^-$	27	42	42	47		0	0	0	0	$\widehat{T}^+$
				$\widehat{\pi}_A^-$	$\widehat{\pi}_0$	$\widehat{\pi}_A^+$				
				14.7%	85.3%	0.0%				
					Low Expense					
$\widehat{FDR}^-$	27.6%	28.2%	30.7%	29.1%		65.1%	57.6%	47.0%	35.3%	$\widehat{FDR}^+$
$\widehat{R}^-$	23	45	62	86		39	33	27	18	$\widehat{R}^+$
$\widehat{F}^-$	6	13	19	25		25	19	13	6	$\widehat{F}^+$
$\widehat{T}^-$	17	32	43	61		14	14	14	12	$\widehat{T}^+$
				$\widehat{\pi}_A^-$	$\widehat{\pi}_0$	$\widehat{\pi}_A^+$				
				26.3%	69.6%	4.1%				
					High minus Low					
$\Delta\widehat{FDR}^-$	-5.6%	-1.2%	4.8%	10.4%		34.9%	42.4%	53.0%	64.7%	$\Delta\widehat{FDR}^+$
$\Delta\widehat{R}^-$	12	12	3	-8		-8	-10	-12	-10	$\Delta\widehat{R}^+$
$\Delta\widehat{F}^-$	2	2	4	6		6	4	2	2	$\Delta\widehat{F}^+$
$\Delta\widehat{T}^-$	10	10	-1	-14		-14	-14	-14	-12	$\Delta\widehat{T}^+$
				$\Delta\widehat{\pi}_A^-$	$\Delta\widehat{\pi}_0$	$\Delta\widehat{\pi}_A^+$				
				-11.6%	15.7%	-4.1%				

Table 5

Performance and Fund Characteristics

Panel C Total Net Asset (TNA)

	Worst funds					Best funds				
γ	0.05	0.10	0.15	0.20		0.20	0.15	0.10	0.05	γ
					High TNA					
$\widehat{FDR}^{-}$	21.3%	24.0%	29.7%	32.9%		65.1%	58.6%	52.6%	35.9%	$\widehat{FDR}^{+}$
$\widehat{R}^{-}$	32	57	68	83		42	34	26	19	$\widehat{R}^{+}$
$\widehat{F}^{-}$	7	14	20	27		27	20	14	7	$\widehat{F}^{+}$
$\widehat{T}^{-}$	25	43	48	56		15	14	12	12	$\widehat{T}^{+}$
				$\widehat{\pi}_A^{-}$	$\widehat{\pi}_0$	$\widehat{\pi}_A^{+}$				
				20.7%	75.1%	4.2%				
					Low TNA					
$\widehat{FDR}^{-}$	24.6%	27.9%	31.9%	35.2%		100.0%	100.0%	100.0%	100.0%	$\widehat{FDR}^{+}$
$\widehat{R}^{-}$	29	51	67	81		28	21	14	7	$\widehat{R}^{+}$
$\widehat{F}^{-}$	7	14	21	28		28	21	14	7	$\widehat{F}^{+}$
$\widehat{T}^{-}$	22	37	46	53		0	0	0	0	$\widehat{T}^{+}$
				$\widehat{\pi}_A^{-}$	$\widehat{\pi}_0$	$\widehat{\pi}_A^{+}$				
				21.2%	78.8%	0.0%				
					High minus Low					
$\Delta\widehat{FDR}^{-}$	-3.3%	-3.9%	-2.2%	-2.3%		-34.9%	-41.4%	-47.4%	-64.1%	$\Delta\widehat{FDR}^{+}$
$\Delta\widehat{R}^{-}$	3	6	1	2		14	13	12	12	$\Delta\widehat{R}^{+}$
$\Delta\widehat{F}^{-}$	0	0	-1	-1		-1	-1	0	0	$\Delta\widehat{F}^{+}$
$\Delta\widehat{T}^{-}$	3	6	2	3		15	14	12	12	$\Delta\widehat{T}^{+}$
				$\Delta\widehat{\pi}_A^{-}$	$\Delta\widehat{\pi}_0$	$\Delta\widehat{\pi}_A^{+}$				
				-0.5%	-3.7%	4.2%				

Table 6

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%	63.1%	71.9%	$\widehat{FDR}^{+}$	100%	100%	57.5%	73.2%
$\widehat{FDR}^{-}$	38.9%	67.2%	13.7%	28.7%	$\widehat{FDR}^{-}$	44.8%	75.9%	13.2%	24.4%

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%	68.4%	85.3%	$\widehat{FDR}^{+}$	100%	100%	59.1%	74.1%
$\widehat{FDR}^{-}$	45.2%	67.2%	20.6%	34.6%	$\widehat{FDR}^{-}$	48.5%	76.5%	17.0%	29.6%

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%
$\widehat{FDR}^{-}$	50.7%	69.9%	10.1%	17.6%	$\widehat{FDR}^{-}$	55.4%	73.8%	6.5%	12.8%

The results for All funds (*All*), Growth funds (*G*), Aggressive Growth funds (*AG*), and Growth and Income funds (*GI*) are presented in Panels A, B, C, and D. The left part contains the estimated FDR among the best and worst funds ($\widehat{FDR}^{+}$ and $\widehat{FDR}^{-}$) at $\gamma = 0.05$ and 0.20 computed with the unconditional CAPM and Fama-French (FF) models . The right part contains the same information for the conditional versions of the CAPM and Fama-French (FF) models.

Table 7

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^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.701	0.078	0.078	0.211	0.211
		0.9	0.901	0.246	0.247	0.508	0.509
	0.10	0.7	0.699	0.143	0.143	0.321	0.321
		0.9	0.902	0.393	0.393	0.646	0.646
0.7	0.05	0.7	0.707	0.165	0.166	0.103	0.104
		0.9	0.899	0.433	0.434	0.307	0.307
	0.10	0.7	0.706	0.281	0.283	0.168	0.170
		0.9	0.899	0.602	0.602	0.439	0.439

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

q^-	γ	π_0	$\widehat{\pi}_0$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.711	0.103	0.105	0.211	0.214
		0.9	0.899	0.307	0.307	0.508	0.508
	0.10	0.7	0.710	0.168	0.170	0.321	0.325
		0.9	0.899	0.439	0.439	0.646	0.646
0.7	0.05	0.7	0.710	0.211	0.214	0.103	0.105
		0.9	0.900	0.508	0.512	0.307	0.307
	0.10	0.7	0.710	0.321	0.325	0.168	0.171
		0.9	0.899	0.646	0.648	0.439	0.439

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

q^-	γ	π_0	$\widehat{\pi}_0$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.706	0.103	0.104	0.165	0.166
		0.9	0.899	0.307	0.308	0.433	0.433
	0.10	0.7	0.705	0.168	0.170	0.281	0.283
		0.9	0.899	0.439	0.440	0.602	0.604
0.7	0.05	0.7	0.700	0.211	0.211	0.078	0.078
		0.9	0.901	0.508	0.511	0.246	0.247
	0.10	0.7	0.700	0.321	0.321	0.143	0.143
		0.9	0.902	0.646	0.647	0.393	0.394

The monthly returns are simulated according to a one-factor model for 1,456 funds and 336 periods. A proportion π_0 of funds have zero alphas. A proportion π_A of funds have differential performance. Among these funds, a proportion $\pi_A^- = \pi_A \cdot q^-$ of funds have a negative alpha α_A^-, and a proportion $\pi_A^+ = \pi_A \cdot (1 - q^-)$ of funds have a positive alpha α_A^-. $q^- \in [0, 1]$ is a positive scalar. The FDR^+ and FDR^- correspond to the true false discovery rates. $\widehat{\pi}_0$, $\widehat{FDR}^+$, and $\widehat{FDR}^-$ stand for the average values of the estimators over 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 7

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^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.704	0.078	0.078	0.211	0.212
		0.9	0.901	0.246	0.247	0.508	0.512
	0.10	0.7	0.704	0.143	0.144	0.321	0.323
		0.9	0.900	0.393	0.393	0.646	0.647
0.7	0.05	0.7	0.711	0.165	0.167	0.103	0.105
		0.9	0.902	0.433	0.435	0.307	0.310
	0.10	0.7	0.712	0.281	0.285	0.168	0.171
		0.9	0.902	0.602	0.604	0.439	0.441

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

q^-	γ	π_0	$\widehat{\pi}_0$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.717	0.103	0.105	0.211	0.215
		0.9	0.906	0.307	0.310	0.508	0.513
	0.10	0.7	0.716	0.168	0.172	0.321	0.328
		0.9	0.904	0.439	0.442	0.646	0.651
0.7	0.05	0.7	0.717	0.211	0.216	0.103	0.105
		0.9	0.905	0.508	0.513	0.307	0.310
	0.10	0.7	0.715	0.321	0.328	0.168	0.172
		0.9	0.905	0.646	0.651	0.439	0.443

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

q^-	γ	π_0	$\widehat{\pi}_0$	FDR^+	$\widehat{FDR}^+$	FDR^-	$\widehat{FDR}^-$
0.3	0.05	0.7	0.710	0.103	0.104	0.165	0.166
		0.9	0.902	0.307	0.309	0.433	0.435
	0.10	0.7	0.711	0.168	0.171	0.281	0.284
		0.9	0.903	0.439	0.442	0.602	0.604
0.7	0.05	0.7	0.704	0.211	0.212	0.078	0.078
		0.9	0.902	0.508	0.512	0.246	0.248
	0.10	0.7	0.706	0.321	0.324	0.143	0.144
		0.9	0.901	0.646	0.651	0.393	0.394

Table 8

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	0.212	0.09	0.087
	0.9	0.07	0.074	0.03	0.024
0.7	0.7	0.09	0.093	0.21	0.200
	0.9	0.03	0.034	0.07	0.067

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

q^-	π_0	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.204	0.09	0.086
	0.9	0.07	0.068	0.03	0.032
0.7	0.7	0.09	0.085	0.21	0.205
	0.9	0.03	0.031	0.07	0.069

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

q^-	γ	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.194	0.09	0.100
	0.9	0.07	0.067	0.03	0.033
0.7	0.7	0.09	0.085	0.21	0.214
	0.9	0.03	0.031	0.07	0.068

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	0.069	0.03	0.027
0.7	0.7	0.09	0.088	0.21	0.187
	0.9	0.03	0.029	0.07	0.063

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

q^-	π_0	π_A^+	$\widehat{\pi}_A^+$	π_A^-	$\widehat{\pi}_A^-$
0.3	0.7	0.21	0.187	0.09	0.080
	0.9	0.07	0.063	0.03	0.027
0.7	0.7	0.09	0.079	0.21	0.187
	0.9	0.03	0.027	0.07	0.063

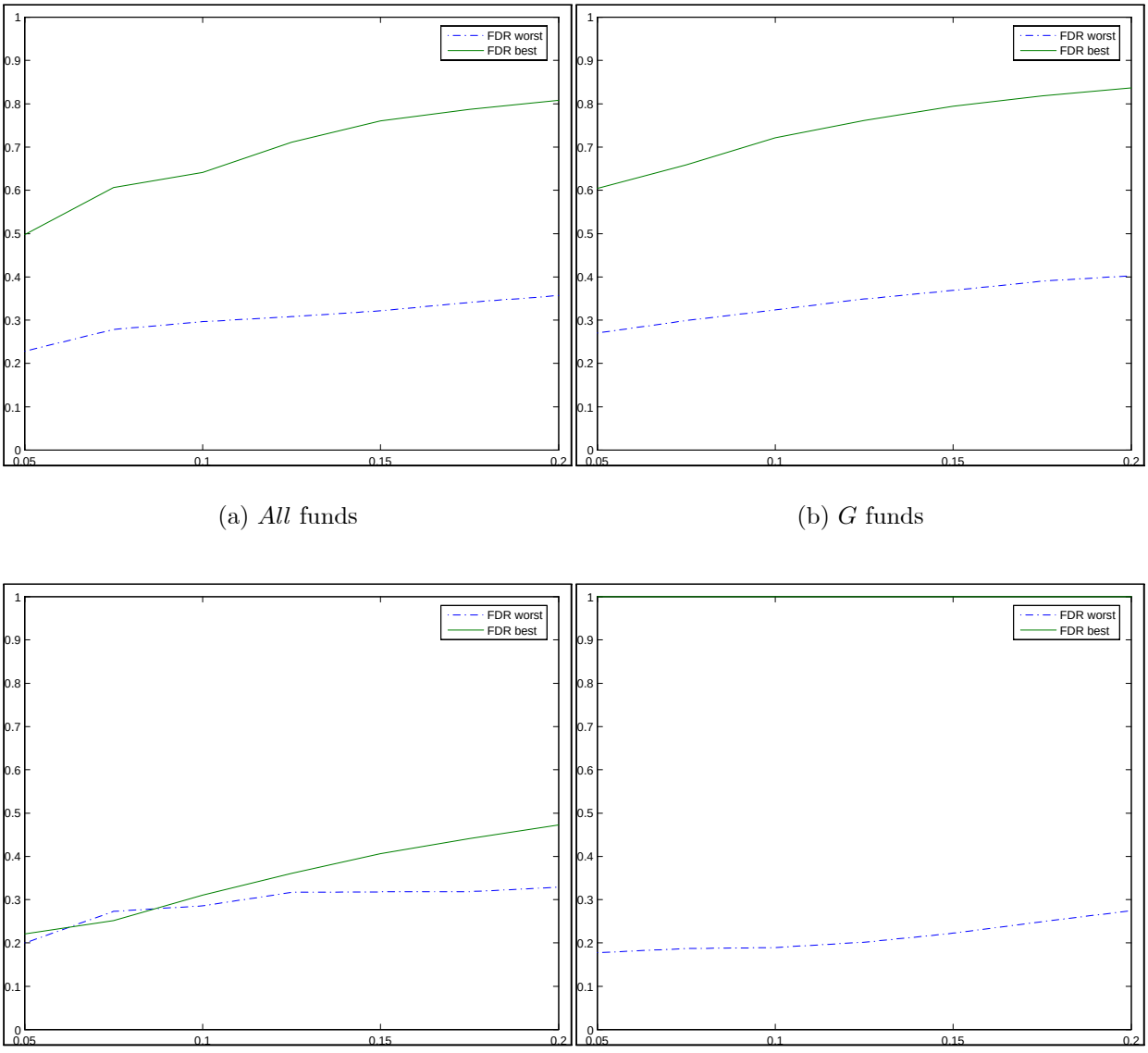
$\alpha_A^+ = 5\%, \alpha_A^- = -8\%$

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	0.029
0.7	0.7	0.09	0.081	0.21	0.209
	0.9	0.03	0.026	0.07	0.070

The monthly returns are simulated according to a one-factor model for 1,456 funds and 336 periods. A proportion π_0 of funds have zero alphas. A proportion π_A of funds have differential performance. Among these funds, a proportion $\pi_A^- = \pi_A \cdot q^-$ of funds have a negative alpha α_A^-, and a proportion $\pi_A^+ = \pi_A \cdot (1 - q^-)$ of funds have a positive alpha α_A^-. $q^- \in [0, 1]$ is a positive scalar. π_A^+ and π_A^+ correspond to the true proportions of funds with positive and negative alphas, respectively. $\widehat{\pi}_A^+$ and $\widehat{\pi}_A^-$ stand for the average values of the estimators over 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.

Figure 1

False Discovery Rates among the Best and the Worst Funds



(a) *All* funds

(b) *G* funds

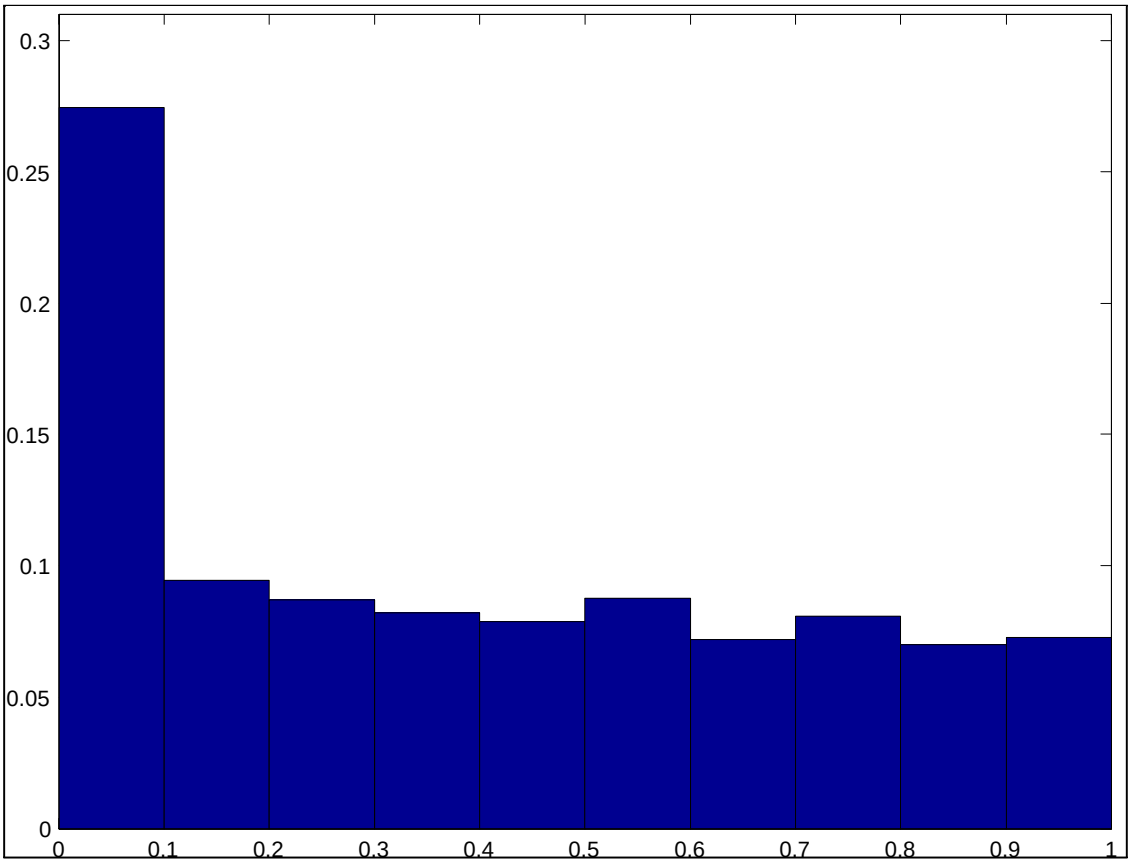
(c) *AG* funds

(d) *GI* funds

The figure plots the estimated FDR among the best and the worst funds as a function of the significance level γ. The $\widehat{FDR}^{+}$, represented by the solid line, corresponds to the FDR among the best funds (i.e. funds with significant positive estimated alphas). Note that the $\widehat{FDR}^{+}$ for GI funds is equal to 1 at any given γ. The $\widehat{FDR}^{-}$, represented by the dashed line, corresponds to the FDR among the worst funds (i.e. funds with significant negative estimated alphas). The alphas of all funds are computed with the unconditional Carhart model.

Figure 2

Histogram of the Fund Estimated p-values



We simulate fund excess returns for 1,456 funds and 336 observations with a one-factor market model (see the Monte-Carlo study for the details). From these simulated time-series, the fund alphas and p-values are estimated. The proportion π_0 of funds with zero alphas is equal to 80%. Among the 20%, 10% of the funds yield a negative alpha of -5% per year, and 10% of them have a positive alpha of 5% per year.