Informative Covariates, False Discoveries and Mutual Fund Performance

Po-Hsuan Hsu[a,*], Ioannis Kyriakou[b], Tren Ma[c], Georgios Sermpinis[c]

[a] *College of Technology Management, National Tsing Hua University, Taiwan*
[b] *Bayes Business School, City, University of London, London EC1Y 8TZ, UK*
[c] *Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ, UK*

Abstract

We introduce a novel multiple hypothesis testing method named the functional False Discovery Rate "plus" ($fFDR^+$). The method incorporates informative covariates (and new information they carry) in estimating the False Discovery Rate (FDR) of predictive models' "conditional" performance. In our simulation based on mutual fund returns, the $fFDR^+$ controls well the FDR and gains considerable power over prior methods that do not account for extra information. Its advantage remains under different alpha distributions, balanced and unbalanced data structure, and cross-sectional dependent and independent error terms. It is also robust to estimation errors in the covariates. In further empirical analyses, we construct portfolios based on several covariates (five well-known and four new ones) and show that they enhance the performance of mutual fund portfolios, highlighting the value of extra information in the multiple hypothesis testing framework.

Keywords: Multiple testing, Functional false discovery rate, Informative covariates, Mutual funds, Alphas

*Corresponding author.

Email addresses: pohsuanhsu@mx.nthu.edu.tw (Po-Hsuan Hsu), ioannis.kyriakou.2@city.ac.uk (Ioannis Kyriakou), DinhTren.Ma@glasgow.ac.uk (Tren Ma), georgios.sermpinis@glasgow.ac.uk (Georgios Sermpinis)

1. Introduction

Aiming to identify models with genuine predictive power from a large set of potential candidates, researchers have to resort to a multiple hypothesis testing framework to appropriately address the "data-snooping" or "p-hacking" bias that is a major challenge to social science (Sullivan *et al.*, 1999, 2001; White, 2000; Hansen, 2005). To address this challenge, researchers propose the concept of the False Discovery Rate (FDR) of Benjamini and Hochberg (1995), Storey (2002), Storey (2003), and Romano and Wolf (2005), i.e., the ratio of models that are mistakenly identified as having predictive power. Testing methods based on FDR has gained considerable attention in the literature and has been successfully applied to many areas of social science.[1]

One common feature of the methodologies in this framework is that the rejection criterion *only* depends on information of raw data and predictive models' performance metrics. However, in economics and finance research, the economic agents use all available information in assessing models' performance. Extra information sources can assist researchers to more accurately estimate FDR. Recently, Chen *et al.* (2021) introduced the functional FDR method that embeds the role of informative covariates (i.e., variables that carry extra information) in forming null hypotheses. This advancement is important in the sense that it enables us to test the "conditional" performance of predictive models, which is more consistent with the rational expectation hypothesis. To illustrate the importance of extra information in multiple testing problems, we can use mutual fund performance assessment as an example. If we use prior testing methods that do not account for extra information, we are testing an unconditional zero hypothesis, which corresponds to investors not updating their information in assessing mutual fund performance. This approach appears inappropriate because mutual funds and their managers are routinely reviewed by investors based on updated information. In other words, a more suitable null hypothesis for a mutual fund's performance should be zero conditional on the updated information set.

Our main contribution is the introduction of the functional False Discovery Rate "plus" ($fFDR^+$). Compared to the work of Chen *et al.* (2021), it has two distinguishing features. First, it allows us to focus on the right or left tail of the distribution and detect the significant outperformers/under-performers, which is important for decision makers (see Barras *et al.*, 2010,

[1] For instance, Fan and Fan (2011) employ FDR in testing and detecting jumps; Lan *et al.* (2016) utilize such a framework to control FDR in testing coefficients in high-dimensional linear models; see also Lan and Du (2019) for extensions and applications in mutual fund selection; or Barbaglia *et al.* (2022) for applications in detecting significant sentiment variables in forecasting with economic news.

hereafter BSW). Second, it is robust to cross-sectional dependencies among predictive models, which is common for most problems in economics and finance. For example, in mutual funds, the alphas are likely dependent due to herding and correlated trading behaviour (Wermers, 1999).

Compared to all earlier methods in the economics literature on control of the FDR, our $fFDR^+$ method incorporates extra information, has higher power, and controls for noise. It is easy to implement, does not rely on any strong assumption and can handle any continuous informative covariate. In examining our method, we use simulated mutual fund performance similarly to BSW and Andrikogiannopoulou and Papakonstantinou (2019) (AP henceforth). We show that, when an informative covariate is available, our $fFDR^+$ approach detects more true positive alpha funds under different alpha distributions, balanced and unbalanced data, and both cross-sectional independence and dependence in the error terms. The gap in power between $fFDR^+$ and prior FDR methods, depending on the distribution of the fund alpha population, can be up to about 30%. Our approach is also robust to estimation errors in the covariates.

We then apply our method and construct portfolios in order to evaluate it empirically in selecting outperforming mutual funds. In particular, we explore nine informative covariates: the first set contains five covariates that have been shown in prior studies to convey information on mutual fund performance, and the second set contains four new covariates that are inspired by asset pricing models. The first set includes the R-square of the asset pricing model (e.g., Carhart four-factor model) as suggested by Amihud and Goyenko (2013), the Return Gap of Kacperczyk *et al.* (2008), the Active Weight of Doshi *et al.* (2015), the Fund Size of Harvey and Liu (2017), and the Fund Flow suggested by Zheng (1999). The second set includes the Sharpe ratio, the Beta and Treynor ratio based on the Capital Asset Pricing Model (CAPM), and the idiosyncratic volatility of the Carhart four-factor model (Sigma).

We find that the set of mutual funds selected as out-performers by $fFDR^+$ is usually larger and different from the one obtained by prior FDR methods. As already discussed, earlier studies do not account for information other than mutual funds' returns and performance metrics; thus, their null hypotheses are unconditional and neglect investors' time-varying expectation. The fact that our $fFDR^+$ discovers more outperforming funds suggests that, with more information updating, there may exist more profitable mutual funds than researchers would have expected.

Based on the funds selected by $fFDR^+$, we build portfolios that consistently outperform the one generated by prior methods. Our results highlight the economic value of extra information.

In particular, the $fFDR^+$ portfolios with the R-square and Beta covariates are found to be the best with annualized alphas of 1.7%, followed by the $fFDR^+$ portfolios with the Active Weight, Fund Flow, Sigma, Treynor ratio, Fund Size, Sharpe Ratio and Return Gap covariates, separately achieving annualized alphas of at least 0.77%. We note that this profitability is persistent in our sample and is even strengthened over the recent period, a finding that disagrees with part of the recent literature which suggests otherwise (see Jones and Mo, 2021). All our $fFDR^+$ portfolios outperform the one generated by prior FDR methods and a set of portfolios created by single- and double-sorting the covariates under study.

In additional analysis, we also consider the $fFDR^+$ portfolio based on various ways of combining the nine covariates, such as the first principal component of the nine covariates (PC 1), the ordinary least squares (OLS), the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996), the ridge regression and the elastic net of Zou *et al.* (2005). We find that the elastic net delivers the best performance with an annualized alpha of 1.25%. The investors may also benefit from such combinations as they result in lower volatility in portfolio performance. This is advantageous as, in reality, investors do not know ex-ante what covariate is the best.

The rest of the paper is organized as follows. In Section 2, we introduce and explain our methodology. In Section 3, we provide a description of our data. Section 4 is devoted to our simulation experiment descriptions, whereas in Section 5 we present in detail our simulation results. Section 6 focuses on the empirical part of our analysis. Section 7 concludes the paper.

2. Methods for Controlling of Luck with Informative Covariate

2.1. Functional False Discovery Rate (fFDR)

Throughout this paper, we use mutual funds to represent predictive models. We define funds' performance based on their net return, that is, the return net of trading cost, fees and other expenses except loads and taxes. A fund is deemed out-performing if it distributes to investors a net return that generates a positive alpha (i.e., a part of a return series that is unexplained by systematic risk). If the alpha is negative (zero), the fund is said to be under-performing (zero-alpha). These definitions of out-performing and under-performing funds coincide with skilled and unskilled funds in BSW, respectively, and reflect the interest of investors.

Suppose that we are assessing m funds and each of them has a net return time series. We also assume that there exists a covariate X, with observed values $(x_1, \ldots, x_m)$, that conveys information about the alpha of each fund. Associated with X, we define Z whose observed value for fund i is $z_i = rank(x_i)/m$, where $rank(x_i)$ is the ranking of x_i in the set of observed values $(x_1, \dots, x_m)$. As X to Z is an one-one mapping and we work based on Z, we call that the covariate from now on. We introduce our notation by means of a single test, conditional on Z, for the alpha of a mutual fund:

$$H_0 : \alpha = 0, \qquad H_1 : \alpha \neq 0. \tag{1}$$

We denote by h the status of the null hypothesis, that is, $h = 0$ if the hypothesis $\alpha = 0$ is true and $h = 1$ if otherwise. In addition, P is the random variable representation of the p-value of the test, Z, as mentioned above, is the covariate which is uniformly distributed on $[0, 1]$, and $T = (P, Z)$. We suppose that $(h|Z = z) \sim \text{Bernoulli}(1 - \pi_0(z))$, that is, conditional on $Z = z$, the fund possesses a zero alpha with probability $\pi_0(z)$; this can be constant if Z does not convey any information about the probability of the fund's alpha being zero. The estimation procedure for $\pi_0(z)$ will be discussed later on. We require that under the true null, $(P|h = 0, Z = z)$ is uniformly distributed on $[0, 1]$ regardless of the value of z; when the null hypothesis is false, the conditional density function of $(P|h = 1, Z = z)$ is $f_1(.|z)$.

To assess the performance of m funds in terms of α within our framework, we consider m conditional hypothesis tests like (1):

$$H_{0,i} : \alpha_i = 0, \qquad H_{1,i} : \alpha_i \neq 0, \qquad i = 1, \dots, m, \tag{2}$$

where α_i is the alpha of fund i. For each i we have $T_i = (P_i, Z_i)$, and we assume that all the pairs are independent and each of them has the same distribution as (T, h).[2] Finally, we denote by $f(p, z)$ the joint density function of (P, Z). We have that

$$\mathbb{P}(h = 0|T = (p, z)) = \frac{\pi_0(z)}{f(p, z)} =: r(p, z) \tag{3}$$

is the posterior probability of the null hypothesis being true given that we observe $T = (p, z)$.[3]

To control the type I error, Storey (2003) introduces the "positive false discovery rate"

$$pFDR = \mathbb{E}\left(\frac{V}{R}\middle| R > 0\right), \tag{4}$$

where R is the number of rejected hypotheses in m tests and V the wrongly rejected ones. Chen *et al.* (2021), CRS henceforth, show that, with a fixed set Γ in $[0, 1]^2$, if we reject hypothesis

[2] In Section IB of the Internet Appendix, we show that this requirement can be eased for a typically cross-sectional dependence in mutual fund data.

[3] For more details about the role of $Z \sim \text{Uniform}(0, 1)$ and the derivation of (3), see Chen *et al.* (2021).

$H_{0,i}$ whenever $T_i \in \Gamma$, then

$$pFDR(\Gamma) = \mathbb{P}(h = 0|T \in \Gamma) = \int_{\Gamma} r(p, z)dpdz. \tag{5}$$

To maximize the number of rejections, we reject the hypotheses with the smallest statistic $r(p, z)$. Thus, the significance region $\{\Gamma_\theta : \theta \in [0, 1]\}$ is defined as

$$\Gamma_\theta = \{(p, z) \in [0, 1]^2 : r(p, z) \leq \theta\}, \tag{6}$$

where a larger θ implies more rejected hypotheses. Finally, we recall from Storey (2003) and CRS the definition of the q-value for the observed (p, z):

$$q(p, z) = \inf_{\{\Gamma_\tau | (p,z) \in \Gamma_\tau\}} pFDR(\Gamma_\tau) = pFDR(\Gamma_{r(p,z)}). \tag{7}$$

Given a target $\tau \in [0, 1]$, a procedure that rejects a hypothesis if and only if its q-value $\leq \tau$ guarantees that $pFDR$ is controlled at τ.

Empirically, let $\hat{\pi}_0(z)$ and $\hat{f}(p, z)$ be the estimated functions $\pi_0(z)$ and $f(p, z)$, respectively.[4] We calculate $\hat{r}(p, z) = \hat{\pi}_0(z)/\hat{f}(p, z)$ and estimate the q-value function as

$$\hat{q}(p_i, z_i) = \frac{1}{S_i} \sum_{k \in S_i} \hat{r}(p_k, z_k), \tag{8}$$

where $S_i = \{j|\hat{r}(p_j, z_j) \leq \hat{r}(p_i, z_i)\}$ and p_i is the p-value of test i. Then, given a target $pFDR$ level $\tau \in [0, 1]$, the null hypothesis $H_{0,i}$ is rejected if and only if $\hat{q}(p_i, z_i) \leq \tau$. CRS call this procedure Functional False Discovery Rate ($fFDR$).

2.2. The $fFDR^+$: application in selecting out-performing funds

By applying the $fFDR$ methodology to mutual funds at a given target $pFDR$ level τ, we obtain a set that includes both significantly out-performing and under-performing funds. To further improve mutual fund selection, we propose a $fFDR$-based method that selects a group of significantly out-performing funds with control of luck. In the following section, we introduce our $fFDR^+$ and discuss its application in a mutual fund context.

Consider a selection of R^+ out-performing funds including V^+ wrongly selected zero-alpha or under-performing funds. We define the positive false discovery rate in those significantly out-performing funds as

$$pFDR^+ = \mathbb{E}\left(\frac{V^+}{R^+}\middle| R^+ > 0\right). \tag{9}$$

For m tests, let A^+ be the set of hypotheses with positive estimated alpha, i.e., $A^+ = \{i|\hat{\alpha}_i > 0\}$,

[4] See Appendix A for more details.

where $\hat{\alpha}_i$ is the estimated alpha of fund i. At a given target τ of $pFDR^+$, by implementing the $fFDR$ procedure to control $pFDR$ at the target τ on the funds in set A^+, we obtain all the funds with positive estimated alphas (referred to as significant alphas).[5] Hence, the $fFDR$ selects positive-alpha funds with control of $pFDR$ at the given target; we call this procedure the functional FDR "plus" ($fFDR^+$).

Next, we highlight the differences between our and BSW's approaches. The starting point of both is the control of the type I error as in Benjamini and Hochberg (1995):

$$FDR = \mathbb{E}\left(\frac{V}{\max\{R,1\}}\right) = \mathbb{E}\left(\frac{V}{R}\bigg| R>0\right)\mathbb{P}(R>0) = pFDR \cdot \mathbb{P}(R>0), \tag{10}$$

where the last equality follows from (4). This implies that controlling for $pFDR$ at a given target τ, also controls for FDR at the same target. Furthermore, for a large number of tests, controlling for $pFDR$ and FDR is equivalent (see Storey, 2002, 2003).

Consider the m tests (2) in the absence of the covariate Z and let t_i be the test statistic of test i. Storey (2002) assumes that $t_1,\ldots,t_m$ are independent and the statuses of the null hypotheses $h_1,\ldots,h_m$ are independent Bernoulli random variables with $\mathbb{P}(h_i = 0) = \pi_0$. Additionally, across i, $(t_i|h_i = 0)$ and $(t_i|h_i = 1)$ are identically distributed. When we reject based on the p-values, for some $\lambda \in [0,1)$, π_0 can be estimated by

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i | p_i > \lambda, i = 1,\ldots,m\}}{(1-\lambda)m} \tag{11}$$

where $\#$ returns the number of elements in the set; this estimate is conservative biased.[6] BSW choose $\lambda = \lambda^*$ on the grid $\{0.3, 0.35, \ldots, 0.7\}$ such that the mean square error (MSE) of $\hat{\pi}_0(\lambda)$ is minimal.[7] We set $\hat{\pi}_0 = \hat{\pi}_0(\lambda^*)$.

To select out-performing funds with controlling for the FDR, BSW define the concept FDR^+ to measure the FDR in a group of funds selected as significant and positive estimated alphas as

$$FDR^+ = \mathbb{E}\left(\frac{V^+}{\max\{R^+,1\}}\right). \tag{12}$$

With a significant threshold γ and a procedure which selects a fund with a positive estimated

[5] In doing so, we assume that the number of funds that are out-performing but exhibit a negative estimated alpha is negligible. This is sensible as in practice we will not select those funds anyway. In BSW, as discussed next, having a positive estimated alpha is a necessary condition for a fund to be selected as out-performer.

[6] To have the estimate of π_0, first, under independence, there are $m\pi_0$ funds with truly zero alpha and their p-values have a uniform distribution in $[0,1]$. Hence, we expect $m\pi_0(1-\lambda)$ p-values in the set to fall in $[\lambda, 1]$. Second, this number can be conservatively approximated by $\#\{p_i | p_i > \lambda\}$, thus we have (11). With a larger λ, the estimate $\hat{\pi}_0$ is less conservative, as there are fewer p-values under the alternative belonging to $[\lambda, 1]$, but its variance is higher.

[7] In MSE $= \mathbb{E}(\hat{\pi}_0(\lambda) - \pi_0)^2$, the unknown π_0 is replaced by $\min_\lambda \hat{\pi}_0(\lambda)$ over the λ grid.

alpha whenever its p-value $\leq \gamma$, BSW estimate FDR^+ by

$$\widehat{FDR}_\gamma^+ = \frac{\hat{\pi}_0 \gamma/2}{\hat{R}^+/m}, \tag{13}$$

where $\hat{R}^+$ is the empirical number of funds selected as out-performers, i.e., $\hat{R}^+ = \#\{i|p_i \leq \gamma, \hat{\alpha}_i > 0\}$. When using this approach to determine out-performing funds with controlling for FDR^+ at a given target τ, the threshold γ is raised gradually until the $\widehat{FDR}^+$ estimate in (13) reaches the target τ. We refer to this procedure as FDR^+.

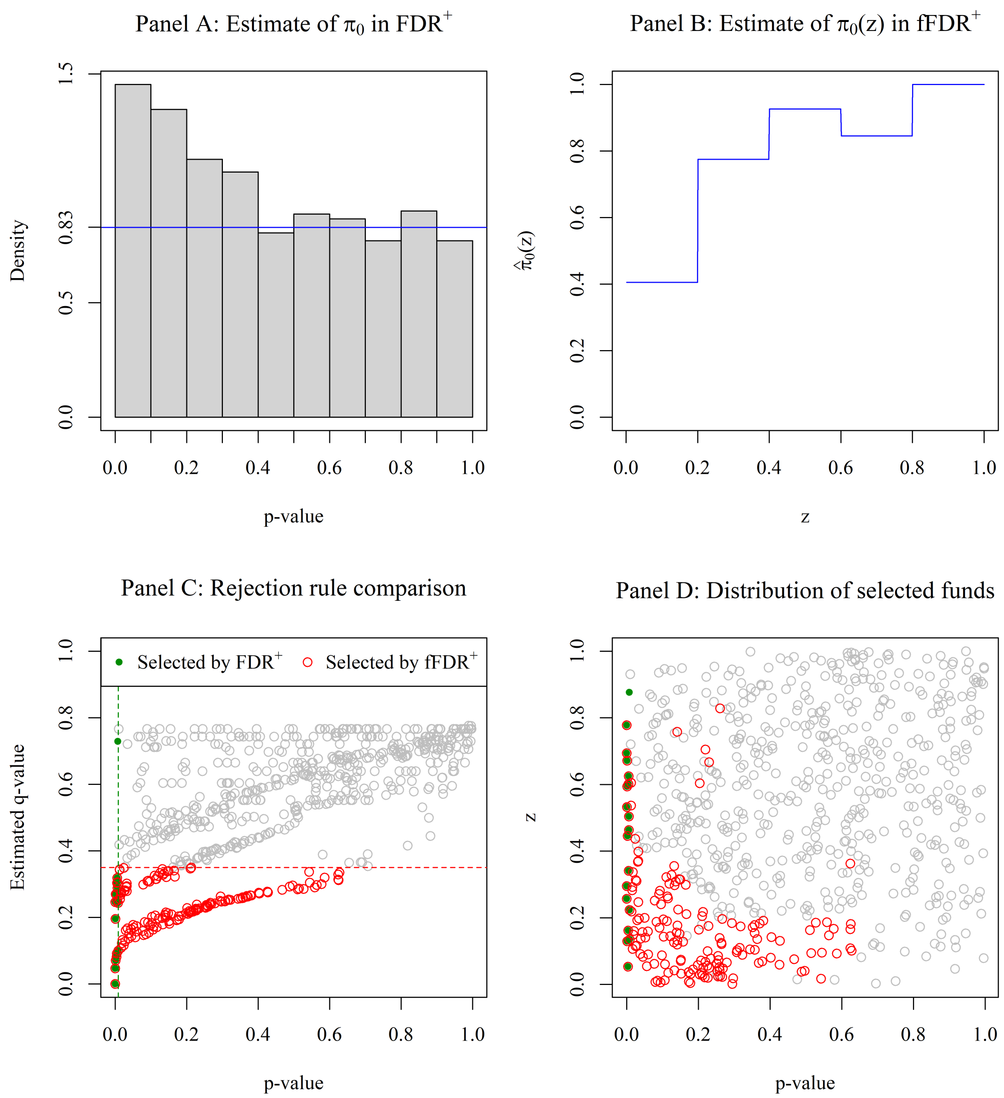
To illustrate the differences between our and BSW's procedures, we opt for a sub-period of five years from 2001 to 2004 and implement the FDR^+ and $fFDR^+$ to detect positive alpha funds, with the alphas determined by the four-factor model of Carhart (1997). In this case, the R-square of the model is used as the covariate for $fFDR^+$.[8] In Figure 1, we demonstrate how the two procedures work. Based on the p-values of all the considered funds, the FDR^+ estimates the proportion of zero-alpha funds in the whole sample, as a first step, giving $\hat{\pi}_0 \approx 0.83$. It then selects the positive estimated alpha funds with smallest p-values until the estimated $\widehat{FDR}_\gamma^+$ reaches a given FDR target. For illustration, we choose the FDR target $\tau = 35\%$, so that both methods select a substantial number of funds.[9] Here, all the funds with p-values less than or equal to $\gamma = 0.0086$ are selected by the FDR^+. The threshold γ is depicted by the green dashed line in Panel C and all the funds corresponding to the points on the left of the vertical line are selected. By contrast, the $fFDR^+$ considers only the set of positive estimated alpha funds and estimates the proportion of zero-alpha funds in this set as a step function of z (the quantiles of R-square).

In this experiment, we split the sample into five bins based on the ranking of the covariate z; thus, $\hat{\pi}_0(z)$ is a function with five "steps". The procedure continues with the estimation of the density function $f(p, z)$ and of the functional q-value $q(p, z)$. The $fFDR^+$ selects all the funds with estimated q-value less than or equal to 0.35: those funds correspond to the points below the red dashed line (the q-value = 0.35 line) in Panel C. This clearly shows that, for the same target, the $fFDR^+$ selects significantly more funds than FDR^+ (170 versus 19). More importantly, the funds selected by the FDR^+ are not merely a subset of those selected by $fFDR^+$. Panel D displays the distribution of the selected funds with respect to the p-value and z. We observe that the $fFDR^+$ assigns more weight to some funds with smaller z (thus,

[8]The details of the funds and the calculation of the p-values are deferred to Section 6. Here, we focus only on illustrating the differences.

[9]If we choose any target $\tau \leq 30\%$, the FDR^+ selects no funds.

Figure 1: Comparison of FDR^+ and $fFDR^+$. The graphs show the differences between the two procedures with respect to their null proportion estimations and their rejection rules. Panels A and B show that π_0 is estimated as a fixed number in FDR^+ procedure (see (11)) but as a step function in $fFDR^+$ procedure (see Appendix A). Panel C shows the rejection rules of the FDR^+ and $fFDR^+$: the former selects all the funds corresponding to the points on the left of the vertical green dashed line which consists of all funds with positive estimated alphas and p-values less than 0.0086, whereas the latter all the funds corresponding to the points below the horizontal red dashed line which consists of all funds with estimated q-value (see (8)) less than 0.35. Panel D shows the distribution of the selected funds in Panel C with respect to the p-value and the covariate z. In Panels C and D, only funds with positive estimated alpha are shown as ultimately both methods select funds from this set. The solid green points represent funds selected by the FDR^+, whereas the red circles the funds selected by the $fFDR^+$; the green points with a red ring are the commonly selected funds.



smaller R-square), but the weight is not equally distributed across the funds with the same level of z. As the rejection rule of $fFDR^+$ is based on the functional q-value, which is based on the estimates of $\pi_0(z)$ and $f(p, z)$, it is not possible to explain this merely by the ranking of the p-value and the covariate z, as evidenced in Panel D: the $fFDR^+$ selects some funds with p-values around 0.6 while skipping many funds with a smaller p-value at roughly the same level of z.

As shown in AP, the FDR^+ relies on an over-conservative estimate of the null proportion and utilizes only p-values and the estimated alphas. On the other hand, the $fFDR^+$ additionally uses an informative covariate about the performance of the funds and expresses the null proportion as a function of it, while accounting for the joint distribution of the p-value and the covariate. This results in a more accurately estimated FDR and, therefore, an increased power in detecting out-performing funds. We are illustrating the prominent power of the $fFDR^+$ via a set of simulation studies in the next sections. In the empirical section, we will show the actual profitability that the five covariates can bring to investors while controlling for luck.

3. Data

We use monthly mutual fund data from January 1975 to December 2019 collected from the CRSP database.[10] As CRSP reports funds at the share class level, we use MFLINKS to acquire fund data at the portfolio level. For a fund at a given point in time with multiple share classes, we average the share classes' net return weighted by the total net asset (TNA) value at the beginning of the month.[11] The TNA at the fund level is estimated by the sum of the share classes' TNA. We omit the following month return after a missed return observation as CRSP fills this with the accumulated returns since the last non-missing month. To obtain the holdings data of the funds, which will be used to calculate our covariates, we merge the CRSP and Thomson/CDA databases by utilizing MFLINKS. The holdings database provides us with stock identifiers, which we use to link the funds' position with the CRSP equity files. From this equity database, we obtain information such as the price and number of shares outstanding of the stocks that the funds hold on their reported portfolio date. We use these to calculate the return gap and the active weight, which are described in more detail later.

We consider only funds with an investment objective belonging to the categories Growth, Aggressive Growth and Growth & Income. Both CRSP and CDA provide this information; CDA is more consistent over time, hence we choose that. As the funds' investment objective can change, we collect first all the funds in these categories. If at some point a fund misses its investment objective, we fill this in by its prior non-missing objective. If a fund's objective changes, we remove those return observations corresponding to periods when its objective does

[10] We are aware of the possible biases in the CRSP mutual fund data before 1984 (Fama and French, 2010) and thus conduct a robustness check using a sample from 1984 to 2019 in Section IC of the Internet Appendix.

[11] Since 1991, we use the monthly TNA of the fund's share classes. Before 1991, most of the funds report their TNA on a quarterly basis. For this, we follow Amihud and Goyenko (2013) to fill in the missing TNA of each fund (at the share class level) by its most recently available one.

not belong to the three aforementioned categories. To obtain a precise four-factor alpha estimate, we select only funds with at least 60 monthly observations. Overall, we gather a sample of 2,224 funds which provides the empirical metrics for our simulation study.

In the empirical part, when calculating the related covariates, we additionally require each fund to hold at least 10 stocks; this is consistent with Kacperczyk *et al.* (2008) and Doshi *et al.* (2015) and is needed here as we use the return gap and active weight from their studies as two of our covariates. The number of funds used when constructing our covariate-based portfolios varies over years and will be reported in detail in the empirical section.

4. Simulation Setup

In this section, we present the details of our simulation design consisting of the choice of the model, the distributions of the alpha population, the data-generating process and the metrics that we will use to gauge the performance of the methods.

4.1. The model

Following the majority of the existing literature on mutual fund performance, we use the four-factor model of Carhart (1997) to compute the fund performance:

$$r_{i,t} = \alpha_i + b_i r_{m,t} + s_i r_{smb,t} + h_i r_{hml,t} + m_i r_{mom,t} + \varepsilon_{i,t}, \qquad i = 1, \ldots, m, \tag{14}$$

where $r_{i,t}$ is the excess net return of fund i over the risk-free rate (i.e., the one-month Treasury bill rate), $r_{m,t}$ the market's excess return on the CRSP NYSE/Amex/NASDAQ value-weighted market portfolio, $r_{smb,t}$ the Fama–French small minus big factor, $r_{hml,t}$ the high minus low factor, $r_{mom,t}$ the momentum factor and $\varepsilon_{i,t}$ the noise of fund i at time t. All factors and the one-month Treasury bill rate are obtained from French's website.

Our simulations are designed similarly to BSW and AP in terms of the data-generating process accounting, in addition, for an informative covariate and considering more distribution types of the fund alpha population. Whereas BSW and AP focus on the estimated proportions of the out-performing, under-performing and zero-alpha funds, we consider the performance of the FDR^+ and $fFDR^+$. More specifically, for a given fund alpha distribution, we first generate in each iteration the true fund alpha population and a covariate that conveys information about the alpha of each fund. Second, we simulate the Fama–French factors (factors loadings) by drawing from a normal distribution with parameters equal to their sample counterparts (obtained from estimations of model (14)). Next, the noise is generated under both cross-sectional independence and dependence. In the first case, the noise is drawn cross-sectionally independent from a normal distribution, that is, $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ where, as in Barras *et al.* (2020), σ_ε is is set equal to the median of its real-data counterpart, that is, approximately 0.0183 for our sample. The results under this assumption are reported in the next section. In the dependent case, the noise is generated as in BSW and the simulation results are deferred to Section IB of the Internet Appendix. The simulated data are then used to generate the net return for each fund. Subsequently, by carrying out regression (14) of the generated net return on the simulated Fama–French factors, we estimate the alpha and calculate the related p-values for the tests (2). Finally, based on these estimated alphas, p-values and the covariate, we implement the $fFDR^+$ and FDR^+, for a given FDR target, to obtain the significantly out-performing funds. We estimate the actual false discoveries rate of the $fFDR^+$ and check if it meets the given target. We then compare the two methods in terms of power, defined as the expected ratio of the number of true positive alpha funds detected to the total number of true positive alpha funds in the population.

4.2. The distribution of fund alphas

We consider three different types for the distribution of fund alphas: a discrete, a discrete-continuous mixture and a continuous. A covariate Z conveys information about the alpha of each fund in the population; more specifically, a fund with $Z = z$ has a probability $\pi_0(z)$ of being zero-alpha. Also, without loss of generality, we assume that, for non-zero alpha funds, their covariates and alphas are positively correlated.[12]

First, in the discrete type, we draw alphas from three mass points $-\alpha^* < 0$, 0 and $\alpha^* > 0$ with probabilities π^-, π_0 and π^+. Thus,

$$\alpha \sim \pi^- \delta_{\alpha=-\alpha^*} + \pi_0 \delta_{\alpha=0} + \pi^+ \delta_{\alpha=\alpha^*}. \tag{15}$$

We consider five values for $\alpha^* \in \{1.5, 2, 2.5, 3, 3.5\}$ (the values are annualized and in %) together with six combinations of the proportions (π^+, π_0, π^-) based on $\pi^+ \in \{0.1, 0.13\}$, $\pi^-/\pi^+ \in \{1.5, 3, 6\}$ and $\pi_0 = 1 - \pi^- - \pi^+$, i.e., a total of thirty cases.[13]

In the mixed discrete-continuous distribution, we draw alphas from two components including the mass point 0 and the normal distribution $\mathcal{N}(0, \sigma^2)$ with, respectively, probabilities

[12] If the correlation is negative, we use instead $-Z$.

[13] The chosen π^+ values are close to those used in the recent literature: $\pi^+ = 10.6\%$ (see Harvey and Liu, 2018) and $\pi^+ = 13\%$ (see Andrikogiannopoulou and Papakonstantinou, 2016). The ratio $\pi^-/\pi^+ = 6$ is studied in AP. Aiming to extend the range of our study, we consider also the ratios 1.5 and 3.

$\pi_0 \in (0,1)$ and $1-\pi_0$. We have, therefore, that

$$\alpha \sim \pi_0 \delta_{\alpha=0} + (1-\pi_0)\mathcal{N}(0,\sigma^2). \tag{16}$$

We consider five values for $\sigma \in \{1,2,3,4,5\}$ (the values are annualized and in %) and the same six π_0 values as in the discrete distribution earlier.

Finally, in the continuous case, we draw alphas from a mixture of two normal distributions $\mathcal{N}(\mu_1,\sigma_1^2)$ and $\mathcal{N}(\mu_2,\sigma_2^2)$ with, respectively, probabilities $\pi_1 \in [0,1]$ and $\pi_2 = 1-\pi_1$, i.e.,

$$\alpha \sim \pi_1\mathcal{N}(\mu_1,\sigma_1^2) + \pi_2\mathcal{N}(\mu_2,\sigma_2^2). \tag{17}$$

When π_1 and π_2 are positive, we have indeed a mixture; we adopt from Harvey and Liu (2018) $\pi_1 = 0.3$ and $\pi_2 = 0.7$ and, to point up the performance of our method, we consider fifteen combinations based on $(\mu_1,\mu_2) \in \{(-2.3,-0.7), (-2,-0.5), (-2.5,0)\}$ and $(\sigma_1,\sigma_2) \in \{(1,0.5), (1.5,0.6), (2,1), (2.5,1.25), (3,1.5)\}$ (the values of the pairs are annualized and in %).[14]

In (17) $\pi_0 = 0$, whereas in (15) and (16) $\pi_0 > 0$. When $\pi_0 > 0$, we study an up-and-down shape of $\pi_0(z)$. Specifically, to guarantee $\pi_0(z) \in [0,1]$ for all z, we choose $\pi_0(z) = \min\{1, \max(f(z),0)\} \in [0,1]$, where

$$f(z) = 3.5(z-0.5)^3 - 0.5(z-0.5) + c \tag{18}$$

and c is chosen to satisfy $\int_0^1 \pi_0(z)dz = \pi_0$. This way we are able to investigate the effect of π_0 on the power of the methods by varying c while keeping the shape of $\pi_0(z)$ roughly unchanged.[15]

Suppose the distribution of alpha and the form of $\pi_0(z)$ are determined. We generate the covariate vector $(z_1, z_2, \ldots, z_m)$ with each element drawn from the uniform distribution $[0,1]$ and assign them to the funds satisfying the descriptions mentioned at the beginning of this section. The noise in equation (14) is generated cross-sectionally independent or dependent. In the former case it is drawn from a normal distribution $\mathcal{N}(0,\sigma_\varepsilon^2)$, where, as in Barras *et al.* (2020), σ_ε is set equal to the median of its real-data counterpart, that is, approximately 0.0183 for our sample. For each replication, we implement the $fFDR^+$ and FDR^+ and compute the rate of falsely selected funds among those classified as out-performers and the rate of truly out-performing funds detected. The two metrics are averaged across 1,000 replications to obtain

[14] Our choices are intended to be wide enough to encompass the cases of Harvey and Liu (2018): $(\pi_1,\pi_2) = (0.283, 0.717)$, $(\mu_1,\mu_2) = (-2.277,-0.685)$ and $(\sigma_1,\sigma_2) = (1.513, 0.586)$. In Section IB of the Internet Appendix, we additionally present results of the case $\pi_2 = 0$, i.e., when the mixture becomes a single normal distribution.

[15] The alternative choices of a decreasing function $\pi_0(z)$ with $f(z) = -1.5(z-0.5)^3 + c$, an increasing function $\pi_0(z)$ with $f(z) = 1.5(z-0.5)^3 + c$ or a constant function $\pi_0(z) = c$ result in some discrepancies, without affecting, though, our main conclusions.

estimates for the actual FDR and the power of each procedure.[16]

5. Analysis of $fFDR^+$ and power

We set the number of funds for simulations at 2,000 which is close to our sample of 2,224 funds. We demonstrate the ability of the $fFDR^+$ to control the FDR for balanced panel data, where the number of observations per fund is equal to 274, under cross-sectional independence. In the interest of space, we refer to Section IB of the Internet Appendix for the results under cross-sectional dependence as well as the unbalanced panel data cases. We then compare the powers of the $fFDR^+$ and the FDR^+ in controlling the FDR at the 10% level; we extend to higher levels and highlight the differences between the two procedures. In each simulation study, we analyze the relationship between the powers of the two methods and: i) the proportion of zero-alpha funds in the sample; ii) the magnitude and proportion of positive alpha funds in the sample. We also study the impact of the number of funds in the sample and the number of observations per fund on the power. Finally, we examine the impact of estimation errors in the covariates, in the power of our procedure.

In general, the results show that the $fFDR^+$ controls well the FDR at any given targets. When the FDR target is set at 10%, the $fFDR^+$ detects more positive alpha funds than the FDR^+ with a difference in power up to 30%, depending on cases and parameters of the distributions. When we raise the FDR target to higher levels, the difference is even higher in favour of the $fFDR^+$. The results are consistent regardless of the number of funds in the sample, the structure of the panel data and the dependence of the cross-sectional error terms.

In an empirical setting, the informative covariates are estimated quantities. This is translated to an estimation noise that may affect the power of our procedure. Our simulations reveal that our method is robust in terms of power up to moderate to high estimation noise.

5.1. False discovery rate control of $fFDR^+$

For varying targets of FDR $\in \{5\%, 10\%, \ldots, 90\%\}$, we implement the simulation procedure in Section 4 with balanced panel data. Figures 2, 3 and 4 exhibit our results for the generated data under cross-sectional independence.

In Figure 2, we show our results for the discrete distribution (15) for varying α^*. The upper three subplots correspond to $\pi^+ = 0.1$, whereas the lower three subplots to $\pi^+ = 0.13$. From left to right, the ratio π^-/π^+ increases from 1.5 to 6 (with the null proportion π_0 decreasing accordingly). For example, the top-left subplot exhibits the actual FDR (vertical axis) and the

[16]We refer to Section IA of the Internet Appendix a detailed description of the simulation procedure.

given targets of FDR (horizontal axis) with the alphas drawn from a discrete population of which 75%, 10% and 15% are, respectively, zero-, positive- and negative-alpha funds. A point on or below the 45°-line indicates that the $fFDR^+$ controls FDR well for the given level; this is the case for $\alpha^* = 1.5$ at all the FDR targets. For $\alpha^* = 3.5$, the FDR is slightly not met for targets in the interval $(0.1, 0.8)$. In general, we witness slight failure of the $fFDR^+$ to control for FDR when α^* is abnormally high. In the last case with smallest π_0, the FDR is controlled well. In Figure 3, we study the case of the fund alpha population described by the mixed discrete-continuous distribution (16). We organize our results based on the same null proportions π_0 as in Figure 2 and present these for varying σ. We observe that the FDR target is slightly unmet only for extreme values of σ when the null proportion is very high and this effect is also milder compared to the discrete distribution cases. Finally, in Figure 4, we report the results for the continuous distribution (17) for varying μ or (μ_1, μ_2) and σ or (σ_1, σ_2). We find that the $fFDR^+$ controls FDR well at all targets.

Figure 2: Performance of $fFDR^+$ for discrete distribution of α. The graphs show the performance of the $fFDR^+$ in terms of FDR control when alphas are drawn from a discrete distribution. The simulated data are balanced panels with cross-sectional independence.

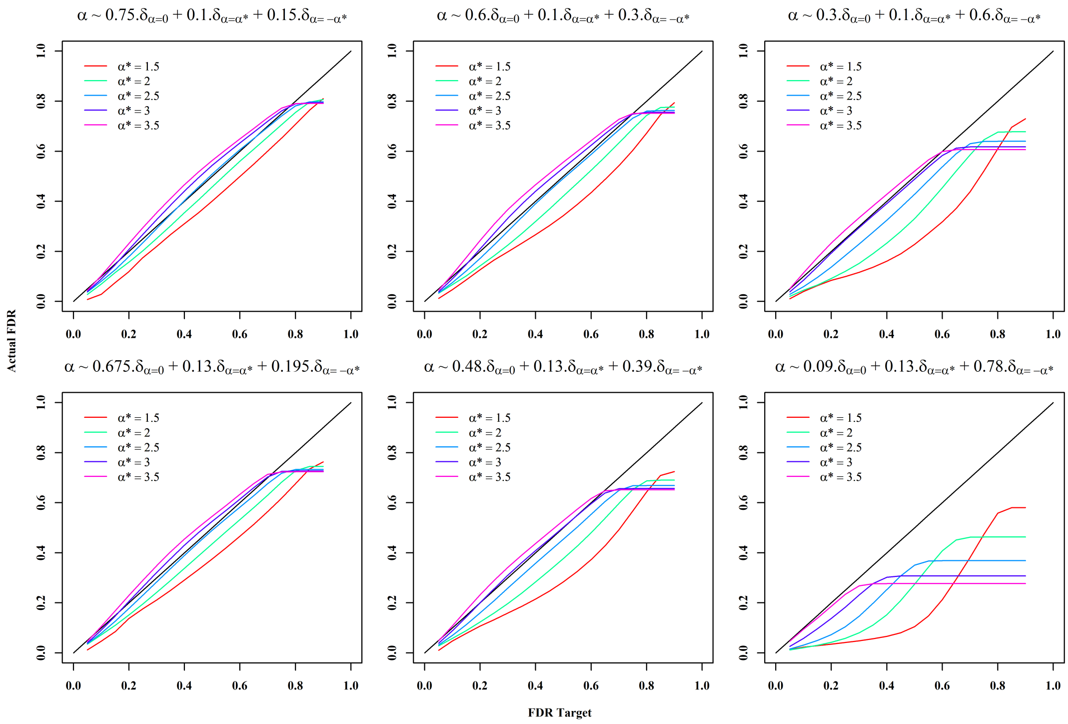


Figure 3: Performance of $fFDR^+$ for discrete and normal distribution mixture of α. The graphs show the performance of the $fFDR^+$ in terms of FDR control when alphas are drawn from a mixture of discrete and normal distributions. The simulated data are balanced panels with cross-sectional independence.

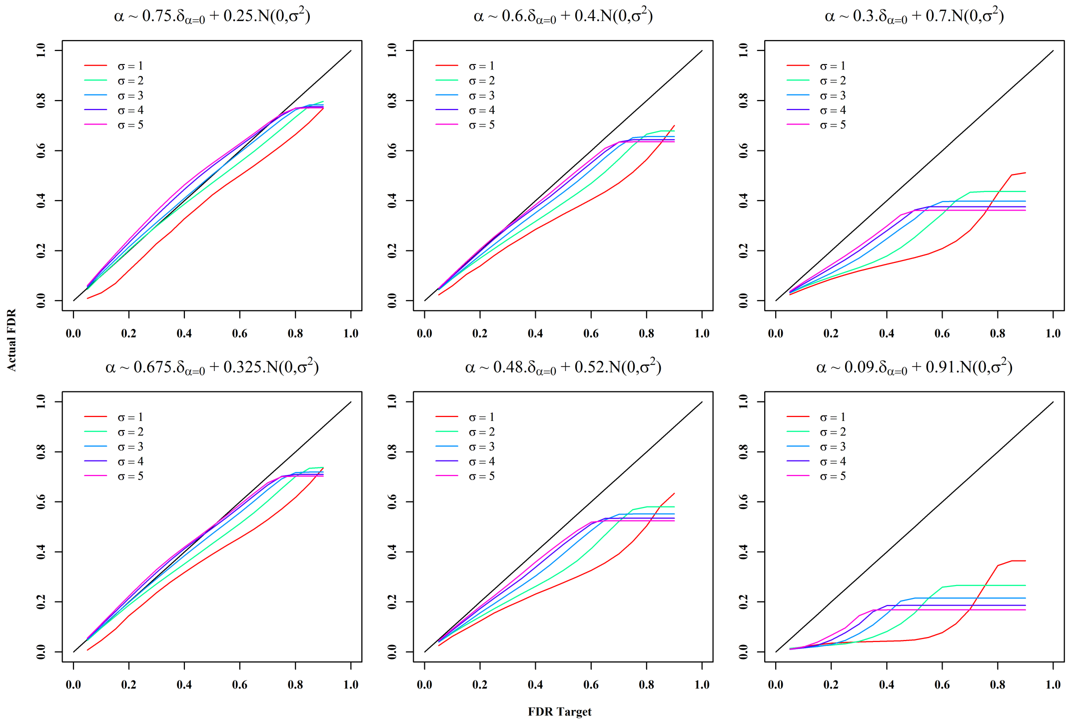
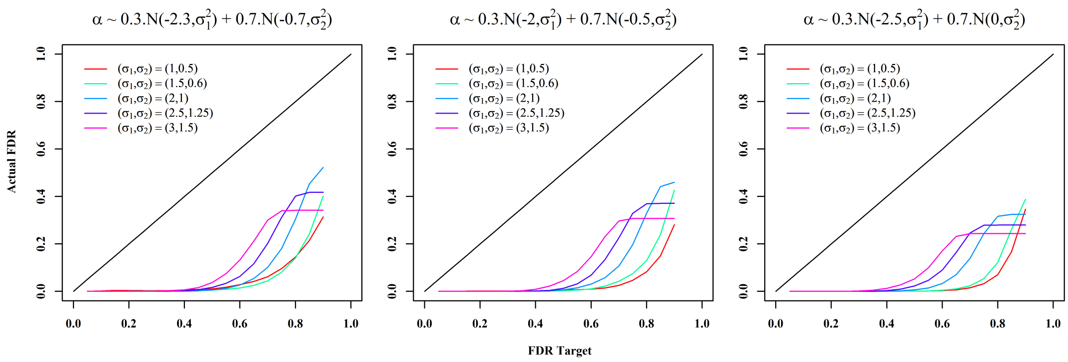


Figure 4: Performance of $fFDR^+$ for continuous distribution of α. The graphs show the performance of the $fFDR^+$ in terms of FDR control when alphas are drawn from a continuous distribution which is a mixture of two normals. The simulated data are balanced panels with cross-sectional independence.



In summary, our simulations are based on proposed fund alpha distributions from the recent literature, from the least realistic cases, with all the out-performing and under-performing funds assumed to have the same mass alpha value, to the more realistic ones, where the alpha is drawn from a continuous distribution, in which no fund has exact zero but rather mostly negative alpha. Our results suggest that, for the continuous distribution, the proposed $fFDR^+$ approach controls well for FDR at any given target.

In Section IB of the Internet Appendix we repeat the exercise for balanced data under cross-sectional dependence and unbalanced data under both cross-sectional independence and dependence. Our findings remain robust.

5.2. Power analysis

Next, we study the power of our $fFDR^+$ approach in detecting truly positive alpha funds, calculated as described in Section 4, and compare it with the FDR^+ of BSW for FDR control at 10%. Although the magnitude of our results varies with different targets of FDR, our main conclusion of the power superiority of the $fFDR^+$ remains.

In Panel A of Table 1, we report for the discrete distribution (15). For $(\pi^+, \pi_0, \pi^-) = (10, 75, 15)\%$ with highest π_0 and smallest $\alpha^* = 1.5$, both the $fFDR^+$ and FDR^+ achieve similar powers, i.e., 0.3% and 0.4%, respectively. This is expected in this particular case as the number and magnitude of the true positive alphas are small, while we are controlling for FDR at 10%.[17] The superiority of the $fFDR^+$ is more perceptible and stabler for larger α^*. This discrepancy depends not only on the magnitude and proportion of positive alphas, but also on the proportion of zero alphas. This is because both procedures use the null proportion (π_0 in FDR^+ and $\pi_0(z)$ in $fFDR^+$) to estimate the FDR. With the same magnitude and proportion of positive alphas, the small proportion of zero alphas implies the higher power of both the $fFDR^+$ and FDR^+. The effect of the null proportion on the gap of $fFDR^+$ over FDR^+ is stronger when the magnitude of positive alphas is not too high. The gap varies by case and may even exceed 30% (when $\pi^+ = 10\%, \pi_0 = 30\%$ and $\alpha^* = 2.5$).[18]

Panel B exhibits the power upshots for the case of the fund alpha population described by the distribution mixture (16). This implies the dependence of the proportion and magnitude of positive alphas on the proportion of the zero-alpha funds and the σ value for non-zero alphas. We expect a higher power for both methods for a smaller zero-alpha proportion and/or a higher value of σ. We find that the $fFDR^+$ is more powerful than FDR^+. More specifically, for the balanced data under cross-sectional independence and $\pi_0 = 75\%$, the power of the $fFDR^+$ (FDR^+) increases from 0.3% to 60.8% (0.2% to 52.2%) with increasing σ from 1 to 5. For given, say, $\sigma = 2$, the power of the $fFDR^+$ (FDR^+) increases from 15.4% to 38% (8.2% and 22%) with reducing π_0. The gap is generally evident for $\sigma > 1$ with power differences around 10% but which can also reach up to 16%.

[17] As will be shown later, with a higher FDR target (such as 30%), the power of the $fFDR^+$ exceeds that of FDR^+ by 6%. Considering a higher target than 10% is sensible for trading and diversification purposes as otherwise very few or no out-performing funds are selected. In the study of BSW, the estimated FDR in the empirical application is at least 41.5% on average (depending on portfolio).

[18] As shown in Section IB of the Internet Appendix, the relevant reports vary slightly when the simulated data are generated with alternative forms of $\pi_0(z)$ mentioned in footnote 15, with unbalanced panel or with cross-sectional dependence, however the overall picture remains the same.

Table 1: Performance comparison in terms of power (%). The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when the alphas of 2,000 funds are drawn from a discrete distribution, i.e. $\alpha \sim \pi^+\delta_{\alpha=\alpha^*} + \pi_0\delta_{\alpha=0} + \pi^-\delta_{\alpha=-\alpha^*}$ (Panel A), a discrete-normal distribution mixture, i.e. $\alpha \sim \pi_0\delta_{\alpha=0} + (1-\pi_0)\mathcal{N}(0,\sigma^2)$ (Panel B), and a mixture of two normal distributions, i.e. $\alpha \sim 0.3\mathcal{N}(\mu_1,\sigma_1^2) + 0.7\mathcal{N}(\mu_2,\sigma_2^2)$ (Panel C) with various setting of parameters. The simulated data are a balanced panel with 274 observations per fund and are generated with cross-sectional independence.

Panel A: discrete distribution.						
(π^+,π_0,π^-)	Procedure	$\alpha^*=1.5$	$\alpha^*=2$	$\alpha^*=2.5$	$\alpha^*=3$	$\alpha^*=3.5$
$(10,75,15)\%$	$fFDR^+$	0.3	5.1	21.8	45.7	67.3
	FDR^+	0.4	2.1	12.1	32.3	53.5
$(10,60,30)\%$	$fFDR^+$	1.1	10.3	33.1	58.5	77.5
	FDR^+	0.4	2.3	13.8	35.9	57.4
$(10,30,60)\%$	$fFDR^+$	3.5	22.9	52.9	76.6	89.7
	FDR^+	0.4	3.3	21.4	47.8	69.6
$(13,67.5,19.5)\%$	$fFDR^+$	0.8	8.8	30.1	55.1	75.1
	FDR^+	0.4	3.1	17.6	39.7	60.9
$(13,48,39)\%$	$fFDR^+$	2.3	16.4	43	68.1	84.3
	FDR^+	0.5	4	21.8	46.1	66.8
$(13,9,78)\%$	$fFDR^+$	6.4	34	67.6	89.2	97.5
	FDR^+	0.5	6.9	37.2	69.2	88
Panel B: discrete-normal distribution mixture.						
π_0	Procedure	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$
75%	$fFDR^+$	0.3	15.4	36.1	51.1	60.8
	FDR^+	0.2	8.2	26.7	41.7	52.4
60%	$fFDR^+$	1.2	21.6	42.8	57.1	66.1
	FDR^+	0.2	11.4	31.5	46.6	56.9
30%	$fFDR^+$	4	31.6	54	67.2	74.8
	FDR^+	0.4	17.5	40.5	55.6	65.4
67.5%	$fFDR^+$	0.8	18.9	40	54.5	63.7
	FDR^+	0.2	9.9	29.6	44.5	55
48%	$fFDR^+$	2.3	25.9	47.8	61.6	70.4
	FDR^+	0.3	13.9	35.4	50.5	60.5
9%	$fFDR^+$	6	37.9	60.6	73.6	80.9
	FDR^+	0.5	22	47.1	62.7	72.2
Panel C: mixture of two normal distributions.						
				(σ_1,σ_2)		
(μ_1,μ_2)	Procedure	$(1,0.5)$	$(1.5,0.6)$	$(2,1)$	$(2.5,1.25)$	$(3,1.5)$
		$\pi^+=6\%$	$\pi^+=10.4\%$	$\pi^+=20.7\%$	$\pi^+=25.5\%$	$\pi^+=29.1\%$
$(-2.3,-0.7)$	$fFDR^+$	0	0.3	4.5	12.9	22.5
	FDR^+	0	0	0.3	1.9	7.1
		$\pi^+=11.8\%$	$\pi^+=16.9\%$	$\pi^+=26.4\%$	$\pi^+=30.5\%$	$\pi^+=33.4\%$
$(-2,-0.5)$	$fFDR^+$	0	0.4	5.9	15.1	24.8
	FDR^+	0	0.1	0.4	2.9	9
		$\pi^+=35.2\%$	$\pi^+=36.4\%$	$\pi^+=38.2\%$	$\pi^+=39.8\%$	$\pi^+=41.1\%$
$(-2.5,0)$	$fFDR^+$	0.1	0.6	8.3	17.8	27.6
	FDR^+	0	0	0.6	4.2	11.4

Finally, in Panel C, we study the outcome from using the mixture of normals (17) with $\pi_1 = 0.3$, $\pi_2 = 0.7$ and non-positive means (μ_1, μ_2) to limit the likelihood of a positive alpha. The proportion of positive alphas ranges from 6% to 41.1%. For small (σ_1, σ_2) values, the positive alphas are also small in magnitude and, consequently, the power is negligible. When

(σ_1, σ_2) are higher than $(2, 1)$, the power of both methods as well as their discrepancy increase significantly and favourably for $fFDR^+$ reaching up to 16%.

Our analysis has shown that, when controlling for FDR at an as low level as 10%, both the $fFDR^+$ and FDR^+ have good power for large (in magnitude) alphas. When this happens, the gain in power of the $fFDR^+$ over FDR^+ can vary depending on the underlying fund alpha distribution: 10% to 16% (continuous distribution) and 20% to 30% (discrete distribution). On the other hand, when the zero-alpha proportion is high and the proportion and magnitude of positive alphas is small, the power of both methods reduces.

Finally, as we demonstrate in Section IB of the Internet Appendix that our conclusions are not affected by the data structure (balanced versus unbalanced panel) or dependencies.

5.3. Power and FDR trade-off

In what follows, we study the impact on power when controlling for FDR at different (higher than 10% level) targets. Our results show clear differences between the $fFDR^+$ and FDR^+ and, in support of the former, even for cases of negligible power for a 10% target. Constructing mutual fund portfolios at higher FDR levels is sensible as otherwise we may end up with empty portfolios. Investors have to face a trade-off between the power in detecting out-performing funds and the FDR threshold, which we discuss next.

Table 2: Power comparison (in %) for varying FDR targets (%). The table presents some selected cases of low powers of the $fFDR^+$ and FDR^+ at FDR target of 10%. We consider a discrete distribution: $\alpha \sim 0.75\delta_{\alpha=0} + 0.1\delta_{\alpha=1.5} + 0.15\delta_{\alpha=-1.5}$; a discrete-normal mixture: $\alpha \sim 0.75\delta_{\alpha=0} + 0.2\mathcal{N}(0, 1.5^2)$; and a two-normal mixture: $\alpha \sim 0.3\mathcal{N}(-2.3, 1^2) + 0.7\mathcal{N}(-0.7, 0.5^2)$. The simulated data are balanced panels with cross-sectional independence.

Distribution	Procedure	FDR target								
		10	20	30	40	50	60	70	80	90
Discrete	$fFDR^+$	0.3	2.5	8	18.1	32.3	48.5	64.3	76.3	85
	FDR^+	0.4	0.9	2	3.9	7.4	14	24.7	41.5	65.1
Mixture of discrete and normal	$fFDR^+$	0.3	1.3	3.2	6.5	11.8	19.8	31.3	46.3	64.1
	FDR^+	0.2	0.4	0.7	1.1	1.7	2.7	4.9	10.4	26.5
Mixture of normals	$fFDR^+$	0	0.1	0.4	1.2	2.7	5.9	11.7	21.3	35.3
	FDR^+	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1

We focus on cases of very low power when the FDR is controlled at 10%. For brevity, we present in Table 2 our results for only balanced data under cross-sectional independence and FDR targets up to 90%, noting that these are largely unchanged for unbalanced data. In particular, for the underlying discrete fund alpha distribution, the $fFDR^+$ gains rapidly power with increasing FDR targets, peaking at 40% in excess of the FDR^+ when the target is set at 70%. For the continuous distribution, the power of the FDR^+ changes very slowly and is persistently negligible (mixture of normals) even for FDR controlled at 90%. On the other hand, the $fFDR^+$ detects abundant positive alpha funds with a power that can reach up to 46% in excess of the FDR^+ (mixture of two normal distributions with 90% target).

5.4. *Varying the number of observations and funds*

Hitherto, we have assumed a sample with $m = 2{,}000$ funds, which reflects our actual dataset for the whole period from 1975 to 2019. When constructing a portfolio, we usually use sub-periods of five years and the number of alive funds in these sub-periods naturally varies. In this section, we investigate the impact of varying number of observations T per fund and the number of funds m on the power.

In Table 3, we present the outcomes for different underlying distributions of fund alphas, when we control FDR at a 10% target and use balanced panel data with cross-sectional independence. We vary m from 500 to 3,000 and T from 120 months (i.e., 10 years) to 420 months

Table 3: Power comparison (in %) for varying sample size and observation length. The table compares the power of the $fFDR^+$ and FDR^+ in a balanced panel data with varying number of observations per fund (T) and number of funds (m). We present three cases where alphas of m funds are drawn from i) discrete distribution: $\alpha \sim 0.1\delta_{\alpha=2} + 0.3\delta_{\alpha=0} + 0.6\delta_{\alpha=-2}$ (Panel A); ii) discrete-normal mixture: $\alpha \sim 0.3\delta_{\alpha=0} + 0.7\mathcal{N}(0, 2^2)$ (Panel B); and mixture of two normal distributions: $\alpha \sim 0.3\mathcal{N}(-2, 2^2) + 0.7\mathcal{N}(-0.5, 1)$ (Panel C). The simulated data are balanced panels with cross-sectional independence.

		Number of observations per fund					
m	Procedure	$T = 120$	$T = 180$	$T = 240$	$T = 300$	$T = 360$	$T = 420$
		Panel A: Discrete distribution					
500	$fFDR^+$	2.7	8.5	19.6	31.8	44.6	54.8
	FDR^+	0.6	1.4	3	5.3	10.6	18.4
1000	$fFDR^+$	1.5	6	16.3	29.4	42.4	52.9
	FDR^+	0.4	0.8	2.1	4.9	10.6	19.1
2000	$fFDR^+$	1.2	5.7	15.4	28	40.6	51.4
	FDR^+	0.2	0.6	1.5	4.8	11.2	20.4
3000	$fFDR^+$	1.1	5.4	15	27.6	39.3	50.8
	FDR^+	0.2	0.5	1.6	4.9	11.8	20.7
		Panel B: Mixture of Discrete and Normal distributions					
500	$fFDR^+$	12.4	21.3	29.1	35.2	40.5	44.9
	FDR^+	2.4	7.5	14.1	20	25.3	29.8
1000	$fFDR^+$	11.7	21	28.1	34.7	40	44.5
	FDR^+	2.1	7.8	14.1	20.1	25.2	29.7
2000	$fFDR^+$	11.4	20.5	28.1	34.1	39.3	43.7
	FDR^+	2.2	7.9	14.2	19.9	25.1	29.7
3000	$fFDR^+$	11.2	20.4	27.8	33.9	39	43.6
	FDR^+	2.3	8	14.1	20	25.2	29.7
		Panel C: Mixture of Normal distributions					
500	$fFDR^+$	1.3	3	5.3	8	10.9	13.4
	FDR^+	0.2	0.3	0.5	0.8	1.3	1.8
1000	$fFDR^+$	0.9	2.4	4.8	7.6	10.1	12.8
	FDR^+	0.1	0.2	0.4	0.6	1.1	1.6
2000	$fFDR^+$	0.7	2.2	4.5	6.9	9.6	12
	FDR^+	0.1	0.1	0.3	0.5	1	1.6
3000	$fFDR^+$	0.7	2.2	4.3	6.8	9.3	11.9
	FDR^+	0	0.1	0.2	0.4	0.9	1.5

(i.e., 35 years). It is evident from the reports that the power of the $fFDR^+$ increases at a much faster pace with increasing T. With rising m, the power of the $fFDR^+$ slightly decreases, whereas such is observed for the FDR^+ mainly in Panel C. This is not a substantial concern though, as in reality we do not have a very large number of alive funds in a given sub-period. Overall, the power of the $fFDR^+$ in excess of the FDR^+ can reach 30%.

Apparently for $T = 120$, both procedures have low power. Empirically, when constructing a portfolio of mutual funds, we usually use in-sample sub-periods of 5 years. In these cases, the investors may have to raise the FDR target to a higher level as explained in the previous section.[19] In Table 4, we focus the spotlight on (small) $m = 500$ and $T = 60$ (i.e., 5 years). It is shown there that both methods yield even lower power at the FDR target of 10%. By increasing the target, the power of the $fFDR^+$ in detecting out-performing funds rises faster than that of the FDR^+, especially for the discrete and mixed normal distributions.

Table 4: Power comparison (in %) for varying FDR targets for sample with small size and small number of observations. In this table, we consider three distributions as in Table 3 for samples consisting of $m = 500$ funds (balanced panels with cross-sectional independence) with $T = 60$ observations per fund (5 years).

		FDR target								
Distribution	Procedure	10	20	30	40	50	60	70	80	90
Discrete	$fFDR^+$	0.5	2.2	5.8	12.2	20.9	30.8	41.5	53.5	66.3
	FDR^+	0.2	0.5	0.7	0.9	1.3	1.7	2.1	2.6	3.6
Mixture of discrete and normal	$fFDR^+$	2.4	7.4	14.4	23	32.7	42.9	53.2	63.5	68.4
	FDR^+	0.4	0.9	1.6	3	5.6	10.4	18.9	32.2	47.3
Mixture of normals	$fFDR^+$	0.2	1	2.9	6.2	11.1	18	26.7	37.5	51
	FDR^+	0.1	0.1	0.2	0.3	0.4	0.5	0.8	1	1.5

5.5. Estimation errors in the covariate

In the main settings of simulations, we consider a simple covariate where in the set of *non-zero* alpha funds, the ranking of funds' alpha is the same as that of funds' covariate. This does not hold in the whole population. Put differently, one cannot simply rank the funds based on a covariate to distinguish the out-performing ones from the zero-alpha and the under-performing ones. In this section, we further study the behaviour of our $fFDR^+$ approach by adding to the original covariate a noise reflecting potential estimation biases, as all covariates in the real data are calculated based on a certain sample period. More specifically, instead of using the covariate Z as in our previous simulations, we use $Z' = (z'_1, \ldots, z'_m)$ given by

$$z'_i = z_i + \eta_i, \tag{19}$$

[19] In fact, in order to construct non-empty FDR^+ portfolios based on five-year in-samples, BSW introduce a procedure where they allow the estimate of FDR^+ to be above 70% for several years.

where η_i denotes the noise and is generated independently from a normal distribution $N(0, \sigma_\eta^2)$. Alternatively Z' can be viewed as a realization of some informative covariate which aims to capture Z. Depending on the scale of the estimation error, the realized covariate could have different levels of information. We do not know actual estimation errors in covariates in reality. Thus, we simulate low to high noise in our covariates. More specifically, we consider three different values of σ_η including $\sigma_1 = 0.5/\sqrt{12}$ and $\sigma_2 = 1/\sqrt{12}$. These values are based on the fact that the covariate $Z \sim U[0, 1]$, which has a standard deviation of $1/\sqrt{12}$. We confirm that the $fFDR^+$ controls well for the FDR in this setting and the figures are virtually the same as those presented in the previous sections in the original setting. This is the most important characteristic of $fFDR^+$ we should expect, that is, ability to control well for the risk even when the new information contains noise.

In Table 5 we provide further information by presenting the power (at FDR target of 10%) of the $fFDR^+$. Comparing with Table 1, the power is lower but still remarkably higher than that of the FDR^+ with a varying gap across cases of the alpha distribution and the choice of σ_η.As will be shown in our empirical analysis, the $fFDR^+$ with use of each covariate gains significant power over the FDR^+. Therefore, we could assume that covariates in our application

Table 5: Power (in %) of $fFDR^+$ under noised covariate. The data are generated as in tables 1–3 except the use of a new covariate containing a noise: $Z' = Z + \eta$ instead of Z. The noise is drawn independently form normal distribution $\eta \sim N(0, \sigma_\eta^2)$ where σ_η taking value in $\{\sigma_1 = 0.5/\sqrt{12}, \sigma_2 = 1/\sqrt{12}\}$.

Panel A: Discrete distribution.										
	$\alpha^* = 1.5$		$\alpha^* = 2$		$\alpha^* = 2.5$		$\alpha^* = 3$		$\alpha^* = 3.5$	
(π^+, π_0, π^-)	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
(10,75,15)%	0.3	0.3	4.7	3.9	19.6	16.9	41.9	37.5	63.7	59.3
(10,60,30)%	1	0.7	8.7	6.5	28	23.1	52	45.7	71.8	66.3
(10,30,60)%	2.6	1.5	16.4	12	43.7	36.1	69.8	61.8	85.7	79.9
(13,67.5,19.5)%	0.7	0.6	8.2	6.7	27.5	23.6	50.8	45.9	71.2	66.6
(13,48,39)%	1.9	1.3	14	10.7	38.2	32.1	62.8	56	80.6	75.2
(13,9,78)%	5.1	3.3	27.8	21.7	62.3	55.2	87.6	82.4	96.6	94.2
Panel B: Mixture of a discrete and a normal distributions.										
	$\sigma = 1$		$\sigma = 2$		$\sigma = 3$		$\sigma = 4$		$\sigma = 5$	
π_0	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
75%	0.2	0.1	14	11.9	33.9	31.7	48.3	46.4	58.1	56.4
60%	0.6	0.3	19.3	16.4	39.6	36.9	53.8	51.3	62.5	60.5
30%	2.2	1.2	28.2	23.9	49.2	45.4	62	59	70.5	68
67.5%	0.4	0.2	16.8	14.3	36.8	34.4	51.2	49	60.7	58.7
48%	1.2	0.7	22.9	19.4	43.6	40.3	57.1	54.4	65.6	63.3
9%	3.6	2.1	34	29	56.1	51.7	68.5	65.4	75.9	73.8
Panel C: Mixture of two normal distributions.										
					(σ_1, σ_2)					
	$(1, 0.5)$		$(1.5, 0.6)$		$(2, 1)$		$(2.5, 1.25)$		$(3, 1.5)$	
(μ_1, μ_2)	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
(-2.3,-0.7)	0	0	0	0	1.2	0.6	6.5	4.2	15	11.2
(-2,-0.5)	0	0	0	0	2.1	1.1	8.9	6	18.2	13.9
(-2.5,0)	0	0	0.1	0.1	4.3	2.4	12.2	8.3	22.1	17.1

have relatively less noise than ones in this simulation.

Concluding this section, we recollect that the simulated power of $fFDR^+$ in detecting out-performing funds is found to be larger than FDR^+'s. This persists for different fund alpha distributions, balanced and unbalanced data, cross-sectional dependence of error terms accounted for or not. This power advantage depends on the magnitude and proportion of positive alphas as well as the proportion of zero alpha in the population, the number of funds in the sample, estimation errors in the covariates, and the average number of observations per fund. Especially when the last factor is small, leading to a diminished power for both procedures, we can recover that for the $fFDR^+$ by uplifting the FDR level. In our empirical application of the next section, we show how the investors can benefit from this.

6. Empirical Results

6.1. Five covariates proposed in the literature

We start our empirical investigation of the $fFDR^+$ approach by considering five covariates that may convey information about the performance of mutual funds. They are shown to be persistent and, therefore, can predict the performance of mutual funds. We also propose four new covariates based on asset pricing models.

First, we study the R-square of Amihud and Goyenko (2013), which is estimated from the Carhart four-factor model and measures the activeness of a fund. If a fund replicates the market, the R-square will be close to one; if, instead, it is more active, it will have a small R-square and in this case, according to the authors, funds tend to perform better.

The second covariate is the Fund Size of Harvey and Liu (2017). This takes into account both the fund size, which is the total net assets under management (TNA) of a fund, and the industry size, which is the total assets under management of all active mutual funds in the sample (sum of TNA). More specifically, for fund i at time t, it is defined as

$$\text{Fund Size}_{i,t} = \ln \frac{\text{TNA}_{i,t}}{\text{IndustrySize}_t} - \ln \frac{\text{TNA}_{i,0^*}}{\text{IndustrySize}_{0^*}}, \tag{20}$$

where $t = 0^*$ corresponds to the time of the first TNA observation in our sample. The Fund Size reflects the growth in scale of a fund relative to the whole active mutual fund market. Harvey and Liu (2017) show a significant negative relationship between Fund Size and funds' performance.[20]

[20] Pastor *et al.* (2015) and Chen *et al.* (2004) as well as Zhu (2018), respectively, argue that the industry size and the fund size (approximated by the logarithm of the fund's TNA) have a negative impact on the funds' performance. We use the Fund Size of Harvey and Liu (2017) as it incorporates information of both covariates.

The third covariate is the Return Gap of Kacperczyk *et al.* (2008), which is intended to reflect the unobserved actions of the funds. Mutual funds usually disclose their portfolio holdings and return periodically, e.g., quarterly or semi-annually. The investors are unaware of the funds' trading activities in the period of consecutive reports. The Return Gap of a fund is defined as the difference between the return that is disclosed by the fund and the return that the fund would have based on disclosure of its last portfolio holdings. Kacperczyk *et al.* (2008) show that the funds' performance can be predicted by their past return gaps; mutual funds with higher past return gap tend to perform better in the future.

Our fourth covariate is the Active Weight of Doshi *et al.* (2015), which aims to gauge the fund's activeness level and is given by the sum of the absolute differences of the stock value weights and the actual weights that the fund assigns to the stocks in its portfolio holdings. In their research, they show that funds with higher active weight tend to perform better. To obtain meaningful values for the active weight and the return gap, as in Kacperczyk *et al.* (2008) and Doshi *et al.* (2015), we require each mutual fund to hold at least 10 stocks in its portfolio at any time.

The fifth covariate is the Fund Flow. The interaction of fund flow and funds' performance has been studied quite extensively such as in Sirri and Tufano (1998), Berk and Green (2004), Harvey and Liu (2017) and Capponi *et al.* (2020), among others. Zheng (1999), in particular, discovers that funds receiving money perform better than those that lose money. The author also shows that investors can earn abnormal returns by using small funds' flow information. Here, we follow Bris *et al.* (2007) and define Fund Flow at time t as

$$\text{Fund Flow}_t = \frac{\text{TNA}_t - (1 + r_t)\text{TNA}_{t-1}}{(1 + r_t)\text{TNA}_{t-1}}, \tag{21}$$

where r_t is the return of the fund in the period $t - 1$ to t.

In addition to the aforementioned well-known covariates, we propose four new covariates that are based on asset pricing models and are available for all funds in our sample. These are the Sharpe ratio, the Beta and Treynor ratio obtained from the Capital Asset Pricing Model, and the idiosyncratic volatility (Sigma) of the Carhart four-factor model. The Sharpe and Treynor ratios are risk-adjusted performance measures of funds, whereas the Beta and Sigma reflect systematic and idiosyncratic risk, respectively. These metrics reveal aspects of the past mutual funds' performance and, thus, may assist in identifying out-performing and under-performing

Other studies on the relationship between fund size and performance and funds' holding liquidity (e.g., Yan, 2008) or funds' merger (i.e., McLemore, 2019) document the same conclusion.

funds. Asset pricing metrics are regularly used by wealth managers and academics in the fields of trading, asset pricing and investors' performance, but are overlooked in the mutual funds literature.[21]

6.2. The FDR^+ and $fFDR^+$ portfolios

In this section, we illustrate how $fFDR^+$ helps to identify out-performing mutual funds using a portfolio approach following BSW. More specifically, at the end of year t, we select a group of funds to invest in year $t + 1$ based on historical information from the last five years ($t - 4$ to t). In order to implement $fFDF^+$ and FDR^+, we require the observed values of the covariates of each fund, the estimated alpha and the p-value of each test. We execute, first, the Carhart four-factor model over the 5-year period to estimate the alpha.

The informative value of the Return Gap, Active Weight, Fund Flow and Fund Size on funds' performance is persistent, i.e., the choice between using the most recent (final-year) observations for these covariates or their average values over the whole in-sample (five years) is of less importance, as demonstrated by our robustness check in Section IF of the Internet Appendix.[22] Although the predictability of the covariates may last for a long horizon of up to five years, we expect their informative values to decrease with time; hence, forming portfolios based on their recent realizations is preferred to their average values of the whole last five years' time. Because of this, Return Gap, Active Weight, Fund Flow and Fund Size are calculated based on data in the final year of the in-sample (i.e., we use the exposure of the fund flow in year t for the Fund Flow, the value at the end of year t for the Fund Size, whereas for the Active Weight and the Return Gap we use their average exposures in year t). The R-square, Sharpe Ratio, Beta, Sigma and Treynor ratio are based on the whole five years. We calculate our p-values in a similar fashion to BSW. For the funds that suffer from heteroskedasticity or autocorrelation, we calculate the t-statistics based on the heteroskedasticity and autocorrelation-consistent standard deviation estimator of Newey and West (1987). For each fund, we implement 10,000 bootstrap replications to estimate the distribution of the t-statistic and subsequently calculate the bootstrapped p-value for the fund.[23]

As required by our method, the p-values of any truly zero-alpha funds, given a covariate

[21] For instance, Clifford *et al.* (2021) study the relation between idiosyncratic volatility and mutual funds flows but they do not focus on using this informative covariate as a factor for funds selection.

[22] Readers may refer to Kacperczyk *et al.* (2008), Doshi *et al.* (2015), Zheng (1999) and Harvey and Liu (2017) for the studies of the persistence of the Return Gap, Active Weight, Fund Flow and Fund Size, respectively. It should also be noted, that in our $fFDR$ framework, all covariates are transformed to uniform with only the ranking of the covariates across the funds counting.

[23] The bootstrapping procedure may result in duplicated bootstrapped p-values. For this, we use an adequate number of replications to reduce that effect and obtain good estimates of $\pi_0(z)$ and $f(p, z)$.

value, should be uniformly distributed. Although it is difficult for us to validate this requirement in reality as we never know which funds are truly zero-alpha, it appears intuitive for us to assume that this condition is satisfied. Consider, for example, the R-square. We expect the truly zero-alpha funds to invest randomly in the stock market, thus they should possess an R-square value of roughly equal to one. Conditional on a specific R-square value that a truly zero-alpha fund could have, i.e., close to one, if the fund is truly zero-alpha then its p-value should follow a uniform distribution like any usual true null hypothesis test.[24]

Next, we describe the selection process of out-performing funds to invest in year $t+1$ given a FDR target τ in $(0,1)$. First, we recall the relevant selection process for BSW's "$FDR\tau$" portfolio. For each γ on the grid $\{0.01, 0.02, ..., 0.6\}$, we calculate the $\widehat{FDR}^+_\gamma$ given by (13). Then, we find γ^* such that $\widehat{FDR}^+_{\gamma^*}$ is closest to τ; this is the significant threshold for BSW's portfolio, that is, all the positively estimated alpha funds in the in-sample window with p-values $\leq \gamma^*$ will be included in the $FDR\tau$ portfolio. This guarantees the non-empty property of the portfolio but does not always meet the FDR target τ, thereby $\widehat{FDR}^+_{\gamma^*}$ may be much higher than τ.

Second, we select out-performing funds for a $fFDR$-based portfolio, namely, "$fFDR\tau$". To establish comparable $fFDR\tau$ and $FDR\tau$ portfolios, we implement the $fFDR^+$ (with a particular covariate) to control $pFDR^+$ at a target τ^* that reflects the FDR level controlled by the $FDR\tau$ portfolio but has to be less than one.[25] As the FDR of the $FDR\tau$ portfolio is controlled at level $\widehat{FDR}^+_{\gamma^*}$ which may be greater than one or less than τ, we set: $\tau^* = \tau$ if $\widehat{FDR}^+_{\gamma^*} \leq \tau < 1$; $\tau^* = \widehat{FDR}^+_{\gamma^*}$ if $\tau < \widehat{FDR}^+_{\gamma^*} < 1$.[26] If $\widehat{FDR}^+_{\gamma^*} \geq 1$, we just select all the funds in the $FDR\tau$ portfolio.

For both the $fFDR\tau$ and $FDR\tau$ portfolios, we invest equally in the selected funds in the following year. If a selected fund does not survive for a month during the year, then its weights are redistributed to the remaining (surviving) funds.

As aforementioned, at the beginning of each year we select funds in to a portfolio by using

[24] Indeed, the p-value of each test i is defined as $p_i = 1 - F(|t_i|)$, where $F(|t_i|) = \mathbb{P}(|\mathcal{T}_i| < |t_i| | \alpha_i = 0)$ and $\mathcal{T}_i$ is the conventional t statistic of test i and t_i its estimated value. If hypothesis $\alpha_i = 0$ is true, conditional on a specific covariate value, the p-value of test i is uniformly distributed since $\mathbb{P}(P_i < p_i) = \mathbb{P}(1 - F(|\mathcal{T}_i|) < p_i) = \mathbb{P}(|\mathcal{T}_i| > F^{-1}(1-p_i)) = 1 - \mathbb{P}(|\mathcal{T}_i| < F^{-1}(1-p_i)) = 1 - F(F^{-1}(1-p_i)) = p_i$.

[25] If we implement the $fFDR^+$ and FDR^+ to strictly control FDR at a target, say, $\tau = 10\%$ or $\tau = 20\%$, both result in empty portfolios for many years. With BSW's $FDR\tau$ portfolios, the problem is solved. In BSW's study, for the $FDR10\%$ portfolio, the empirical $\widehat{FDR}^+_{\gamma^*}$ is always greater than 10% with an average of 41.5%. For our data, among the thirty eight times of portfolio construction, with target $\tau = 20\%$ (10%) the $\widehat{FDR}^+_{\gamma^*}$ is less than τ on eight (zero) occasions and greater than one on five occasions for both targets.

[26] We could have set $\tau^* = \widehat{FDR}^+_{\gamma^*}$ for both cases. However, it seems fairer to set $\tau^* = \tau$ if $\widehat{FDR}^+_{\gamma^*} \leq \tau$ since both portfolios initially aim to control FDR at τ.

the previous five consecutive years as in-sample. To be eligible for this, a fund needs to have 60 observations in the in-sample. We start constructing our portfolios from December 1981.[27]

6.3. Performance comparison

In this section, we assess the portfolios' performance based on their alphas. We demonstrate the advantage of the $fFDR^+$ in picking out-performing funds and the efficient use of the covariates' information. We estimate the alpha evolution and the average alphas of our $fFDR\tau$ portfolios based on the nine covariates and compare with those of the $FDR\tau$ portfolio. We also explore the performance of $fFDR\tau$ portfolios after linearly combining the nine covariates and using their first principal component, an ordinary least squares regression, a least absolute shrinkage and selection operator, a ridge regression and an elastic net.[28]

We focus on portfolios with small FDR targets of $\tau = 10\%$. We repeat all estimations with $\tau = 20\%$ in Section IE of the Internet Appendix. Our results remain unchanged for all exercises.

6.3.1. The alpha evolution

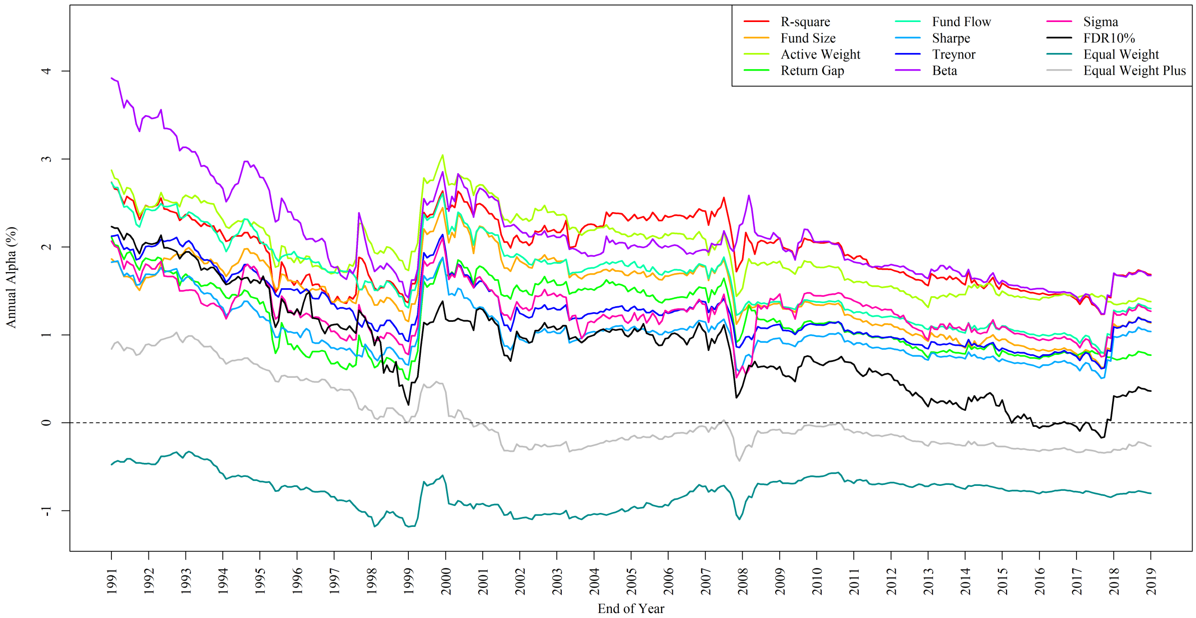
For each portfolio, we obtain its alpha evolution by calculating the Carhart four-factor alpha using its returns from January 1982 up to the end of each month from December 1991 onwards. In addition to the aforementioned portfolios, we construct two naive benchmark equally weighted portfolios, without control for the FDR: one that simply includes all the mutual funds in the in-sample window to be invested in the following year; and, another that contains only those with positive estimated alphas. We name these two portfolios Equal Weight and Equal Weight Plus.

We present all the alpha evolution in Figure 5. It is obvious from it that the $FDR10\%$ portfolio gains higher alphas than the equally weighted portfolio and all the $fFDR10\%$ portfolios outperform the $FDR10\%$. Ultimately, at the end of 2019, the $fFDR10\%$ portfolios with the R-square and Beta covariates are found to be the best with annualized alphas of about 1.7%, followed by the $fFDR10\%$ portfolios with the Active Weight, Fund Flow, Sigma, Treynor ratio, Fund Size, Sharpe ratio and Return Gap covariates achieving annualized alphas of at least 0.77%. By contrast, the $FDR10\%$, without the use of covariate information, winds up with a small positive alpha of 0.36%. It is noteworthy that all $fFDR10\%$ and the $FDR10\%$ portfolios seem to rebounce in terms of performance over the last two years of our sample.

[27] As Fama and French (2010) point out possible biases in the CRSP mutual fund data before 1984, we conduct a robustness check using a sample from 1984 to 2019; based on our results, presented in Section IC of the Internet Appendix, our conclusions remain unchanged.

[28] In Appendix B we provide a detailed comparison of all the $fFDR\tau$ portfolios in regard to several trading metrics, whereas in Section ID of the Internet Appendix the performance in terms of wealth evolution is presented.

Figure 5: Alpha evolution of $fFDR10\%$ and $FDR10\%$ portfolios over time. The graph presents the evolution of annualized alphas (in %) of the nine $fFDR10\%$ portfolios corresponding to the nine covariates, the portfolio $FDR10\%$ of BSW and the two equally weighted portfolios.



6.3.2. The average alpha

The alpha evolution in the previous section is calculated based on the portfolio returns from the start of 1982 up to a time point of interest. This may represent limited information in the case of investors with a different investment period of, say, five or ten years. For this, in Table 6, we report the average alpha that the investors will gain if they invest for $n \in \{5, 10, 15, 20, 30, 35, 38\}$ consecutive years: for each portfolio, we calculate its "n-year" alpha based on the portfolio returns over a period of $12n$ consecutive months, we repeat by shifting every time one month forward, and eventually present the average alpha. We report the $fFDR10\%$ for each covariate and the $FDR10\%$. We note that the last case, $n = 38$, corresponds to the alphas for the whole period from January 1982 to December 2019 and are the last points in the plots in Figure 5.

We find that the $fFDR10\%$ portfolios outperform the $FDR10\%$ for all considered covariates and for all n. Although these results should be interpreted with caution (some covariates were not well known in the literature at the start of our sample, such as the Active Weight and the Fund Size which were published in 2015 and 2017, respectively), they do indicate the stability of our approach for different investment horizons.

Table 6: Comparison of portfolios' performances for varying time lengths of investing. In this table, we consider 10 portfolios including nine $fFDR10\%$ portfolios corresponding to the nine covariates and the $FDR10\%$ portfolio of BSW. We compare the average alphas of the portfolios that are kept in periods of exactly n consecutive years. For example, consider $n = 5$. For each portfolio, we calculate the alpha for the first 5 years based on the portfolios' returns from January 1982 to December 1986. Then, we roll forward by a month and calculate the second alpha. The process is repeated and the last alpha is estimated based on the portfolios' returns from January 2015 to December 2019. The average of these alphas is presented in the first rows of the table.

	$fFDR10\%$									$FDR10\%$
n	R-square	Fund Size	Active Weight	Return Gap	Fund Flow	Sharpe	Treynor	Beta	Sigma	
5	1.49	0.87	1.24	0.56	0.92	0.57	0.73	1.09	1.19	0.12
10	1.48	0.85	1.18	0.51	0.93	0.65	0.76	1.2	1.06	0.05
15	1.7	0.94	1.4	0.72	1.06	0.79	0.88	1.2	1.09	0.14
20	1.84	1.05	1.59	0.91	1.15	0.91	0.96	1.31	1.17	0.26
25	1.61	0.9	1.36	0.67	0.99	0.8	0.86	1.24	1.09	0.13
30	1.41	0.78	1.23	0.54	0.95	0.78	0.86	1.2	1.01	0.01
38	1.69	1.14	1.38	0.77	1.3	1.04	1.15	1.67	1.27	0.36

6.3.3. Sub-period performance

In the alpha evolution in Figure 5, we note that the performance of our portfolios varies over time. By construction, this figure contain returns which start from January 1982 and are not representative of the recent mutual fund performance. In order to investigate the contribution of the returns in different periods to the performance of the portfolios, we split the whole period into four non-overlapping sub-periods: 1982–1991 (P1), 1992–2001 (P2), 2002–2011 (P3) and 2012–2019 (P4). We repeat the exercise for each sub-period and present in Table 7 the average 5-year alpha and alpha of portfolios (with a FDR target $\tau = 10\%$) in the sub-period.

Table 7: Performance of portfolios in sub-periods. The table displays the performance of the nine $fFDR10\%$ portfolios corresponding to the nine covariates, the $FDR10\%$ and equally weighted portfolios in sub-periods (P1: 1982–1991, P2: 1992–2001, P3: 2002–2011 and P4: 2012–2019) in terms of the average 5-year alpha (annualized, in %), the annualized alpha (in %) of the whole sub-period, the corresponding t-statistic (with use of Newey–West heteroskedasticity and autocorrelation-consistent standard error) and the annual Sharpe ratio.

	Average 5-year alpha				Whole sub-period alpha				Whole sub-period t-statistic				Annual Sharpe Ratio			
Portfolio	P1	P2	P3	P4	P1	P2	P3	P4	P1	P2	P3	P4	P1	P2	P3	P4
R-square	3.18	2.37	1.29	1.97	2.74	2.74	2.14	3.21	2.81	1.64	0.71	1.59	0.65	0.7	0.26	1.37
Fund Size	2.01	1.74	0.23	1.44	1.86	2.27	0.53	3.07	2.18	1.18	-0.29	1.56	0.62	0.61	0.2	1.37
Active Weight	3.01	3.1	-0.48	1.19	2.87	3.11	-0.01	0.56	2.47	1.85	-0.54	0.53	0.65	0.74	0.19	1.17
Return Gap	2.29	0.91	-0.43	0.55	2.11	1.78	0.17	0.09	2.3	1.04	-0.3	0.09	0.6	0.61	0.2	1.12
Fund Flow	2.65	0.73	0.06	1.82	2.73	1.32	0.54	3.44	2.22	0.74	-0.06	1.77	0.66	0.62	0.22	1.42
Sharpe	1.45	0.7	0.57	1.11	1.83	0.87	0.94	2.99	1.97	0.59	0.25	1.46	0.64	0.72	0.25	1.37
Treynor	1.77	0.73	0.62	1.37	2.12	0.98	0.93	3.19	2.03	0.63	0.19	1.61	0.64	0.69	0.24	1.38
Beta	3.52	0.72	0.45	2.02	3.92	1.58	1.33	3.65	2.15	0.64	0.06	1.94	0.65	0.45	0.21	1.43
Sigma	2.19	1.66	1.6	2.36	2.07	1.66	2.03	3.63	1.88	0.91	0.84	1.93	0.59	0.64	0.29	1.38
$FDR10\%$	2.7	0.6	-0.47	-0.35	2.23	1.2	0.09	1.63	2.01	0.83	-0.33	0.69	0.6	0.65	0.19	1.09
Equal Weight	-0.45	-1.65	0.29	-1.56	-0.48	-1.28	0.2	-1.34	-1.11	-1.53	-0.36	-2.65	0.48	0.54	0.23	1.01
Equal Weight Plus	0.76	-0.96	0.26	-0.65	0.84	-1.01	0.4	-0.38	1.17	-1.12	-0.36	-0.62	0.55	0.54	0.21	1.11

In terms of alphas and average 5-year alphas, it is clear that all the portfolios perform well in the first two sub-periods before suffering a decline in the third sub-period. On P3, we observe negative average 5-year alphas for the $FDR10\%$ portfolio and the $fFDR10\%$ portfolios with Active Weight and Return Gap covariates. On the last sub-period, this decrease continues for $FDR10\%$, whilst all of the $fFDR10\%$ portfolios witness rebounds. We note that all the $fFDR10\%$, except the ones with Return Gap and Active Weight covariates, achieve both positive alpha and average 5-year alpha in all the sub-periods. The t-statistic columns for the whole sub-period alpha, show that most portfolios have significantly positive alphas in the first sub-period. Interestingly, for the Sharpe ratio, we witness the highest reports in the last sub-period (which is also slightly shorter), whereas the lowest ones appear in the third sub-period which covers the global financial crisis of 2007–2008. From the realizations of the equally weighted portfolio, that is, the portfolio that selects all the eligible funds in the in-sample windows and invests them equally in the following year, we infer that the high Sharpe ratio in the final sub-period partially comes from the whole mutual fund market. The Equal Weight Plus portfolio, which invests in all funds with positive estimated alphas in the previous five years, is always better than the Equal Weight one. This simple screening portfolio even outperforms the $FDR10\%$ in the last two sub-periods. The alphas of the $fFDR10\%$ portfolios, by contrast, are nuanced depending on the covariate used; most of them beat the equally weighted one in all the sub-periods and for all the metrics (with notable exceptions of the Active Weight and Return Gap covariates in the third sub-period).

The implications of these results are as follows. First, we note that the R-square, Return Gap, Active Weight, Fund Flow and Fund Size retain their predictive abilities for mutual fund performance in recent years. From the five traditional covariates, the R-square, Fund Size and Fund Flow still have predictive abilities even after their respective publication dates.[29] Our results disagree partly with the findings of Jones and Mo (2021) who argue that published predictors are losing value in the recent period due to increases in arbitrage activities. Second, we note that our four new covariates contain valuable information on mutual funds' performance that in recent years can surpass the conventional covariates in some cases (see, for example, the performance of the $fFDR10\%$ portfolios in P4 with the Sigma and the Return Gap). Third, they further verify that our approach can resolve the identification issues in mutual funds due to noise/luck where other approaches (such as BSW) fail to.

[29] Appendix C shows that three out of the five covariates still gain significant alphas in the post-published period.

To further support the aforementioned argument on identification issues, we compare the performance of the portfolios formed in the $fFDR$ framework with a traditional sorting portfolio formation. If a covariate has a highly linear relation with the performance of mutual funds, then forming a portfolio based on sorting the funds on the covariate should be sufficient. We construct single- and double-sorting portfolios similarly to Kacperczyk *et al.* (2008) and Doshi *et al.* (2015), and Amihud and Goyenko (2013), respectively.[30]

The performance in terms of alpha of those portfolios from 1982 to 2019 is presented in Table 8. Our results show that most of the sorting portfolios, except the Active Weight and Sharpe ratio, have negative or negligible positive alphas at the end of 2019, which contrasts to the assumption of a linear relationship between the covariate and the funds' performance. Obviously, sorted portfolios perform better if they are based on the correct sign of the correlation between the underlying covariate and our funds' performance.

Table 8: Performance comparison of portfolios based on $fFDR$ and portfolios based on sorting on covariates (single-sorting) as well as based on both covariates and past alpha (double-sorting). The table shows the portfolios' annual Carhart four-factor alpha (in %) for the period January 1982 to December 2019. At the end of each year, for the single-sorting 10% portfolio, funds are sorted by the covariate. Depending on whether the relationship of the covariate and the fund performance is positive or negative, the funds in the top or bottom 10% are chosen to invest in the following year. For the double-sorting 10% portfolio, the funds chosen in the single-sorting 10% are ranked based on the past five-year alpha and then only 10% of the funds in the top are selected. *Note.* As documented in the literature, the R-square and Fund Size (Fund flow, Return Gap and Active Weight) have a negative (positive) effect on the mutual funds' performance. The single- and double-sorting portfolios constructed based on this assumption appear italicized.

Portfolio	R-square	Fund Size	Active Weight	Return Gap	Fund Flow	Sharpe	Treynor	Beta	Sigma
Panel A: Performance of $fFDR$10% and $fFDR$20% portfolios									
$fFDR$10%	1.69	1.14	1.38	0.77	1.30	1.04	1.15	1.67	1.27
$fFDR$20%	1.84	1.16	1.45	0.82	1.28	1.02	1.10	1.77	1.61
Panel B: Assuming a positive effect of the covariate on performance of the fund									
Single sort 10%	-1.07	-0.64	*-0.63*	*-1.46*	*-1.02*	0.13	-0.07	-2.11	-2.40
Double sort 10%	-1.03	0.03	*1.43*	*-0.40*	*0.33*	0.18	0.44	0.30	0.97
Single sort 20%	-1.17	-0.75	*-0.67*	*-1.15*	*-0.75*	-0.17	-0.28	-1.80	-1.69
Double sort 20%	-0.60	-0.18	*1.15*	*-0.07*	*0.11*	0.01	-0.10	-0.64	-0.53
Panel C: Assuming a negative effect of the covariate on performance of the fund									
Single sort 10%	*-0.89*	*-0.83*	-1.40	-1.45	-1.00	-1.96	-2.28	0.49	-0.50
Double sort 10%	*-1.72*	*0.30*	-1.39	-0.37	0.31	1.86	0.80	0.18	0.47
Single sort 20%	*-0.86*	*-1.01*	-1.14	-1.34	-1.04	-1.49	-1.49	0.21	-0.67
Double sort 20%	*-0.34*	*0.25*	-1.20	0.04	-0.01	0.47	0.16	0.19	-0.03

The portfolios based on $fFDR$ gain significant positive alphas and beat the corresponding sorted portfolios. These results further validate the advantage of our method in exploiting the non-linear relationship of the covariates, luck and funds' performance. The inability of the traditional sorted portfolios, that dominate the related literature, to reflect the predictive value of the covariates under study is thus noteworthy.

In Section IG of the Internet Appendix, we implement an exercise to combine the covariates

[30] For further details on the construction of these portfolios we refer the reader to Appendix D.

to a new one via linear regression and shrinkage method. We see that these simple linear combinations of the covariates does not improve the performance of the $fFDR$ based portfolios. This result further supports the assumption of the non-linear relationship between the considered covariates and the performance of mutual funds. As further robustness checks, in Section IH of the Internet Appendix, we demonstrate that our findings are robust with respect to a data subset where we require a minimum of $15 million in TNA for a fund to be considered. In Section IH of the Internet Appendix, we construct a similar set of portfolios, namely $fFDR^{-}\tau$, that aim to select the under-performing funds. We see that these portfolios successfully pick the unprofitable funds and are consistently beaten by the equally weighted portfolios.

7. Concluding Discussion

In this paper, we introduce the $fFDR^{+}$, a novel multiple hypothesis testing framework, that incorporates informative covariates to raise the power of detecting outperformers, and apply it to mutual fund investing. First, we conduct simulation experiments to assess how well our method performs in controlling FDR and raising power compared to the FDR^{+} method of BSW. We then construct empirical portfolios based on our new method and nine covariates. We study five covariates, which, based on earlier contributions, convey information about mutual funds' performance and propose four new ones based on asset pricing models. We show how the admixture of control for FDR and incorporated covariates advances the generation of more positive and higher alphas than a portfolio that controls FDR only or a portfolio based on sorting on the covariate and the past funds' performance.

The implications of our study are both methodological and empirical. The methodological literature in the field of selecting out-performing mutual funds is rich and expanding. In addition to the influential and well-cited study of BSW, other notable contributions are due to Kosowski *et al.* (2006), Andrikogiannopoulou and Papakonstantinou (2016), Harvey and Liu (2020) and Grønborg *et al.* (2021). All these have their merits and the authors present several promising empirical findings. In our study we focus on the FDR, whilst we defer an examination of their power relative to ours to future research. Nevertheless, we ought to note three main distinguishing features of our method. First, it allows the use of more data in the form of informative covariates, whilst the vast majority of others are limited to funds' past returns and their cross dependencies. Second, it is simple to implement and computationally less intensive than some of the most recent ones (e.g., the double bootstrap of Harvey and Liu, 2020). Third, our work can be extended to other problems in which statistical power weighs more than conservatism (i.e., the FDR threshold is higher), such as in the selection of hedge funds and bond funds or the assessment of trading strategies.

The empirical implications of our study are also of interest to academics and practitioners. We demonstrate that the five traditional mutual fund covariates can offer substantial profits in more recent periods. However, the relationship between these covariates, luck and funds' performance is non-linear. To fully exploit them, one should rely on powerful methods that control luck and noise. Our method ensures that. We also introduce four new covariates and find that their performance in our context is strong and surpasses that of traditional covariates; a finding that is expected to be of interest to investment managers who are constantly looking for valuable covariates in portfolio selection.

As with any methodological approach, there are caveats with our $fFDR$ procedure. In particular, this requires large datasets and gains higher power as the FDR threshold increases (see Sections 5.3 and 5.4). This implies that our approach should not be applied in problems which require a small FDR target (i.e., when the risk of a false discovery can lead to disastrous outcomes). As in our context of mutual funds' performance, it is difficult to explore covariates that seem promising (see, for example, the list of covariates studied in Jones and Mo, 2021) but with limited data availability.

We aspire that the $fFDR$ and $fFDR^+$ methods will become essential tools for people confronted by multiple competing factors, funds or models. The fields of finance and economics are extending towards big datasets and the literature is filled with predictors that may have value in economic variables of interest. Our approach can contribute to the evaluation of all these predictors and be a valuable arrow in the quiver of both academics and practitioners.

References

Amihud, Y. and Goyenko, R. (2013) Mutual fund's R^2 as predictor of performance. *The Review of Financial Studies*, **26**, 667–694.

Andrikogiannopoulou, A. and Papakonstantinou, F. (2016) Estimating mutual fund skill: A new approach. Working paper.

Andrikogiannopoulou, A. and Papakonstantinou, F. (2019) Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power? *The Journal of Finance*, **74**, 2667–2688.

Barbaglia, L., Consoli, S. and Manzan, S. (2022) Forecasting with economic news. *Journal of Business & Economic Statistics*, **0**, 1–12.

Barras, L., Scaillet, O. and Wermers, R. (2010) False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance*, **65**, 179–216.

Barras, L., Scaillet, O. and Wermers, R. (2020) Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power? A reply. *The Journal of Finance*. URL https://doi.org/10.37214/jofweb.2. Replications and Corrigenda (web-only).

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 289–300.

Berk, J. B. and Green, R. C. (2004) Mutual fund flows and performance in rational markets. *Journal of Political Economy*, **112**, 1269–1295.

Bris, A., Gulen, H., Kadiyala, P. and Rau, P. R. (2007) Good stewards, cheap talkers, or family men? The impact of mutual fund closures on fund managers, flows, fees, and performance. *The Review of Financial Studies*, **20**, 953–982.

Capponi, A., Glasserman, P. and Weber, M. (2020) Swing pricing for mutual funds: Breaking the feedback loop between fire sales and fund redemptions. *Management Science*, **66**, 3581–3602.

Carhart, M. M. (1997) On persistence in mutual fund performance. *The Journal of Finance*, **52**, 57–82.

Chen, J., Hong, H., Huang, M. and Kubik, J. D. (2004) Does fund size erode mutual fund performance? the role of liquidity and organization. *The American Economic Review*, **94**, 1276–1302.

Chen, X., Robinson, D. G. and Storey, J. D. (2021) The functional false discovery rate with applications to genomics. *Biostatistics*, **22**, 68–81.

Clifford, C. P., Fulkerson, J. A., Jame, R. and Jordan, B. D. (2021) Salience and mutual fund investor demand for idiosyncratic volatility. *Management Science*, **67**, 5234–5254.

Doshi, H., Elkamhi, R. and Simutin, M. (2015) Managerial activeness and mutual fund performance. *The Review of Asset Pricing Studies*, **5**, 156–184.

Fama, E. F. and French, K. R. (2010) Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, **65**, 1915–1947.

Fan, Y. and Fan, J. (2011) Testing and detecting jumps based on a discretely observed process. *Journal of Econometrics*, **164**, 331–344.

Geenens, G. (2014) Probit transformation for nonparametric kernel estimation on the unit interval. *Journal of the American Statistical Association*, **109**, 346–358.

Grønborg, N. S., Lunde, A., Timmermann, A. and Wermers, R. (2021) Picking funds with confidence. *Journal of Financial Economics*, **139**, 1–28.

Hansen, P. R. (2005) A test for superior predictive ability. *Journal of Business & Economic Statistics*, **23**, 365–380.

Harvey, C. R. and Liu, Y. (2017) Decreasing returns to scale, fund flows, and performance. Working paper.

Harvey, C. R. and Liu, Y. (2018) Detecting repeatable performance. *The Review of Financial Studies*, **31**, 2499–2552.

Harvey, C. R. and Liu, Y. (2020) False (and missed) discoveries in financial economics. *The Journal of Finance*, **75**, 2503–2553.

Jones, C. S. and Mo, H. (2021) Out-of-sample performance of mutual fund predictors. *The Review of Financial Studies*, **34**, 149–193.

Kacperczyk, M., Sialm, C. and Zheng, L. (2008) Unobserved actions of mutual funds. *The Review of Financial Studies*, **21**, 2379–2416.

Kosowski, R., Timmermann, A., Wermers, R. and White, H. (2006) Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis. *The Journal of Finance*, **61**, 2551–595.

Lan, W. and Du, L. (2019) A factor-adjusted multiple testing procedure with application to mutual fund selection. *Journal of Business & Economic Statistics*, **37**, 147–157.

Lan, W., Zhong, P.-S., Li, R., Wang, H. and Tsai, C.-L. (2016) Testing a single regression coefficient in high dimensional linear models. *Journal of Econometrics*, **195**, 154–168.

McLemore, P. (2019) Do mutual funds have decreasing returns to scale? evidence from fund mergers. *Journal of Financial and Quantitative Analysis*, **54**, 1683–1711.

Newey, W. K. and West, K. D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.

Pastor, L., Stambaugh, R. F. and Taylor, L. A. (2015) Scale and skill in active management. *Journal of Financial Economics*, **116**, 23–45.

Romano, J. P. and Wolf, M. (2005) Stepwise multiple testing as formalized data snooping. *Econometrica*, **73**, 1237–1282.

Sirri, E. R. and Tufano, P. (1998) Costly search and mutual fund flows. *The Journal of Finance*, **53**, 1589–1622.

Storey, J. D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **64**, 479–498.

Storey, J. D. (2003) The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics*, **31**, 2013–2035.

Storey, J. D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, **100**, 9440–9445.

Sullivan, R., Timmermann, A. and White, H. (1999) Data-snooping, technical trading rule performance, and the bootstrap. *The Journal of Finance*, **54**, 1647–1691.

Sullivan, R., Timmermann, A. and White, H. (2001) Dangers of data mining: The case of calendar effects in stock returns. *Journal of Econometrics*, **105**, 249–286.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267–288.

Wermers, R. (1999) Mutual fund herding and the impact on stock prices. *The Journal of Finance*, **54**, 581–622.

White, H. (2000) A reality check for data snooping. *Econometrica*, **68**, 1097–1126.

Yan, X. S. (2008) Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis*, **43**, 741–767.

Zheng, L. (1999) Is money smart? A study of mutual fund investors' fund selection ability. *The Journal of Finance*, **54**, 901–933.

Zhu, M. (2018) Informative fund size, managerial skill, and investor rationality. *Journal of Financial Economics*, **130**, 114–134.

Zou, H., and Hastie, T. (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **67**, 301–320.

Appendix A. Estimating $\pi_0(z)$ and $f(p,z)$

Let $\{(p_i, z_i)\}_{i=1}^m$ be the collection of p-value and covariate realizations of the different funds under consideration, with $\{z_i\}_{i=1}^m$ transformed in uniform distribution $[0,1]$ (see Section 2.1). We create fund bins $\{K_b\}_{b=1}^n$, where K_b contains a fund i if $z_i \in ((b-1)/n, b/n]$ and for each bin K_b we estimate a common $\pi_0(z)$ for all the funds i in the bin. For some common $\lambda \in (0,1)$, we estimate the $\pi_0(z)$ in each bin b by

$$\hat{\pi}_{0,b}(\lambda) = \frac{\#\{p_i > \lambda, z_i \in K_b\}}{(1-\lambda)\#K_b}, \quad b = 1, 2, \ldots, n. \tag{A.1}$$

We determine λ by minimizing the mean integrated square error (MISE):

$$\text{MISE}(\lambda) = \text{bias}^2 + \text{variance} = \left(\int_0^1 \phi(z,\lambda)dz - \pi_0\right)^2 + \int_0^1 [\hat{\pi}_0(z,\lambda) - \phi(z,\lambda)]^2 dz \tag{A.2}$$

We estimate π_0 using the smoothing spline method of Storey and Tibshirani (2003, Remark B).[31] Similarly to CRS, we calculate $\hat{\pi}_0(z_i,\lambda) = \hat{\pi}_{0,b}(\lambda)$ for each grid value $\lambda \in \Lambda = \{0.4, 0.5, \ldots, 0.9\}$, $i = 1, \ldots, m$ and $b = 1, 2, \ldots, n$, the $\hat{\pi}_0(z_i,\lambda)$ and, subsequently, $\int_0^1 \hat{\pi}_0(z,\lambda)dz = \sum_{i=1}^m \hat{\pi}_0(z_i,\lambda)/m$. The unknown $\phi(z,\lambda)$ is estimated by $\hat{\phi}(\lambda,z) = \hat{\pi}_0(z,\Lambda_{\min}) - c_\lambda(1-\hat{\pi}_0(z,\Lambda_{\min}))$, where c_λ is chosen such that $\int_0^1 \hat{\phi}(\lambda,z)dz = \int_0^1 \hat{\pi}_0(\lambda,z)dz$. We then obtain the optimal $\lambda^* = \arg\min_\lambda \text{MISE}(\lambda)$.

To estimate the joint density function $f(p,z)$, CRS use a local likelihood kernel density estimation (KDE) method with a probit transformation (Geenens, 2014). Specifically, let $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-x^2/2}dx$ and Φ^{-1} its inverse. Using $z_i' = \Phi^{-1}(z_i)$ and $p_i' = \Phi^{-1}(p_i)$, we obtain a "pseudo-sample" $\{(p_i', z_i')\}_{i=1}^n$, i.e., we transform the variables (p,z) to (p',z'); we denote by $\tilde{f}(p',z')$ the joint density function of (p',z'), which CRS estimate using the local likelihood KDE method. The bandwidth of the KDE is chosen locally via a k-Nearest-Neighbor approach using generalized cross-validation; this step can be implemented easily via the freely available R package `locfit`. The desired density function is then estimated as $\hat{f}(p,z) = \frac{\tilde{f}(p',z')}{\phi(p')\phi(z')}$ where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

Additionally, $f(p,z)$ may be non-increasing in p for each fixed z. CRS implement one more step which modifies the $\hat{f}(p,z)$ so that monotonicity is ensured. In our simulations, we use all the aforementioned techniques. In the empirical part, the monotonicity is switched off as this property is unknown in our data. For more details, readers are referred to CRS and their R package `fFDR`, Geenens (2014) as well as to the references therein.

[31] On rare occasions when the sample size m is small, the smoothing spline method does not work adequately. In these cases, we use the bootstrap method of Barras *et al.* (2010, Appendix A.1).

Appendix B. A comparison of portfolios' trading metrics

Next, we evaluate our portfolios in regard to a set of trading metrics, including the annualized estimated alpha $\hat{\alpha}$ of the Carhart four-factor model, its bootstrap p-value and t-statistic (with use of heteroskedasticity and autocorrelation-consistent standard error), the annual standard deviation of the four-factor model residuals ($\hat{\sigma}_\varepsilon$), the geometric mean return in excess of the one-month T-bill rate, the annual Sharpe ratio and the annual Information Ratio $\hat{\alpha}/\hat{\sigma}_\varepsilon$. All metrics are presented in Table B.9. We find that the $fFDR10\%$ portfolio based on the R-square covariate is the best for all considered metrics.

Table B.9: Comparison of performance statistics of all considered portfolios with $\tau = 10\%$. The table compares the portfolios with regard to metrics including the annual Carhart four-factor alpha ($\hat{\alpha}$, in %) with its bootstrap p-value and t-statistic (with use of Newey–West heteroskedasticity and autocorrelation-consistent standard error), the annual standard deviation of the four-factor model residuals ($\hat{\sigma}_\varepsilon$, in %), the mean return in excess of the one-month T-bill rate (in %), the annual Sharpe ratio and the annual Information Ratio ($IR = \hat{\alpha}/\hat{\sigma}_\varepsilon$).

Covariate	$\hat{\alpha}$ (p-value)	t-statistic	$\hat{\sigma}_\varepsilon$	Mean Return	Sharpe Ratio	IR
R-square	1.69 (0.06)	1.85	4.42	7.93	0.61	0.38
Fund Size	1.14 (0.2)	1.32	4.02	7.34	0.56	0.28
Active Weight	1.38 (0.1)	1.72	3.79	8	0.6	0.36
Return Gap	0.77 (0.34)	0.99	3.81	7.38	0.55	0.2
Fund flow	1.3 (0.14)	1.56	3.78	7.75	0.6	0.34
Sharpe	1.04 (0.2)	1.33	3.37	7.77	0.62	0.31
Treynor	1.15 (0.15)	1.45	3.49	7.65	0.6	0.33
Beta	1.67 (0.07)	1.78	4.92	7.28	0.55	0.34
Sigma	1.27 (0.26)	1.16	5.01	7.69	0.57	0.25
$FDR10\%$	0.36 (0.72)	0.37	4.75	6.5	0.52	0.08
Equal Weight	-0.8 (0.03)	-2	1.86	6.3	0.5	-0.43
Equal Weight Plus	-0.26 (0.48)	-0.56	2.18	6.7	0.52	-0.12

Appendix C. Performance of $fFDR10\%$ in various periods

In this section, we present the alpha of the $fFDR10\%$ portfolios in periods before and after the covariates were published. The first line of Table C.10 shows that all covariates gain positive alpha for the period January 1982 to the end of the prior-published year. The last line of the table indicates that three of the five previously known covariates still gain significant alpha in the post-published period.

Table C.10: Performance of $fFDR10\%$ portfolios in various periods prior- and post-published year of the covariates. The table shows the annualized alpha (in %) of the $fFDR10\%$ portfolio corresponding to each covariate in specific periods, with $[a, b]$ denoting a period extending from the beginning of year a over to the end of year b. For instance, the first value in the R-square column, that is, 1.75, is the alpha of the $fFDR10\%$ with R-square covariate for the period from the beginning of 1982 to the end of 2012 (i.e., $n - 1 = 2012$, where $n = 2013$ is the published year of the covariate). The middle value in the column is the Carhart four-factor alpha of the portfolio for year n, which contains only 12 months corresponding to 12 data points of returns.

	$fFDR10\%$				
Period	R-square $n = 2013$	Fund Size $n = 2017$	Active Weight $n = 2015$	Return Gap $n = 2008$	Fund flow $n = 1999$
$[1982, n-1]$	1.75	0.82	1.46	1.53	1.6
$[n-10, n-1]$	1.20	-2.04	-1.27	3.00	0.35
$[n-5, n-1]$	-2.11	0.11	-1.54	1.71	-0.65
$[n-4, n-1]$	-1.81	-0.12	-0.76	0.62	-1.01
$[n-3, n-1]$	-2.50	0.22	2.79	0.20	-1.44
$[n-2, n-1]$	-2.44	0.50	3.09	0.82	-1.83
$[n-1, n-1]$	-0.92	-2.1	8.00	-0.04	-0.87
$[n, n]$	-4.77	-2.39	-3.22	2.67	-0.40
$[n+1, n+1]$	4.27	1.33	-1.81	2.70	20.76
$[n+1, n+2]$	-0.21	5.45	-0.91	-0.95	6.85
$[n+1, n+3]$	1.45	-	-0.53	-1.73	4.33
$[n+1, n+4]$	1.82	-	-0.05	-0.81	2.36
$[n+1, n+5]$	3.03	-	-	-2.13	1.90
$[n+1, n+10]$	-	-	-	-0.59	2.47
$[n+1, 2019]$	3.73	5.45	-0.05	-0.3	1.31

Appendix D. The construction of sorting portfolios

Here, we describe the constructions of the single- and double-sorting portfolios which are traditionally conducted in the literature. Specifically, the single-sorting portfolios based on a covariate are as in Kacperczyk *et al.* (2008) and Doshi *et al.* (2015), and the double-sorting based on a covariate and the past alpha are as in Amihud and Goyenko (2013).

To construct the single-sorting portfolio for each covariate, at the end of each year from 1981, all the mutual funds are sorted into deciles (quintiles) according to the given covariate. For the covariate that has a negative/positive relationship with the performance of the funds, the funds in the bottom/top decile (quintile) are selected. These form a portfolio to be invested in the following year. To form the double-sorting portfolio, the funds selected in the single-sorting portfolio are again sorted into decile (quintile) according to the past alpha. The funds in the top decile (quintile) form the portfolio to be invested in the following year. This process is rolled forward until the end of the sample period.

For consistency with the $fFDR$ portfolios, we use the same type of 5-year rolling window, i.e., each time we use the aforementioned observed covariates and the alpha and covariates calculated based on the last five years.

Internet Appendix for
"Informative Covariates, False Discoveries and Mutual Funds Performance"

IA. Simulation execution

We summarize the simulation procedure as follows.

As a first step, we generate the covariate and alpha for each of the m funds. We generate the covariate vector $(z_1, z_2, \ldots, z_m)$ with each element drawn from the uniform distribution $[0, 1]$ and assign them to the funds. For the cases (15) or (16), we determine c in (18) such that $\int_0^1 \pi_0(z)dz = \pi_0$ for a given $\pi_0 > 0$. For each fund i, we draw h_i from the Bernoulli distribution with success probability $1 - \pi_0(z_i)$ and assign a zero alpha to fund i with $h_i = 0$. Finally, for the remaining funds, we draw true non-zero alphas from the given distribution (15) or (16) and assign them such that a fund with a smaller z has a smaller alpha. For the case (17), we draw alphas from the distribution and then assign them to the funds; again, a fund with a smaller z has a smaller alpha.

In the second step, we simulate the return factors from a normal distribution with parameters equal to their sample counterparts. The loadings of these factors are also drawn from a normal distribution with parameters equal to their sample counterparts obtained from the fund level estimation of equation (14). We consider balanced panel data for $2,000$ funds with 274 time-series observations; the number of $2,000$ is chosen to be close to our real sample of 2,224 funds, whereas the number of 274 periods is the median of our sample funds' observations. In unbalanced panel data, the number of observations for each fund is drawn randomly with replacement from the set of the number of observations of the funds in the real-data counterpart. Under cross-sectional independence, the noise term $\varepsilon_{i,t}$ is drawn from a normal distribution $\mathcal{N}(0, \sigma_\varepsilon^2)$, where, as in Barras *et al.* (2020), σ_ε is set equal to the median of its real-data counterpart, that is, approximately 0.0183 for our sample. Under cross-sectional dependence, we follow Barras *et al.* (2010) (BSW henceforth) and assume that all fund residuals load on a common latent factor G_t, whereas the out-performing and under-performing funds load on the specific factors G_t^+ and G_t^-, respectively. Thus,

$$\varepsilon_{i,t} = \gamma G_t + \gamma G_t^+ \mathbb{1}_{\alpha_i > 0} + \gamma G_t^- \mathbb{1}_{\alpha_i < 0} + \eta_{i,t}, \tag{A.1}$$

where $\mathbb{1}_{\alpha_i > 0}$ and $\mathbb{1}_{\alpha_i < 0}$ are, respectively, out-performing and under-performing indicators taking the value 1 if the fund i is out-performing or under-performing, and 0 otherwise. The three latent factors G_t, G_t^+ and G_t^- are assumed to be mutually orthogonal and to the four risk factors and have a normal distribution $\mathcal{N}(0, \sigma_G^2)$, where, from BSW, σ_G is set equal to the average of the monthly standard deviations of the three risk factors (size, book-to-market and momentum). The coefficient γ is set equal to the average of the loading of the three risk factors of the 2,224 funds in our sample. Finally, $\{\eta_{i,t}\}_i$ are uncorrelated and drawn from the normal distribution $\mathcal{N}(0, \sigma_\eta^2)$, where σ_η is chosen such that σ_ε is equated to the median of its real-data counterpart, as in the independent case.

In the last step, we implement the $fFDR^+$ and FDR^+ and compute their performance metrics. More specifically, based on the simulated data from the previous step, we calculate the Carhart four-factor model alpha and the corresponding p-value for each fund. We use the resulting p-value, the estimated alpha and the covariate as inputs to the $fFDR^+$ and FDR^+ procedures. At a given target of FDR, we calculate for each method a rate of falsely classified funds $\widetilde{FDR}^+$ and a detected rate $\widetilde{Power}^+$:

$$\widetilde{FDR}^+ = \frac{\widetilde{V}^+}{\max\left\{\widetilde{R}^+, 1\right\}} \quad \text{and} \quad \widetilde{Power}^+ = \frac{\widetilde{C}^+}{\widetilde{T}^+}, \tag{A.2}$$

where $\widetilde{R}^+$ is the number of classified out-performing funds and, among them, $\widetilde{V}^+$ funds are truly zero-alpha or under-performing funds. $\widetilde{T}^+$ is the number of truly out-performing funds in the population and, among them, $\widetilde{C}^+$ funds are classified correctly.

The previous three steps are repeated 1,000 times and we use the average $\widetilde{FDR}^+$ and $\widetilde{Power}^+$ as estimates for the actual FDR and power.

IB. Additional simulation results

To complement Section 5 of the main manuscript, we show here the performance of the $fFDR^+$ in terms of FDR control and power under several settings. We first present the performance of $fFDR^+$, where $\pi_0(z)$ can take three different forms. We then show the results corresponding to the balanced panel data under cross-sectional dependence. Next, we present results for unbalanced panel data under both cross-sectional independence and dependence. Finally, to cover all distributions studied in the literature, we exhibit simulation results for the case where alphas are drawn from a single normal distribution.

IB.1. Results for alternative forms of $\pi_0(z)$

In this section, we consider three forms of $\pi_0(z)$, including decreasing, increasing and being constant with respect to z. For the decreasing and increasing cases, we specify $\pi_0(z)$ based on $f(z) = -1.5(z - 0.5)^3 + c$ or $f(z) = 1.5(z - 0.5)^3 + c$. In the interest of space, we present results for the mass distribution of alphas with balanced panel data which is generated under cross-sectional independence. For all forms of $\pi_0(z)$, even when this is constant, we conclude similarly to the case of $\pi_0(z)$ with an up-and-down shape presented in the main manuscript.

Results for other distributions as well as under cross-sectional dependence convey the same message and are available upon request.

Table I: Power comparison (in %) for discrete distribution when $\pi_0(z)$ is an increasing function. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when the alphas of 2,000 funds are drawn from a discrete distribution: $\alpha \sim \pi^+\delta_{\alpha=\alpha^*} + \pi_0\delta_{\alpha=0} + \pi^-\delta_{\alpha=-\alpha^*}$ with varying α^* (annualized, in %) and proportions (π^+, π_0, π^-). The simulated data are a balanced panel with 274 observations per fund and are generated with cross-sectional independence.

(π^+, π_0, π^-)	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	0.20	3.60	18.10	40.60	62.50
	FDR^+	0.40	1.90	12.00	32.20	53.20
$(10, 60, 30)\%$	$fFDR^+$	0.70	8.00	27.90	52.20	72.60
	FDR^+	0.40	2.20	13.90	36.10	57.70
$(10, 30, 60)\%$	$fFDR^+$	3.10	20.10	48.60	73.30	88.20
	FDR^+	0.40	3.00	21.40	47.40	69.60
$(13, 67.5, 19.5)\%$	$fFDR^+$	0.50	7.20	26.20	49.90	71.00
	FDR^+	0.40	3.10	17.60	39.50	60.70
$(13, 48, 39)\%$	$fFDR^+$	1.60	13.60	39.00	64.90	82.50
	FDR^+	0.40	3.70	21.60	45.80	66.80
$(13, 9, 78)\%$	$fFDR^+$	6.30	33.20	65.10	88.90	97.10
	FDR^+	0.60	7.10	37.60	69.40	88.10

Table II: Power comparison (in %) for discrete distribution:when $\pi_0(z)$ is a decreasing function. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when the alphas of 2,000 funds are drawn from a discrete distribution: $\alpha \sim \pi^+\delta_{\alpha=\alpha^*} + \pi_0\delta_{\alpha=0} + \pi^-\delta_{\alpha=-\alpha^*}$ with varying α^* (annualized, in %) and proportions (π^+, π_0, π^-). The simulated data are a balanced panel with 274 observations per fund and are generated with cross-sectional independence.

(π^+, π_0, π^-)	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	1.20	10.40	31.00	55.00	74.50
	FDR^+	0.40	2.00	11.80	32.00	53.20
$(10, 60, 30)\%$	$fFDR^+$	2.70	16.60	43.10	68.00	84.20
	FDR^+	0.40	2.20	14.00	35.80	57.60
$(10, 30, 60)\%$	$fFDR^+$	6.50	29.20	67.70	89.30	96.10
	FDR^+	0.40	3.00	21.20	47.40	69.60
$(13, 67.5, 19.5)\%$	$fFDR^+$	2.00	14.10	37.50	61.20	78.70
	FDR^+	0.50	3.10	17.70	39.50	60.50
$(13, 48, 39)\%$	$fFDR^+$	4.40	24.50	54.50	77.20	89.90
	FDR^+	0.40	3.60	21.60	45.60	66.80
$(13, 9, 78)\%$	$fFDR^+$	9.40	39.20	79.20	93.20	97.90
	FDR^+	0.50	6.90	37.50	69.10	88.10

Table III: Power comparison (in %) for discrete distribution when $\pi_0(z)$ is a constant function. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when the alphas of 2,000 funds are drawn from a discrete distribution: $\alpha \sim \pi^+\delta_{\alpha=\alpha^*} + \pi_0\delta_{\alpha=0} + \pi^-\delta_{\alpha=-\alpha^*}$ with varying α^* (annualized, in %) and proportions (π^+, π_0, π^-). The simulated data are a balanced panel with 274 observations per fund and are generated with cross-sectional independence.

(π^+, π_0, π^-)	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	0.50	5.90	22.20	44.90	66.90
	FDR^+	0.40	2.00	12.00	31.80	53.60
$(10, 60, 30)\%$	$fFDR^+$	1.40	12.10	35.60	60.80	78.60
	FDR^+	0.40	2.20	13.80	35.90	57.40
$(10, 30, 60)\%$	$fFDR^+$	4.40	25.90	58.70	80.70	91.80
	FDR^+	0.40	3.30	21.30	47.90	69.50
$(13, 67.5, 19.5)\%$	$fFDR^+$	1.10	9.80	30.20	54.70	74.40
	FDR^+	0.50	3.20	17.60	39.80	60.80
$(13, 48, 39)\%$	$fFDR^+$	2.90	18.20	44.80	69.40	84.90
	FDR^+	0.50	3.60	21.00	46.00	66.70
$(13, 9, 78)\%$	$fFDR^+$	8.30	37.00	73.00	91.00	97.70
	FDR^+	0.60	7.10	37.50	69.40	88.20

IB.2. Results for balanced panel data under cross-sectional dependence

We start by presenting in Figures I–III the cases where the data are generated as balanced panels under cross-sectional dependent errors. The comparisons in terms of power between $fFDR^+$ and FDR^+ are shown in Tables IV–VIII.

Figure I: Performance of $fFDR^+$ for discrete distribution of α. The graphs show the performance of the $fFDR^+$ in terms of FDR control when alphas are drawn from a discrete distribution. The simulated data are balanced panels with cross-sectional dependence.

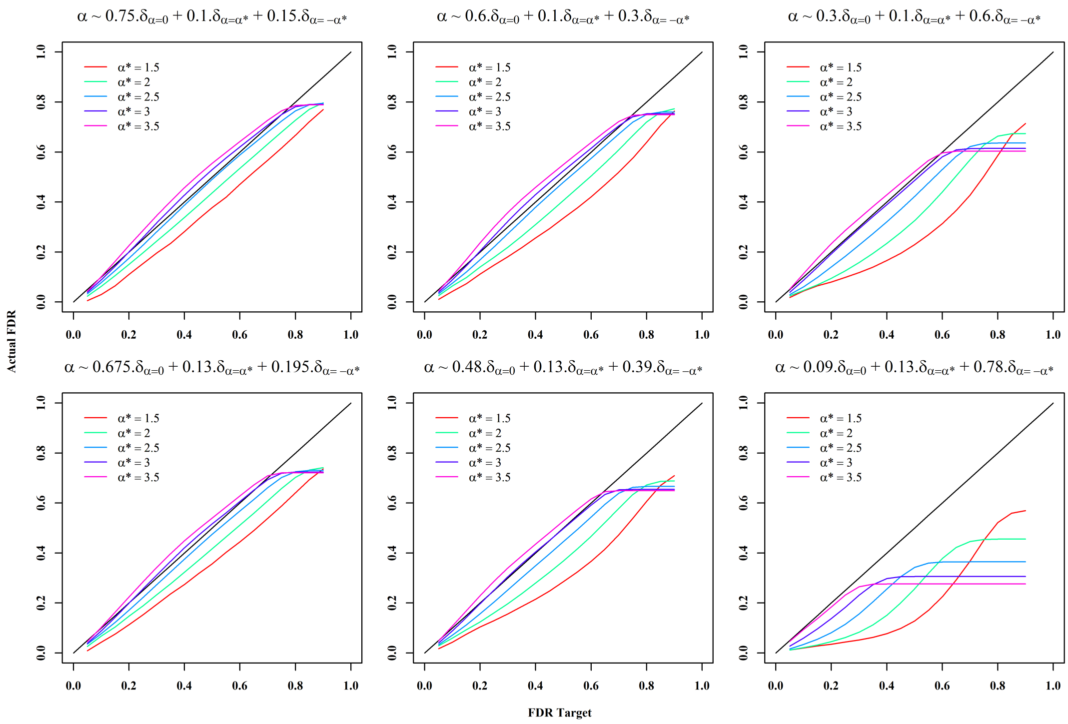


Figure II: Performance of $fFDR^{+}$ for discrete and normal distribution mixture of α. The graphs show the performance of the $fFDR^{+}$ in terms of FDR control when alphas are drawn from a mixture of discrete and normal distributions. The simulated data are balanced panels with cross-sectional dependence.

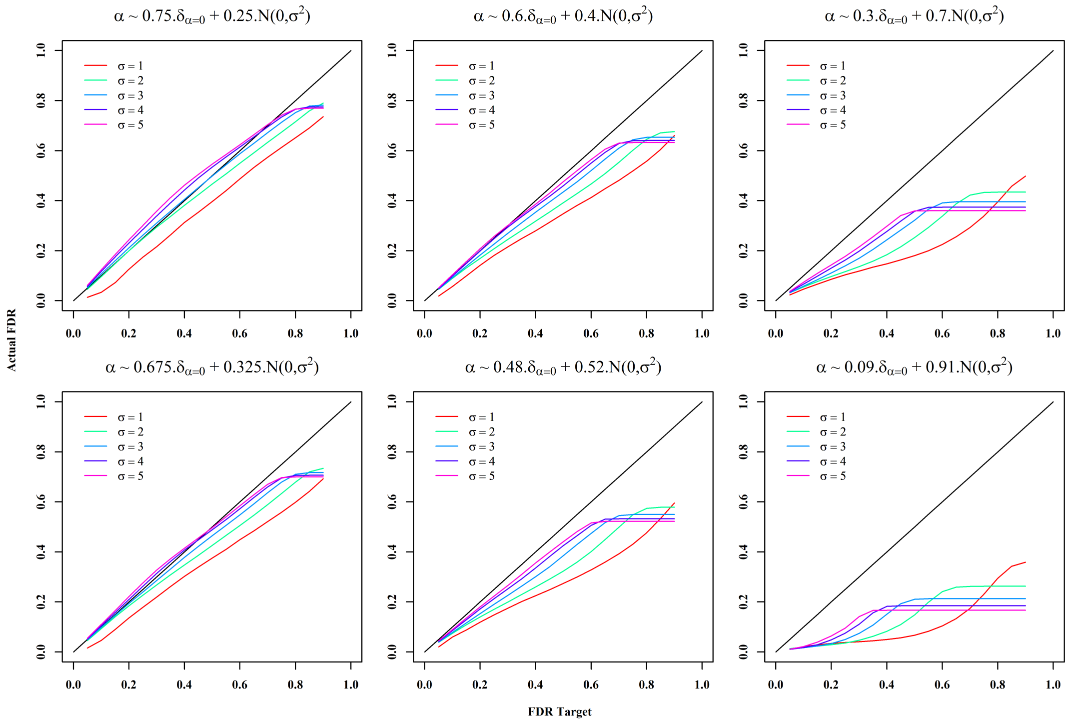


Figure III: Performance of $fFDR^{+}$ for continuous distribution of α. The graphs show the performance of the $fFDR^{+}$ in terms of FDR control when alphas are drawn from a continuous distribution which is a mixture of two normals. The simulated data are balanced panels with cross-sectional dependence.

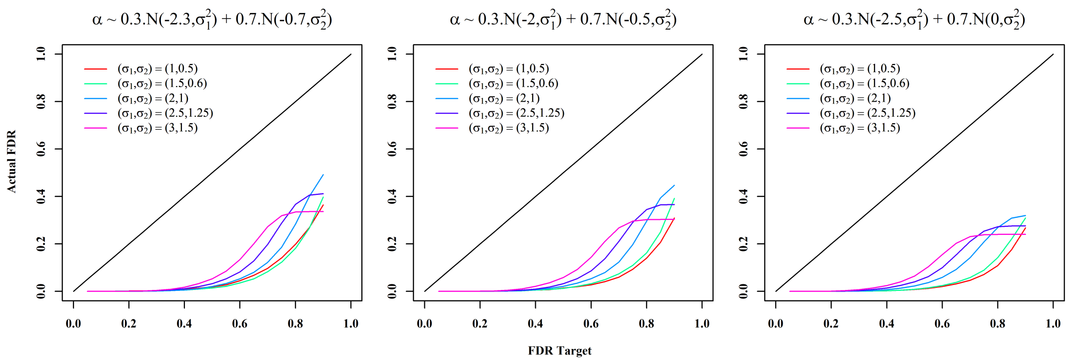


Table IV: Power comparison (in %) for discrete distribution. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when the alphas of 2,000 funds are drawn from a discrete distribution: $\alpha \sim \pi^+\delta_{\alpha=\alpha^*} + \pi_0\delta_{\alpha=0} + \pi^-\delta_{\alpha=-\alpha^*}$ with varying α^* (annualized, in %) and proportions (π^+, π_0, π^-). The simulated data are a balanced panel with 274 observations per fund and generated with cross-sectional dependence.

(π^+, π_0, π^-)	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	0.8	6.1	21.3	43.6	65.5
	FDR^+	0.5	2.6	12.1	30.5	51.9
$(10, 60, 30)\%$	$fFDR^+$	1.9	11.2	32.3	56.6	76
	FDR^+	0.5	3	14.1	34.3	56
$(10, 30, 60)\%$	$fFDR^+$	4.6	23.1	51.5	75.4	89.1
	FDR^+	0.5	4	20.7	46.6	68.8
$(13, 67.5, 19.5)\%$	$fFDR^+$	1.5	9.7	29	53.1	73.7
	FDR^+	0.7	4.1	17	38	59.2
$(13, 48, 39)\%$	$fFDR^+$	3.4	17.1	41.3	66.3	83.3
	FDR^+	0.6	4.6	20.7	44.4	65.3
$(13, 9, 78)\%$	$fFDR^+$	8.5	34.2	67.9	89	97.2
	FDR^+	0.7	8.2	37.1	69.1	87.9

Table V: Power comparison (in %) for discrete-normal distribution mixture. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when alphas of 2,000 funds are drawn from a discrete-normal distribution mixture: $\alpha \sim \pi_0\delta_{\alpha=0} + (1-\pi_0)\mathcal{N}(0, \sigma^2)$ with varying σ (annualized, in %) and null proportion π_0. The simulated data are a balanced panel with 274 observations per fund and generated with cross-sectional dependence.

π_0	Procedure	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
75%	$fFDR^+$	0.5	15.7	36.2	51.2	60.8
	FDR^+	0.2	8.4	26.6	41.7	52.3
60%	$fFDR^+$	1.6	21.5	42.8	57.1	66.3
	FDR^+	0.3	11.4	31.3	46.5	56.8
30%	$fFDR^+$	4.7	32.4	54.5	67.6	75
	FDR^+	0.6	17.9	40.8	55.8	65.4
67.5%	$fFDR^+$	1	18.7	39.4	54	63.3
	FDR^+	0.2	9.8	29	44	54.5
48%	$fFDR^+$	2.5	25.5	47.3	61.3	70.2
	FDR^+	0.3	13.5	34.6	49.8	59.9
9%	$fFDR^+$	6.7	38	60.7	73.6	80.7
	FDR^+	0.7	22	46.9	62.5	72

Table VI: Power comparison (in %) for mixture of two normal distributions. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when alphas of 2,000 funds are drawn from a mixture of two normal distributions: $\alpha \sim 0.3\mathcal{N}(\mu_1, \sigma_1^2) + 0.7\mathcal{N}(\mu_2, \sigma_2^2)$ with varying standard deviation pairs (σ_1, σ_2) and mean pairs (μ_1, μ_2) (both parameters' pairs are annualized and in %). The simulated data are a balanced panel with 274 observations per fund and generated with cross-sectional dependence.

		(σ_1, σ_2)				
(μ_1, μ_2)	Procedure	$(1, 0.5)$	$(1.5, 0.6)$	$(2, 1)$	$(2.5, 1.25)$	$(3, 1.5)$
		$\pi^+ = 6\%$	$\pi^+ = 10.4\%$	$\pi^+ = 20.7\%$	$\pi^+ = 25.5\%$	$\pi^+ = 29.1\%$
$(-2.3, -0.7)$	$fFDR^+$	0.1	0.4	5	13.6	23.3
	FDR^+	0	0	0.3	2.2	7.4
		$\pi^+ = 11.8\%$	$\pi^+ = 16.9\%$	$\pi^+ = 26.4\%$	$\pi^+ = 30.5\%$	$\pi^+ = 33.4\%$
$(-2, -0.5)$	$fFDR^+$	0.1	0.6	6.5	15.8	25.5
	FDR^+	0	0.1	0.5	3.2	9.1
		$\pi^+ = 35.2\%$	$\pi^+ = 36.4\%$	$\pi^+ = 38.2\%$	$\pi^+ = 39.8\%$	$\pi^+ = 41.1\%$
$(-2.5, 0)$	$fFDR^+$	0.4	1	9.2	18.6	28.3
	FDR^+	0	0.1	1	4.7	11.7

Table VII: Power comparison (in %) for varying sample size and observation length. The table compares the power of the $fFDR^+$ and FDR^+ in a balanced panel data with varying number of observations per fund (T) and number of funds (m). We present three cases where alphas of m funds are drawn from i) discrete distribution: $\alpha \sim 0.1\delta_{\alpha=2} + 0.3\delta_{\alpha=0} + 0.6\delta_{\alpha=-2}$ (Panel A); ii) discrete-normal mixture: $\alpha \sim 0.3\delta_{\alpha=0} + 0.7\mathcal{N}(0, 2^2)$ (Panel B); and mixture of two normal distributions: $\alpha \sim 0.3\mathcal{N}(-2, 2^2) + 0.7\mathcal{N}(-0.5, 1)$ (Panel C). For each alpha population, we generate data with cross-sectional dependence.

		Number of observations per fund					
m	Procedure	$T = 120$	$T = 180$	$T = 240$	$T = 300$	$T = 360$	$T = 420$
		Panel A: Discrete distribution					
500	$fFDR^+$	3.7	9.4	19.9	31	43.5	54.5
	FDR^+	0.7	1.4	3.2	6.2	12	18.9
1000	$fFDR^+$	2.2	8.3	17.1	29.8	40.4	52.9
	FDR^+	0.4	1.1	2.6	5.9	11.3	19.9
2000	$fFDR^+$	2.1	7.3	16.5	26.8	40.6	50.6
	FDR^+	0.2	0.9	2.5	5.5	11.9	19.9
3000	$fFDR^+$	1.9	7	16	27.8	39.5	48.9
	FDR^+	0.2	0.7	2.2	5.9	12.3	19.6
		Panel B: Mixture of Discrete and Normal distributions					
500	$fFDR^+$	13	22	29.2	35.8	40.6	45.5
	FDR^+	3	8.1	13.8	20	25.3	29.9
1000	$fFDR^+$	12.5	21.2	29.1	35.1	39.8	44.2
	FDR^+	2.9	8.2	14.6	20.3	24.9	29.6
2000	$fFDR^+$	12.1	20.9	28.4	34.9	39.4	44.3
	FDR^+	2.7	8.2	14.4	20.4	25	29.8
3000	$fFDR^+$	11.8	20.8	28.3	34.4	39.9	43.7
	FDR^+	2.7	8.3	14.4	20.1	25.6	29.6
		Panel C: Mixture of Normal distributions					
500	$fFDR^+$	1.7	3.5	6.4	8.2	11.2	14.2
	FDR^+	0.2	0.3	0.6	0.9	1.4	2
1000	$fFDR^+$	1.2	3.2	5.6	8.6	10.8	13.3
	FDR^+	0.1	0.2	0.4	0.9	1.2	1.9
2000	$fFDR^+$	1.1	2.8	4.9	7.6	10.1	12.8
	FDR^+	0.1	0.2	0.3	0.7	1.1	2
3000	$fFDR^+$	1.1	2.8	5	7.6	10.3	12.6
	FDR^+	0.1	0	0.3	0.6	1.2	1.9

Table VIII: Power comparison (in %) for varying FDR targets (in %) for sample with small size and small number of observations under cross-sectional dependence. In this table, we consider three distributions as in Table VII for samples consisting of $m = 500$ funds (balanced panels) with $T = 60$ observations per fund (5 years).

Distribution	Procedure	FDR target								
		10	20	30	40	50	60	70	80	90
Discrete	$fFDR^+$	0.8	3	7.3	13.6	21.9	31.4	41	51.3	63.3
	FDR^+	0.3	0.5	0.8	1	1.4	1.9	2.8	4	5.9
Mixture of discrete and normal	$fFDR^+$	3.1	8.5	15.4	23.5	32.3	41.4	50.8	60.9	67.2
	FDR^+	0.4	1.2	2.7	5.2	8.6	14.5	22.3	32.5	41.3
Mixture of normals	$fFDR^+$	0.4	1.8	4.3	8.1	13.4	29.8	27.7	37.6	50.7
	FDR^+	0.1	0.1	0.3	0.4	0.5	0.8	1.4	2.5	4.1

IB.3. Results for unbalanced panel data

In this section, we present the performance of the $fFDR^+$ under both cross-sectional independence and dependence. Figures IV–VI depict the FDR control of the $fFDR^+$, while the power comparisons are given in Tables IX–XI.

Figure IV: Performance of $fFDR^{+}$ in terms of FDR control when alphas are drawn from the discrete distribution with unbalanced panel data.

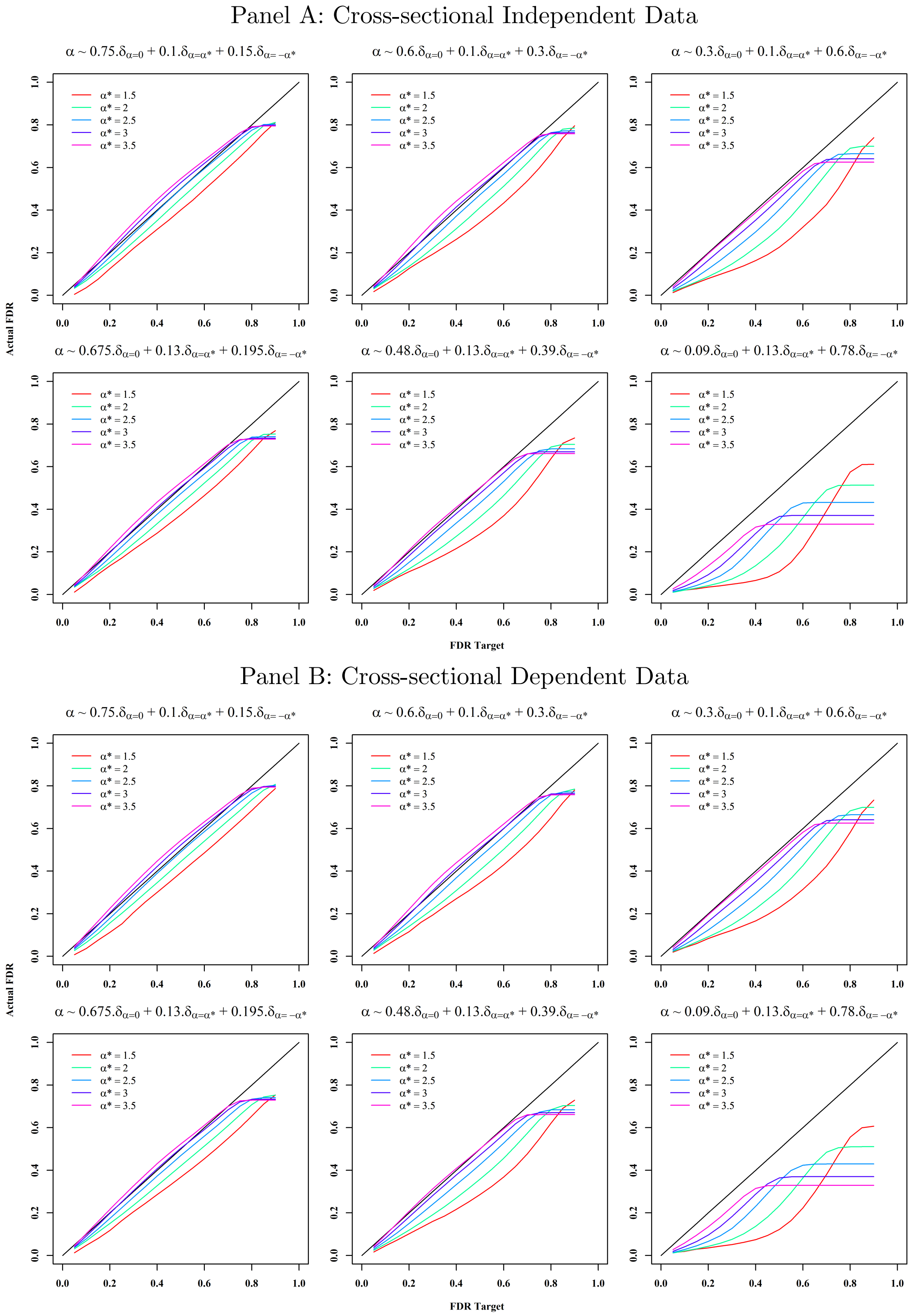


Figure V: Performance of $fFDR^{+}$ in terms of FDR control when alphas are drawn from the discrete-normal distribution mixture with unbalanced panel data.

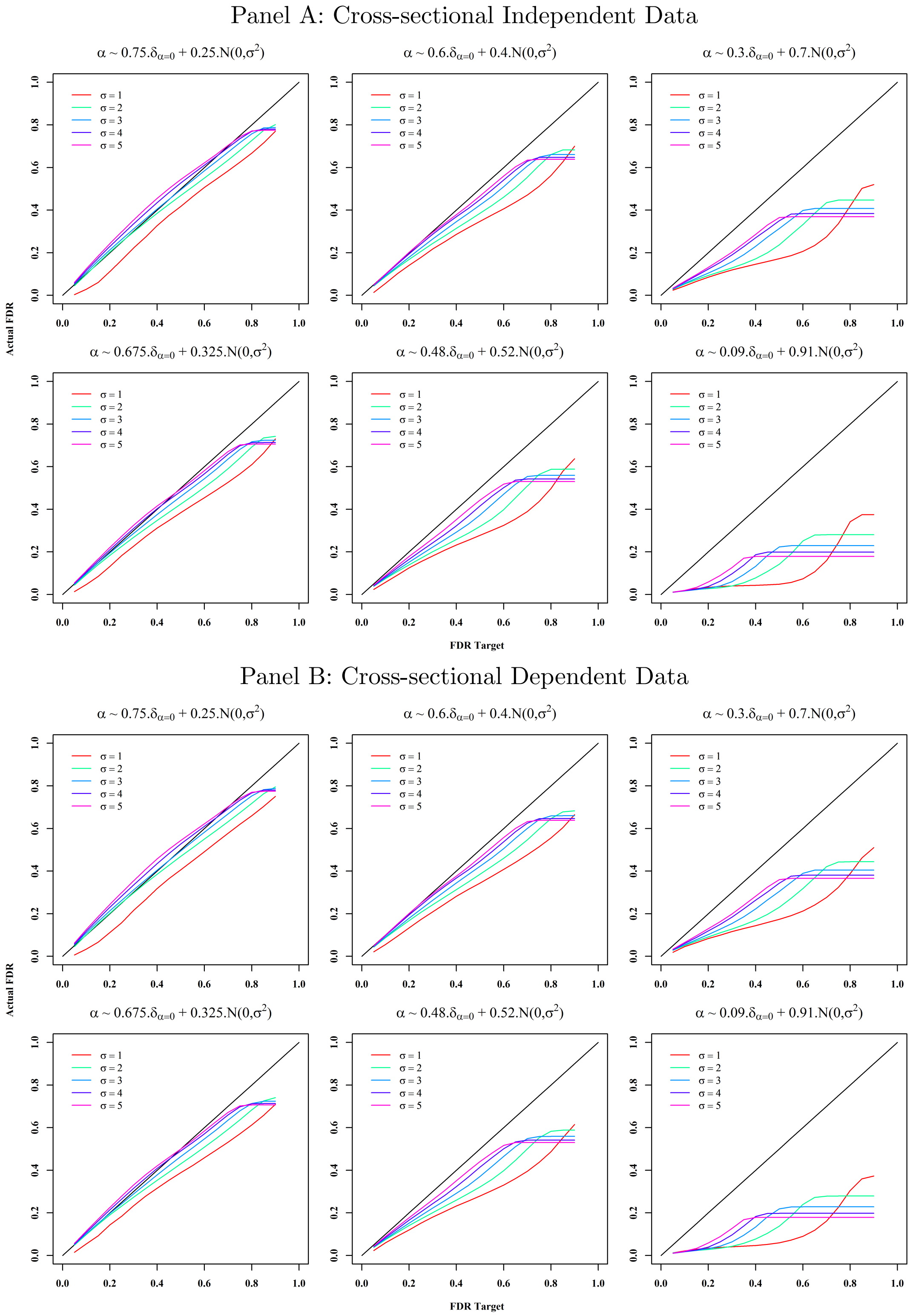


Figure VI: Performance of $fFDR^{+}$ in terms of FDR control when alphas are drawn from the mixture of two normals with unbalanced panel data.

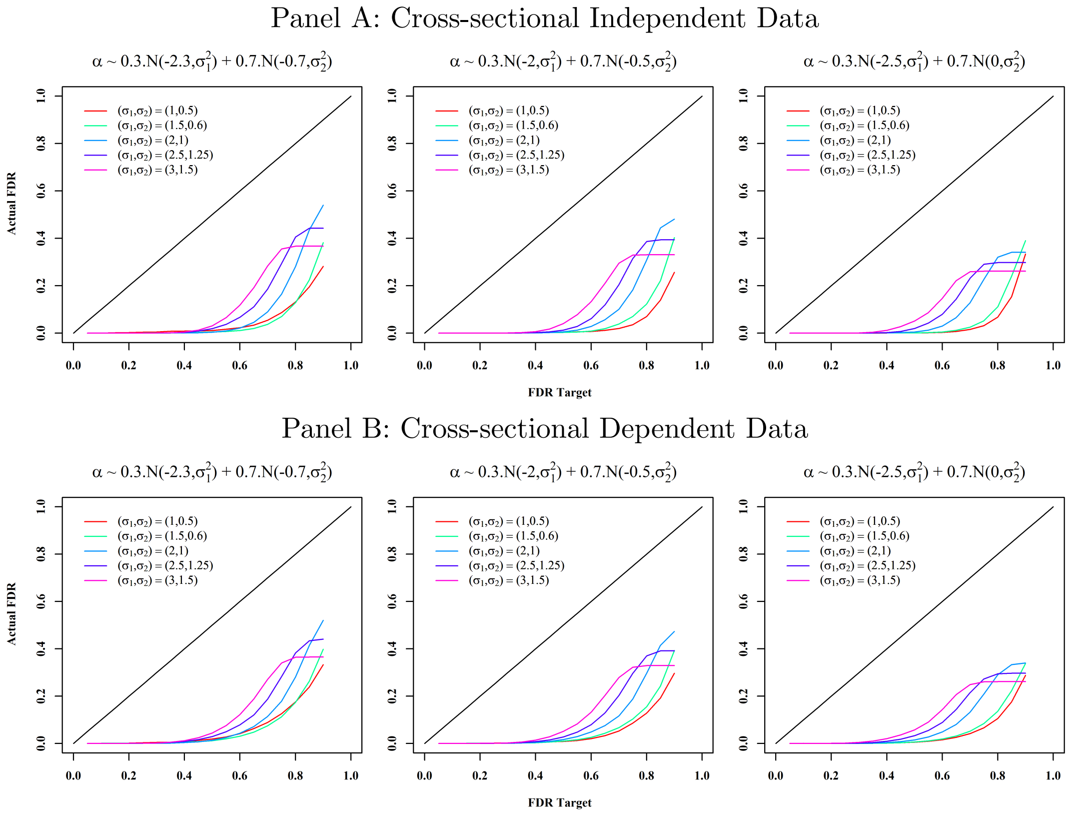


Table IX: Power comparison (in %) for discrete distribution. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when the alphas of 2,000 funds are drawn from a discrete distribution: $\alpha \sim \pi^+ \delta_{\alpha=\alpha^*} + \pi_0 \delta_{\alpha=0} + \pi^- \delta_{\alpha=-\alpha^*}$ with varying α^* (annualized, in %) and proportions (π^+, π_0, π^-). The simulated data are an unbalanced panel with the number of observations of each fund drawn randomly with replacement from the real-data counterpart. We study the simulated data with both cross-sectional independence (left-hand side) and cross-sectional dependence (right-hand side).

		Cross-sectional Independence					Cross-sectional Dependence				
(π^+, π_0, π^-)	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	0.4	5.5	21.1	41.3	58.9	0.6	5.9	20.6	40.3	57.8
	FDR^+	0.4	2.6	12.7	29.9	46.7	0.4	2.9	13	29.3	45.9
$(10, 60, 30)\%$	$fFDR^+$	1.1	10.3	30.6	51.8	68	1.5	10.5	29.8	50.7	66.9
	FDR^+	0.5	2.9	14.6	32.6	49.9	0.5	3.2	14.3	31.9	49
$(10, 30, 60)\%$	$fFDR^+$	3.2	19.8	46.6	66.8	79.8	3.9	19.8	45.6	66	79.4
	FDR^+	0.5	3.6	19.1	40	58.1	0.5	4	18.9	39.5	57.6
$(13, 67.5, 19.5)\%$	$fFDR^+$	0.9	8.9	27.7	48.5	65.1	1.2	9.2	27.1	47.5	64.1
	FDR^+	0.5	3.9	17.4	35.6	52.3	0.6	4.2	17	34.9	51.5
$(13, 48, 39)\%$	$fFDR^+$	2.2	15.5	37.8	58.8	73.7	2.9	15.5	37.1	57.8	73
	FDR^+	0.5	4.5	20.3	39.8	56.9	0.7	4.8	19.5	39	56
$(13, 9, 78)\%$	$fFDR^+$	6.2	27.5	60.2	78.1	88.7	7.5	29.2	60	78.4	88.9
	FDR^+	0.6	6.8	29.5	54.2	72.5	0.8	7.7	30	54.7	72.8

Table X: Power comparison (in %) for discrete-normal distribution mixture. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when alphas of 2,000 funds are drawn from a discrete-normal distribution mixture: $\alpha \sim \pi_0 \delta_{\alpha=0} + (1-\pi_0)\mathcal{N}(0,\sigma^2)$ with varying σ (annualized, in %) and null proportion π_0. The simulated data are an unbalanced panel with the number of observations of each fund drawn randomly with replacement from the real-data counterpart. We study the simulated data with both cross-sectional independence (left-hand side) and cross-sectional dependence (right-hand side).

		Cross-sectional Independence					Cross-sectional Dependence				
π_0	Procedure	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$
75%	$fFDR^+$	0.3	13.9	31.9	45.9	55.7	0.4	13.9	31.7	45.7	55.5
	FDR^+	0.2	7.8	23.5	37.1	47.5	0.3	7.9	23.3	36.9	47.2
60%	$fFDR^+$	1.3	19.2	38	51.8	60.9	1.3	19	37.8	51.7	60.9
	FDR^+	0.3	10.4	27.8	41.9	52.2	0.3	10.3	27.6	41.7	52
30%	$fFDR^+$	3.5	27.6	48.3	61.9	70.3	3.6	27.4	48	61.4	70.1
	FDR^+	0.4	15.4	35.7	50.5	60.6	0.5	15.2	35.4	50.2	60.3
67.5%	$fFDR^+$	0.8	16.8	35.2	48.9	58.4	0.9	16.9	35.1	49	58.5
	FDR^+	0.3	9.2	25.9	39.6	49.8	0.3	9.2	25.7	39.6	49.9
48%	$fFDR^+$	2.1	22.9	42.5	56.1	65.2	2.3	22.9	42.4	56	65.1
	FDR^+	0.3	12.4	31.2	45.6	55.7	0.4	12.5	31.1	45.4	55.4
9%	$fFDR^+$	5.3	33.3	54.9	68.2	76.7	5.6	33.5	55	68.2	76.7
	FDR^+	0.6	19.1	41.6	57	67.2	0.7	19.1	41.5	57	67.2

Table XI: Power comparison (in %) for mixture of two normal distributions. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when alphas of 2,000 funds are drawn from a mixture of two normal distributions: $\alpha \sim 0.3\mathcal{N}(\mu_1, \sigma_1^2) + 0.7\mathcal{N}(\mu_2, \sigma_2^2)$ with varying standard deviation pairs (σ_1, σ_2) and mean pairs (μ_1, μ_2) (both parameters' pairs are annualized and in %). The simulated data are an unbalanced panel with the number of observations of each fund drawn randomly with replacement from the real-data counterpart. We study the simulated data with both cross-sectional independence (left-hand side) and cross-sectional dependence (right-hand side).

		Cross-sectional Independence					Cross-sectional Dependence				
(μ_1, μ_2)	Procedure	σ^1	σ^2	σ^3	σ^4	σ^5	σ^1	σ^2	σ^3	σ^4	σ^5
$(-2.3, -0.7)$	$fFDR^+$	0	0.2	3.8	11.3	19.5	0	0.3	4.2	11.7	20.1
	FDR^+	0	0	0.3	1.8	6.3	0	0	0.3	2	6.4
$(-2, -0.5)$	$fFDR^+$	0	0.4	5	13	21.7	0.1	0.5	5.5	13.7	22.2
	FDR^+	0	0.1	0.4	2.7	7.8	0	0.1	0.5	2.9	8.1
$(-2.5, 0)$	$fFDR^+$	0.1	0.5	7.3	15.4	24	0.3	0.8	7.8	16	24.6
	FDR^+	0	0.1	0.6	3.9	9.9	0	0.1	0.9	4.2	10.3

where $\sigma^1 = (1, 0.5), \sigma^2 = (1.5, 0.6), \sigma^3 = (2, 1), \sigma^4 = (2.5, 1.25), \sigma^5 = (3, 1.5)$.

IB.4. Simulation results for single normal distribution

In this section, we present the simulation results for a special case of continuous distribution where the mixture (17) has only one component. Specifically, we consider the case $\pi_2 = 0$, $\alpha \sim \mathcal{N}(\mu, \sigma^2)$ and, based on Jones and Shanken (2005) and Fama and French (2010), we use $\mu \in \{-0.8, -0.5, 0\}$ and $\sigma \in \{1, 1.5, 2, 2.5, 3\}$ (the presented values of both parameters are annualized and in %).[32]

Figures VII and VIII present the performance of the $fFDR^+$ procedure when the alphas are drawn from balanced and unbalanced panel data, respectively. It is shown that the FDR is controlled at any given target.

[32] Jones and Shanken (2005) assume that the fund alphas are drawn from a normal distribution and their estimates for the mean and standard deviation are based on prior beliefs. They find that the mean is 1.3%-1.4% per annum before expenses (about 2%) and the standard deviation is 1.5%-1.8%. In addition, Fama and French (2010) assume that the fund (gross) alpha population has a normal distribution centered at 0.

Figure VII: Performance of $fFDR^+$ in terms of FDR control when alphas are drawn from the single normal distribution with balanced panel data.

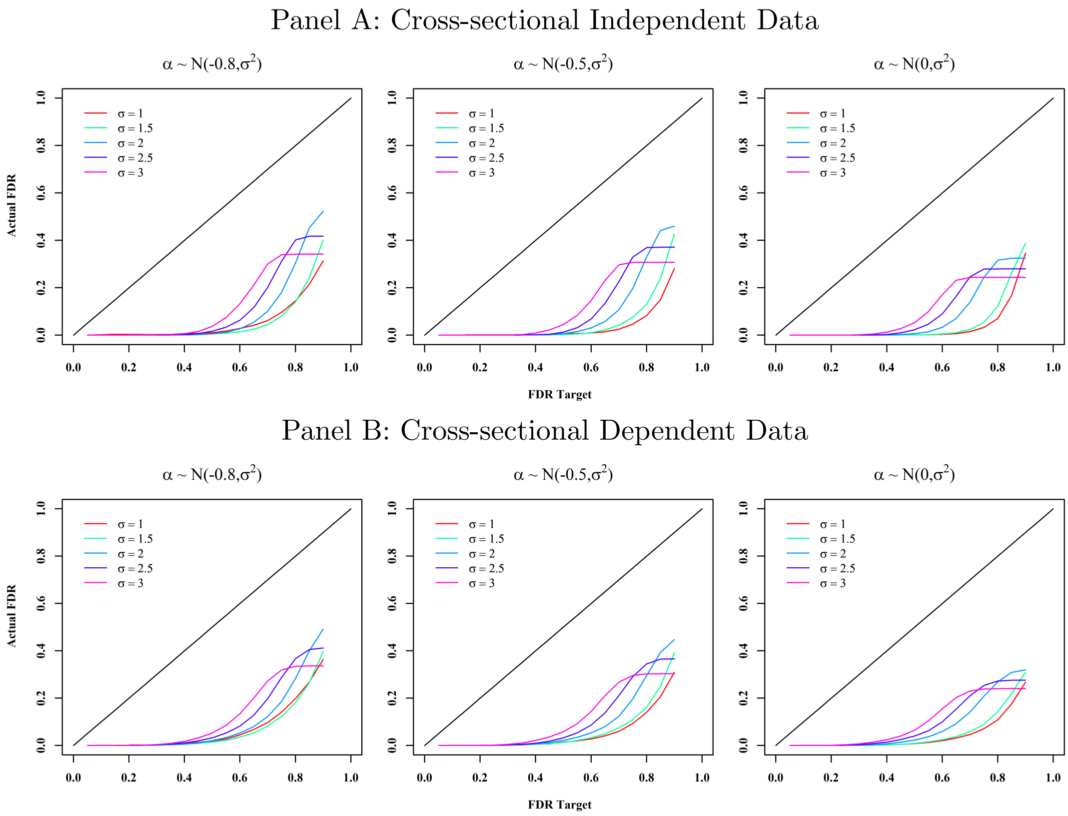
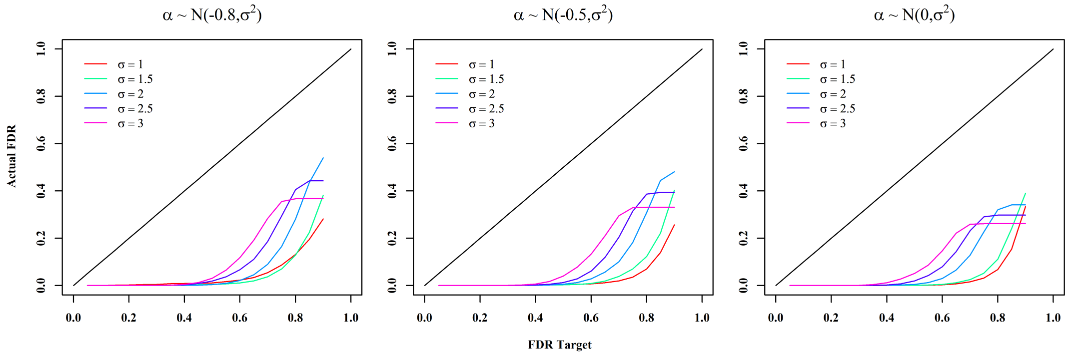
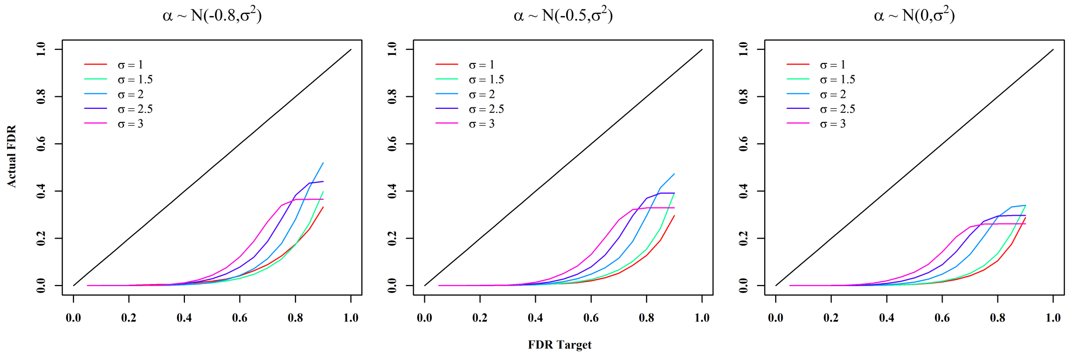


Figure VIII: Performance of $fFDR^+$ in terms of FDR control when alphas are drawn from the single normal distribution with unbalanced panel data.

Panel A: Cross-sectional Independent Data



Panel B: Cross-sectional Dependent Data



In Table XII, we focus on comparing the performance of $fFDR^+$ and FDR^+ in terms of power. As π^+ depends on both the mean μ and variance σ^2 of the distribution, we need to distinguish the value of π^+ from the pairs (μ, σ). We provide in Panel A additional information about π^+, which helps us assess the impact of the magnitude of positive alphas on the power. For instance, for $\pi^+ \approx 40\%$, the power of the two procedures for $(\mu, \sigma) = (-0.8, 3)$ is significantly higher than for $(\mu, \sigma) = (-0.5, 2)$. We observe a boost in power for both methods with increasing σ (for given non-positive μ), resulting in larger proportion and magnitude of positive alphas. In all the cases under consideration, the $fFDR^+$ dominates FDR^+ in terms of power and this gap soon becomes omnipresent for $\sigma \geq 1.5$ reaching up to 18%.

Table XII: Power comparison (in %) for single normal distribution. The table compares the power of the $fFDR^+$ and FDR^+ at FDR target of 10% when alphas of 2,000 funds are drawn from a normal distribution: $\alpha \sim \mathcal{N}(\mu, \sigma^2)$ with varying standard deviation σ and mean μ (both parameters are annualized and in %). In Panel A the simulated data are a balanced panel with 274 observations per fund, whereas in Panel B an unbalanced panel with the number of observations of each fund drawn randomly with replacement from the real-data counterpart. For each type of panel data, we generate data cross-sectional independence (left-hand side) and with cross-sectional dependence (right-hand side).

		Cross-sectional Independence					Cross-sectional Dependence				
		σ					σ				
μ	Procedure	1	1.5	2	2.5	3	1	1.5	2	2.5	3
		Panel A: Balanced Data									
	π^+	21.2	29.7	34.5	37.4	39.5	21.2	29.7	34.5	37.4	39.5
−0.8	$fFDR^+$	1.6	14.1	30.5	44.4	55.1	2.1	14.9	30.9	44.7	55.4
	FDR^+	0.1	1.7	12.6	27.2	40.6	0.1	2.1	12.8	27.4	40.7
	π^+	30.9	36.9	40.1	42.1	43.4	30.9	36.9	40.1	42.1	43.4
−0.5	$fFDR^+$	3	17.6	33.8	47.3	57.7	3.8	18.3	34.5	47.8	58.1
	FDR^+	0.1	3.6	16.5	31.3	44.1	0.2	4	16.7	31.5	44.3
	π^+	50	50	50	50	50	50	50	50	50	50
0	$fFDR^+$	7.9	24.8	40.7	52.8	62.4	8.9	25.7	41.3	53.3	62.7
	FDR^+	0.6	9.1	24.2	38.7	50.3	1	9.5	24.6	38.9	50.5
		Panel B: Unbalanced Data									
−0.8	$fFDR^+$	1.4	12.1	26.5	39.5	50.1	1.7	12.7	27.1	39.8	50.2
	FDR^+	0.1	1.7	10.8	23.2	35.2	0.1	2	11.2	23.5	35.4
−0.5	$fFDR^+$	2.6	15.2	29.8	42.5	52.6	3.1	15.8	30.2	42.8	52.7
	FDR^+	0.2	3.4	14.1	26.8	38.6	0.2	3.7	14.5	27.2	38.8
0	$fFDR^+$	6.8	21.6	36	47.8	56.9	7.4	22.4	36.4	48	57.1
	FDR^+	0.6	8.1	20.8	33.6	44.5	0.9	8.5	21.2	33.9	44.6

IC. Results for data sample period from 1984

Given potential biases in the mutual fund data for the period before 1984, we construct portfolios using a data sample from 1984 as a robustness check. We start by using the first five years' data, spanning from January 1984 to December 1988, to calculate the inputs of the procedures. The detected out-performing funds are equally invested in 1989. Then, the five years of data from January 1985 to December 1989 are used for the recalculation of the inputs of the procedures to detect out-performing funds invested in 1990, and so on. The process is yearly rolled over until the end of the sample. Thus, the OOS returns of the portfolios start from January 1989 to December 2019. At the end of each month from December 1998, i.e. when the portfolios' return series reach a length of at least ten years, we calculate the portfolios' alpha based on the returns from January 1989 to that month and present that in Figure IX. We also report the average n-year alpha with the length of investing n from 5 to 31 years in Table XIII.

Figure IX: Alpha evolution of $fFDR10\%$ and $FDR10\%$ portfolios over time with use of data from 1984. The graph presents the evolution of annualized alphas (in %) of the nine $fFDR10\%$ portfolios corresponding to the nine covariates, the portfolio $FDR10\%$ of BSW and the two equally weighted portfolios.

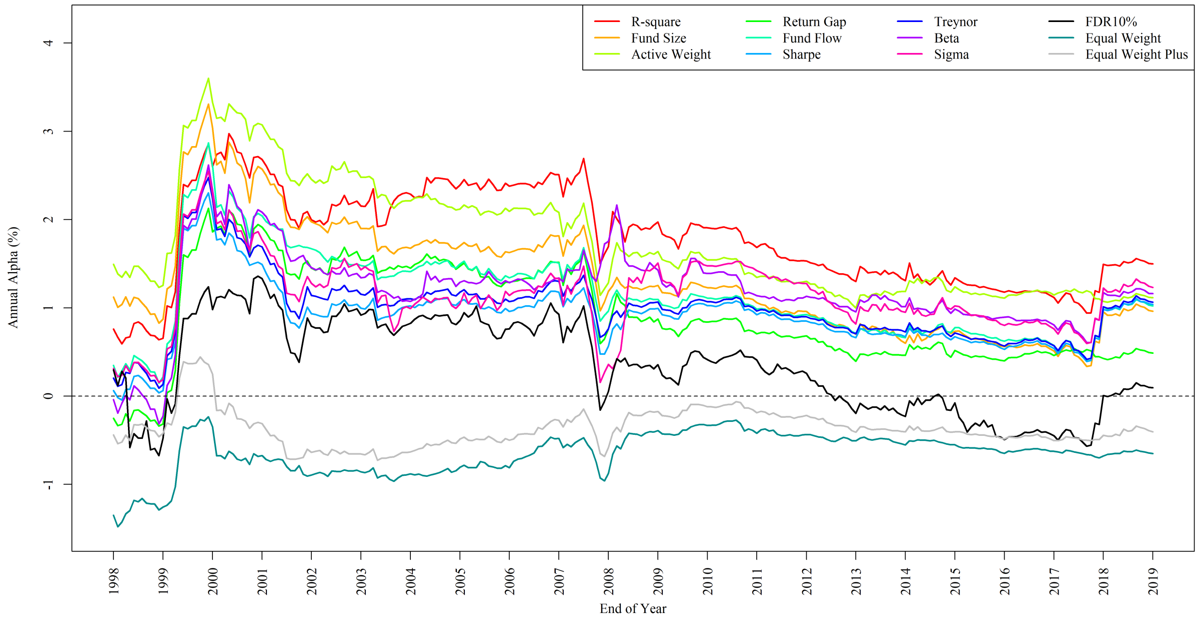


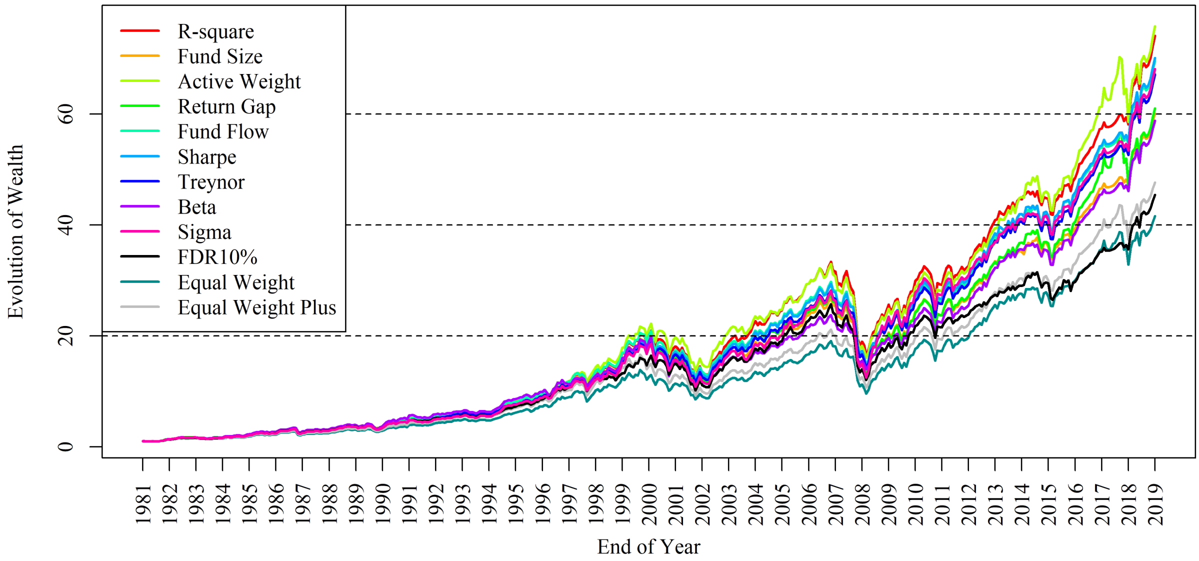
Table XIII: Comparison of portfolios' performances for varying time lengths of investing: results for sample data from 1984 to 2019. In this table, we consider 10 portfolios including nine $fFDR10\%$ portfolios corresponding to the nine covariates and the $FDR10\%$ portfolio of BSW. We compare the average alphas of the portfolios that are kept in periods of exactly n consecutive years. For example, consider $n = 5$. For each portfolio, we calculate the alpha for the first 5 years based on the portfolios' returns from January 1989 to December 1993. Then, we roll forward by a month and calculate the second alpha. The process is repeated and the last alpha is estimated based on the portfolios' returns from January 2015 to December 2019. The average of these alphas is presented in the first rows of the table.

	$fFDR10\%$									$FDR10\%$
n	R-square	Fund Size	Active Weight	Return Gap	Fund Flow	Sharpe	Treynor	Beta	Sigma	
5	1.17	0.62	0.88	0.22	0.5	0.35	0.47	0.53	1.02	-0.45
10	1.43	0.61	0.99	0.46	0.58	0.5	0.51	0.91	1.04	-0.37
15	1.64	0.6	1.09	0.65	0.69	0.65	0.63	0.96	1.03	-0.17
20	1.61	0.65	1.29	0.7	0.77	0.79	0.75	1.07	1.17	-0.12
25	1.28	0.53	1.12	0.43	0.61	0.59	0.57	0.9	0.93	-0.33
30	1.45	0.93	1.07	0.43	1.02	1.05	1.05	1.13	1.21	0.03
31	1.5	0.96	1.11	0.49	1.02	1.04	1.06	1.16	1.23	0.1

ID. Wealth evolution

In Figure 5 in the main manuscript, we study the alpha evolution of the portfolios over time. However, an investor may be interested in the gain in value. Figure X shows the growth of 1 dollar that the investor invests in each portfolio at the beginning of 1982. Ultimately, at the end of 2019, this amount grows to about 74 dollars if she chooses the $fFDR10\%$ portfolio with R-square as the covariate, as opposed to just 45, 47 and 41 dollars with the $FDR10\%$, the equal weight plus and equally weighted portfolios, respectively. This exercise reveals the potential profitability of an investor who had the perfect oracle in 1982 on the methods and the covariate that would be presented over the next 30 years.

Figure X: Evolution of wealth. The graph plots the evolution of 1 dollar invested at the beginning of 1982 in the nine $FDR10\%$ portfolios corresponding to the nine covariates, the $fFDR10\%$, the Equal Weight and Equal Weight Plus portfolios.



IE. Results for alternative target of FDR

In this section, we repeat the exercise with the FDR target of 20%. Figure XI presents the alpha evolution of the individual covariates. Table XIV shows the average n-year alpha of those portfolios. Finally, Table XV presents the statistic metrics for all mentioned portfolios.

Figure XI: Alpha evolution of $fFDR20\%$ and $FDR20\%$ portfolios over time. The graph presents the evolution of annualized alpha of the nine $fFDR20\%$ portfolios corresponding to the nine covariates, the $FDR20\%$ of BSW and the two equally weighted portfolios.

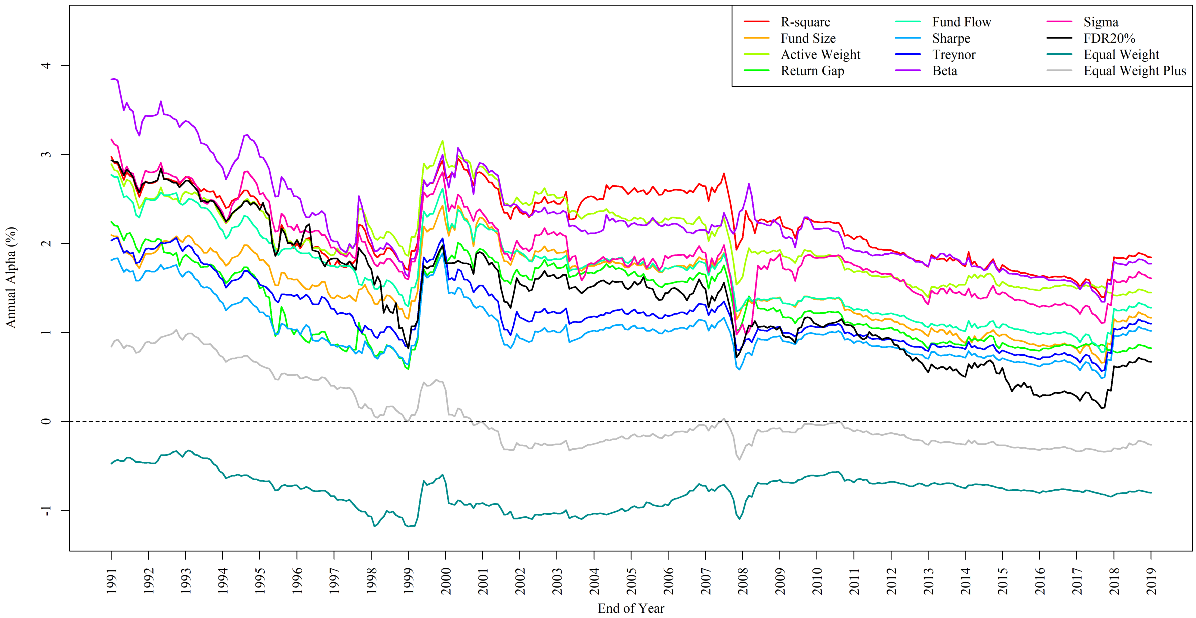


Table XIV: Comparison of portfolios' performances for varying time lengths of investing. In this table, we consider 10 portfolios including nine $fFDR20\%$ portfolios corresponding to the nine covariates and the $FDR20\%$ portfolio of BSW. We compare the average alphas (annualized and in %) of the portfolios that are kept in periods of exactly n consecutive years. For example, consider $n = 5$. For each portfolio, we calculate the alpha for the first 5 years based on the portfolios' returns from January 1982 to December 1986. Then, we roll forward by a month and calculate the second alpha. The process is repeated and the last alpha is estimated based on the portfolios' returns from January 2015 to December 2019. The average of these alphas is presented in the first row in the table.

n	R-square	Fund Size	Active Weight	Return Gap	Fund flow	Sharpe	Treynor	Beta	Sigma	$FDR20\%$
5	1.6	0.8	1.23	0.61	0.89	0.58	0.65	1.15	1.4	0.41
10	1.63	0.82	1.21	0.61	0.93	0.65	0.7	1.33	1.2	0.34
15	1.84	0.92	1.46	0.82	1.06	0.79	0.83	1.34	1.22	0.41
20	1.97	1.05	1.66	1.03	1.15	0.9	0.92	1.44	1.28	0.53
25	1.75	0.9	1.42	0.78	0.99	0.79	0.82	1.37	1.18	0.42
30	1.55	0.81	1.28	0.67	0.95	0.76	0.8	1.35	1.16	0.31
38	1.84	1.16	1.45	0.82	1.28	1.02	1.1	1.77	1.61	0.67

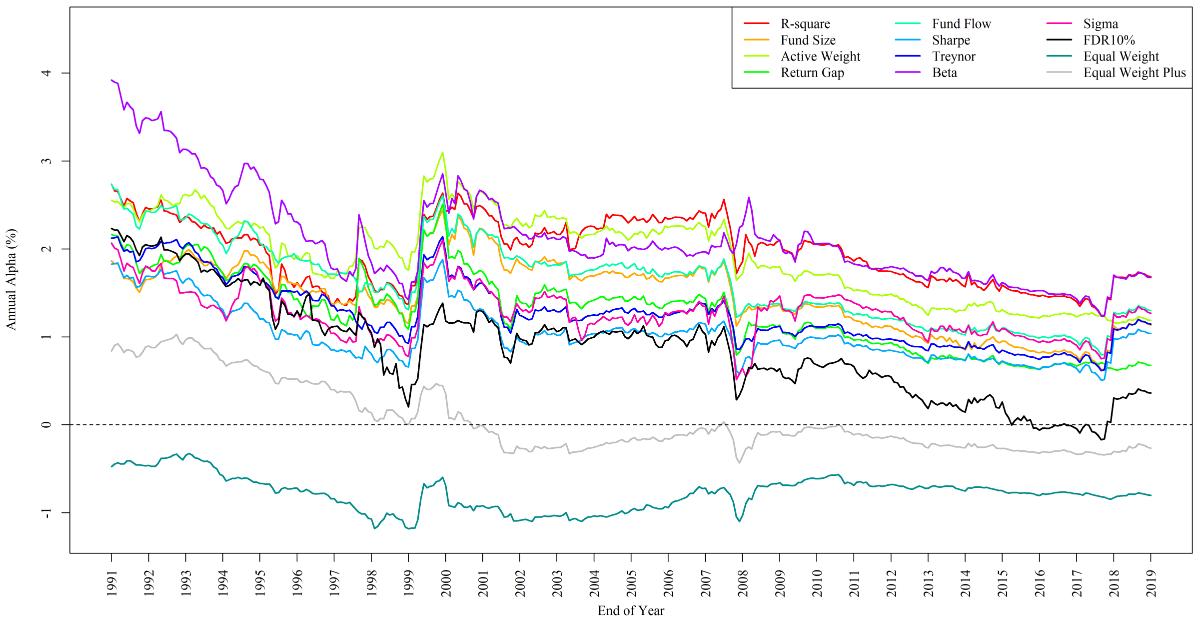
Table XV: Comparison of performance statistics of all considered portfolios with $\tau = 20\%$. The table compares the portfolios with regard to metrics including the annual Carhart four-factor alpha ($\hat{\alpha}$, in %) with its bootstrap p-value and t-statistic (with use of Newey–West heteroskedasticity and autocorrelation-consistent standard error), the annual standard deviation of the four-factor model residuals ($\hat{\sigma}_\varepsilon$, in %), the mean return in excess of the one-month T-bill rate (in %), the annual Sharpe ratio and the annual Information Ratio ($IR = \hat{\alpha}/\hat{\sigma}_\varepsilon$).

Covariate	$\hat{\alpha}$ (p-value)	t-statistic	$\hat{\sigma}_\varepsilon$	Mean Return	Sharpe Ratio	IR
R-square	1.84 (0.04)	2.03	4.41	8.08	0.62	0.42
Fund Size	1.16 (0.18)	1.35	4	7.37	0.56	0.29
Active Weight	1.45 (0.08)	1.81	3.7	8.06	0.6	0.39
Return Gap	0.82 (0.31)	1.05	3.77	7.43	0.55	0.22
Fund flow	1.28 (0.14)	1.54	3.76	7.76	0.59	0.34
Sharpe	1.02 (0.2)	1.31	3.37	7.77	0.61	0.3
Treynor	1.1 (0.17)	1.38	3.5	7.6	0.6	0.31
Beta	1.77 (0.05)	1.93	4.77	7.31	0.56	0.37
Sigma	1.61 (0.18)	1.44	5.02	7.91	0.59	0.32
$FDR10\%$	0.67 (0.5)	0.69	4.79	6.9	0.54	0.14
Equal Weight	-0.8 (0.03)	-2	1.86	6.3	0.5	-0.43
Equal Weight Plus	-0.26 (0.48)	-0.56	2.18	6.7	0.52	-0.12

IF. Results from using an alternative proxy of covariates

In this section, we present in Figure XII the alpha evolution of $fFDR10\%$ portfolios where the proxy for each covariate is based on whole data in the in-sample period instead of the data in final year as in the main manuscript. We see that the performance of the portfolios does not vary significantly.

Figure XII: Alpha evolution of $fFDR\tau$ portfolios over time where the proxy for each covariate (except the R-square and the four covariates obtained from the asset pricing models) is its average realizations in the five years in-sample period. The graph presents the evolution of annualized alpha (in %) of the nine $fFDR10\%$ portfolios (corresponding to the nine covariates), the portfolio $FDR10\%$ of BSW and the two equally weighted portfolios.



IG. Covariate combinations

So far, we have considered the effect from the information brought in by each single covariate. In what follows, we explore the effect from combining the information from the different covariates and potential consequent performance improvement. More specifically, we create a new covariate given by the linear combination of the underlying covariates. More specifically, for each fund i at time t, we have

$$\begin{aligned}\text{New Covariate}_{t,i} &= c_{1t}\text{R-square}_{t,i} + c_{2t}\text{Active Weight}_{t,i} + c_{3t}\text{Return Gap}_{t,i} \\ &\quad + c_{4t}\text{Fund Size}_{t,i} + c_{5t}\text{Fund Flow}_{t,i} + c_{6t}\text{Sharpe Ratio}_{t,i} \\ &\quad + c_{7t}\text{Treynor Ratio}_{t,i} + c_{8t}\text{Sigma}_{t,i} + c_{9t}\text{Beta}_{t,i}. \end{aligned} \tag{G.1}$$

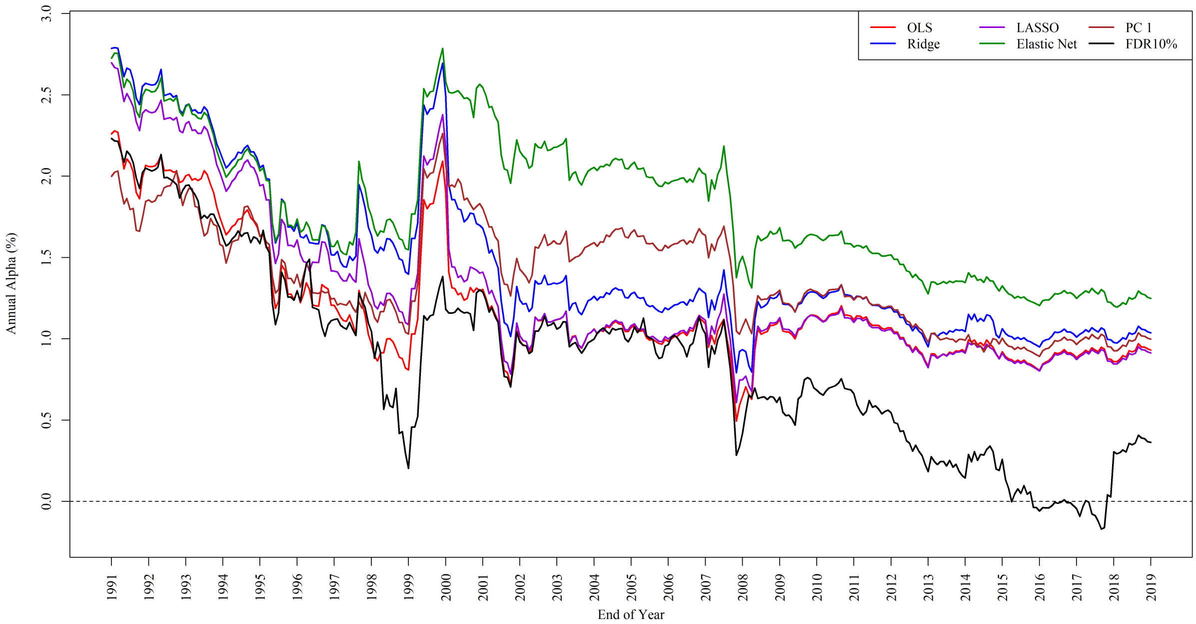
We consider two approaches to estimating the coefficients $c_{1t}, \ldots, c_{9t}$ in (G.1). First, we use as our new covariate the first principal component of all nine (standardized) covariates. By transforming the covariates to their principal components, their information about the performance of a fund is preserved and conveyed. We use the first principal component as it captures most of the variation of the covariates. Second, we use a linear model that regresses the fund returns for year k on the observed value of the covariates in year $k-1$, where $k \in \{t, t-1, t-2, t-3\}$. Then, we predict the return for year $t+1$ based on the estimated regression model and the covariates in year t. This is equivalent to using equation (G.1) with the regression's estimated coefficients as the $c_{1t}, \ldots, c_{9t}$. We use ordinary least squares (OLS), the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996) , ridge regression and the elastic net of Zou *et al.* (2005).[33]

Figure XIII exhibits the performance of the $fFDR\tau$ portfolios with the newly created covariates in terms the alpha evolution.[34] We find that the portfolio based on the covariate obtained from the elastic net performs best amongst the combined covariates at $\tau = 10\%$.

[33] For each method (except OLS), the covariates are standardized before being used in the estimation. We use cross-validation to determine the parameters in the LASSO, ridge and elastic net methods.

[34] There are a few years where LASSO (two years) and the elastic net (three years) shrink all the regression coefficients to zero. In these cases, the new covariate is equal to zero for all funds and, to avoid an empty portfolio, we simply select all the funds in the $FDR\tau$ portfolio.

Figure XIII: Alpha evolution of $fFDR10\%$ portfolios with combined covariates. The graph shows the alpha evolution of the $fFDR10\%$ portfolios with each using a covariate obtained from either the principal component method or regression method; for the former, the covariate is the first principal component (PC 1) of the five covariates, whereas for the latter the new covariate is a linear combination of the five underlying covariates with the weights obtained based on one of the OLS, LASSO, Ridge and elastic net regressions.



Aiming to acquire a more complete portrayal of the various covariates combinations, we study also the portfolios' alphas for various time lengths of investing. Table XVI shows the average n-year alphas of the $fFDR10\%$ portfolios.

Table XVI: Performance of $fFDR10\%$ portfolios with combined covariates for varying time lengths of investing. The table displays the average n-year alpha (annualized and in %) of the $fFDR10\%$ portfolios which use covariates obtained by the first principal component (PC 1), the OLS, LASSO, Ridge and elastic net (see descriptions in Figure XIII). The average n-year alpha of each portfolio is calculated as per the description in Table 6.

n	OLS	Ridge	LASSO	Elastic Net	PC 1
5	0.78	1.02	0.8	1.2	0.76
10	0.81	1.03	0.81	1.36	0.94
15	0.91	1.07	0.89	1.5	1.17
20	1.06	1.15	1	1.67	1.31
25	0.96	1.07	0.9	1.44	1.13
30	0.94	1.05	0.89	1.32	1.02
38	0.93	1.04	0.91	1.25	1

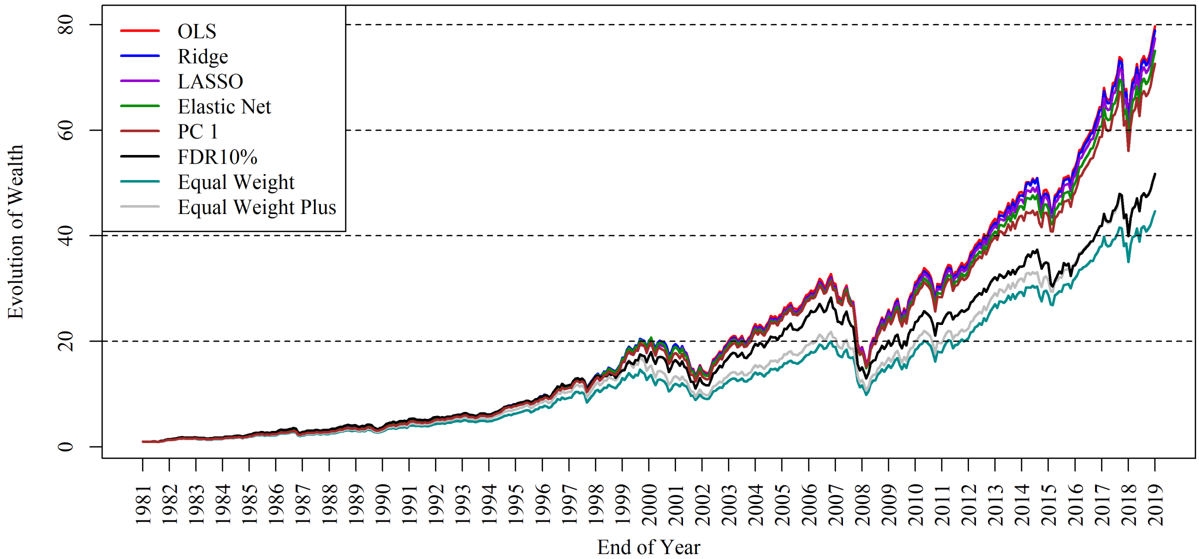
The elastic net performs also better for all time lengths. However, this best combined covariate does not beat the R-square and Beta under the $fFDR$ framework. When we partition the sample into four sub-periods, as mentioned above, Tables 7 and XVII show that for any sub-period there is always an underlying individual covariate that beats all the combined covariates. Nevertheless, since investors do not know which covariate will perform best in advance, the combination of covariates is still advantageous in prediction in practice.

Table XVII: Performance of portfolios in sub-periods. The table displays the performance of the $fFDR10\%$, $FDR10\%$ and equally weighted portfolios in sub-periods (P1: 1982–1991, P2: 1992–2001, P3: 2002–2011 and P4: 2012–2019) in terms of the average 5-year alpha, the alpha of the whole sub-period (both metrics are annualized and in %), the corresponding t-statistic (with use of Newey–West heteroskedasticity and autocorrelation-consistent standard error) and the annual Sharpe ratio.

	Average 5-year alpha				Whole sub-period alpha				t-statistic				Sharpe Ratio			
Portfolio	P1	P2	P3	P4	P1	P2	P3	P4	P1	P2	P3	P4	P1	P2	P3	P4
OLS	2.26	1.34	0.47	-0.1	2.26	1.15	0.99	-0.35	1.75	0.76	0.53	-0.41	0.64	0.8	0.26	1.11
Ridge	2.33	1.99	0.49	0.46	2.79	1.65	1.02	0.04	1.96	1.06	0.23	0.04	0.63	0.81	0.24	1.13
LASSO	2.65	1.16	0.49	-0.1	2.7	1.05	1.02	-0.35	1.83	0.67	0.3	-0.41	0.64	0.78	0.24	1.11
Elastic Net	2.45	2.63	0.39	-0.14	2.72	3	0.89	-0.37	1.68	1.88	0.2	-0.43	0.62	0.79	0.23	1.11
PC 1	1.69	1.48	0.62	-0.21	2	1.8	1.11	-0.47	1.75	1.19	0.56	-0.54	0.62	0.76	0.25	1.1

Similarly to Figure X, in Figure XIV we depict the wealth evolution of one dollar invested in the $fFDR10\%$ portfolios based on the combined covariates. At the end of 2019, 1 dollar grows to about 73 to 80 dollars if the investor invests in one of the $fFDR10\%$ portfolios with the covariates obtained from OLS, LASSO, Ridge and elastic net regressions.

Figure XIV: Evolution of wealth of $fFDR\tau$ portfolios with combined covariates. The graph plots the evolution of 1 dollar invested at the beginning of 1982 in the nine $FDR10\%$ portfolios corresponding to the nine covariates, $fFDR10\%$, Equal Weight and Equal Weight Plus portfolios. In this graph, the $fFDR10\%$ portfolios are the ones described in Figure XIII.



IH. Restricted data

As supplementary to our empirical study of Section 6, we repeat here our experiments for a data subset where a mutual fund enters the sample when its TNA reaches $15 million (adjusted for inflation as of January 2019). This choice of threshold is consistent with Pastor *et al.* (2015) and Zhu (2018). Table XVIII shows the average n-year alpha for the $fFDR10\%$ and $fFDR20\%$ portfolios based on each individual covariate. We then present in Table XIX our results for the $fFDR10\%$ and $fFDR20\%$ portfolios based on combinations of the covariates.

Table XVIII: Comparison of portfolios' performances for varying time lengths of investing: restricted data. We consider 20 portfolios including nine $fFDR10\%$ portfolios, nine $fFDR20\%$ portfolios, the $FDR10\%$ and $FDR20\%$ portfolios of BSW. We compare the average alphas (annualized, in %) of the portfolios that are kept for periods of exactly n consecutive years. For more details, refer to Table 6 of the main manuscript.

n	R-square	Fund Size	Active Weight	Return Gap	Fund flow	Sharpe	Treynor	Beta	Sigma	FDR
				Panel A: $fFDR10\%$ versus $FDR10\%$						
5	1.5	0.81	1.39	0.62	0.93	0.57	0.73	1.09	1.19	0.13
10	1.48	0.68	1.36	0.63	0.93	0.65	0.75	1.2	1.06	0.06
15	1.7	0.73	1.6	0.85	1.06	0.8	0.87	1.2	1.1	0.15
20	1.84	0.82	1.79	1.07	1.14	0.91	0.96	1.31	1.18	0.27
25	1.62	0.71	1.56	0.82	0.98	0.81	0.86	1.24	1.09	0.14
30	1.42	0.63	1.41	0.69	0.95	0.79	0.86	1.2	1.01	0.02
38	1.69	1.01	1.52	0.94	1.3	1.04	1.14	1.68	1.27	0.37
				Panel B: $fFDR20\%$ versus $FDR20\%$						
5	1.61	0.74	1.37	0.67	0.91	0.58	0.65	1.15	1.41	0.42
10	1.63	0.67	1.37	0.72	0.96	0.65	0.7	1.33	1.2	0.35
15	1.85	0.72	1.63	0.93	1.08	0.79	0.82	1.34	1.22	0.42
20	1.98	0.82	1.83	1.16	1.17	0.91	0.92	1.44	1.28	0.54
25	1.76	0.71	1.59	0.91	1.01	0.8	0.81	1.37	1.18	0.43
30	1.56	0.66	1.43	0.8	0.98	0.77	0.8	1.35	1.16	0.32
38	1.85	1.04	1.57	0.98	1.3	1.02	1.1	1.78	1.61	0.68

Table XIX: Performance of $fFDR\tau$ portfolios with combined covariates for varying time lengths of investing: restricted data. The table displays the average n-year alpha of the $fFDR10\%$ (Panel A) and $fFDR20\%$ (Panel B) portfolios using the covariates given by the first principal component (PC 1), the OLS, ridge, LASSO and elastic net (see descriptions in Figure XIII of the main manuscript). The average n-year alpha (annualized, in %) of each portfolio is calculated as described in Table 6 of the main manuscript.

n	OLS	Ridge	LASSO	Elastic Net	PC 1
			Panel A: $\tau = 10\%$		
5	0.76	1.02	0.84	0.94	0.78
10	0.73	1.04	0.96	0.99	0.99
15	0.82	1.09	1.08	1.07	1.22
20	0.95	1.19	1.25	1.17	1.4
25	0.83	1.07	1.1	1.03	1.19
30	0.8	1.01	1.06	0.97	1.08
38	0.79	0.97	1.07	0.96	1.05
			Panel B: $\tau = 20\%$		
5	0.73	1	0.68	0.96	0.77
10	0.73	1.01	0.79	1.01	0.96
15	0.81	1.06	0.93	1.08	1.2
20	0.93	1.17	1.1	1.18	1.38
25	0.82	1.05	0.95	1.04	1.17
30	0.8	1	0.89	1	1.06
38	0.77	0.96	0.88	1.01	1.02

II. Selecting unprofitable funds with fFDR

In this section, we obtain, by analogy with the $fFDR\tau$ portfolio, a selection of unprofitable funds. First, consider a selection of R^- under-performing funds including V^- wrongly selected zero-alpha or out-performing funds. We define

$$FDR^- = \mathbb{E}\left(\frac{V^-}{\max\{R^-, 1\}}\right) \tag{I.1}$$

and

$$pFDR^- = \mathbb{E}\left(\frac{V^-}{R^-}\middle| R^- > 0\right). \tag{I.2}$$

If a fund i with p-value p_i and negative estimated alpha ($\hat{\alpha}_i < 0$) is selected as under-performing fund whenever $p_i < \gamma$, then FDR^- is estimated by

$$\widehat{FDR}_\gamma^- = \frac{\hat{\pi}_0\gamma/2}{\hat{R}^-/m} \tag{I.3}$$

where $\hat{R}^- = \#\{i|p_i < \gamma, \hat{\alpha}_i < 0\}$ and $\hat{\pi}_0$ is calculated as in equation (11) in the main manuscript.

At a given target τ of FDR^-, we form the $FDR^-\tau$ ($fFDR^-\tau$) portfolio of under-performing funds similarly to the $FDR\tau$ ($fFDR\tau$) portfolio of out-performing funds. Specifically, we establish the $FDR^-\tau$ portfolio using the same γ grid as for the $FDR\tau$ and form the $fFDR^-\tau$ portfolio by implementing the $fFDR$ procedure (with a specific covariate) on the set of non-positive estimated alpha funds to control $pFDR^-$ at the same level as the portfolio $FDR^-\tau$.

The following tables present the average n-year alpha of the portfolios at target $\tau = 10\%$ (Table XX) and their trading metrics (Table XXI). We also construct a portfolio, namely Equal Weight Minus, which includes all funds with negative estimated in-sample alpha invested in the following year.

Table XX: Comparison of portfolios' performance for varying time lengths of investing: portfolios of unprofitable funds. We consider 11 portfolios including the equal weight minus (EW^-), the $FDR^-10\%$ and the $fFDR^-10\%$ with the nine individual covariates. The table compares the average alphas (annualized, in %) of portfolios that are kept in periods of exactly n consecutive years. For more details, refer to Table 6.

n	R-square	Fund Size	Active Weight	Return Gap	Fund flow	Sharpe	Treynor	Beta	Sigma	EW^-	$FDR^-10\%$
5	-3.96	-4.56	-2.85	-4.12	-3.43	-2.29	-2.16	-4.35	-4.13	-1.36	-4.77
10	-3.82	-4.37	-2.83	-3.85	-3.1	-2.05	-1.91	-4.18	-3.86	-1.24	-4.41
15	-3.59	-4.07	-2.62	-3.52	-2.86	-1.81	-1.65	-3.88	-3.62	-1.09	-4.09
20	-3.45	-3.89	-2.53	-3.33	-2.73	-1.7	-1.54	-3.72	-3.51	-1.01	-3.93
25	-3.61	-4.07	-2.73	-3.56	-2.93	-1.81	-1.66	-3.94	-3.69	-1.04	-4.17
30	-3.83	-4.29	-2.92	-3.83	-3.17	-2.05	-1.91	-4.22	-3.99	-1.1	-4.5
38	-4.12	-4.51	-3.21	-4.17	-3.74	-2.53	-2.41	-4.6	-4.31	-1.31	-4.91

Table XXI: Compe numbarison of performance statistics of all considered portfolios of unprofitable funds with $\tau = 10\%$. The table compares the portfolios with regard to metrics including the annual Carhart four-factor alpha ($\hat{\alpha}$, in %) with its bootstrap p-value and t-statistic (with use of Newey–West heteroskedasticity and autocorrelation-consistent standard error), the annual standard deviation of the four-factor model residuals ($\hat{\sigma}_\varepsilon$, in %), the mean return in excess of the one-month T-bill rate (in %), the annual Sharpe ratio and the annual Information Ratio ($IR = \hat{\alpha}/\hat{\sigma}_\varepsilon$).

Covariate	$\hat{\alpha}$ (p-value)	t-statistic	$\hat{\sigma}_\varepsilon$	Mean Return	Sharpe Ratio	IR
R-square	-4.12 (< 0.01)	-5.63	3.21	3.33	0.3	-1.29
Fund Size	-4.51 (< 0.01)	-6.26	3	2.86	0.27	-1.51
Active Weight	-3.21 (< 0.01)	-4.95	3.15	4.1	0.35	-1.02
Return Gap	-4.17 (< 0.01)	-5.89	3.32	3.2	0.29	-1.26
Fund flow	-3.74 (< 0.01)	-5.35	3.11	3.63	0.32	-1.2
Sharpe	-2.53 (< 0.01)	-4.2	2.68	4.48	0.38	-0.94
Treynor	-2.41 (< 0.01)	-4.04	2.68	4.68	0.39	-0.9
Beta	-4.6 (< 0.01)	-5.3	4.19	3.08	0.28	-1.1
Sigma	-4.31 (< 0.01)	-5.18	3.86	2.89	0.27	-1.12
$FDR^{-}10\%$	-4.91 (< 0.01)	-6.08	3.48	2.3	0.23	-1.41
Equal Weight	-0.8 (0.03)	-2	1.86	6.3	0.5	-0.43
Equal Weight Minus	-1.31 (< 0.01)	-3	1.98	5.9	0.48	-0.66

References

Barras, L., Scaillet, O. and Wermers, R. (2010) False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance*, **65**, 179–216.

Fama, E. F. and French, K. R. (2010) Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, **65**, 1915–1947.

Jones, C. S. and Shanken, J. (2005) Mutual fund performance with learning across funds. *Journal of Financial Economics*, **78**, 507–552.

Pastor, L., Stambaugh, R. F. and Taylor, L. A. (2015) Scale and skill in active management. *Journal of Financial Economics*, **116**, 23–45.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267–288.

Zou, H., and Hastie, T. (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **67**, 301–320.