Mutual Funds' Conditional Performance Free of Data Snooping Bias

Po-Hsuan Hsu, Ioannis Kyriakou, Tren Ma, and Georgios Sermpinis *

Abstract

We introduce a test to assess mutual funds' "conditional" performance that is based on updated information and corrects data snooping bias. Our method, named the functional False Discovery Rate "plus" ($fFDR^+$), incorporates fund characteristics in estimating fund performance free of data snooping bias. Simulations suggest that the $fFDR^+$ controls well the ratio of false discoveries and gains considerable power over prior methods that do not account for extra information. Portfolios of funds selected by the $fFDR^+$ outperform other tests not accounting for information updating, highlighting the importance of evaluating mutual funds from a conditional perspective.

* Po-Hsuan Hsu, pohsuanhsu@mx.nthu.edu.tw, College of Technology Management, National Tsing Hua University, Taiwan. Ioannis Kyriakou, ioannis.kyriakou.2@city.ac.uk, Bayes Business School, City, University of London, UK. Tren Ma, tren.ma@nottingham.ac.uk, Nottingham University Business School, University of Nottingham, Nottingham, UK. Georgios Sermpinis, Adam Smith Business School, University of Glasgow, UK. We would like to thank Keith Cuthbertson, Jens Perch Nielsen, Zheng Sun, Qifei Zhu, and the audience of AFA student poster session and CICF for their comments. We appreciate the replies and discussions of Gery Geenens, Kuangyu Wen and Ximing Wu on their works on density estimations, as well as Laurent Barras and Russ Wermers on the mutual fund data. All errors are our own.

I. Introduction

It is well-known that luck plays an important role in mutual funds' performance (Kosowski, Timmermann, Wermers, and White (2006)). In order to appropriately assess fund performance, investors should rely on a multiple hypothesis testing framework to correct for "data snooping" bias or "p-hacking", a major challenge to social science (Sullivan, Timmermann, and White (1999, 2001); White (2000); Hansen (2005); Hsu, Taylor, and Wang (2016); Chordia, Goyal, and Saretto (2020)).

Prior research has developed multiple hypothesis testing frameworks to correct such bias and controls the number of false discoveries which are, in our context, seemingly profitable funds that are just due to luck. In particular, 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. 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. This feature appears too restrictive or even unrealistic because, in economics and finance research, the economic agents use all available and routinely update information in assessing models' performance. Extra information sources can assist researchers to more accurately estimate the FDR. Recently, Chen, Robinson, and Storey (2021), CRS henceforth, 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.[1] In the context of mutual funds, 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 collecting external 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.

In this paper, we introduce the functional False Discovery Rate "plus" ($fFDR^{+}$) method. Compared to the work of CRS it has two distinguishing features in assessing mutual fund performance. First, it allows us to focus on the right tail of the distribution and detect the significantly outperforming funds, which is important for investors (see Barras et al. (2010), hereafter BSW). Second, it is robust to cross-sectional dependencies among performance measures, which are common for mutual funds because their alphas are likely dependent due to common shareholdings and herding in 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. In addition, it is easy to implement, does not rely on any strong assumption and can handle any continuous fund characteristic.

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

[1] Since White (2000), several multiple testing procedures have been proposed to correct for luck in the past (Hansen (2005); Romano and Wolf (2005); Barras, Scaillet, and Wermers (2010); Hsu, Hsu, and Kuan (2010); Bajgrowicz and Scaillet (2012)), however, they only consider unconditional null hypotheses.

informative covariate (i.e., fund characteristics) 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 the $fFDR^+$ 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 out-performing mutual funds. In particular, we explore ten fund characteristics as informative covariates: the first set contains six fund attributes that have been shown in prior studies to convey information on mutual fund performance, and the second set contains four new attributes 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, Sialm, and Zheng (2008), the Active Weight of Doshi, Elkamhi, and Simutin (2015), the Fund Size of Harvey and Liu (2017), the Fund Flow suggested by Zheng (1999), and the expense ratio. 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 out-performing funds suggests that, with more information input, there may exist more profitable mutual funds than researchers have detected.

Based on the funds selected by the $fFDR^+$, we build portfolios that consistently outperform the one generated by prior FDR methods. Our results highlight the economic value of extra information. In particular, the $fFDR^+$ portfolios based on Beta is found to be the best with annualized alphas of 1.1%, followed by the $fFDR^+$ portfolios based on Expense ratio, R-square, Active Weight, Sigma, Fund Flow, Return Gap, Treynor ratio, Fund Size and Sharpe Ratio, separately achieving annualized alphas of at least 0.17%. We note that this profitability is persistent in our sample and is even strengthened over the recent period prior to Covid-19 pandemic, 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. This finding suggests that the relationship between luck and funds' performance with the mentioned fund characteristics is non-linear and that traditional portfolio approaches that do not control of luck may be inadequate.

In additional analysis, we also consider the $fFDR^+$ portfolio based on various ways of combining the ten covariates, such as the first principal component of the ten 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 and Hastie (2005). We find that the ridge and elastic net deliver the best performance with an annualized alpha of at least 0.86%. The investors may also benefit from such combinations as they result in lower volatility in portfolio performance. In fact, we find that our $fFDR^+$ portfolios based on the combined covariates gain highest Sharpe ratio in the most recent decade. This is advantageous as, in reality, investors do not know ex-ante what covariate is the best.

The literature on mutual funds' performance has two main strands: one that tries to model the distribution of mutual funds in terms of alphas and identify the out-performing funds and managers, and another that focuses on identifying covariates that explain mutual funds' performance.[2] Consequently, to identify out-performing funds researchers simply rank the funds' alphas and covariates.[3] Our study contributes to the mutual fund literature as follows. First, our $fFDR^+$ approach estimates the FDR as a function of a fund characteristic that is related to a fund's performance. By designing and implementing suitable Monte Carlo simulation experiments, we illustrate that our $fFDR^+$ approach actually controls for the FDR and delivers higher power than its benchmark under cross-sectional dependence and different distributions of the fund alpha population.[4] In our empirical analyses, fund portfolios based on $fFDR^+$

[2]In one of the earliest studies, Jensen (1968) documents that the majority of active equity mutual fund managers are unable to beat passive investment strategies net of fee. More recent research incorporates cross-sectional information and assesses funds' performance via a Bayesian approach with some prior beliefs about the distribution of fund alphas. For example, Jones and Shanken (2005) assume that the fund alpha population has a normal distribution and use the Gibbs sampling technique to estimate the parameters of the distribution, whereas Harvey and Liu (2018) adopt a mixture of normals and introduce an expectation maximization technique to estimate the weights and the parameters of the component distributions. Others assess funds in different aspects such as the horizon of the return used to estimated alpha (see, e.g., Bessembinder, Cooper, and Zhang (2023)).

[3]For instance, Carhart (1997) constructs a portfolio by sorting mutual funds according to their past performance (e.g., lagged one-year return and three-year past four-factor model alpha). Kacperczyk et al. (2008) discover that the Return Gap, defined as the difference between the fund's reported return and the return based on previous holdings, can predict the fund's future performance. Similarly, Doshi et al. (2015) present the Active Weight metric that conveys information about the fund's future performance and demonstrate predictability. Other researchers do multiple sorting on variables related to funds' performance. For example, Amihud and Goyenko (2013) show that a fund's R-square can predict its performance.

[4]We consider a discrete distribution as in BSW, a mixture of discrete and normal distributions as in AP, a single normal distribution as in Fama and French (2010) and Jones and Shanken (2005), and a mixture of two normals studied in Harvey and Liu (2018).

consistently dominate benchmark portfolios in terms of generating positive alphas. Our simulations and empirical evidence collectively highlight the importance of evaluating the conditional performance of mutual funds and the persistence of out-performing funds identified by the $fFDR^{+}$.

Second, our research adds to the mutual fund literature by exploring different information contents of fund characteristics. Based on our $fFDR^{+}$, we construct portfolios that persistently produce positive alphas for decades. Our portfolios based on four new covariates perform well and outrank, in the context of our method, those based on the traditional six covariates on several metrics and sub-samples. Finally, we move one step further and combine the ten covariates into single ones via linear combinations with the weights obtained from a principal component analysis and shrinkage regression methods. We find that investors might benefit from such combinations as they offer lower volatility in portfolio performance.

The rest of the paper is organized as follows. In Section II, we introduce and explain our methodology. In Section III, we provide a description of our data. Section IV is devoted to our simulation experiment descriptions, whereas in Section V we present in detail our simulation results. Section VI focuses on the empirical part of our analysis. Section VII concludes the paper. All technical details, simulations, and robustness checks are provided in the Internet Appendix.

II. Methods for Controlling of Luck with Fund Characteristics

A. 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.[5]

Suppose that we are assessing m funds and each of them has a net return time series. We also assume that there exists a continuous fund characteristic X, with observed values $(x_1, \ldots, x_m)$, that conveys information about the alpha of each fund. Our fund characteristic

[5]We note that, traditional approaches, such as the studies of Carhart (1997), Kosowski et al. (2006) and Fama and French (2010), define fund skill by the alpha that the fund delivers to investor. However, recent literature in mutual fund propose differently. Berk and van Binsbergen (2015) provide convincing arguments that ones should not use the net alpha nor the gross alpha that the fund delivers to investors as a measure of skill. They show that the value added, i.e., the value that a mutual fund extracts from capital markets, always measures skill. Subsequently, other studies such as Barras, Gagliardini, and Scaillet (2022) further separate the effects of scale from the measure of skill. Thereby, the skill is defined as the gross alpha earned on the first dollar invested in the fund. More specifically, Barras et al. (2022) model a time-varying gross alpha (the alpha calculated based on gross return) of a fund and express it as $a - bq_{t-1}$ where a and b are defined as skill and scale coefficients and the q_{t-1} is lagged fund size. In this study, we consider the net alpha as a measurement of performance.

corresponds to the informative covariate in the statistic literature (e.g., CRS and Ignatiadis and Huber (2021)). 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 $(x_1, \ldots, x_m)$. 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):

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

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 triples (T_i, h_i) are independent and each of them has the same distribution as (T, h).[6] Finally, we denote by $f(p, z)$ the joint density function of (P, Z). We thus have

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

as the posterior probability of the null hypothesis being true given that we observe $T = (p, z)$.[7]

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

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

where R is the number of rejected hypotheses in m tests and V the wrongly rejected ones. CRS show that, with a fixed set Γ in $[0, 1]^2$, if we reject hypothesis $H_{0,i}$ whenever $T_i \in \Gamma$, then

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

To maximize the number of rejections, we reject the hypotheses with the smallest statistic $r(p, z)$.

[6] In Sections IV.A and IV.B of the Internet Appendix, we show that this requirement can be eased for a typically cross-sectional dependence in mutual fund data. We also note that the FDR framework of Storey (2002) and the FDR^+ of BSW also work outside the independent and identically distributed framework (see Storey, Taylor, and Siegmund (2004) and Bajgrowicz and Scaillet, 2012).

[7] For more details about the role of $Z \sim \text{Uniform}(0, 1)$ and the derivation of (3), see CRS.

Thus, the significance region $\{\Gamma_\theta : \theta \in [0,1]\}$ is defined as

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

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):

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

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.[8] We calculate $\hat{r}(p,z) = \hat{\pi}_0(z)/\hat{f}(p,z)$ and estimate the q-value function as

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

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 and $\#S_i$ returns the number of elements in the set S_i.[9] 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 ($fFDR$).

[8] See Appendix A for more details.

[9] The $\#S_i$ is the number of discoveries given θ while the numerator is the expected number of false discoveries. This estimation is proposed by Newton, Noueiry, Sarkar, and Ahlquist (2004) and Storey, Akey, and Kruglyak (2005) and subsequently adopted in CRS.

B. 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

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

For m tests, let A^+ be the set of hypotheses with positive estimated alpha, i.e., $A^+ = \{i|\hat{\alpha}_i > 0\}$, 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).[10] Hence, the $fFDR$ selects positive-alpha funds with control of pFDR at the given target; we call this procedure the functional FDR "plus" ($fFDR^+$).

Different from our approach, BSW propose a procedure to estimate the FDR in detecting

[10]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 in Section I of the Internet Appendix, having a positive estimated alpha is a necessary condition for a fund to be selected as out-performer.

out-performing mutual funds, namely the FDR^+, which utilizes only $p-$value and alpha of funds. For the sake of space, we present details of the FDR^+ and a comparison between it and our $fFDR^+$ in Section I of the Internet Appendix.

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, our $fFDR^+$ additionally uses a fund characteristic and expresses the null proportion as a function of it, while accounting for the joint distribution of the p-value and the fund characteristic. As documented in CRS, this results in a better estimate of FDR, in terms of both bias and variance, and 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 following sections. In the empirical section, we will show the actual profitability that the fund characteristics can bring to investors while controlling for luck.

III. Data

We use monthly mutual fund data from January 1975 to December 2022 collected from the CRSP database. 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

[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.

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 all the funds in these 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 2,291 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.

IV. 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.

A. 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:

$$r_{i,t} = \alpha_i + b_i r_{m,t} + s_i r_{smb,t} + h_i r_{hml,t} + m_i r_{mom,t} + \varepsilon_{i,t}, \qquad i = 1, \dots, m, \tag{10}$$

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 (10)). 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, Scaillet, and Wermers (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 IV.A of the Internet Appendix. The simulated data are then used to generate the net return for each fund.[12] Subsequently, by carrying out regression (10) 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 identify 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.

B. 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

[12]We consider both balanced and unbalanced panel data. For the interest of space, the simulation results of the unbalanced panel data case are deferred to Section IV.B of the Internet Appendix.

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.[13]

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,

$$\alpha \sim \pi^- \delta_{\alpha=-\alpha^*} + \pi_0 \delta_{\alpha=0} + \pi^+ \delta_{\alpha=\alpha^*}. \tag{11}$$

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.[14]

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

$$\alpha \sim \pi_0 \delta_{\alpha=0} + (1 - \pi_0)\mathcal{N}(0, \sigma^2). \tag{12}$$

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

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

[14] 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.

$\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ with, respectively, probabilities $\pi_1 \in [0, 1]$ and $\pi_2 = 1 - \pi_1$, i.e.,

$$\alpha \sim \pi_1 \mathcal{N}(\mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(\mu_2, \sigma_2^2). \tag{13}$$

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 %).[15]

In (13) $\pi_0 = 0$, whereas in (11) and (12) $\pi_0 > 0$. 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

$$f(z) = 3.5(z - 0.5)^3 - 0.5(z - 0.5) + c \tag{14}$$

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.[16]

Suppose the distribution of alpha and the form of $\pi_0(z)$ are determined. We generate the

[15]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 IV.C of the Internet Appendix, we additionally present results of the case $\pi_2 = 0$, i.e., when the mixture becomes a single normal distribution.

[16]In Section IV.D of the Internet Appendix, we show that 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.

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 (10) 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 0.0182 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 estimates for the actual FDR and the power of each procedure.[17]

V. Analysis of False Discovery Rate and Power

We set the number of funds for simulations at $2{,}000$ which is close to our sample of $2{,}291$ 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 284, under cross-sectional independence. For the interest of space, we refer to Sections IV.A and IV.B 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

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

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 $fFDR^{+}$ 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.[18] 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 fund characteristics 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.

A. False discovery rate control of $fFDR^{+}$

For varying targets of FDR $\in \{5\%, 10\%, \ldots, 90\%\}$, we implement the simulation procedure in Section IV with balanced panel data. Figures 1, 2 and 3 exhibit our results for the generated data under cross-sectional independence.

In Figure 1, we show our results for the discrete distribution (11) 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

[18]In Section V of the Internet Appendix, we additionally study the $fFDR^{+}$ with use of an *non-informative* covariate, which is a covariate generated randomly and independently from the tests. We find that the $fFDR^{+}$ controls well for the FDR and its power is similar to that of the FDR^{+}.

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.

[Insert Figure 1 approximately here]

In Figure 2, we study the case of the fund alpha population described by the mixed discrete-continuous distribution (12). We organize our results based on the same null proportions π_0 as in Figure 1 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.

[Insert Figure 2 approximately here]

Finally, in Figure 3, we report the results for the continuous distribution (13) 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.

[Insert Figure 3 approximately here]

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 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 III of the Internet Appendix, we show that the variance of the estimated actual FDR of the $fFDR^+$ is smaller than that of the FDR^+. This means that the reported estimated actual FDR curves of the $fFDR^+$ are less varying than that of the FDR^+. In other words, if the estimated actual FDR of the two methods are the same and lie below or on the $45°$ line, there are less chance that the actual FDR of the $fFDR^+$ to lie above the $45°$ line than that of the FDR^+.

In Sections IV.A and IV.B 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.

B. Power analysis

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

In Panel A of Table 1, we report for the discrete distribution (11). 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., 1% and 0.6%, 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%.[19] The superiority of the $fFDR^+$ is more perceptible and stabler for larger α^*.

[19] 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).

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 $\alpha^* = 2.5$).[20]

[Insert Table 1 approximately here]

Panel B exhibits the power upshots for the case of the fund alpha population described by the distribution mixture (12). 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 0.6% to 61.3% (0.3% to 52.9%) with increasing σ from 1 to 5. For given, say, $\sigma = 2$, the power of the $fFDR^+$ (FDR^+) increases from 16.8% to 39.8% (9.2% and 23.5%) 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%.

Finally, in Panel C, we study the outcome from using the mixture of normals (13) 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

[20] As shown in Section IV 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 16, with unbalanced panel or with cross-sectional dependence, however the overall picture remains the same.

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 the $fFDR^+$ 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 the 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 IV.A and IV.B of the Internet Appendix that our conclusions are not affected by the data structure (balanced versus unbalanced panel) or dependencies.

C. 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.

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 increasing power with increasing FDR targets, peaking at 38% 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 35% in excess of the FDR^{+} (mixture of two normal distributions with 90% target).

[Insert Table 2 approximately here]

In Section VI of the Internet Appendix, we conduct a set of simulations to investigate the impact of varying number of observations T per fund and the number of funds m on the power. We see that the power of the $fFDR^{+}$ increases at a much faster pace, compared to the FDR^{+}, with increasing T, and slightly decreases with rising m.

In Section VII, we design simulations to study impact of using a covariate with noise on the power of the $fFDR^{+}$ by adding to the original covariate a noise reflecting potential estimation biases. We see that the $fFDR^{+}$ controls well for FDR and the power of the $fFDR^{+}$ is lower than that in Table 1, but still remarkably higher than that of the FDR^{+}.[21]

Concluding this section, we recollect that the simulated power of the $fFDR^{+}$ in detecting out-performing funds is found to be larger than the 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

[21]In Section V of the Internet Appendix, we further show that if a non-informative is used instead, the $fFDR^{+}$ controls well FDR and gains the same power as the FDR^{+}.

well as the proportion of zero alpha in the population, the number of funds in the sample, estimation errors in the covariates (fund characteristics), 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.

VI. Empirical Results

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

In this section, we illustrate how the $fFDR^+$ helps to identify out-performing mutual funds using a portfolio approach following BSW. We use as covariates six fund characteristics including the R-square of Amihud and Goyenko (2013), the Fund Size of Harvey and Liu (2017), the Return Gap of Kacperczyk et al. (2008), the Active Weight of Doshi et al. (2015), the Fund Flow and Expense ratio. For the interest of space, we refer the details of the construction of the six covariates to Section VIII of the Internet Appendix. 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 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.[22]

Similar to BSW, 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 the $fFDF^{+}$ 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.[23]

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 XI of the Internet Appendix.[24] 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, Fund Size and expense ratio are calculated based on data in the final year of the in-sample (i.e., we use the exposure of the fund

[22]For instance, Clifford, Fulkerson, Jame, and Jordan (2021) study the relation between idiosyncratic volatility and mutual funds flows but they do not focus on using this fund characteristic as a factor for funds selection.

[23]In Section IX of the Internet Appendix, we further validate the performance of the methods with use of simulated data (i.e., the return and ten covariates) resembling the real sample.

[24]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.

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 note that each of the covariates is converted into interval $[0, 1]$ via the formula described in Section A.

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).[25] For each fund, we implement $1{,}000$ bootstrap replications to estimate the distribution of the t -statistic and subsequently calculate the bootstrapped p-value for the fund.[26]

As required by our method, the p-values of any truly zero-alpha funds, given a covariate 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.[27]

Next, we describe the selection process of out-performing funds to invest in year $t + 1$

[25] We check heteroskedasticity, autocorrelation and ARCH effect by using White, Ljung-box and Engle tests, respectively. We see that a half of funds in our sample suffer from at least one of the mentioned effects.

[26] 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 $f(p, z)$.

[27] 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

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 equation (4) in the Internet Appendix. 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.[28] 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 $\tau < \widehat{FDR}^+_{\gamma^*} < 1$.[29] If $\widehat{FDR}^+_{\gamma^*} \geq 1$, we just select all the funds in the $FDR\tau$ portfolio.

Similar to BSW, for both the $fFDR\tau$ and $FDR\tau$ portfolios, we invest equally in the

---

covariate value, the p-value of test i is uniformly distributed since

$$\mathbb{P}(P_i < p_i) = \mathbb{P}(1-F(|\mathcal{T}_i|) < p_i) = \mathbb{P}(|\mathcal{T}_i| > F^{-1}(1-p_i)) = 1-\mathbb{P}(|\mathcal{T}_i| < F^{-1}(1-p_i)) = 1-F(F^{-1}(1-p_i)) = p_i.$$

[28]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.

[29]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.

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 into a portfolio by using 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.

B. 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 information contained in fund characteristics. We estimate the alpha evolution and the average alphas of our $fFDR\tau$ portfolios based on the ten covariates and compare with those of the $FDR\tau$ portfolio. We also explore the performance of the $fFDR\tau$ portfolios after linearly combining the ten 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.

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

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 4. It is obvious that the $FDR10\%$ portfolio gains higher alphas than the equally weighted portfolio and that all the $fFDR10\%$ portfolios outperform the $FDR10\%$. Ultimately, at the end of 2022, the $fFDR10\%$ portfolios with the Beta covariate is found to be the best with annualized alphas of about 1.1%, followed by the $fFDR10\%$ portfolios with the Expense Ratio, R-square, Active Weight, Sigma, Return Gap, Fund Flow, Treynor ratio, Fund Size and Sharpe ratio covariates achieving annualized alphas of at least 0.17%. By contrast, the $FDR10\%$, without using the information of fund characteristics, winds up with a small negative alpha of -0.05%. It is noteworthy that the performance of the $fFDR10\%$ and the $FDR10\%$ portfolios are affected by the Covid-19 pandemic period, which is marked by shaded area. Prior to the event, we observe that all portfolios seem to rebounce and gain alphas ranging from 0.53% to 1.58% for the $fFDR10\%$ portfolios and 0.32% for the $FDR10\%$ one. For this reason, in the followings, we report results of two samples: one ends in 2019 and one ends in 2022.

[Insert Figure 4 approximately here]

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 Panel A of Table 3, we report the average alpha that the investors will gain if they invest for $n \in \{5, 10, 15, 20, 30, 35, 40, 41\}$ 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 fund characteristic and the $FDR10\%$. We note that the last case, $n = 41$, corresponds to the alphas for the whole period from January 1982 to December 2022 and are the last points in the plots in Figure 4. Panel B of the same table reports similar metrics but for period from 1982 to the end of 2019 and $n \in \{5, 10, 15, 20, 30, 35, 38\}$.

[Insert Table 3 approximately here]

In both two panels, we find that the $fFDR10\%$ portfolios outperform the $FDR10\%$ portfolio 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.

C. Combined covariates

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 fund characteristics and potential consequent performance improvement. More specifically, we create a new covariate given by the linear combination of the underlying fund characteristics. More specifically, for each fund i at time t, we have

$$\begin{aligned}\text{New Covariate}_{t,i} &= c_{1,t}\text{R-square}_{t,i} + c_{2,t}\text{Active Weight}_{t,i} + c_{3,t}\text{Return Gap}_{t,i} \\ &\quad + c_{4,t}\text{Fund Size}_{t,i} + c_{5,t}\text{Fund Flow}_{t,i} + c_{6,t}\text{Expense Ratio}_{t,i} \\ &\quad + c_{7,t}\text{Sharpe Ratio}_{t,i} + c_{8,t}\text{Treynor Ratio}_{t,i} \\ &\quad + c_{9,t}\text{Sigma}_{t,i} + c_{10,t}\text{Beta}_{t,i}. \end{aligned} \tag{15}$$

We consider two approaches to estimate the coefficients $c_{1,t}, \ldots, c_{10,t}$ in (15). First, we use as our new covariate the first principal component of all ten (standardized) fund characteristics. By transforming the fund characteristics 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 (15) with the regression's estimated coefficients as the $c_{1,t}, \ldots, c_{10,t}$. Thereby, we consider four linear regression models including ordinary least squares (OLS), the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996), ridge regression and the elastic net of Zou and Hastie (2005).[30]

Figure 5 exhibits the performance of the $fFDR\tau$ portfolios with the newly created

[30] 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.

covariates in terms the alpha evolution.[31] We find that the portfolio based on the combined covariate obtained from the ridge and elastic net perform best amongst the combined covariates at $\tau = 10\%$.

[Insert Figure 5 approximately here]

In Table 4 we show the average n-year alphas of the $fFDR10\%$ portfolios from January 1982 to December 2022 (Panel A) and to December 2019 (Panel B). The elastic net performs also better for all time lengths except the longest ones of each considering the full sample periods. However, the best combined covariate does not beat the Beta under the $fFDR$ framework (as shown in Table 3).[32]

[Insert Table 4 approximately here]

D. Comparison with single- and double-sorting portfolios

We also 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

[31]There are a few years where LASSO and the elastic net 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.

[32]In Section XII of the Internet Appendix we provide a detailed comparison of all the $fFDR\tau$ portfolios in regard to several trading metrics, whereas in Section XIII of the same appendix the performance in terms of wealth evolution is presented. In Section XIV of the Internet Appendix, we further partition our sample into four non-overlapping sub-periods including the first three decades 1982–1991, 1992–2001, 2002–2011, and the remainder. Overall, we see that all portfolios perform well in terms of alpha in first two sub-periods, then decline in third sub-period. Two-third of the $fFDR^+$ portfolios rebound in the remaining period up to 2019, but all portfolios are decreasing if the pandemic years are included. In terms of Sharpe ratio, however, all portfolios gain the highest reports in the final sub-period.

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. For the interest of space, we present the results in Section XV of the Internet Appendix. Thereby, the portfolios based on $fFDR$ gain positive alphas and beat the corresponding sorted portfolios in most cases. These results further validate the advantage of our method in exploiting the non-linear relationship of the fund characteristics, 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 XVI of the Internet Appendix, we further examine a combination of the FDR^{+} procedure and covariates via constructing $FDR10\%$ portfolios in each quintile based on the covariates. We see that such combination cannot substitute our $fFDR^{+}$ approach.

As further robustness checks, in Section XVII 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 XVIII 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.

VII. Concluding Discussion

In this paper, we introduce the $fFDR^{+}$, a novel multiple hypothesis testing framework, that incorporates fund characteristics to assess the conditional performance of mutual funds by controlling data snooping bias. We conduct simulation experiments to assess how well our method performs in controlling the FDR and raising power compared to prior FDR methods. We then construct empirical portfolios based on our new method and ten fund characteristics as informative covariates. We study six characteristics, 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 the FDR and incorporated characteristics advances the generation of more positive and higher alphas than a portfolio that controls the 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–such as Kosowski et al. (2006), Andrikogiannopoulou and Papakonstantinou (2016), Harvey and Liu (2020) and Grønborg, Lunde, Timmermann, and Wermers (2021)–all these have their merits and present 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 fund characteristics, 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 six traditional mutual fund characteristics can offer important information. 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 for luck and noise. Our method ensures that. We also introduce four new characteristics and find that their information in our context is important and surpasses that of traditional ones; a finding that is expected to be of interest to investment managers who are concerned with portfolio performance in a timely manner.

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 Section V.C). 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.

References

Amihud, Y., and R. Goyenko. "Mutual Fund's R^2 as Predictor of Performance." *The Review of Financial Studies*, 26 (2013), 667–694.

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

Andrikogiannopoulou, A., and F. Papakonstantinou. "Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power?" *The Journal of Finance*, 74 (2019), 2667–2688.

Bajgrowicz, P., and O. Scaillet. "Technical trading revisited: False discoveries, persistence tests, and transaction costs." *Journal of Financial Economics*, 106 (2012), 473–491.

Barras, L.; P. Gagliardini; and O. Scaillet. "Skill, Scale, and Value Creation in the Mutual Fund Industry." *The Journal of Finance*, 77 (2022), 601–638.

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

Barras, L.; O. Scaillet; and R. Wermers. "Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? A Reply." *The Journal of Finance*. Replications and Corrigenda (web-only).

Benjamini, Y., and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)*, 57 (1995), 289–300.

Berk, J. B., and J. H. van Binsbergen. “Measuring skill in the mutual fund industry.” *Journal of Financial Economics*, 118 (2015), 1–20.

Bessembinder, H.; M. J. Cooper; and F. Zhang. “Mutual fund performance at long horizons.” *Journal of Financial Economics*, 147 (2023), 132–158.

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

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

Chordia, T.; A. Goyal; and A. Saretto. “Anomalies and False Rejections.” *The Review of Financial Studies*, 33 (2020), 2134–2179.

Clifford, C. P.; J. A. Fulkerson; R. Jame; and B. D. Jordan. “Salience and Mutual Fund Investor Demand for Idiosyncratic Volatility.” *Management Science*, 67 (2021), 5234–5254.

Doshi, H.; R. Elkamhi; and M. Simutin. “Managerial Activeness and Mutual Fund Performance.” *The Review of Asset Pricing Studies*, 5 (2015), 156–184.

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

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

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

Hansen, P. R. “A Test for Superior Predictive Ability.” *Journal of Business & Economic Statistics*, 23 (2005), 365–380.

Harvey, C. R., and Y. Liu. “Decreasing Returns to Scale, Fund Flows, and Performance.” (2017). Working paper.

Harvey, C. R., and Y. Liu. “Detecting Repeatable Performance.” *The Review of Financial Studies*, 31 (2018), 2499–2552.

Harvey, C. R., and Y. Liu. “False (and Missed) Discoveries in Financial Economics.” *The Journal of Finance*, 75 (2020), 2503–2553.

Hsu, P.-H.; Y.-C. Hsu; and C.-M. Kuan. “Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias.” *Journal of Empirical Finance*, 17 (2010), 471–484.

Hsu, P.-H.; M. P. Taylor; and Z. Wang. “Technical trading: Is it still beating the foreign exchange market?” *Journal of International Economics*, 102 (2016), 188–208.

Ignatiadis, N., and W. Huber. “Covariate powered cross-weighted multiple testing.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83 (2021), 720–751.

Jensen, M. C. “The performance of mutual funds in the period 1945–1964.” *The Journal of Finance*, 23 (1968), 389–416.

Jones, C. S., and H. Mo. “Out-of-Sample Performance of Mutual Fund Predictors.” *The Review of Financial Studies*, 34 (2021), 149–193.

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

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

Kosowski, R.; A. Timmermann; R. Wermers; and H. White. “Can Mutual Fund “Stars” Really Pick Stocks? New Evidence from a Bootstrap Analysis.” *The Journal of Finance*, 61 (2006), 2551–595.

Newey, W. K., and K. D. West. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, 55 (1987), 703–708.

Newton, M. A.; A. Noueiry; D. Sarkar; and P. Ahlquist. “Detecting differential gene expression with a semiparametric hierarchical mixture method.” *Biostatistics*, 5 (2004), 155–176.

Romano, J. P., and M. Wolf. “Stepwise Multiple Testing as Formalized Data Snooping.” *Econometrica*, 73 (2005), 1237–1282.

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

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

Storey, J. D.; J. M. Akey; and L. Kruglyak. “Multiple Locus Linkage Analysis of Genomewide Expression in Yeast.” *PLOS Biology*, 3 (2005), 1380–1390.

Storey, J. D.; J. E. Taylor; and D. Siegmund. “Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66 (2004), 187–205.

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

Sullivan, R.; A. Timmermann; and H. White. “Data-Snooping, Technical Trading Rule Performance, and the Bootstrap.” *The Journal of Finance*, 54 (1999), 1647–1691.

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

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

Wen, K., and X. Wu. “An Improved Transformation-Based Kernel Estimator of Densities on the Unit Interval.” *Journal of the American Statistical Association*, 110 (2015), 773–783.

Wermers, R. “Mutual Fund Herding and the Impact on Stock Prices.” *The Journal of Finance*, 54 (1999), 581–622.

White, H. “A Reality Check for Data Snooping.” *Econometrica*, 68 (2000), 1097–1126.

Zheng, L. “Is Money Smart? A Study of Mutual Fund Investors’ Fund Selection Ability.” *The Journal of Finance*, 54 (1999), 901–933.

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

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 284 observations per fund and are generated with cross-sectional independence.

Panel A: discrete distribution.

(π^+,π_0,π^-)	Procedure	$\alpha^*=1.5$	$\alpha^*=2$	$\alpha^*=2.5$	$\alpha^*=3$	$\alpha^*=3.5$
$(10,75,15)\%$	$fFDR^+$	1	6.8	23.9	46.6	68.7
	FDR^+	0.6	2.9	13.9	33.6	55.3
$(10,60,30)\%$	$fFDR^+$	2	12.6	35.5	59.6	77.8
	FDR^+	0.5	3.4	16.2	37.7	58.5
$(10,30,60)\%$	$fFDR^+$	5.5	26	54	77.6	90.2
	FDR^+	0.6	5.3	23.3	49.9	71.3
$(13,67.5,19.5)\%$	$fFDR^+$	1.8	11.5	32.8	56.7	76.7
	FDR^+	0.7	5	19.9	41.7	62.8
$(13,48,39)\%$	$fFDR^+$	3.8	19.3	44.6	70	85.1
	FDR^+	0.7	5.5	23.5	48.5	68.3
$(13,9,78)\%$	$fFDR^+$	9.7	37.6	70.7	91.5	97.8
	FDR^+	0.9	10	41	73.4	89.8

Panel B: discrete-normal distribution mixture.

π_0	Procedure	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$
75%	$fFDR^+$	0.6	16.8	37.3	51.8	61.3
	FDR^+	0.3	9.2	27.7	42.4	52.9
60%	$fFDR^+$	1.8	22.6	44.2	58.1	67.2
	FDR^+	0.4	12.3	32.8	47.5	57.8
30%	$fFDR^+$	5.1	32.9	54.9	68.1	75.5
	FDR^+	0.6	18.7	41.3	56.5	66.1
67.5%	$fFDR^+$	1.1	20.1	40.9	55.3	64.2
	FDR^+	0.3	11	30.4	45.3	55.7
48%	$fFDR^+$	3.2	27.9	49.1	62.8	71.6
	FDR^+	0.4	15.4	36.4	51.5	61.4
9%	$fFDR^+$	7.5	39.8	62.2	74.6	81.4
	FDR^+	0.9	23.5	48.7	63.9	73.1

Panel C: mixture of two normal distributions.

				(σ_1,σ_2)		
(μ_1,μ_2)	Procedure	$(1,0.5)$	$(1.5,0.6)$	$(2,1)$	$(2.5,1.25)$	$(3,1.5)$
		$\pi^+=6\%$	$\pi^+=10.4\%$	$\pi^+=20.7\%$	$\pi^+=25.5\%$	$\pi^+=29.1\%$
$(-2.3,-0.7)$	$fFDR^+$	0.1	0.5	5.8	14.4	24.5
	FDR^+	0	0	0.4	2.4	8.1
		$\pi^+=11.8\%$	$\pi^+=16.9\%$	$\pi^+=26.4\%$	$\pi^+=30.5\%$	$\pi^+=33.4\%$
$(-2,-0.5)$	$fFDR^+$	0.1	0.7	7	16.5	26.5
	FDR^+	0	0	0.6	3.6	10.1
		$\pi^+=35.2\%$	$\pi^+=36.4\%$	$\pi^+=38.2\%$	$\pi^+=39.8\%$	$\pi^+=41.1\%$
$(-2.5,0)$	$fFDR^+$	0.5	1.1	9.9	19.3	29.4
	FDR^+	0	0.1	1.1	5.1	12.7

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.

		FDR target								
Distribution	Procedure	10	20	30	40	50	60	70	80	90
Discrete	$fFDR^+$	0.4	2.9	9	19.6	34	50.6	66	77.6	85.9
	FDR^+	0.4	1.1	2.2	4.3	8.4	15.6	27.4	45.6	68.5
Mixture of discrete and normal	$fFDR^+$	0.4	1.5	3.7	7.3	12.7	20.8	32.4	47.4	64.4
	FDR^+	0.2	0.5	0.7	1.2	1.9	3.3	5.7	11.4	27.4
Mixture of normals	$fFDR^+$	0	0.2	0.5	1.4	3.1	6.5	12.3	21.6	35.3
	FDR^+	0	0	0	0.1	0.1	0.1	0.1	0.1	0.2

TABLE 3

Comparison of portfolios' alphas (in %) for varying time lengths of investing

In this table, we consider 11 portfolios including 10 $fFDR10\%$ portfolios corresponding to the ten 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 2018 to December 2022. The average of these alphas is presented in the first rows in Panel A of the table. Panel B report similar metrics with use of portfolios' return series from January 1982 to December 2019.

	$fFDR10\%$										$FDR10\%$
n	R-square	Fund Size	Active Weight	Return Gap	Fund Flow	Expense	Sharpe	Treynor	Beta	Sigma	
Panel A: Whole sample											
5	0.95	0.10	0.78	0.57	0.19	1.37	-0.03	0.10	1.13	0.59	-0.06
10	0.88	0.01	0.72	0.41	0.30	1.07	0.08	0.18	1.25	0.49	-0.21
15	0.98	0.04	0.75	0.35	0.35	0.93	0.20	0.29	1.33	0.53	-0.23
20	1.21	0.26	0.93	0.50	0.50	0.96	0.39	0.44	1.45	0.80	-0.03
25	1.13	0.20	0.81	0.43	0.45	0.87	0.38	0.44	1.44	0.67	-0.06
30	0.88	0.10	0.62	0.33	0.36	0.74	0.29	0.36	1.30	0.41	-0.20
35	0.77	0.07	0.67	0.38	0.28	0.94	0.22	0.27	1.12	0.25	-0.23
40	0.76	0.12	0.65	0.22	0.24	0.90	0.17	0.22	1.01	0.26	-0.16
41	0.84	0.20	0.73	0.31	0.33	0.94	0.17	0.24	1.11	0.34	-0.05
Panel B: Sample period prior to December 2019											
5	1.08	0.16	0.94	0.70	0.26	1.33	0.03	0.17	1.29	0.68	-0.01
10	1.13	0.23	0.91	0.56	0.48	1.12	0.30	0.40	1.55	0.72	0.02
15	1.38	0.43	1.07	0.67	0.64	1.18	0.51	0.59	1.58	0.97	0.16
20	1.52	0.61	1.22	0.80	0.77	1.21	0.65	0.72	1.70	1.16	0.30
25	1.27	0.41	0.91	0.56	0.60	0.95	0.51	0.59	1.62	0.85	0.12
30	1.08	0.30	0.78	0.51	0.51	0.85	0.44	0.52	1.54	0.55	-0.02
35	0.86	0.12	0.84	0.51	0.28	1.05	0.17	0.23	1.19	0.28	-0.21
38	1.29	0.57	1.05	0.62	0.70	1.39	0.53	0.60	1.58	0.72	0.32

TABLE 4

Average n-year alpha of $fFDR10\%$ portfolios with combined covariates

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 5). The average n-year alpha of each portfolio is calculated as per the description in Table 3.

n	OLS	Ridge	LASSO	Elastic Net	PC 1
Panel A: Whole sample					
5	0.49	1.07	0.89	1.14	0.89
10	0.55	1.05	0.82	1.15	0.95
15	0.68	1.03	0.80	1.18	0.96
20	0.90	1.19	0.96	1.31	1.07
25	0.83	1.10	0.86	1.24	1.02
30	0.68	0.97	0.70	1.07	0.92
35	0.69	0.99	0.69	1.02	0.97
40	0.51	0.84	0.55	0.82	0.71
41	0.55	0.89	0.61	0.86	0.69
Panel B: Sample period prior to December 2019					
5	0.61	1.24	1.05	1.33	1.03
10	0.70	1.23	1.00	1.38	1.05
15	0.92	1.35	1.13	1.49	1.17
20	1.09	1.52	1.30	1.66	1.28
25	0.87	1.27	1.03	1.42	1.09
30	0.77	1.16	0.87	1.24	1.03
35	0.72	1.15	0.84	1.15	1.00
38	0.81	1.20	0.89	1.17	0.99

FIGURE 1

Performance of the $fFDR^{+}$ 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.

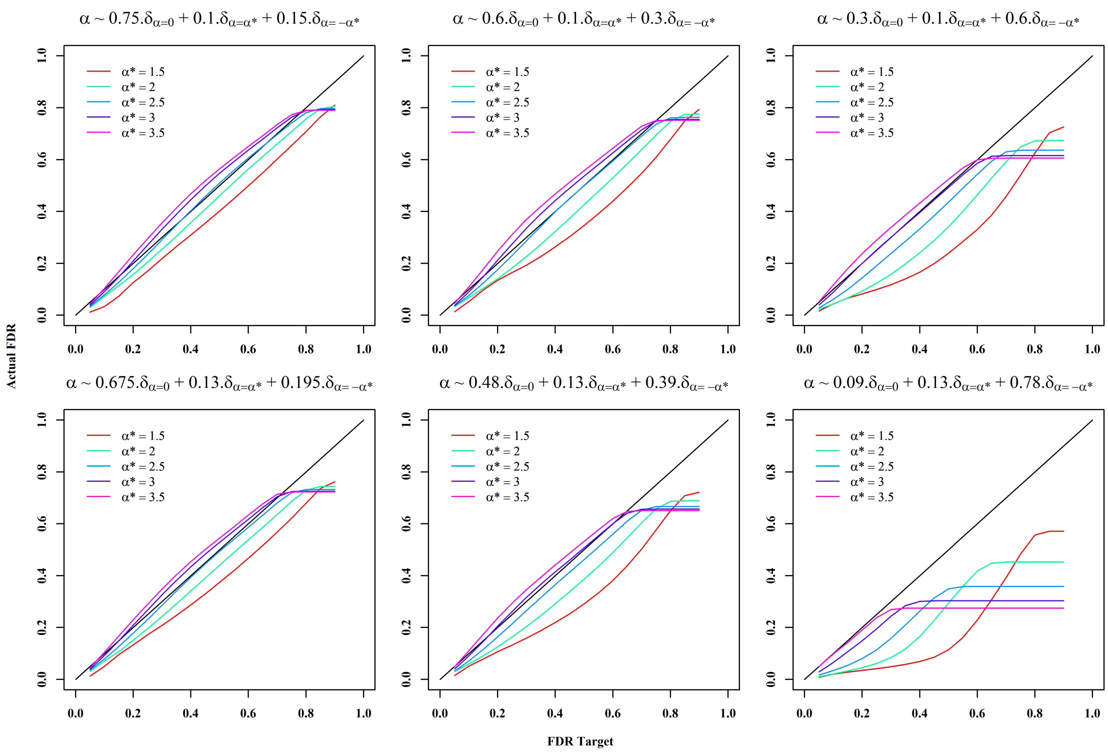


FIGURE 2

Performance of the $fFDR^{+}$ 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.

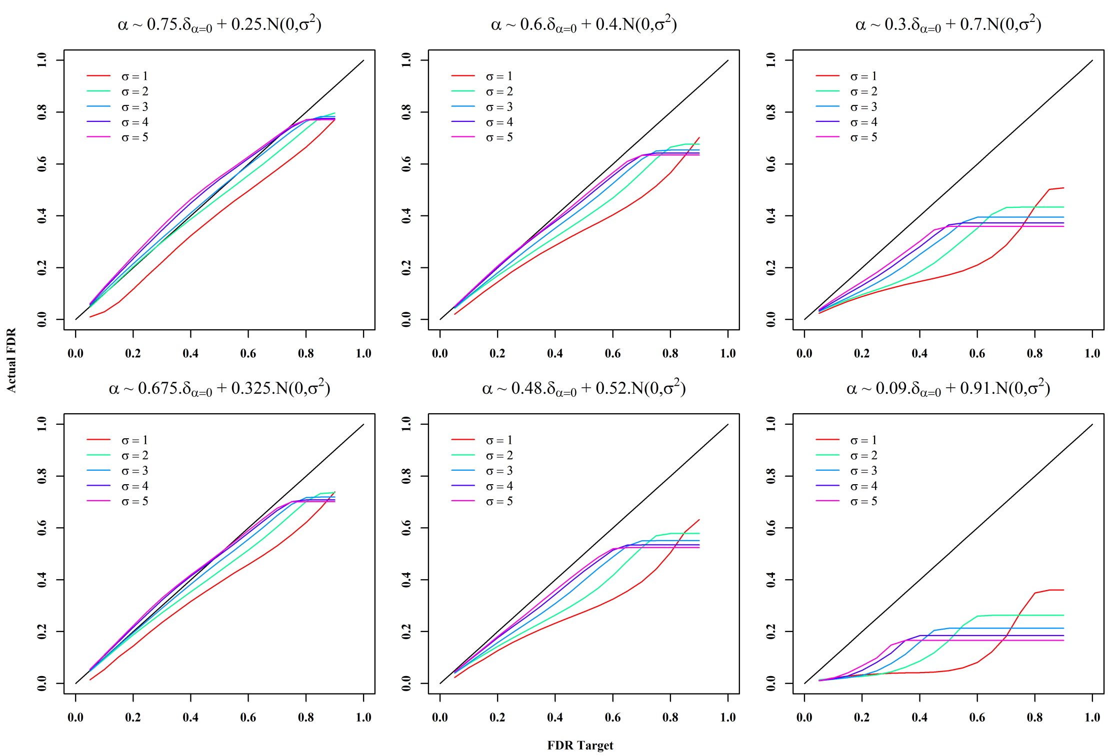


FIGURE 3

Performance of the $fFDR^{+}$ 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.

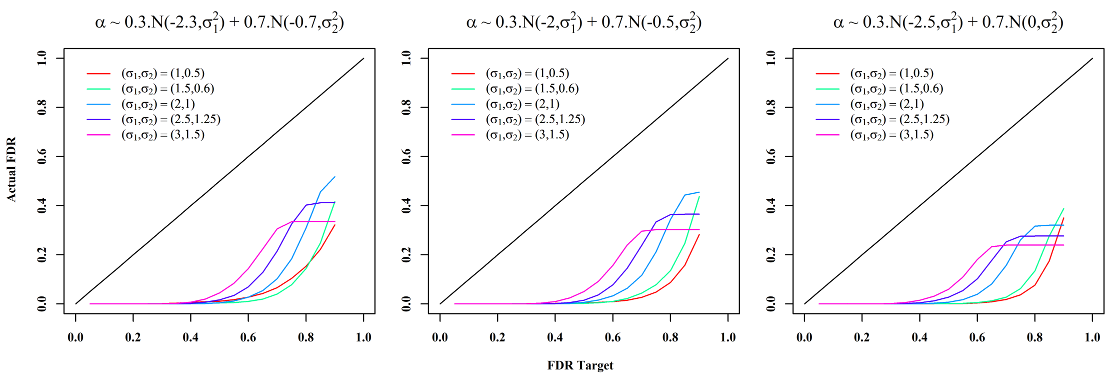


FIGURE 4

Alpha evolution of $fFDR10\%$ and $FDR10\%$ portfolios over time

The graph presents the evolution of annualized alphas (in $\%$) of the ten $fFDR10\%$ portfolios corresponding to the ten covariates, the $FDR10\%$ portfolio of BSW and the two equally weighted portfolios.

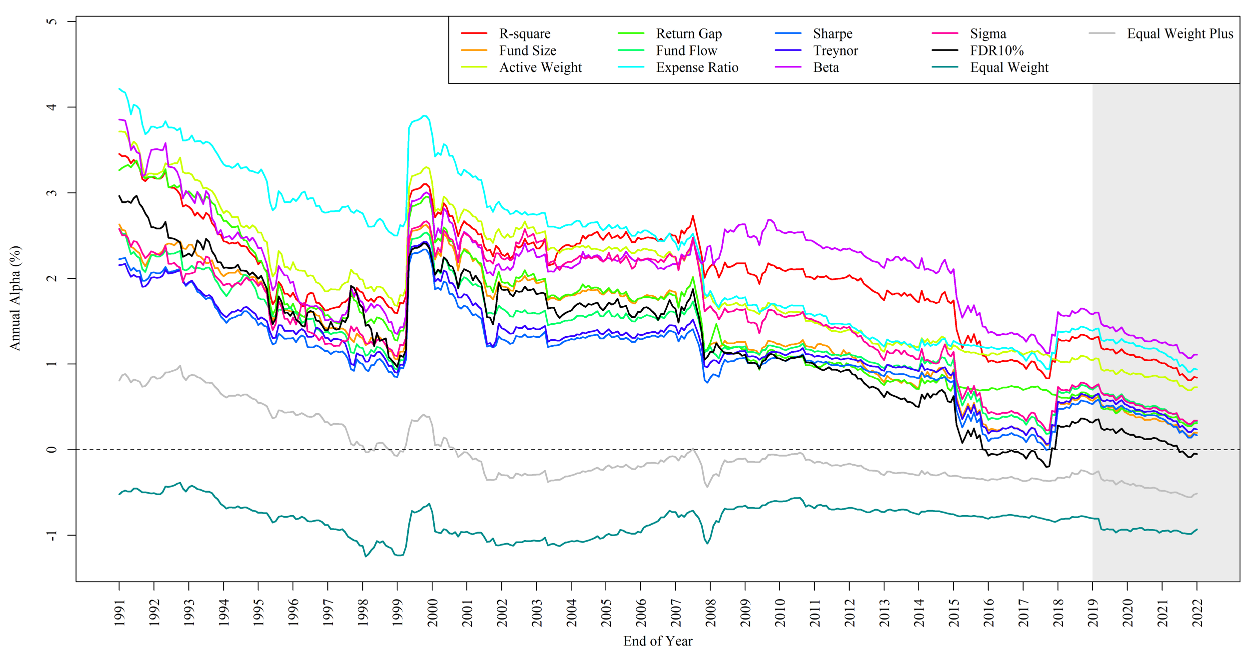
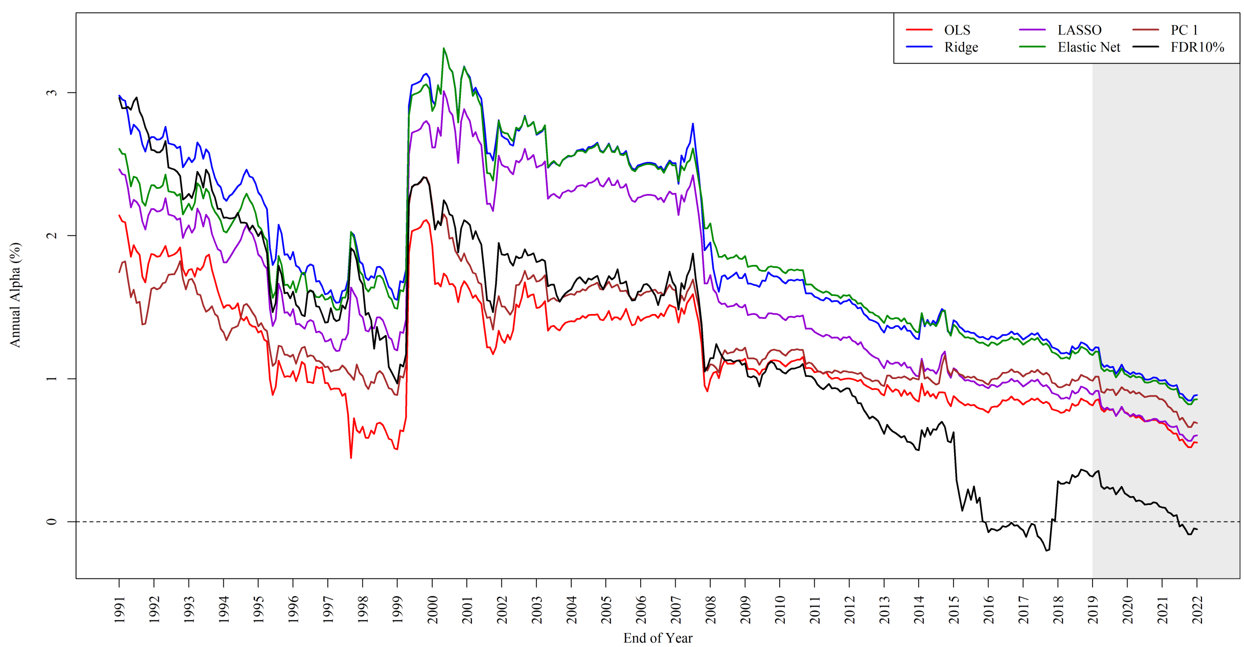


FIGURE 5

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.



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 A). 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

$$\hat{\pi}_{0,b}(\lambda) = \frac{\#\{p_i > \lambda, z_i \in K_b\}}{(1-\lambda)\#K_b}, \quad b = 1, 2, \ldots, n. \tag{16}$$

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

$$\text{MISE}(\lambda) = \text{bias}^2 + \text{variance} = \left(\int_0^1 \phi(z, \lambda)dz - \pi_0\right)^2 + \int_0^1 [\hat{\pi}_0(z, \lambda) - \phi(z, \lambda)]^2 dz \tag{17}$$

We estimate π_0 using the smoothing spline method of Storey and Tibshirani (2003, Remark B).[33] 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 $\hat{\phi}(\lambda, z) = \hat{\pi}_0(z, \Lambda_{\text{min}}) - c_\lambda(1 - \hat{\pi}_0(z, \Lambda_{\text{min}}))$, 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

[33] 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).

$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$ 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.[34] 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.

[34]This approach is to overcome the erratic behaviour at boundary observed when using other popular approaches. Recently, Wen and Wu (2015) propose an alternative method to the issue. We additionally conduct simulations to compare the two approaches and see that the local likelihood KDE approach is computationally more efficient and performs better under our framework (FDR control and power gain). For the sake of space, the results are available upon request.

Internet Appendix for

"Mutual Funds' Conditional Performance Free of Data Snooping Bias"

I. Comparison of Procedures' Estimation

In this section, we recall the procedure of BSW in estimating FDR for selecting out-performing mutual funds and illustrate the differences with our $fFDR^+$ procedure.

The starting point for both procedures is to controlling for the type I error as in Benjamini and Hochberg (1995):

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

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) and Storey (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 $t_1,\dots,t_m$ are independent and the statuses of the null hypotheses $h_1,\dots,h_m$ 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 as

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i|p_i>\lambda, i=1,\dots,m\}}{(1-\lambda)m} \tag{2}$$

where $\#$ returns the number of elements in the set; this estimate is conservative biased.[35] BSW

[35]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 $[\lambda,1]$. This number can be conservatively

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.[36] We set $\hat{\pi}_0 = \hat{\pi}_0(\lambda^*)$.

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

$$\text{FDR}^+ = \mathbb{E}\left(\frac{V^+}{\max\{R^+, 1\}}\right). \tag{3}$$

With a significant threshold γ and a procedure which selects a fund with a positive estimated alpha whenever its p-value $\leq \gamma$, BSW estimate FDR^+ as

$$\widehat{\text{FDR}}^+_\gamma = \frac{\hat{\pi}_0 \gamma/2}{\hat{R}^+/m}, \tag{4}$$

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, we estimate the FDR$^+$ based on a grid of the threshold γ and use as the rejection threshold the one that producing $\widehat{\text{FDR}}^+$ closest to the target τ. We refer to this procedure as the FDR^+.

Next, we conduct an illustration to show the differences in estimation between the $fFDR^+$ and FDR^+. For this, 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

approximated by $\#\{p_i | p_i > \lambda\}$, from which we get (2). For a larger λ, the estimate $\hat{\pi}_0$ becomes less conservative, as there are fewer p-values under the alternative belonging to $[\lambda, 1]$, but its variance gets higher.

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

covariate for $fFDR^+$.[37] In Figure IA1, we demonstrate how the two procedures work. Based on the p-values of all considered funds, the FDR^+ estimates the proportion of zero-alpha funds in the whole sample, as a first step, giving $\hat{\pi}_0 \approx 0.84$. It then selects the positive estimated alpha funds with smallest p-values until the estimated $\widehat{\text{FDR}}_\gamma^+$ reaches a given FDR target. For the sake of exemplification, we choose the FDR target $\tau = 45\%$, so that both methods select a substantial number of funds.[38] Here, all funds with p-values less than or equal to $\gamma = 0.008$ are selected by the FDR^+. The threshold γ is depicted by the green dashed line in Panel C and all 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).

[Insert Figure IA1 approximately here]

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". In this particular case, $\hat{\pi}_0(z)$ is a non-decreasing function of z. 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 funds with estimated q-value less than or equal to 0.45: those funds correspond to the points below the red dashed line (the q-value $= 0.45$ line) in Panel C. This clearly shows that, for the same target, the $fFDR^+$ selects significantly more funds than FDR^+ (185 versus 16). 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,

[37]The details of the funds and the calculation of the p-values are deferred to Section VI.

[38]If we choose any target $\tau \leq 40\%$, the FDR^+ selects no funds.

smaller R-square), but the weight is not equally distributed across the funds with the same level of z. To explain this, we investigate further the second component of the posterior $r(p, z)$, the density function $f(p, z)$. We produce in Figure IA2 the heatmap of the density function and see that the value of $f(p, z)$ is higher where z is small. This combines with the low value of $\pi_0(z)$ for small z, which is presented in Panel B of Figure IA1, implying a low value of the posterior probability of being null $r(p, z)$ – which is used to determine the threshold for rejecting the null. Consequently, more funds with small z are selected by the $fFDR^+$ as profitable, regardless of the fact that the $p-$value may get some relatively high values (0.2–0.6) as shown in Panel D of Figure IA1.

[Insert Figure IA2 approximately here]

II. 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 (11) or (12), we determine c in (14) 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 (11) or (12) and assign them such that a fund with a smaller z has a smaller alpha. For the case (13), 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 (10). We consider balanced panel data for $2,000$ funds with 284 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 284 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,

$$\varepsilon_{i,t} = \gamma G_t + \gamma G_t^+ \mathbb{1}_{\alpha_i>0} + \gamma G_t^- \mathbb{1}_{\alpha_i<0} + \eta_{i,t}, \tag{5}$$

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 equates 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 ratio of falsely classified funds $\widetilde{FDR}^+$ and a detected ratio $\widetilde{Power}^+$:

$$\widetilde{FDR}^+ = \frac{\widetilde{V}^+}{\max\left\{\widetilde{R}^+, 1\right\}} \quad \text{and} \quad \widetilde{Power}^+ = \frac{\widetilde{C}^+}{\widetilde{T}^+}, \tag{6}$$

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.

III. Variance Comparison of FDR Estimation

In this section, we investigate the performance of the two methods in terms of FDR estimation variance. As described in Section II, the actual FDR of the two methods is estimated by the average of the ratio of falsely classified funds $\widetilde{FDR}^+$. As the iterations are independent, the variance of the estimated actual FDR is proportionate to the variance of the $\widetilde{FDR}^+$. In

Figures IA3, IA4 and IA5, we report the gap in variance of the $\widetilde{FDR}^{+}$ of the FDR^{+} over the $fFDR^{+}$. We observe that the gap curves are either varying close to zero or positive for most of the cases of the distributions. This implies that the variance of the estimated actual FDR of the $fFDR^{+}$ is less than that of the FDR^{+}.

[Insert Figure IA3 approximately here]

[Insert Figure IA4 approximately here]

[Insert Figure IA5 approximately here]

IV. Additional Simulation Results

In supplementing Section V of the main manuscript, we present 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 balanced panel data under cross-sectional dependence, before proceeding to the results for unbalanced panel data under both cross-sectional independence and dependence. Finally, we exhibit simulation results for the case where alphas are drawn from a single normal distribution.

A. Results for balanced panel data under cross-sectional dependence

We present in Figures IA6–IA8 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 reported in Tables IA2–IA6.

[Insert Figure IA6 approximately here]

[Insert Figure IA7 approximately here]

[Insert Figure IA8 approximately here]

[Insert Table IA2 approximately here]

[Insert Table IA3 approximately here]

[Insert Table IA4 approximately here]

[Insert Table IA6 approximately here]

B. Results for unbalanced panel data

We present the performance of the $fFDR^{+}$ under both cross-sectional independence and dependence. Figures IA9–IA11 depict the FDR control of the $fFDR^{+}$, while the power comparisons are given in Tables IA7–IA9.

[Insert Figure IA9 approximately here]

[Insert Figure IA10 approximately here]

[Insert Figure IA11 approximately here]

[Insert Table IA7 approximately here]

[Insert Table IA8 approximately here]

[Insert Table IA9 approximately here]

C. Simulation results for single normal distribution

We present the simulation results for a special case of continuous distribution where the mixture (13) 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 $\sigma \in \{1, 1.5, 2, 2.5, 3\}$ (both parameters' values are annualized and in % terms).[39]

Figures IA12 and IA13 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.

[Insert Figure IA12 approximately here]

[Insert Figure IA13 approximately here]

In Table IA10, we compare 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

[39] 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.

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%.

[Insert Table IA10 approximately here]

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

We consider three forms of $\pi_0(z)$, including decreasing, increasing and constant with respect to z. For the first two 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 in Tables IA11–IA13 results in terms of power 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.

[Insert Table IA11 approximately here]

[Insert Table IA12 approximately here]

[Insert Table IA13 approximately here]

V. Performance of $fFDR^+$ when Using a Non-informative Covariate

In what follows, we present the simulation results when a non-informative covariate is used instead of the informative as in the simulations in the main paper. The simulated data is the same as in the main paper, except that for each iteration a covariate is drawn randomly from the uniform distribution on $[0, 1]$ and is used as covariate input of the $fFDR^+$. This covariate is non-informative in that it has no connection to the true alpha of funds and, thus, no information for detecting truly positive alpha funds. We see that the $fFDR^+$ controls well FDR under all alpha distributions, similarly to Figures 1, 2 and 3 main paper. In the interest of space, these results are not reported but can be made available upon request. In terms of power, the $fFDR^+$ with use of the non-informative covariate performs very similarly to the FDR^+ as exhibited in Table IA14.

[Insert Table IA14 approximately here]

VI. Varying Number of Observations and Funds

In the simulations in the main paper, we have assumed a sample of $m = 2{,}000$ funds, which reflects our actual dataset for the whole period from 1975 to 2022. 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 IA15, 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 (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. The power of the $fFDR^+$ slightly decreases with rising m, 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 difference between the $fFDR^+$ and the FDR^+ can reach 34%.

[Insert Table IA15 approximately here]

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.[40] In Table IA16, 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.

[Insert Table IA16 approximately here]

VII. Estimation Errors in Covariates

In the simulations in the main paper, we consider a simple covariate where in the set of *non-zero* alpha funds, the ranking of the funds' alpha is the same as that of the 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 funds from the zero-alpha and the under-performing ones.

Here, we further study the behaviour of our $fFDR^+$ approach by adding a noise to the original covariate that reflects potential estimation biases, as all covariates in the real data are

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

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

$$z'_i = z_i + \eta_i, \tag{7}$$

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 fund characteristic which aims to capture Z. Depending on the scale of the estimation error, the realized covariate can have different levels of information. In reality, we do not know the actual estimation errors in the covariates. Thus, we simulate low to high noise in our covariates. In particular, we consider two 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 Section V.A in the main paper. This is the most important characteristic of the $fFDR^+$ we would expect, that is, the ability to control well for the risk even when the new information contains noise.

In Table IA17, 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 can assume that the covariates in our application have relatively less noise than the ones in this simulation.

[Insert Table IA17 approximately here]

VIII. Fund Characteristics as Informative Covariates

As part of our empirical investigation of the $fFDR^{+}$ approach, we consider six 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. When a fund replicates the market, the R-square is close to one; if, instead, it is more active, it has 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

$$\text{Fund Size}_{i,t} = \ln \frac{\text{TNA}_{i,t}}{\text{IndustrySize}_t} - \ln \frac{\text{TNA}_{i,0^*}}{\text{IndustrySize}_{0^*}}, \tag{8}$$

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.[41]

[41] Pastor, Stambaugh, and Taylor (2015) and Chen, Hong, Huang, and Kubik (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. Other studies on the relationship between fund size and performance and funds'

The third covariate is the Return Gap of Kacperczyk et al. (2008), which aims 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. They show that funds with higher active weight tend to perform better. We note that the active weight is also related to the fund's turnover, which plays a role in explaining performance as pointed out in Pastor, Stambaugh, and Taylor (2017) and BSW. 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), Capponi, Glasserman, and Weber (2020) and Bessembinder, Chen, Cooper, Xue, and Zhang (2023), among others. Zheng (1999), in particular, discovers that funds receiving money perform better than those that lose money. The author also shows that investors

holding liquidity (e.g., Yan (2008)) or funds' merger (i.e., McLemore (2019)) document the same conclusion. Fund size is also strongly related to skills as great investment idea is difficult to scaled up (see Barras et al. (2022)).

can earn abnormal returns using small funds' flow information. Here, we follow Bris, Gulen, Kadiyala, and Rau (2007) and define Fund Flow at time t as

$$\text{Fund Flow}_t = \frac{\text{TNA}_t - (1 + r_t)\text{TNA}_{t-1}}{(1 + r_t)\text{TNA}_{t-1}}, \tag{9}$$

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

Finally, we study the information carried by expenses and fees, which is reflected in expense ratio. The impact and informativeness of this funds' characteristic on active mutual fund performance has been discussed by Berk and Green (2004), BSW, and Berk and van Binsbergen (2015).

IX. Data-based Simulations

In this simulation experiment, we design a setting close to our later empirical exercises, in which the simulation data retains the dependencies among alphas and covariates as in the real data.[42] Since we construct portfolios based on in-sample of five-year data (as will be presented in Section VI in the main paper), we opt data on returns and covariates of all funds in a five-year period from January 1999 to December 2003. This is the period that offers us the largest number of funds (1,567) having all ten covariates.

[42] We additionally conduct a simulation set which is similar to BSW and Andrikogiannopoulou and Papakonstantinou (2019). Our conclusions on the advantages of the covariate-augmented method remain under this setting regardless of different alpha distributions, balanced and unbalanced data structures, and cross-sectional dependent or independent error terms. For the sake of space, the results are available upon request.

A. Data generating process

The data generating process is as follows.

- First, we calculate each fund's alpha, beta coefficients (b, s, h, m) and residuals of the Carhart four-factor model. We then calculate the correlation coefficients between alpha and all betas (b, s, h, m) across all 1,567 funds. For each fund, we also calculate the covariates mentioned in Section VIII where R-square, Beta, Sigma, Sharpe and Treynor ratios are from the asset pricing models, whereas Active Weight, Return Gap, FundSize, Expense Ratio and Fund Flow are obtained from averaging the available values realised over the five years.[43] Similarly, we calculate the correlation coefficient matrix of 11 vectors for alpha and ten covariates across 1,567 funds. To determine the probability of being truly zero-alpha of each fund, we estimate $\pi_0(z)$ corresponding to each of the 10 covariates, then average across the covariates to have an empirically representative $\hat{\pi}_0(z)$.

- Second, in each iteration of the simulation,

  - we generate simulated alphas for funds such that the correlation coefficients between the alpha and the b, s, h, m are the same as in the real counterpart. We then assign zero for zero-alpha funds based on the representative $\hat{\pi}_0(z)$. We denote the simulated alpha as α^s.

  - we generate 10 simulated covariates such that the matrix of correlation coefficients

[43]The results are similar if we instead use the average over the final year. Averaging available values of each covariate over the five years gives us a higher number of funds as there are a number of funds missing covariate values in the final year.

among the 11 vectors including covariates and α^s is the same as the real counterpart calculated above.[44]

- we generate the simulated return of each fund via the following formula

$$R_{i,t} = \alpha_i^s + b_i.R_{mkt,t} + s_i.SMB_t + h_i.HML_t + m_i.Mom_t + \epsilon_{it}, \tag{10}$$

  where the noise $\epsilon_{i,t}$ is randomly drawn from the collected residuals via the stationary bootstrap procedure of Politis and Romano (1994) with an average block length of 10 following literature.

- we regress the simulated returns of each fund on the four factors to obtain $\hat{\alpha}$ and calculate the related $p-$value (based on two-sided $t-$tests).

- then, generated covariates are transformed to a unit interval $[0, 1]$ as the formula described in Section II.A, and we implement the FDR^+ and $fFDR^+$ procedures, control for FDR at predetermined targets, to detect truly positive Carhart alpha funds with use of the $\hat{\alpha}$s, calculated $p-$values and simulated covariates.

- In each iteration, by comparing the simulated α^s and the selected out-performing funds, we 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 estimates for the actual FDR and the power of each procedure.

We present our analyses of the performance of the $fFDR^+$ with use of each covariate. Figure IA14 depicts the performance of the $fFDR^+$ in terms of FDR control in its left panel and

[44] We present details in Section B the procedure generating the correlated vectors.

power in its right panel (presented by thin dotted or dashed lines) and compare with that of the FDR^+ (presented by a thick red solid line). First, it is clear that all procedures asymptotically control for FDR at any given target (from 0.05 to 0.95), and the $fFDR^+$ with use of any one of the covariates, gains higher power than the FDR^+. Second, the covariates differ in informativeness level, for example, the Sigma gains higher power with a gap of around 10% compared to the FDR^+, while the Expense Ratio and Fund Flow are the least powerful among the $fFDR^+$ ones.

[Insert Figure IA14 approximately here]

B. Dependent data generating process

In this section, we formally present our procedure to generate dependent data described in Section IX. Given a set of linearly independent vectors $\{X_1, \ldots, X_k\}$ in vector space $\mathbb{R}^n$, $n > k$ and (column) vector of correlation coefficients $\rho = (\rho_1, \ldots, \rho_k)'$, we generate a vector Y in $\mathbb{R}^n$ such that correlation coefficient of Y and X_i is ρ_i, $i = 1, \ldots, k$. The mechanism to generate Y is designed as follows.[45]

First, we scale $X_1, \ldots, X_k$ so that each of them has mean zero and standard deviation of one and denote by X the matrix with columns the X_is. To ease notation, we keep using these notations after scaling. Note that the correlation coefficient of Y and X_i is not effected by the aforementioned scaling.

Next, we generate a vector U of length n from the Gaussian distribution. We denote the residuals of the multivariate regression of U on X by e, which is in $\mathbb{R}^n$. We find

[45] Readers can easily implement the mechanism using the freely available `R` package `faux` of DeBruine (2021).

$X^* = \{X_1^*, \ldots, X_k^*\}$ such that the scalar product $X_i \cdot X_j^* = 1$ if $i = j$ and 0 if otherwise. This can be done via singular value decomposition $X = TDV'$, where T and V are orthogonal matrices of sizes $n \times n$ and $k \times k$, respectively, and D a $n \times k$ matrix with zeros everywhere except the main diagonal elements which are positive. The columns of $W = T\tilde{D}V'$ form the X^*, where $\tilde{D}$ satisfies $D'\tilde{D} = I_k$. Indeed, $X'W = VD'T'T\tilde{D}V' = I_k$ which is identity matrix of size k.

Finally, set $Y = W\rho + \sigma e$ where $\sigma^2 = \frac{1-\rho' Cov(W)\rho}{Var(e)}$ and Cov and Var are covariance and variance, respectively. We have that $Cov(Y, X_i) = Cov(W\rho + \sigma e, X_i) = Cov(W\rho, X_i) = \sum_{j=1}^{k} \rho_j Cov(X_j^*, X_i) = \sum_{j=1}^{k} \rho_j [\mathbb{E}(X_j^* X_i) - \mathbb{E}X_j^* \mathbb{E}X_i] = \rho_i$ since $\mathbb{E}X_i = 0$ and $\mathbb{E}(X_j^* X_i) = 0$ for $i \neq j$ and 1 otherwise. From the scaling step $Var(X_i) = 1$, and it is easy to see that $Var(Y) = Var(W\rho) + \sigma^2 Var(e) = \rho' Cov(W)\rho + \sigma^2 Var(e) = 1$. Thus, the correlation coefficient of Y and X_i is $\frac{Cov(Y,X_i)}{\sqrt{Var(Y)Var(X_i)}} = \rho_i$.

When generating the simulated α^s, the X consists of b, s, h, m, and ρ consists of correlation coefficients of α and each of the b, s, h, m.

When generating the 10 simulated covariates, given α^s and the real correlation coefficient matrix of 10 vectors, the first covariate Z_1 is generated such that its correlation coefficient with α^s equals the real one. The second covariates Z_2 is generated such that its correlation coefficients with Z_1 and α^s are the same as the real ones, and so on. The final covariate Z_9 is generated such that its correlation coefficients with $Z_1, \ldots, Z_9$ and α^s are the same those in the real correlation coefficient matrix. It is also noted that, in our simulation, n =1,567 which is a very large number comparing to 11. Thus, the k' generated vectors (α^s and $k' - 1$ covariates) are likely to be linear independent in $\mathbb{R}^n$, thus the $(k'+1)^{th}$ vector can be generated, $k' = 1, \ldots, 10$.

X. Results for Alternative Target of FDR

In this section, we repeat the exercise with the FDR target of 20%. Figure IA17 presents the alpha evolution of the individual covariates. Table IA18 shows the average n-year alpha of those portfolios. Finally, Table IA19 presents the statistic metrics for all mentioned portfolios.

[Insert Figure IA17 approximately here]

[Insert Table IA18 approximately here]

[Insert Table IA19 approximately here]

XI. Results from Using an Alternative Proxy of Covariates

Here, we present in Figure IA18 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 the final year as in the main manuscript. We see that the performance of the portfolios does not vary significantly.

[Insert Figure IA18 approximately here]

XII. Comparison of Portfolios' Trading Metrics

Here, 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 IA1.

[Insert Table IA1 approximately here]

XIII. Wealth Evolution

In Figure 4 in the main paper, we study the alpha evolution of the portfolios over time. However, an investor may be interested in the gain in value. Figure IA15 shows the growth of 1 dollar that the investor invests in each portfolio at the beginning of 1982. Ultimately, at the end of 2022, this amount grows to about 68 dollars if she chooses the $fFDR10\%$ portfolio with Active Weight as the covariate, as opposed to just 41, 51 and 48 dollars with the $FDR10\%$, the equal weight plus and equally weighted portfolios, respectively. This exercise reveals the potential profitability of an investor who had a perfect oracle in 1982 about the methods and the covariate that would play out over the next 41 years.

[Insert Figure IA15 approximately here]

Similarly to Figure IA15, in Figure IA16 we depict the wealth evolution of one dollar invested in the $fFDR10\%$ portfolios based on the combined covariates. At the end of 2022, 1 dollar grows to about 64 to 76 dollars if the investor invests in one of the $fFDR10\%$ portfolios with the covariates obtained from LASSO, elastic net, Ridge, PC1 and OLS regressions.

[Insert Figure IA16 approximately here]

XIV. Sub-period Performance

By construction, Figure 4 in the main paper contains 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 over different periods to the performance of the portfolios, we split the whole period into four non-overlapping sub-periods including the first three decades 1982–1991 (P1), 1992–2001 (P2), 2002–2011 (P3), and the remainder. For the final sub-period, we consider separately the part that does not cover the Covid-19 pandemic, that is, 2012–2019 (P4a), and the part that does, that is, 2012–2022 (P4b). We present in Table IA20 the alpha, its $t-$statistic and Sharpe ratio of portfolios (with a FDR target $\tau = 10\%$) in the sub-periods.

[Insert Table IA20 approximately here]

In terms of alphas, it is clear that all portfolios performing well in the first two sub-periods suffer a decline in the third sub-period. In P3, we observe negative alphas for the $FDR10\%$ portfolio and the $fFDR10\%$ portfolios with FundSize, Active Weight, Return Gap, Expense Ratio covariates and combined covariates based on shrinkage methods. Note that this sub-period suffers from the global financial crisis 2007–2008. In the sub-period P4a, this decrease continues for 6 out of 15 $fFDR10\%$ portfolios (including those with combined covariates) while the others rebounce. Moving from P3 to P4b, we find that alphas of all portfolios are decreasing and this reflects the severe impact of the pandemic. The t-statistic columns 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 sub-period P4a and even in P4b. This also occurs in the Equal Weight portfolio, that is, the portfolio that selects all the eligible funds in the in-sample windows and invests them equally in the following year; thus, 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. The alphas of the $fFDR10\%$ portfolios, by contrast, are nuanced depending on the covariate used; except in the sub-period P4b, most of them beat the equally weighted one in all the other sub-periods in terms of alpha (with notable exceptions of the FundSize, Active Weight and Return Gap covariates in the third sub-period).

XV. Comparison to Sorting Portfolios

First, we describe the single- and double-sorting portfolios which are traditionally constructed 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 deciles (quintiles) 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.

The performance in terms of alpha of those portfolios from 1982 to 2022 is presented in Table IA21. Our results suggest that most of the sorting portfolios, except the Active Weight and Sharpe ratio, have negative or negligible positive alphas at the end of 2022, which contrasts with 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. We see that the portfolios based on $fFDR$ gain significant positive alphas and beat the corresponding sorted portfolios in most cases.

[Insert Table IA21 approximately here]

XVI. $FDR\tau$ Portfolios Conditional on Covariates

In what follows, we combine fund characteristics and FDR^+ to construct $FDR\tau$ portfolios conditional on covariates. For each covariate, we first partition the sample into quintiles based on the value of each mutual fund's covariate (in ascending order). We then implement the BSW approach to select funds in each quintile, generating five $FDR10\%$ portfolios. The portfolios are constructed in the same fashion as the $FDR10\%$ portfolios in the main paper. This way we are able to examine the performance of the portfolio of selected funds in each quintile.

Table IA22 presents the portfolios' alphas based on monthly fund returns from 1982 to 2022. For convenience, suppose we look at the R-square covariate. Starting from the end of December 1981, we use past five years historical data to calculate inputs including the $p-$value, estimated alpha and R-square. We partition the funds into quintiles based on the covariate. For each quintile $q = 1, \ldots, 5$, we implement the FDR^+ procedure of BSW at FDR target of 10% to select a group of funds invested in the year 1982. At the end of December 1982, we repeat the process: we calculate the $p-$value, the alpha estimate and the R-square based on the past five years' data; we partition funds into quintiles, and for each of them we implement the FDR^+ to select funds to invest in 1983. We repeat this process until the end of December 2021 to select funds for each quintile q portfolio. For ten covariates, we obtain $10 \times 5 = 50$ portfolios which we present in Table IA22.

[Insert Table IA22 approximately here]

We observe that among the five well-known fund characteristics (R-square, Active Weight, FundSize, Return Gap and Fund Flow), only Active Weight shows a clear pattern in which the portfolio of funds from the group of higher active weight performs better. Our experiment suggests that, except from Fund Flow case, the two-step procedure (i.e., first ranking funds based on a covariate and then implementing the $FDR10\%$) does not exploit the informativeness of the covariate or offer a clear strategy. This implies that our $fFDR$ method cannot be substituted by or be considered as overlapping with the two-step procedure.

XVII. Restricted Data

As supplementary to our empirical study of Section VI, 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 2022). This choice of threshold is consistent with Pastor et al. (2015) and Zhu (2018). Figure IA19 exhibits the alpha evolution while Table IA23 shows the average n-year alpha for the $fFDR10\%$ portfolios based on each individual covariate. Similarly, we present in Figure IA20 and Table IA24 the $fFDR10\%$ results of the portfolios based on combinations of the covariates.

[Insert Figure IA19 approximately here]

[Insert Table IA23 approximately here]

[Insert Figure IA20 approximately here]

[Insert Table IA24 approximately here]

XVIII. Selecting Unprofitable Funds with fFDR

In this part, we obtain, by analogy to 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

$$\text{FDR}^- = \mathbb{E}\left(\frac{V^-}{\max\{R^-, 1\}}\right) \tag{11}$$

and

$$\text{pFDR}^- = \mathbb{E}\left(\frac{V^-}{R^-}\middle| R^- > 0\right). \tag{12}$$

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 as

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

where $\hat{R}^- = \#\{i|p_i < \gamma, \hat{\alpha}_i < 0\}$ and $\hat{\pi}_0$ is calculated as in equation (2) 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 IA25) and their trading metrics (Table IA26). We also construct the Equal Weight Minus portfolio, which includes all funds with negative estimated in-sample alpha invested in the following year.

[Insert Table IA25 approximately here]

[Insert Table IA26 approximately here]

TABLE IA1

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 ($IR = \hat{\alpha}/\hat{\sigma}_\varepsilon$). Panel A reports the metrics calculated based on the portfolios' return from January 1982 to December 2022 while Panel B reports those based on return from January 1982 to December 2019.

Covariate	$\hat{\alpha}$ (p-value)	t-statistic	$\hat{\sigma}_\varepsilon$	Mean Return	Sharpe Ratio	IR
Panel A: Whole sample						
R-square	0.84 (0.41)	0.83	4.97	6.71	0.51	0.17
Fund Size	0.20 (0.89)	0.19	4.98	6.13	0.46	0.04
Active Weight	0.73 (0.41)	0.84	4.16	7.09	0.51	0.18
Return Gap	0.31 (0.75)	0.34	4.46	6.70	0.49	0.07
Fund Flow	0.33 (0.75)	0.33	4.57	6.54	0.49	0.07
Expense Ratio	0.94 (0.31)	0.97	4.24	6.95	0.54	0.22
Sharpe	0.17 (0.91)	0.17	4.34	6.53	0.51	0.04
Treynor	0.24 (0.84)	0.25	4.42	6.48	0.50	0.05
Beta	1.11 (0.30)	1.02	5.60	6.57	0.48	0.20
Sigma	0.34 (0.78)	0.31	5.48	6.25	0.46	0.06
OLS	0.55 (0.52)	0.66	3.96	7.34	0.53	0.14
Ridge	0.89 (0.33)	0.97	4.50	7.18	0.51	0.20
LASSO	0.61 (0.49)	0.70	4.28	6.90	0.50	0.14
Elastic Net	0.86 90.30)	1.01	4.29	7.11	0.51	0.20
PC 1	0.69 (0.38)	0.86	3.65	7.33	0.55	0.19
$FDR10\%$	-0.05 (0.94)	-0.05	5.27	5.77	0.45	-0.01
Equal Weight	-0.93 (0.02)	-2.36	1.92	6.12	0.48	-0.49
Equal Weight Plus	-0.51 (0.22)	-1.14	2.22	6.30	0.49	-0.23
Panel B: Sample period until December 2019						
R-square	1.29 (0.27)	1.19	5.06	7.16	0.55	0.25
Fund Size	0.57 (0.62)	0.51	5.08	6.53	0.50	0.11
Active Weight	1.05 (0.26)	1.14	4.24	7.41	0.54	0.25
Return Gap	0.62 (0.55)	0.64	4.55	7.01	0.52	0.14
Fund Flow	0.70 (0.54)	0.67	4.65	6.98	0.53	0.15
Expense Ratio	1.39 (0.15)	1.35	4.29	7.41	0.59	0.32
Sharpe	0.53 (0.66)	0.52	4.40	6.96	0.55	0.12
Treynor	0.60 (0.60)	0.59	4.49	6.91	0.54	0.13
Beta	1.58 (0.16)	1.37	5.72	7.00	0.52	0.28
Sigma	0.72 (0.56)	0.61	5.60	6.66	0.50	0.13
OLS	0.81 (0.37)	0.91	4.03	7.69	0.57	0.20
Ridge	1.20 (0.22)	1.21	4.59	7.51	0.54	0.26
LASSO	0.89 (0.36)	0.96	4.36	7.22	0.53	0.20
Elastic Net	1.17 (0.18)	1.29	4.37	7.44	0.54	0.27
PC 1	0.99 (0.22)	1.16	3.70	7.68	0.58	0.27
$FDR10\%$	0.32 (0.81)	0.28	5.39	6.14	0.48	0.06
Equal Weight	-0.80 (0.03)	-2.00	1.85	6.26	0.50	-0.43
Equal Weight Plus	-0.29 (0.44)	-0.61	2.18	6.62	0.52	-0.13

TABLE IA2

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 $(\pi^{+}, \pi_0, \pi^{-})$. The simulated data are a balanced panel with 284 observations per fund and generated with cross-sectional dependence.

$(\pi^{+}, \pi_0, \pi^{-})$	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^{+}$	1	6.8	23.9	46.6	68.7
	FDR^{+}	0.6	2.9	13.9	33.6	55.3
$(10, 60, 30)\%$	$fFDR^{+}$	2	12.6	35.5	59.6	77.8
	FDR^{+}	0.5	3.4	16.2	37.7	58.5
$(10, 30, 60)\%$	$fFDR^{+}$	5.5	26	54	77.6	90.2
	FDR^{+}	0.6	5.3	23.3	49.9	71.3
$(13, 67.5, 19.5)\%$	$fFDR^{+}$	1.8	11.5	32.8	56.7	76.7
	FDR^{+}	0.7	5	19.9	41.7	62.8
$(13, 48, 39)\%$	$fFDR^{+}$	3.8	19.3	44.6	70	85.1
	FDR^{+}	0.7	5.5	23.5	48.5	68.3
$(13, 9, 78)\%$	$fFDR^{+}$	9.7	37.6	70.7	91.5	97.8
	FDR^{+}	0.9	10	41	73.4	89.8

TABLE IA3

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 284 observations per fund and generated with cross-sectional dependence.

π_0	Procedure	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$
75%	$fFDR^+$	0.6	16.8	37.3	51.8	61.3
	FDR^+	0.3	9.2	27.7	42.4	52.9
60%	$fFDR^+$	1.8	22.6	44.2	58.1	67.2
	FDR^+	0.4	12.3	32.8	47.5	57.8
30%	$fFDR^+$	5.1	32.9	54.9	68.1	75.5
	FDR^+	0.6	18.7	41.3	56.5	66.1
67.5%	$fFDR^+$	1.1	20.1	40.9	55.3	64.2
	FDR^+	0.3	11	30.4	45.3	55.7
48%	$fFDR^+$	3.2	27.9	49.1	62.8	71.6
	FDR^+	0.4	15.4	36.4	51.5	61.4
9%	$fFDR^+$	7.5	39.8	62.2	74.6	81.4
	FDR^+	0.9	23.5	48.7	63.9	73.1

TABLE IA4

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 284 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	0.5	5.8	14.4	24.5
	FDR^+	0	0	0.4	2.4	8.1
		$\pi^+ = 11.8\%$	$\pi^+ = 16.9\%$	$\pi^+ = 26.4\%$	$\pi^+ = 30.5\%$	$\pi^+ = 33.4\%$
$(-2, -0.5)$	$fFDR^+$	0.1	0.7	7	16.5	26.5
	FDR^+	0	0	0.6	3.6	10.1
		$\pi^+ = 35.2\%$	$\pi^+ = 36.4\%$	$\pi^+ = 38.2\%$	$\pi^+ = 39.8\%$	$\pi^+ = 41.1\%$
$(-2.5, 0)$	$fFDR^+$	0.5	1.1	9.9	19.3	29.4
	FDR^+	0	0.1	1.1	5.1	12.7

TABLE IA5

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	$fFDR^+$	3.9	10.3	20.8	33.3	45.1	54.3
	FDR^+	0.7	1.7	3.4	6.7	12.3	18.8
1000	$fFDR^+$	2.4	7.8	18.0	30.8	41.6	52.7
	FDR^+	0.4	1.1	2.7	6.6	12.1	19.8
2000	$fFDR^+$	2.2	7.4	17.7	28.9	41.2	50.6
	FDR^+	0.3	0.9	2.7	6.6	12.9	19.7
3000	$fFDR^+$	2.2	6.8	16.2	28.1	39.2	50.4
	FDR^+	0.2	0.7	2.3	6.0	12.5	20.7
		Panel B: Mixture of Discrete and Normal distributions					
500	$fFDR^+$	12.7	22.3	30.4	36.4	40.7	46.2
	FDR^+	2.9	8.5	14.8	20.4	25.3	30.6
1000	$fFDR^+$	12.7	21.7	29.1	35.6	40.7	44.8
	FDR^+	2.9	8.5	14.6	20.6	25.6	30.1
2000	$fFDR^+$	12.1	21.4	28.7	35.3	39.9	44.4
	FDR^+	2.8	8.5	14.5	20.6	25.2	30.0
3000	$fFDR^+$	12.2	21.0	28.2	34.7	39.6	43.8
	FDR^+	2.9	8.4	14.4	20.4	25.4	29.8
		Panel C: Mixture of Normal distributions					
500	$fFDR^+$	1.8	3.8	6.0	9.3	11.8	15.1
	FDR^+	0.2	0.4	0.6	1.0	1.4	2.3
1000	$fFDR^+$	1.4	3.1	5.2	8.3	11.1	13.6
	FDR^+	0.1	0.2	0.4	0.8	1.3	1.9
2000	$fFDR^+$	1.1	2.7	5.2	7.7	10.6	13.2
	FDR^+	0.1	0.1	0.3	0.6	1.2	2.0
3000	$fFDR^+$	1.2	2.8	5.0	7.9	10.6	12.8
	FDR^+	0.0	0.1	0.3	0.7	1.2	2.0

TABLE IA6

Power comparison (in %): small size and small number of observations

In this table, we consider three distributions as in Table IA5 for samples consisting of $m = 500$ funds (balanced panels) with $T = 60$ observations per fund (5 years) under cross-sectional dependence.

		FDR target								
Distribution	Procedure	10	20	30	40	50	60	70	80	90
Discrete	$fFDR^+$	0.7	3.4	8	14.7	22.9	32.4	42.3	53.2	65.2
	FDR^+	0.2	0.4	0.7	1	1.4	2.1	3	4.5	6.5
Mixture of discrete and normal	$fFDR^+$	3.3	8.9	16	24.2	33.1	42.2	51.6	61.4	67.3
	FDR^+	0.5	1.3	2.7	5.3	9.1	14.9	22.6	32.8	43.3
Mixture of normals	$fFDR^+$	0.6	1.9	4.3	8.3	13.4	20	28.2	38.4	50.9
	FDR^+	0.1	0.2	0.3	0.5	0.9	1.2	1.8	3.1	5.3

TABLE IA7

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	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	0.5	7	24.2	44.6	61.2	0.8	7.1	23.4	43.1	60.6
	FDR^+	0.5	3.2	15.6	33	49.5	0.6	3.6	14.9	31.8	48.6
$(10, 60, 30)\%$	$fFDR^+$	1.4	12	33.5	54.2	69.3	1.8	12	32.4	53.2	68.9
	FDR^+	0.5	3.5	16.9	35.3	52.1	0.6	4	16.3	34.5	51.4
$(10, 30, 60)\%$	$fFDR^+$	4.2	22.4	49.3	68.3	80.8	4.6	21.8	48.4	67.8	80.5
	FDR^+	0.6	4.4	22	43.1	60.3	0.7	4.9	21.4	42.1	59.8
$(13, 67.5, 19.5)\%$	$fFDR^+$	1.1	10.7	30.6	51.1	67.2	1.5	11.2	29.8	50.1	66.2
	FDR^+	0.6	4.8	20.1	38.4	55	0.7	5.4	19.4	37.4	53.8
$(13, 48, 39)\%$	$fFDR^+$	2.9	17.7	40.9	61	75.1	3.5	17.9	40.3	60.2	74.3
	FDR^+	0.7	5.5	23.2	42.6	59	0.8	6.2	22.5	42	58.2
$(13, 9, 78)\%$	$fFDR^+$	7.6	31.2	63.2	79.6	89.7	8.2	31.4	62.3	79.8	89.4
	FDR^+	0.8	8.8	32.9	56.9	74.4	0.9	9.3	32.9	57.2	74.3

TABLE IA8

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%	$fFDR^{+}$	0.5	15	33.1	47	56.5	0.5	14.8	32.5	46.6	56.1
	FDR^{+}	0.3	8.7	24.6	38.3	48.4	0.3	8.5	24.1	37.9	48.1
60%	$fFDR^{+}$	1.5	20.4	39.3	52.7	61.7	1.6	20.2	39	52.5	61.5
	FDR^{+}	0.4	11.5	29.1	43	53.1	0.4	11.3	28.7	42.6	52.8
30%	$fFDR^{+}$	4.1	28.7	49.4	62.7	71	4.3	28.9	49.7	62.7	71.1
	FDR^{+}	0.5	16.6	37.1	51.8	61.5	0.6	16.5	37	51.6	61.6
67.5%	$fFDR^{+}$	0.9	17.9	36.2	50	58.9	0.9	17.9	36	49.5	59
	FDR^{+}	0.3	10.1	26.8	40.7	50.7	0.3	10.2	26.6	40.3	50.6
48%	$fFDR^{+}$	2.4	23.9	43.3	56.9	65.9	2.7	23.8	43.6	56.9	65.6
	FDR^{+}	0.4	13.6	32.3	46.6	56.4	0.4	13.4	32.3	46.3	56.2
9%	$fFDR^{+}$	6.1	34.6	55.8	69.2	77.4	6.1	34.7	56	69	77.2
	FDR^{+}	0.7	20.3	42.7	58.3	68.1	0.9	20.5	42.8	58.1	68

TABLE IA9

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	0.3	4.2	12	20.7	0	0.4	4.6	12.3	20.8
	FDR^+	0	0	0.4	2.2	6.9	0	0.1	0.4	2.3	7
$(-2, -0.5)$	$fFDR^+$	0	0.5	5.7	14	22.7	0.1	0.6	5.9	14.1	23.2
	FDR^+	0	0.1	0.5	3.2	8.6	0	0.1	0.6	3.2	8.9
$(-2.5, 0)$	$fFDR^+$	0.2	0.6	8	16.2	25.4	0.3	0.9	8.5	16.8	25.6
	FDR^+	0	0.1	0.9	4.6	10.8	0	0.1	1	4.9	11.2

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

TABLE IA10

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 284 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	$fFDR^{+}$	1.8	15.1	31.6	45.5	56.2	2.4	15.5	32.6	46.1	56
	FDR^{+}	0.1	2.2	13.4	28.5	42	0.1	2.4	14.4	29.1	41.5
	π^{+}	30.9	36.9	40.1	42.1	43.4	30.9	36.9	40.1	42.1	43.4
-0.5	$fFDR^{+}$	3.3	18.5	35.5	48.7	58.9	4.2	19.4	35.7	49.3	58.9
	FDR^{+}	0.1	4	17.7	32.8	45.5	0.2	4.5	17.9	33.1	45.5
	π^{+}	50	50	50	50	50	50	50	50	50	50
0	$fFDR^{+}$	8.3	26	41.9	54	63.5	9.3	27	42.3	54.4	63.9
	FDR^{+}	0.7	9.9	25.6	40.2	51.7	1.2	10.5	25.7	40.3	51.8
Panel B: Unbalanced Data											
-0.8	$fFDR^{+}$	1.7	13.3	27.6	40.8	50.7	1.9	13.5	27.9	40.8	50.7
	FDR^{+}	0.1	2.3	11.9	24.5	36.4	0.1	2.4	11.9	24.7	36.1
-0.5	$fFDR^{+}$	3	16.2	30.6	43.5	53.4	3.5	16.8	31.4	43.8	53.6
	FDR^{+}	0.2	4	15.2	28.3	39.8	0.2	4.3	15.5	28.4	39.9
0	$fFDR^{+}$	7.5	22.5	37.2	48.9	58.2	7.7	22.9	37.2	48.3	58.2
	FDR^{+}	0.8	9	22.1	35.1	45.8	1	9.1	22.1	34.4	45.9

TABLE IA11

Power comparison (in %) 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 284 observations per fund and are generated with cross-sectional independence.

(π^+, π_0, π^-)	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	0.3	4.7	20.3	43.8	65.8
	FDR^+	0.4	2.6	14.1	35.2	56.5
$(10, 60, 30)\%$	$fFDR^+$	0.9	9.1	30.6	55.5	75.1
	FDR^+	0.4	2.7	16.2	39.3	60.6
$(10, 30, 60)\%$	$fFDR^+$	3.7	22.7	51.9	76.5	89.7
	FDR^+	0.5	4.1	24.3	51.3	72.6
$(13, 67.5, 19.5)\%$	$fFDR^+$	0.7	8.3	28.8	53.8	73.4
	FDR^+	0.5	3.8	19.9	43.5	63.5
$(13, 48, 39)\%$	$fFDR^+$	2.1	15.9	42.7	68.1	84.6
	FDR^+	0.5	4.7	24.7	49.5	69.8
$(13, 9, 78)\%$	$fFDR^+$	7.6	37.4	68.8	91	97.7
	FDR^+	0.6	8.9	41.7	72.7	90.1

TABLE IA12

Power comparison (in %) 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 284 observations per fund and are generated with cross-sectional independence.

(π^+, π_0, π^-)	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	1.5	11.8	33.7	58.4	77.2
	FDR^+	0.4	2.4	13.9	35.4	56.8
$(10, 60, 30)\%$	$fFDR^+$	3.4	19	46.2	70.8	86.2
	FDR^+	0.4	2.7	16.1	39.4	60.8
$(10, 30, 60)\%$	$fFDR^+$	7.3	33.2	72.2	91	96.7
	FDR^+	0.5	3.9	24.9	51.2	72.4
$(13, 67.5, 19.5)\%$	$fFDR^+$	2.4	16	40.3	64.3	81
	FDR^+	0.5	3.7	20.2	43.1	63.7
$(13, 48, 39)\%$	$fFDR^+$	5.1	27.2	57.7	79.8	91.1
	FDR^+	0.6	4.5	24.5	49.5	69.9
$(13, 9, 78)\%$	$fFDR^+$	11.1	43.6	81.6	94.3	98.3
	FDR^+	0.6	9	41.4	72.8	89.9

TABLE IA13

Power comparison (in %) 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 284 observations per fund and are generated with cross-sectional independence.

(π^+, π_0, π^-)	Procedure	$\alpha^* = 1.5$	$\alpha^* = 2$	$\alpha^* = 2.5$	$\alpha^* = 3$	$\alpha^* = 3.5$
$(10, 75, 15)\%$	$fFDR^+$	0.6	6.9	24.6	48.3	69.7
	FDR^+	0.4	2.4	13.9	35.1	56.4
$(10, 60, 30)\%$	$fFDR^+$	1.8	14.1	39	64.1	81.1
	FDR^+	0.5	2.7	16.7	39.3	60.6
$(10, 30, 60)\%$	$fFDR^+$	5.4	29.4	61.8	82.9	93
	FDR^+	0.5	4	24.3	51.2	72.4
$(13, 67.5, 19.5)\%$	$fFDR^+$	1.3	10.9	32.8	58	77.2
	FDR^+	0.5	3.7	19.8	43.1	64
$(13, 48, 39)\%$	$fFDR^+$	3.4	20.6	48.5	72.4	86.9
	FDR^+	0.6	4.6	24.4	49.2	69.8
$(13, 9, 78)\%$	$fFDR^+$	9.6	41.5	76	92.4	98.1
	FDR^+	0.6	9.4	41.8	72.8	89.8

TABLE IA14

Performance comparison in terms of power (%): non-informative covariate

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) under different combinations of parameter values. The simulated data are a balanced panel with 284 observations per fund and are generated with cross-sectional independence. The covariate input of the $fFDR^+$ is a random variable drawn randomly from the standard uniform distribution, Uniform$(0,1)$, without any connections to the alpha.

		Panel A: discrete distribution				
(π^+,π_0,π^-)	Procedure	$\alpha^*=1.5$	$\alpha^*=2$	$\alpha^*=2.5$	$\alpha^*=3$	$\alpha^*=3.5$
$(10,75,15)\%$	$fFDR^+$	0.2	2.7	14.1	34.5	57
	FDR^+	0.4	2.3	13.8	35.1	56.8
$(10,60,30)\%$	$fFDR^+$	0.2	3.5	16.9	39.1	61.4
	FDR^+	0.5	2.7	15.9	39.5	61.1
$(10,30,60)\%$	$fFDR^+$	0.4	6.3	26.2	52.3	73.4
	FDR^+	0.5	3.9	24.1	51.3	72.1
$(13,67.5,19.5)\%$	$fFDR^+$	0.3	4.7	19.6	42.8	63.9
	FDR^+	0.5	3.8	19.7	43.4	63.8
$(13,48,39)\%$	$fFDR^+$	0.5	6.4	24.9	49.1	70.3
	FDR^+	0.5	4.5	24.5	49.3	69.7
$(13,9,78)\%$	$fFDR^+$	1.1	14	46.9	77.2	92.1
	FDR^+	0.6	9.2	41.8	72.9	89.9
		Panel B: discrete-normal distribution mixture				
π_0	Procedure	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$
75%	$fFDR^+$	0	9.5	30.1	45.6	55.9
	FDR^+	0.2	8.9	27.9	43	53.3
60%	$fFDR^+$	0.1	13.2	34.6	49.9	59.7
	FDR^+	0.3	12.4	32.9	48	58.1
30%	$fFDR^+$	0.3	19.5	42.9	57.8	67.3
	FDR^+	0.4	18.7	42	56.9	66.4
67.5%	$fFDR^+$	0.1	11.4	32.5	47.8	57.8
	FDR^+	0.2	10.6	30.5	45.6	55.7
48%	$fFDR^+$	0.2	15.8	38	52.9	62.5
	FDR^+	0.3	14.9	36.5	51.5	61.4
9%	$fFDR^+$	0.6	24.2	49.3	64.5	73.5
	FDR^+	0.6	23.4	48.6	63.8	73
		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	0.2	2.8	8.8
	FDR^+	0	0.1	0.3	2.2	7.7
		$\pi^+=11.8\%$	$\pi^+=16.9\%$	$\pi^+=26.4\%$	$\pi^+=30.5\%$	$\pi^+=33.4\%$
$(-2,-0.5)$	$fFDR^+$	0	0	0.5	3.9	10.7
	FDR^+	0	0.1	0.4	3.2	9.7
		$\pi^+=35.2\%$	$\pi^+=36.4\%$	$\pi^+=38.2\%$	$\pi^+=39.8\%$	$\pi^+=41.1\%$
$(-2.5,0)$	$fFDR^+$	0	0	1	5.4	13.2
	FDR^+	0	0.1	0.7	4.8	12.4

TABLE IA15

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	20.6	33.3	46.2	55.2
	FDR^+	0.6	1.4	3.2	6.1	11.6	18.7
1000	$fFDR^+$	1.7	6.4	17.1	30.4	42.7	53.8
	FDR^+	0.4	0.9	2.2	5.2	11.3	19.8
2000	$fFDR^+$	1.2	6.2	15.8	29.3	41.5	52.0
	FDR^+	0.2	0.7	1.6	5.4	12.1	21.1
3000	$fFDR^+$	1.1	5.9	15.3	28.4	40.6	52.1
	FDR^+	0.2	0.5	1.6	5.3	12.7	21.7
		Panel B: Mixture of Discrete and Normal distributions					
500	$fFDR^+$	12.6	22.1	29.4	35.8	41.1	45.3
	FDR^+	2.6	7.9	14.1	20.3	25.6	30.2
1000	$fFDR^+$	12.0	21.1	28.8	35.1	40.7	44.4
	FDR^+	2.4	8.0	14.5	20.3	25.7	30.0
2000	$fFDR^+$	11.4	20.6	28.2	34.5	39.6	44.0
	FDR^+	2.3	8.2	14.4	20.4	25.4	29.8
3000	$fFDR^+$	11.4	20.5	28.0	34.2	39.2	43.6
	FDR^+	2.4	8.1	14.5	20.3	25.3	29.9
		Panel C: Mixture of Normal distributions					
500	$fFDR^+$	1.2	3.2	5.2	8.4	11.1	13.6
	FDR^+	0.2	0.3	0.5	0.9	1.3	1.8
1000	$fFDR^+$	0.9	2.4	4.9	7.6	10.2	13.3
	FDR^+	0.1	0.2	0.4	0.7	1.1	1.7
2000	$fFDR^+$	0.7	2.3	4.4	6.9	9.7	12.2
	FDR^+	0.1	0.1	0.3	0.5	0.9	1.6
3000	$fFDR^+$	0.7	2.1	4.4	6.8	9.6	12.1
	FDR^+	0.0	0.1	0.2	0.5	0.9	1.7

TABLE IA16

Power comparison (in $\%$) for small size and small number of observations

In this table, we consider three distributions as in Table IA15 for samples consisting of $m = 500$ funds (balanced panels with cross-sectional independence) with $T = 60$ observations per fund (5 years).

Distribution	Procedure	FDR target								
		10	20	30	40	50	60	70	80	90
Discrete	$fFDR^+$	0.4	2	5.5	11.8	20.3	30.4	41.4	53.1	66.6
	FDR^+	0.2	0.4	0.6	0.9	1.1	1.5	1.9	2.5	3.4
Mixture of discrete and normal	$fFDR^+$	2.7	7.9	14.8	23.3	33	43.4	53.7	63.8	68.3
	FDR^+	0.4	1	1.9	3.4	6	11	19.4	32.7	48.9
Mixture of normals	$fFDR^+$	0.3	1.2	3.1	6.4	11.4	18.3	27	38.2	52.1
	FDR^+	0.1	0.2	0.3	0.3	0.5	0.6	0.8	1.1	1.5

TABLE IA17

Power (in %) of the $fFDR^+$ under covariate with noise

The data are generated as in Tables 1 except the use of a new covariate containing a noise: $Z' = Z + \eta$ instead of Z. The noise is drawn independently from normal distribution $\eta \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta \in \{\sigma_1 = 0.5/\sqrt{12}, \sigma_2 = 1/\sqrt{12}\}$.

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)%	0.5	0.4	5.7	4.7	21.9	19	45.4	41	67	62.7
(10,60,30)%	1.2	0.9	9.7	7.3	30.9	25.6	55.6	49.4	75	69.6
(10,30,60)%	3.2	1.9	18.9	14.1	47.6	39.5	72.8	65	87.4	82
(13,67.5,19.5)%	0.9	0.7	9.3	7.7	30.1	26	54.6	49.6	73.8	69.5
(13,48,39)%	2.3	1.7	16.1	12.5	41.7	35.2	66	59.3	82.7	77.7
(13,9,78)%	6	4	31.1	24.9	66.1	59.4	89.6	84.8	97.3	95.2
Panel B: Mixture of discrete and 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	14.8	12.8	35.2	33	49.5	47.7	59.1	57.5
60%	0.8	0.4	20.2	17.4	40.7	37.9	54.8	52.3	63.4	61.3
30%	2.6	1.5	29.2	25	50.5	46.8	63.1	60.1	71	68.7
67.5%	0.5	0.3	17.7	15.2	38.2	35.7	52.2	50	61.3	59.5
48%	1.4	0.8	24	20.5	44.8	41.6	58	55.4	66.5	64.2
9%	4.1	2.4	35.4	30.3	57.1	52.9	69.3	66.3	76.6	74.6
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	1.3	0.6	7.2	4.7	16	12.1
(-2,-0.5)	0	0	0.1	0	2.5	1.3	9.6	6.6	19.1	14.8
(-2.5,0)	0	0	0.2	0.1	4.8	2.7	13.4	9.4	23.6	18.5

TABLE IA18

Comparison of portfolios' performances for varying time lengths of investing

In this table, we consider 10 portfolios including ten $fFDR20\%$ portfolios corresponding to the ten 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 2018 to December 2022. The average of these alphas is presented in the first row of Panel A of the table. Panel B reports similar metrics using the portfolios' return from January 1982 to December 2022.

	$fFDR20\%$										$FDR20\%$
n	R-square	Fund Size	Active Weight	Return Gap	Fund Flow	Expense	Sharpe	Treynor	Beta	Sigma	
Panel A: Whole sample											
5	0.90	0.01	0.55	0.41	0.31	1.42	-0.10	0.12	1.13	0.83	-0.03
10	0.84	-0.06	0.57	0.31	0.37	1.08	0.03	0.22	1.24	0.65	-0.23
15	0.95	-0.02	0.63	0.26	0.42	0.94	0.16	0.32	1.32	0.68	-0.25
20	1.18	0.21	0.83	0.44	0.57	0.96	0.34	0.48	1.45	0.94	-0.06
25	1.10	0.15	0.71	0.37	0.52	0.88	0.34	0.48	1.43	0.82	-0.08
30	0.85	0.04	0.50	0.27	0.43	0.75	0.23	0.40	1.29	0.58	-0.22
35	0.69	-0.00	0.50	0.30	0.39	0.94	0.15	0.31	1.13	0.48	-0.25
40	0.61	0.03	0.42	0.21	0.33	0.86	0.07	0.24	1.07	0.55	-0.14
41	0.70	0.12	0.50	0.29	0.41	0.90	0.08	0.26	1.16	0.62	-0.03
Panel B: Sample period until December 2019											
5	1.03	0.07	0.69	0.52	0.39	1.38	-0.04	0.19	1.28	0.94	0.02
10	1.09	0.15	0.74	0.45	0.56	1.13	0.24	0.44	1.53	0.90	-0.00
15	1.33	0.37	0.93	0.58	0.72	1.18	0.46	0.62	1.57	1.14	0.13
20	1.48	0.55	1.11	0.73	0.84	1.22	0.60	0.75	1.70	1.33	0.27
25	1.23	0.35	0.78	0.49	0.68	0.96	0.46	0.64	1.61	1.02	0.09
30	1.03	0.23	0.62	0.43	0.62	0.87	0.37	0.57	1.54	0.77	-0.05
35	0.74	0.03	0.62	0.44	0.40	1.02	0.08	0.27	1.23	0.59	-0.21
38	1.13	0.48	0.79	0.60	0.80	1.35	0.43	0.63	1.64	1.02	0.34

TABLE IA19

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 ($IR = \hat{\alpha}/\hat{\sigma}_{\varepsilon}$). Panel A presents the metrics with use of the portfolios' return from January 1982 to December 2022 whereas Panel B the portfolios' return from January 1982 to December 2019.

Covariate	$\hat{\alpha}$ (p-value)	t-statistic	$\hat{\sigma}_{\varepsilon}$	Mean Return	Sharpe Ratio	IR
Panel A: Whole sample						
R-square	0.70 (0.49)	0.69	4.96	6.58	0.50	0.14
Fund Size	0.12 (0.96)	0.11	4.97	6.05	0.46	0.02
Active Weight	0.50 (0.57)	0.58	4.11	6.85	0.50	0.12
Return Gap	0.29 (0.81)	0.32	4.31	6.74	0.49	0.07
Fund Flow	0.41 (0.67)	0.42	4.57	6.64	0.50	0.09
Expense Ratio	0.90 (0.31)	0.96	4.15	6.92	0.54	0.22
Sharpe	0.08 (0.97)	0.08	4.33	6.43	0.50	0.02
Treynor	0.26 (0.82)	0.27	4.43	6.53	0.50	0.06
Beta	1.16 (0.27)	1.08	5.49	6.63	0.49	0.21
Sigma	0.62 (0.62)	0.56	5.49	6.54	0.48	0.11
$FDR20\%$	-0.03 (0.94)	-0.02	5.27	5.82	0.45	-0.01
Equal Weight	-0.93 (0.02)	-2.36	1.92	6.12	0.48	-0.49
Equal Weight Plus	-0.51 (0.22)	-1.14	2.22	6.30	0.49	-0.23
Panel B: Sample period until December 2019						
R-square	1.13 (0.31)	1.06	5.06	7.02	0.54	0.22
Fund Size	0.48 (0.67)	0.43	5.07	6.45	0.49	0.10
Active Weight	0.79 (0.40)	0.87	4.18	7.16	0.53	0.19
Return Gap	0.60 (0.55)	0.62	4.39	7.05	0.52	0.14
Fund Flow	0.80 (0.49)	0.76	4.65	7.08	0.54	0.17
Expense Ratio	1.35 (0.16)	1.36	4.19	7.38	0.59	0.32
Sharpe	0.43 (0.72)	0.42	4.39	6.85	0.54	0.10
Treynor	0.63 (0.59)	0.61	4.50	6.96	0.54	0.14
Beta	1.64 (0.14)	1.44	5.60	7.07	0.53	0.29
Sigma	1.02 (0.41)	0.86	5.61	6.97	0.52	0.18
$FDR20\%$	0.34 (0.79)	0.30	5.39	6.20	0.48	0.06
Equal Weight	-0.80 (0.03)	-2.00	1.85	6.26	0.50	-0.43
Equal Weight Plus	-0.29 (0.44)	-0.61	2.18	6.62	0.52	-0.13

TABLE IA20

Performance of all considered portfolios in sub-periods

The table displays the performance of the 15 $fFDR10\%$ portfolios corresponding to the 10 underlying covariates and the 5 combined covariates, the $FDR10\%$ and equally weighted portfolios in sub-periods (P1: 1982–1991, P2: 1992–2001, P3: 2002–2011, P4a: 2012–2019 and P4b: 2012–2022) in terms of 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.

	Whole sub-period alpha					Whole sub-period t-statistic					Annual Sharpe Ratio				
Portfolio	P1	P2	P3	P4a	P4b	P1	P2	P3	P4a	P4b	P1	P2	P3	P4a	P4b
R-square	3.45	1.81	1.44	0.58	-2.01	2.71	0.72	1.08	0.16	-0.80	0.67	0.57	0.28	0.84	0.53
Fund Size	2.63	1.49	-0.46	-0.11	-2.40	3.18	0.56	-0.35	-0.03	-0.98	0.67	0.53	0.19	0.78	0.51
Active Weight	3.72	2.17	-0.38	-0.16	-1.39	2.42	0.84	-0.33	-0.14	-1.51	0.65	0.56	0.19	1.10	0.69
Return Gap	3.26	1.77	-1.00	0.05	-1.30	2.46	0.67	-0.70	0.05	-1.43	0.60	0.55	0.15	1.13	0.70
Fund Flow	2.58	1.04	0.12	0.39	-1.90	2.32	0.39	0.13	0.11	-0.78	0.65	0.56	0.23	0.88	0.56
Expense Ratio	4.22	2.34	-0.26	2.87	-0.45	2.03	0.96	-0.28	1.52	-0.26	0.71	0.63	0.20	1.34	0.70
Sharpe	2.23	0.94	0.20	-0.45	-2.40	2.19	0.38	0.24	-0.12	-0.98	0.66	0.65	0.25	0.82	0.53
Treynor	2.16	1.16	0.11	-0.20	-2.28	2.29	0.46	0.13	-0.05	-0.93	0.65	0.63	0.24	0.83	0.53
Beta	3.86	1.19	2.46	0.44	-2.10	1.87	0.40	1.84	0.12	-0.84	0.66	0.42	0.34	0.84	0.53
Sigma	2.58	2.79	0.61	0.34	-2.31	1.91	1.08	0.42	0.09	-0.91	0.58	0.57	0.25	0.80	0.50
OLS	2.14	1.18	0.34	-0.18	-1.26	1.54	0.45	0.33	-0.19	-1.61	0.61	0.61	0.26	1.14	0.71
Ridge	2.98	3.14	-0.34	0.39	-1.05	2.79	1.12	-0.23	0.34	-1.13	0.65	0.55	0.20	1.16	0.72
LASSO	2.47	3.02	-0.49	0.03	-1.33	2.54	1.09	-0.41	0.03	-1.49	0.63	0.55	0.19	1.12	0.69
Elastic Net	2.61	3.63	-0.15	0.06	-1.30	2.32	1.30	-0.14	0.05	-1.49	0.62	0.58	0.21	1.13	0.70
PC 1	1.74	1.76	0.27	1.10	-0.66	1.79	0.71	0.31	0.82	-0.66	0.61	0.65	0.24	1.23	0.75
$FDR10\%$	2.96	1.53	-0.53	-0.54	-2.87	2.52	0.60	-0.38	-0.15	-1.17	0.65	0.52	0.19	0.75	0.48
Equal Weight	-0.52	-1.33	-0.20	-1.35	-1.46	-1.25	-1.58	-0.30	-2.74	-2.61	0.48	0.54	0.23	1.01	0.69
Equal Weight Plus	0.81	-1.07	-0.27	-0.35	-1.42	1.18	-1.14	-0.37	-0.58	-2.24	0.55	0.52	0.21	1.12	0.69

TABLE IA21

Performance comparison of portfolios based on $fFDR$ and sorting

The table shows the portfolios' annual Carhart four-factor alpha (in %) for the period January 1982 to December 2022. The sorting portfolios are those based on covariates (single-sorting) as well as based on both covariates and past alpha (double-sorting).. 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	Expense Ratio	Sharpe	Treynor	Beta	Sigma
Panel A: Performance of $fFDR10\%$ and $fFDR20\%$ portfolios										
$fFDR10\%$	0.84	0.20	0.73	0.31	0.33	0.94	0.17	0.24	1.11	0.34
$fFDR20\%$	0.70	0.12	0.50	0.29	0.41	0.90	0.08	0.26	1.16	0.62
Panel B: Assuming a positive effect of the covariate on performance of the fund										
Single sort 10%	-1.08	-0.73	*-0.95*	*-1.90*	*-1.18*	-2.56	-0.10	-0.45	-2.29	-2.60
Double sort 10%	-0.67	0.07	*1.13*	*-1.81*	*-0.55*	-2.35	-0.04	-0.40	-1.98	-0.40
Single sort 20%	-1.22	-0.74	*-0.90*	*-1.60*	*-0.84*	-1.89	-0.33	-0.48	-1.98	-1.88
Double sort 20%	-1.20	-0.04	*0.25*	*-1.04*	*-0.43*	-1.04	-0.21	-0.38	-0.71	-0.43
Panel C: Assuming a negative effect of the covariate on performance of the fund										
Single sort 10%	*-1.27*	*-1.01*	-1.25	-1.14	-1.31	-0.58	-2.15	-2.49	0.12	-0.52
Double sort 10%	*-2.51*	*-0.48*	-1.60	-0.53	-0.69	0.94	1.10	0.24	-0.34	0.38
Single sort 20%	*-1.11*	*-1.00*	-1.18	-1.18	-1.29	-0.46	-1.67	-1.66	0.00	-0.71
Double sort 20%	*-0.92*	*-0.11*	-1.18	-0.12	-0.66	-0.02	0.51	-0.33	-0.08	-0.26

TABLE IA22

Performance of $FDR10\%$ portfolios conditional on each of covariates

Covariate	Quintile	$\hat{\alpha}$ ($p-$value)	t-statistic	$\hat{\sigma}_{\varepsilon}$	Mean Return	Sharpe Ratio	IR	#Funds
R-square	1	0.02 (0.97)	0.02	6.98	5.19	0.46	0.00	15
	2	2.50 (0.08)	1.36	8.65	8.26	0.55	0.29	9
	3	-0.25 (0.79)	-0.31	4.62	6.51	0.48	-0.05	7
	4	0.03 (0.99)	0.03	5.26	6.05	0.43	0.01	6
	5	-1.27 (0.24)	-1.25	6.46	5.98	0.41	-0.20	12
Fund Flow	1	-0.33 (0.78)	-0.35	7.17	6.25	0.47	-0.05	12
	2	0.05 (0.99)	0.04	6.01	6.82	0.48	0.01	8
	3	0.18 (0.92)	0.16	6.27	6.20	0.48	0.03	7
	4	-0.53 (0.52)	-0.61	5.21	6.15	0.46	-0.10	10
	5	0.17 (0.88)	0.13	6.85	5.06	0.39	0.03	9
Active Weight	1	3.20 (0.45)	0.55	26.95	6.75	0.32	0.12	14
	2	-1.69 (0.10)	-1.75	5.19	4.85	0.37	-0.32	10
	3	0.81 (0.41)	0.91	5.24	7.13	0.50	0.16	7
	4	0.38 (0.77)	0.31	7.31	6.59	0.47	0.05	6
	5	0.36 (0.71)	0.39	5.54	6.17	0.47	0.06	6
Return Gap	1	1.30 (0.45)	0.79	7.99	6.04	0.45	0.16	7
	2	0.46 (0.60)	0.57	4.51	7.38	0.54	0.10	9
	3	-0.73 (0.52)	-0.59	6.51	6.02	0.43	-0.11	8
	4	-0.64 (0.47)	-0.73	4.78	6.77	0.49	-0.13	9
	5	-1.48 (0.23)	-1.23	6.94	4.70	0.34	-0.21	8
Fund Flow	1	-0.88 (0.57)	-0.59	7.98	4.16	0.33	-0.11	7
	2	-0.21 (0.85)	-0.13	7.27	5.90	0.42	-0.03	11
	3	-0.11 (0.88)	-0.10	6.09	6.62	0.50	-0.02	13
	4	0.23 (0.82)	0.26	4.84	6.37	0.49	0.05	8
	5	0.38 (0.75)	0.34	5.70	6.16	0.47	0.07	14
Expense Ratio	1	-0.47 (0.67)	-0.39	5.79	5.23	0.43	-0.08	11
	2	-0.72 (0.42)	-0.87	4.84	5.40	0.43	-0.15	10
	3	1.59 (0.40)	0.70	10.81	6.52	0.44	0.15	8
	4	0.36 (0.77)	0.31	5.88	6.34	0.47	0.06	6
	5	-0.70 (0.62)	-0.53	7.77	5.07	0.36	-0.09	10
Sharpe Ratio	1	1.03 (0.65)	0.45	12.94	5.51	0.36	0.08	2
	2	-1.47 (0.38)	-0.85	8.19	3.97	0.30	-0.18	10
	3	-0.33 (0.72)	-0.31	6.31	6.77	0.46	-0.05	22
	4	0.93 (0.38)	0.89	6.19	7.02	0.49	0.15	33
	5	0.73 (0.72)	0.33	10.40	6.05	0.43	0.07	58
Treynor Ratio	1	0.33 (0.84)	0.15	11.57	2.34	0.24	0.03	2
	2	1.59 (0.42)	0.77	8.69	8.11	0.51	0.18	8
	3	-0.27 (0.83)	-0.21	6.40	7.02	0.46	-0.04	18
	4	-0.40 (0.66)	-0.45	4.66	6.55	0.46	-0.08	32
	5	0.02 (0.98)	0.02	5.39	5.79	0.46	0	53
Beta	1	-0.63 (0.44)	-0.67	4.64	4.23	0.47	-0.14	8
	2	-0.09 (0.93)	-0.08	6.12	6.65	0.51	-0.01	6
	3	-0.65 (0.52)	-0.55	6.60	6.05	0.43	-0.10	11
	4	0.81 (0.40)	0.82	5.41	8.47	0.54	0.15	9
	5	2 (0.47)	0.59	15.53	6.40	0.36	0.13	9
Sigma	1	-0.18 (0.79)	-0.25	3.77	5.02	0.47	-0.05	11
	2	-0.09 (0.90)	-0.11	4.92	5.67	0.47	-0.02	7
	3	-0.93 (0.21)	-1.27	4.49	5.35	0.43	-0.21	8

4	0.34 (0.80)	0.31	6.01	7	0.51	0.06	7
5	2.02 (0.46)	0.60	16.29	5.74	0.35	0.12	17

TABLE IA23

Comparison of portfolios' performances for varying time lengths of investing: restricted data

We consider 10 portfolios including ten $fFDR10\%$ portfolios and the $FDR10\%$ 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 3 of the main paper.

n	R-square	Fund Size	Active Weight	Return Gap	Fund flow	Expense Ratio	Sharpe	Treynor	Beta	Sigma	$FDR10\%$
Panel A: Whole sample											
5	1.30	0.96	1.06	0.08	0.82	0.85	0.48	0.61	1.42	0.74	0.28
10	1.17	0.82	0.90	-0.05	0.75	0.58	0.45	0.58	1.44	0.43	0.09
15	1.23	0.75	0.91	0.02	0.70	0.41	0.44	0.53	1.40	0.32	-0.02
20	1.40	0.86	1.10	0.22	0.80	0.45	0.55	0.62	1.42	0.45	0.15
25	1.35	0.77	1.04	0.15	0.73	0.40	0.54	0.61	1.35	0.38	0.13
30	1.09	0.63	0.79	-0.07	0.66	0.29	0.50	0.58	1.24	0.27	-0.03
35	1.03	0.67	0.61	-0.16	0.72	0.35	0.56	0.65	1.23	0.36	-0.08
40	0.91	0.57	0.47	-0.19	0.59	0.26	0.43	0.50	1.08	0.36	-0.06
41	0.99	0.63	0.50	-0.07	0.66	0.33	0.40	0.47	1.14	0.47	0.05
Panel B: Sample period until December 2019											
5	1.22	0.85	1.23	0.17	0.70	0.76	0.34	0.48	1.36	0.60	0.25
10	1.22	0.85	1.06	0.08	0.74	0.60	0.43	0.56	1.51	0.42	0.21
15	1.45	0.97	1.22	0.29	0.85	0.63	0.56	0.65	1.53	0.51	0.32
20	1.56	1.09	1.37	0.46	0.95	0.65	0.66	0.73	1.61	0.61	0.46
25	1.34	0.84	1.07	0.19	0.78	0.44	0.53	0.61	1.45	0.41	0.24
30	1.08	0.67	0.83	-0.02	0.70	0.31	0.48	0.57	1.36	0.30	0.04
35	1.02	0.68	0.75	-0.08	0.69	0.31	0.47	0.55	1.32	0.36	-0.07
38	1.43	1.04	0.79	0.19	1.06	0.73	0.77	0.85	1.63	0.85	0.43

TABLE IA24

Performance of $fFDR\tau$ portfolios with combined covariates: restricted data

The table displays the average n-year alpha of the $fFDR10\%$ portfolios using the covariates given by the first principal component (PC 1), the OLS, ridge, LASSO and elastic net (see descriptions in Figure 5 of the main manuscript). The average n-year alpha (annualized, in $\%$) of each portfolio is calculated as described in Table 3 of the main manuscript.

n	OLS	Ridge	LASSO	Elastic Net	PC 1
Panel A: Whole sample					
5	0.58	0.84	0.84	0.64	0.62
10	0.67	1.00	0.93	0.72	0.75
15	0.73	1.08	1.00	0.79	0.85
20	0.98	1.26	1.15	0.96	1.01
25	0.90	1.19	1.03	0.84	0.95
30	0.77	1.01	0.84	0.65	0.83
35	0.79	0.92	0.82	0.59	0.84
40	0.56	0.74	0.74	0.39	0.62
41	0.62	0.81	0.87	0.47	0.63
Panel B: Sample period until December 2019					
5	0.72	1.00	1.01	0.78	0.76
10	0.82	1.24	1.17	0.95	0.90
15	1.00	1.42	1.33	1.13	1.09
20	1.18	1.61	1.49	1.30	1.20
25	0.96	1.38	1.21	1.03	1.00
30	0.88	1.17	1.03	0.81	0.91
35	0.81	1.05	0.96	0.68	0.87
38	0.88	1.14	1.19	0.76	0.91

TABLE IA25

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 ten 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 3 of the main paper.

n	R-square	Fund Size	Active Weight	Return Gap	Fund flow	Expense Ratio	Sharpe	Treynor	Beta	Sigma	EW^{-}	$FDR^{-}10\%$
5	-3.64	-3.91	-2.42	-2.74	-2.87	-3.51	-2.22	-2.09	-3.61	-3.96	-1.41	-3.91
10	-3.51	-3.82	-2.39	-2.65	-2.64	-3.35	-2.02	-1.91	-3.48	-3.82	-1.31	-3.83
15	-3.28	-3.49	-2.23	-2.36	-2.34	-3.02	-1.69	-1.60	-3.15	-3.54	-1.11	-3.54
20	-3.11	-3.23	-2.14	-2.17	-2.15	-2.77	-1.49	-1.43	-2.94	-3.32	-0.97	-3.29
25	-3.16	-3.28	-2.14	-2.20	-2.18	-2.81	-1.53	-1.45	-2.99	-3.35	-0.95	-3.36
30	-3.40	-3.59	-2.34	-2.47	-2.43	-3.12	-1.76	-1.68	-3.28	-3.67	-1.05	-3.65
35	-3.77	-3.95	-2.65	-2.87	-2.85	-3.53	-2.18	-2.08	-3.71	-4.15	-1.22	-3.97
40	-3.87	-3.98	-2.90	-2.99	-3.08	-3.69	-2.41	-2.32	-3.91	-4.28	-1.42	-3.98
41	-3.76	-3.86	-2.79	-2.91	-3.08	-3.57	-2.38	-2.28	-3.92	-4.17	-1.35	-3.93

TABLE IA26

Performance statistics of the 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	-3.76 (< 0.01)	-5.14	3.37	3.49	0.31	-1.12
Fund Size	-3.86 (< 0.01)	-5.33	3.11	3.51	0.31	-1.24
Active Weight	-2.79 (< 0.01)	-4.59	3.16	4.56	0.37	-0.88
Return Gap	-2.91 (< 0.01)	-4.06	3.33	4.50	0.37	-0.87
Fund Flow	-3.08 (< 0.01)	-4.51	3.21	4.24	0.35	-0.96
Expense Ratio	-3.57 (< 0.01)	-4.88	3.40	3.56	0.31	-1.05
Sharpe	-2.38 (< 0.01)	-3.48	3.01	4.81	0.39	-0.79
Treynor	-2.28 (< 0.01)	-3.52	2.90	4.99	0.40	-0.78
Beta	-3.92 (< 0.01)	-5.20	3.93	3.88	0.33	-1.00
Sigma	-4.17 (< 0.01)	-5.06	3.59	2.96	0.28	-1.16
$FDR^{-}10\%$	-3.93 (< 0.01)	-5.33	3.46	3.35	0.30	-1.14
Equal Weight	-0.93 (0.02)	-2.36	1.92	6.12	0.48	-0.49
Equal Weight Minus	-1.35 (< 0.01)	-3.07	2.08	5.86	0.46	-0.65

FIGURE IA1

Comparison of FDR^+ and $fFDR^+$.

The graphs show the differences between the two procedures with respect to their null proportion estimations and rejection rules. Panels A and B show that π_0 is estimated as a fixed number in the FDR^+ (see (2)) but as a step function in the $fFDR^+$ (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 0.008, 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 0.45. 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.

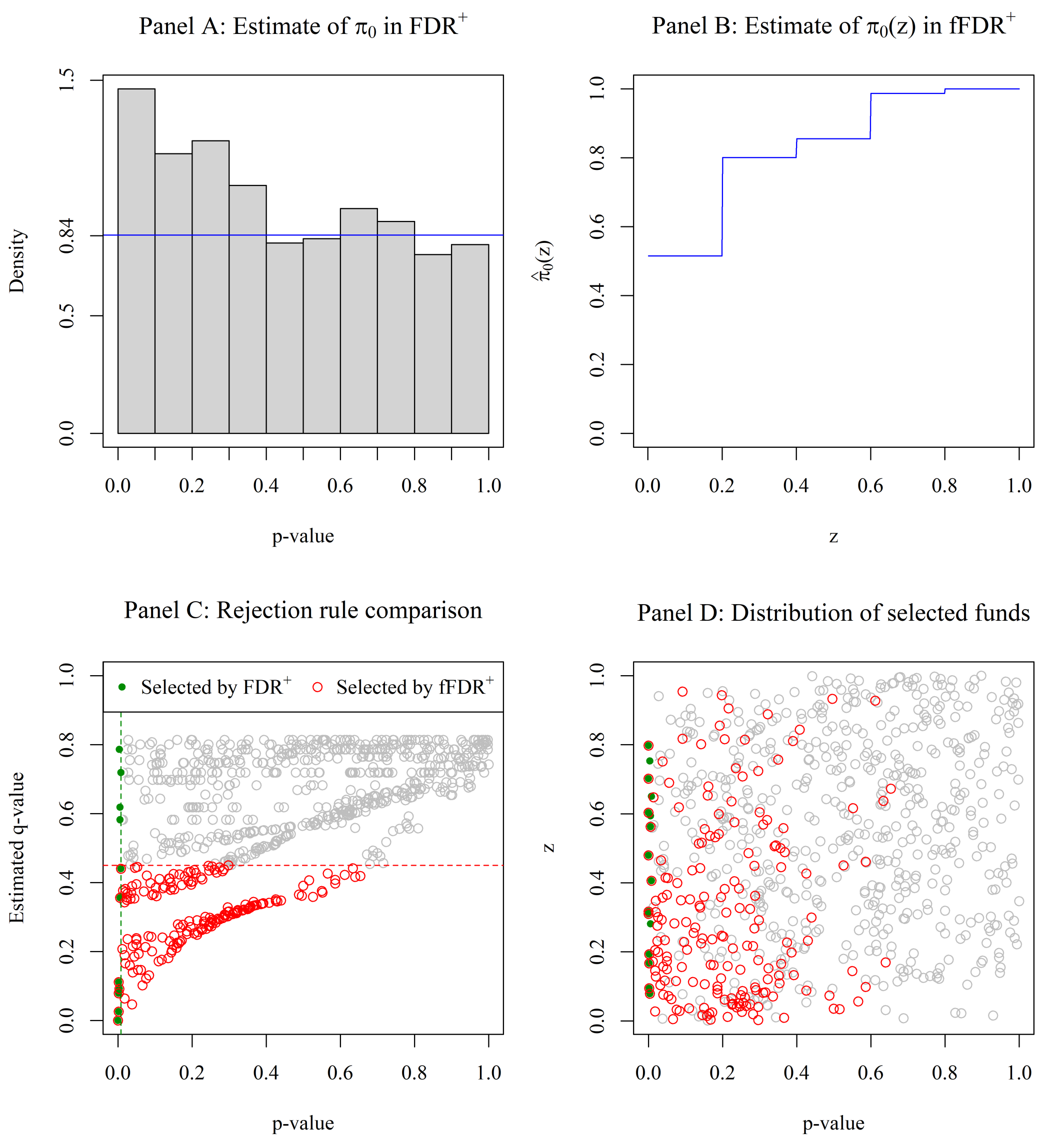


FIGURE IA2

Joint density function $f(p, z)$

The graph shows the heatmap of the density function $f(p, z)$.

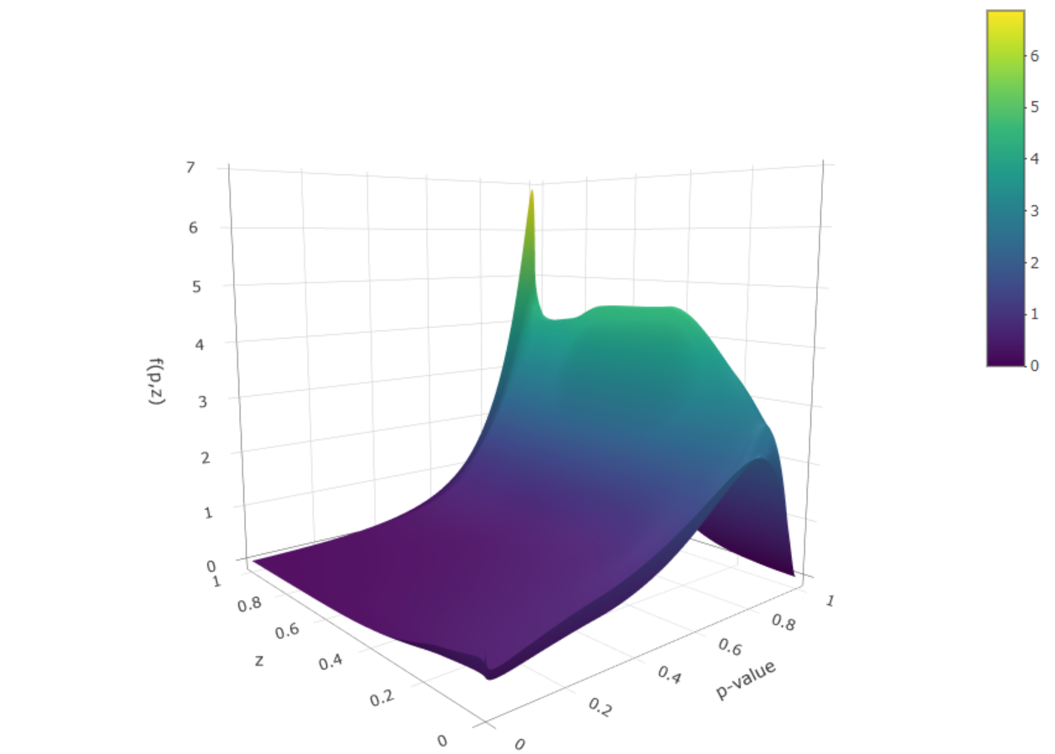


FIGURE IA3

Variance of falsely classified fund ratio: discrete distribution of α

The graphs show the gap in variance of the falsely classified fund ratio of the FDR^+ over the $fFDR^+$. The simulated data are balanced panels with cross-sectional independence.

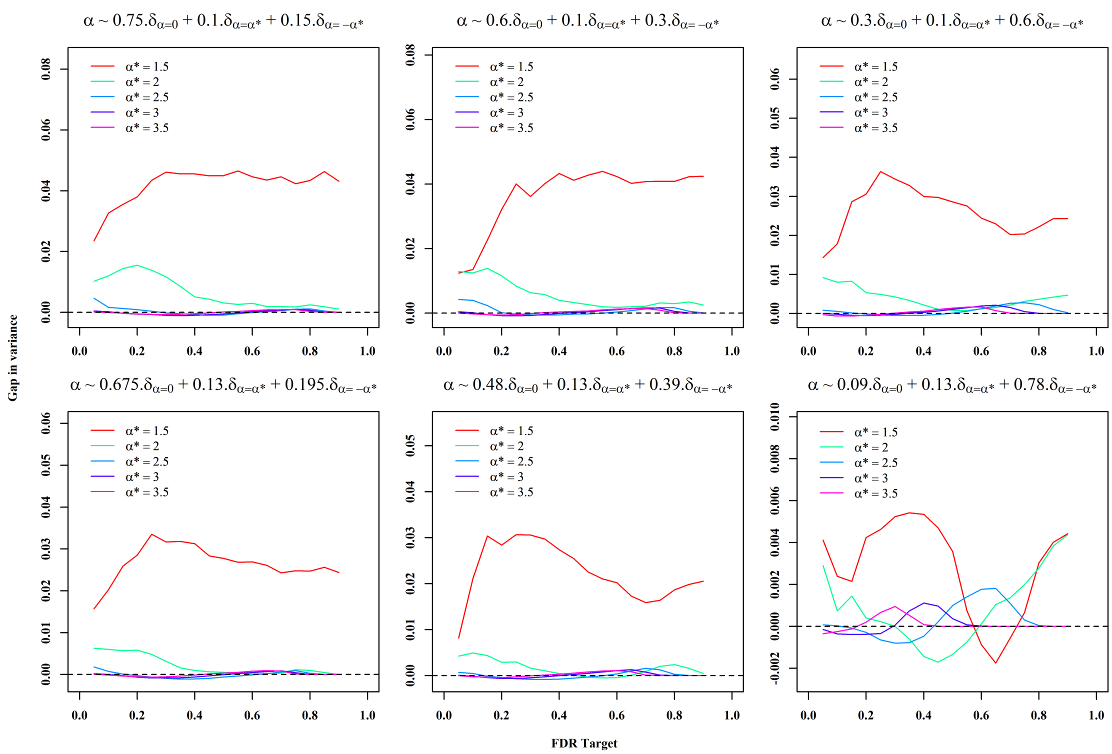


FIGURE IA4

Variance of falsely classified fund ratio: mixed discrete and normal α

The graphs show the gap in variance of the falsely classified fund ratio of the FDR^+ over the $fFDR^+$. The simulated data are balanced panels with cross-sectional independence.

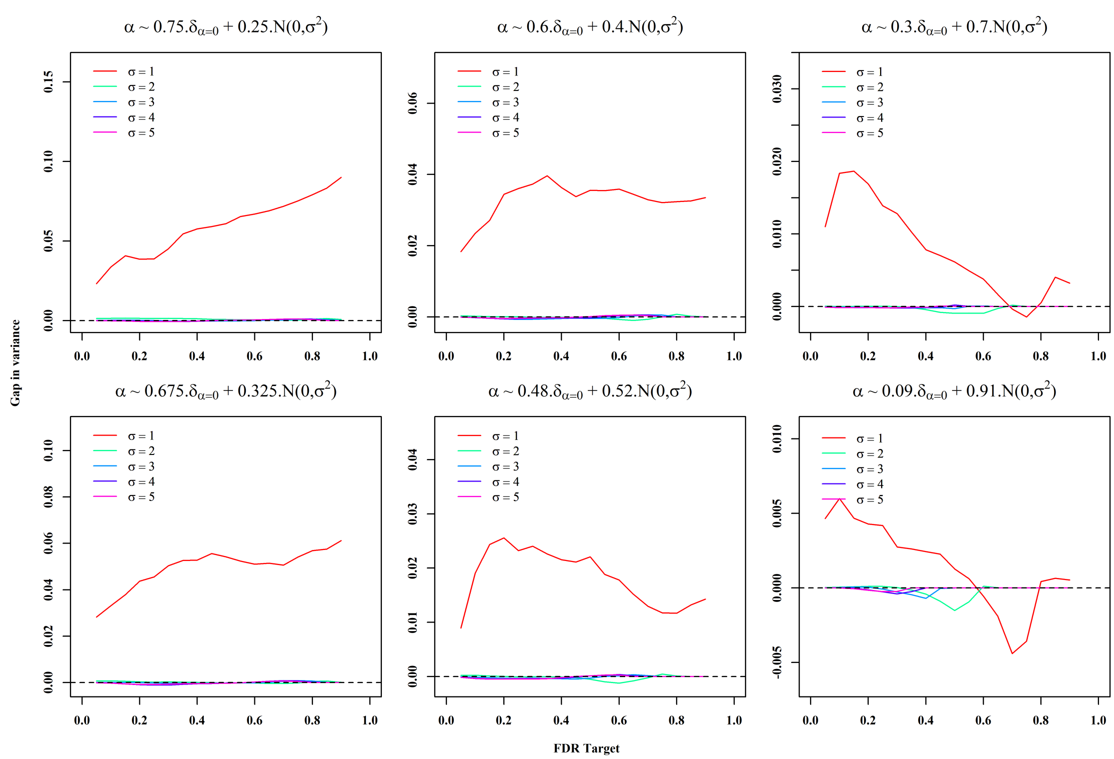


FIGURE IA5

Variance of falsely classified fund ratio: two normal distributions of α

The graphs show the gap in variance of the falsely classified fund ratio of the FDR^{+} over the $fFDR^{+}$. The simulated data are balanced panels with cross-sectional independence.

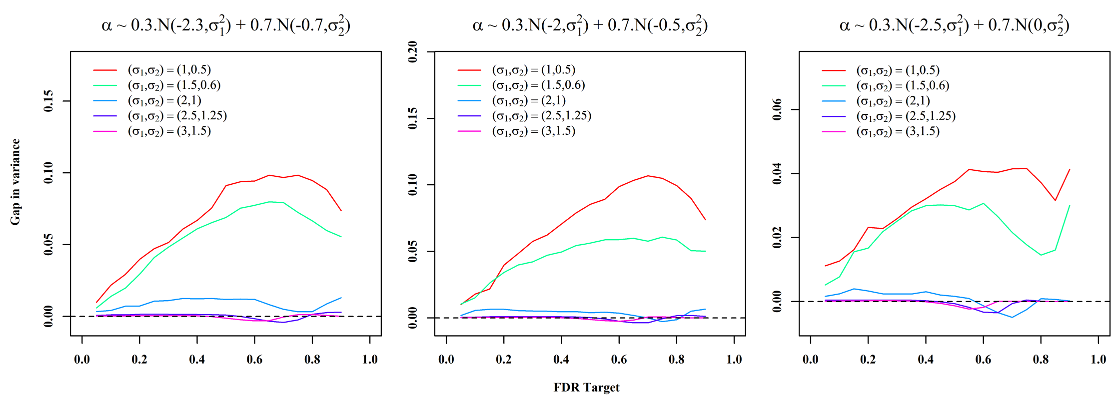


FIGURE IA6

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.

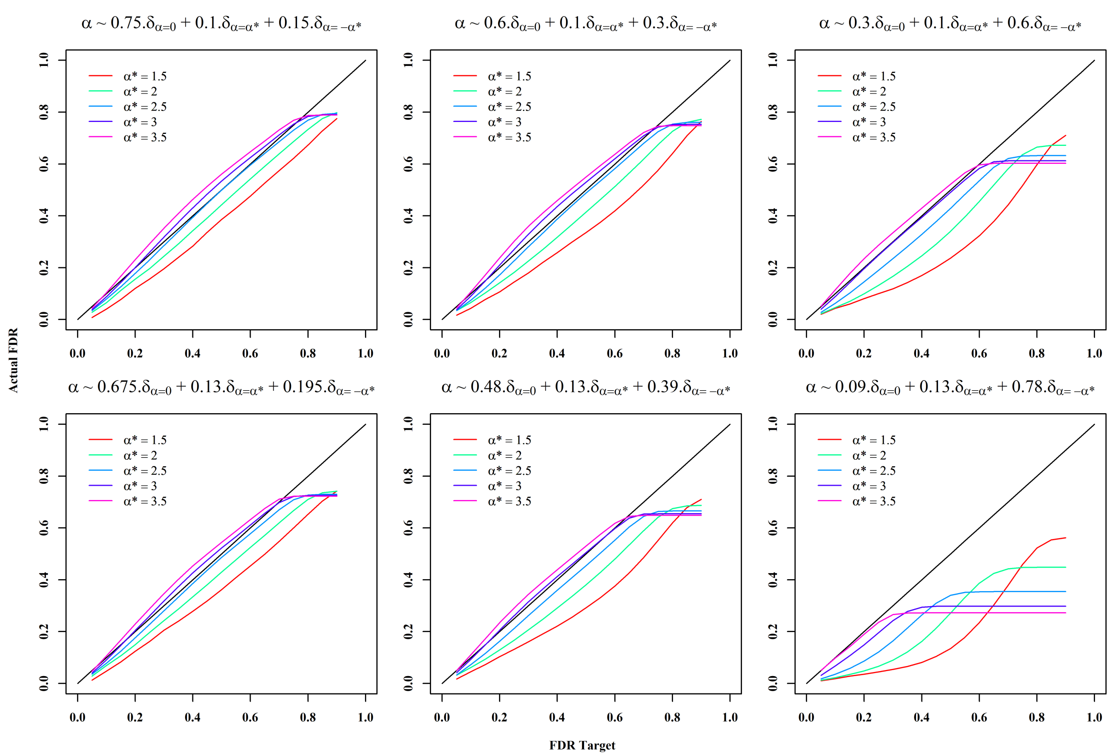


FIGURE IA7

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.

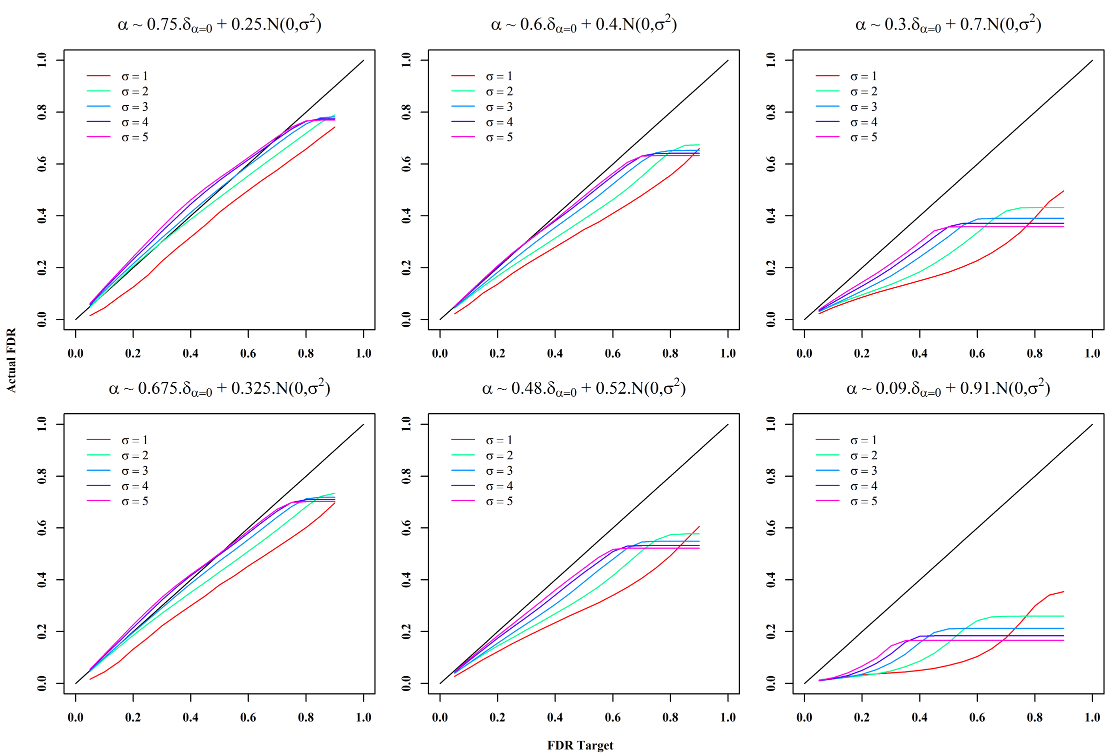


FIGURE IA8

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.

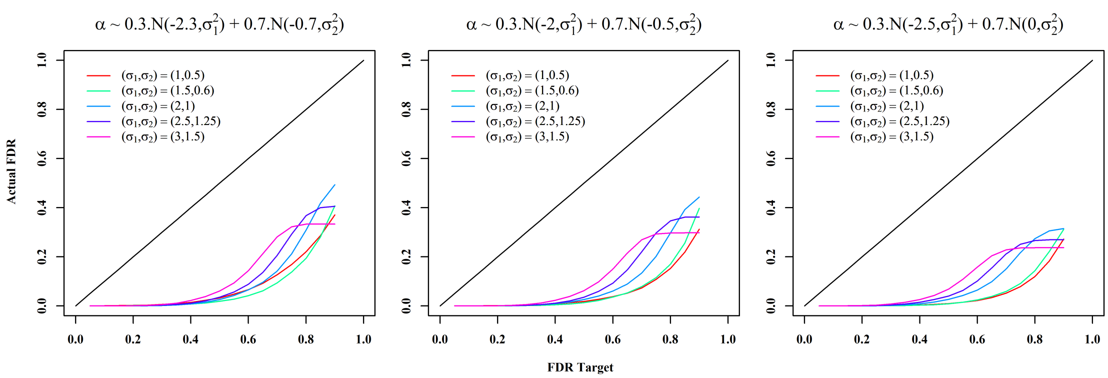
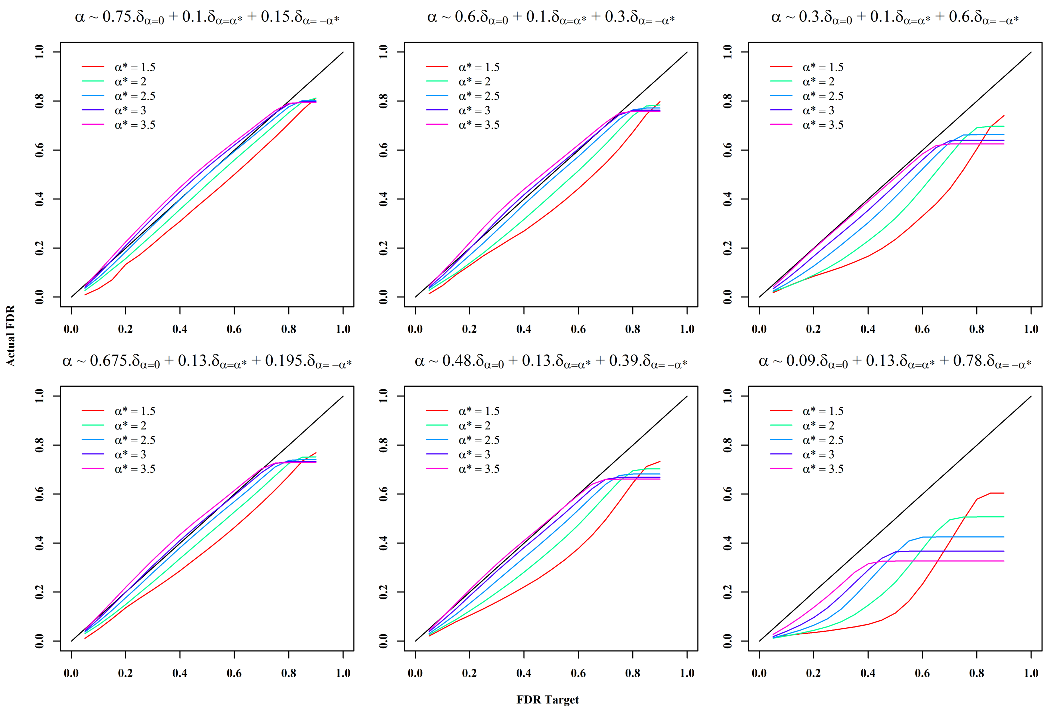


FIGURE IA9

Performance of $fFDR^{+}$ for discrete α with unbalanced panel data

The graphs show the performance of the $fFDR^{+}$ in terms of FDR control when alphas are drawn from the discrete distribution with unbalanced panel data.

Panel A: Cross-sectional Independent Data



Panel B: Cross-sectional Dependent Data

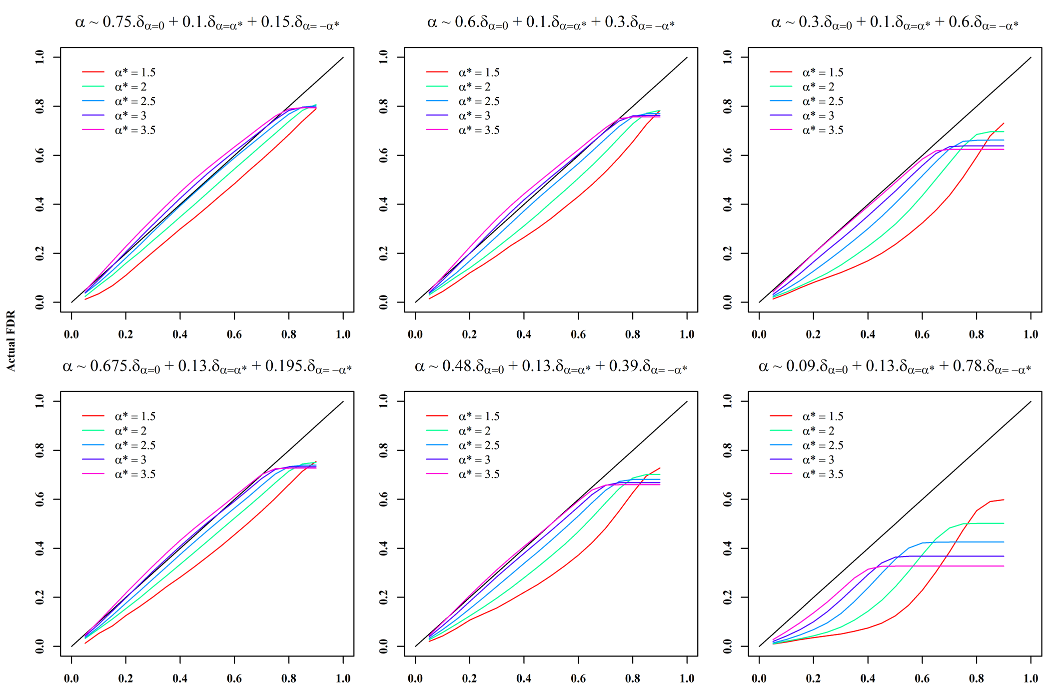
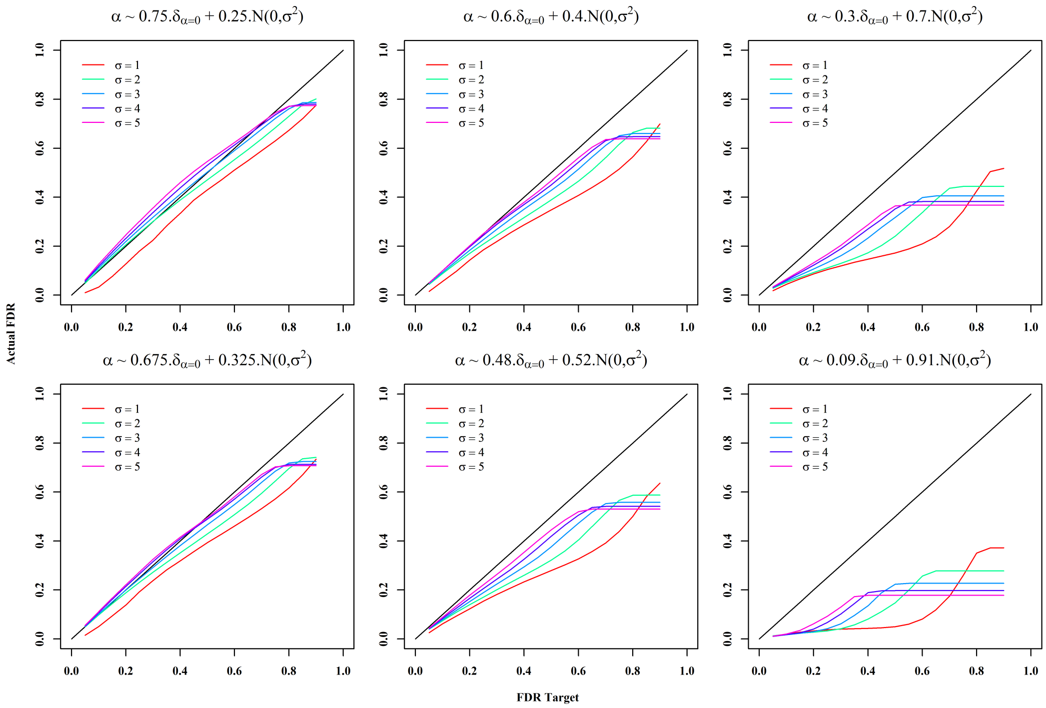


FIGURE IA10

Performance of $fFDR^+$ for discrete-normal α with unbalanced panel data.

The graphs show the performance of the $fFDR^+$ in terms of FDR control when alphas are drawn from the discrete-normal distribution with unbalanced panel data.

Panel A: Cross-sectional Independent Data



Panel B: Cross-sectional Dependent Data

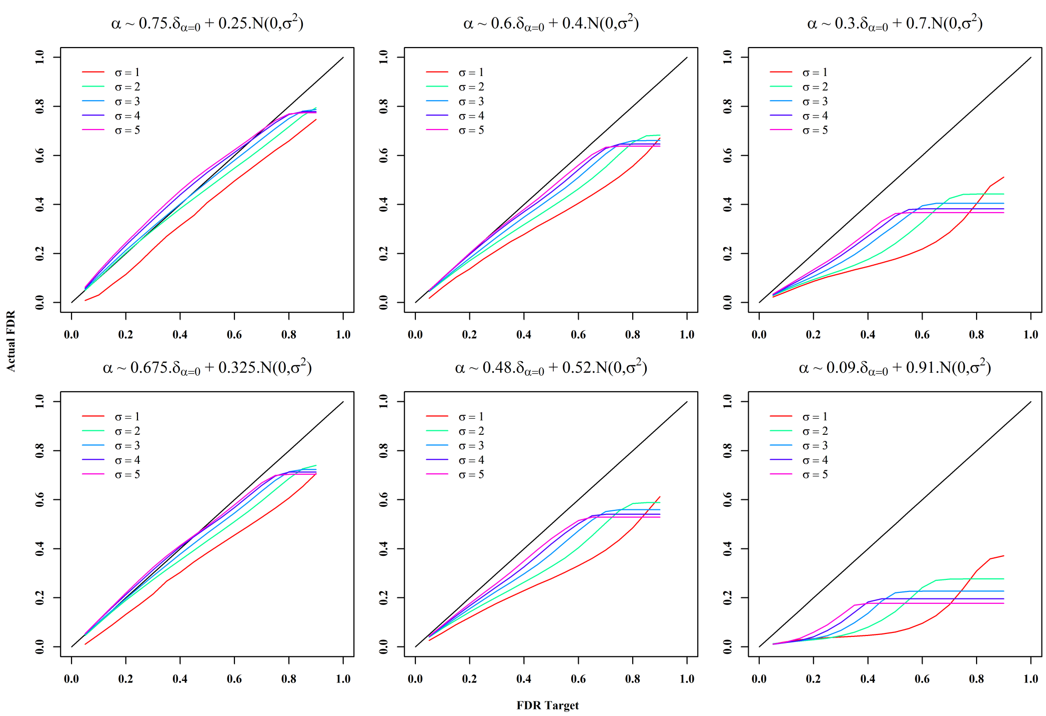
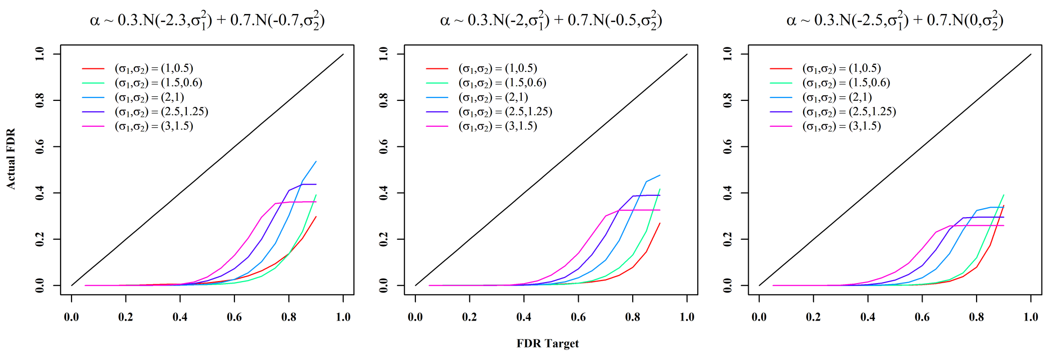


FIGURE IA11

Performance of $fFDR^{+}$ for two-normal α with unbalanced panel data

The graphs show the performance of the $fFDR^{+}$ in terms of FDR control when alphas are drawn from the mixture of two normal distributions with unbalanced panel data.

Panel A: Cross-sectional Independent Data



Panel B: Cross-sectional Dependent Data

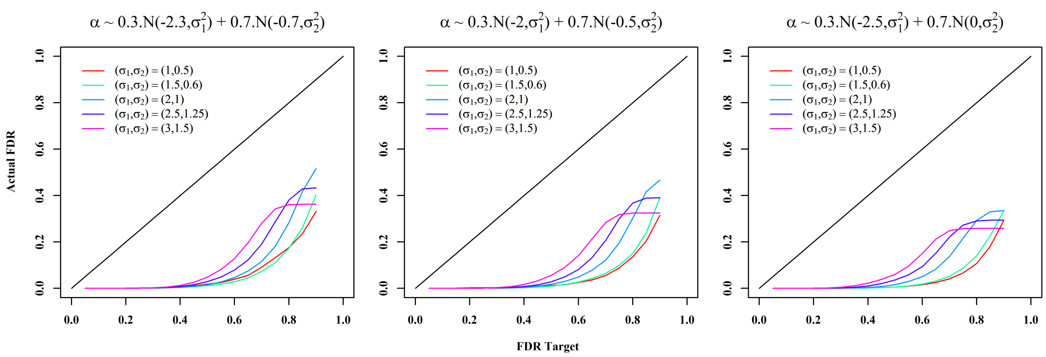
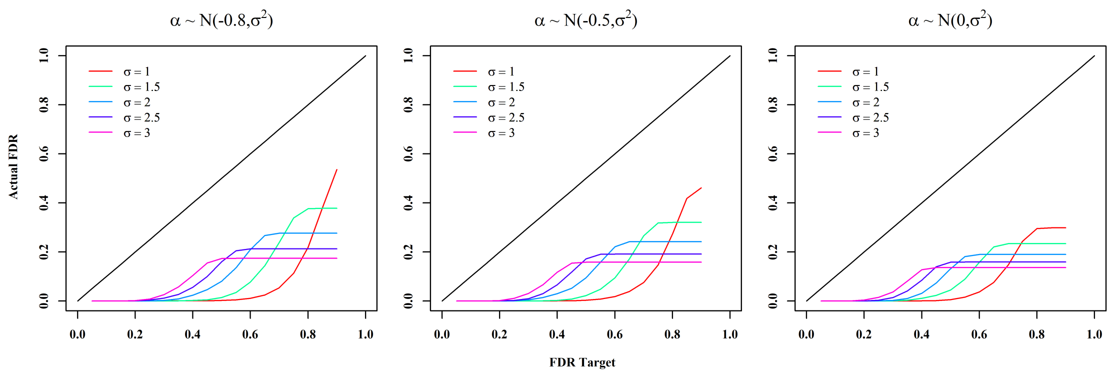


FIGURE IA12

Performance of $fFDR^{+}$ for single normal α with balanced panel data

The graphs show the performance of the $fFDR^{+}$ in terms of FDR control when alphas are drawn from the single normal distribution with balanced panel data.

Panel A: Cross-sectional Independent Data



Panel B: Cross-sectional Dependent Data

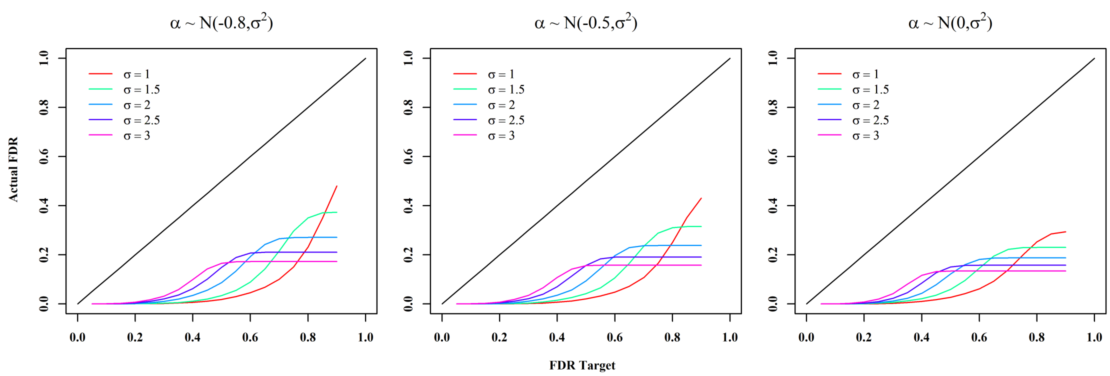
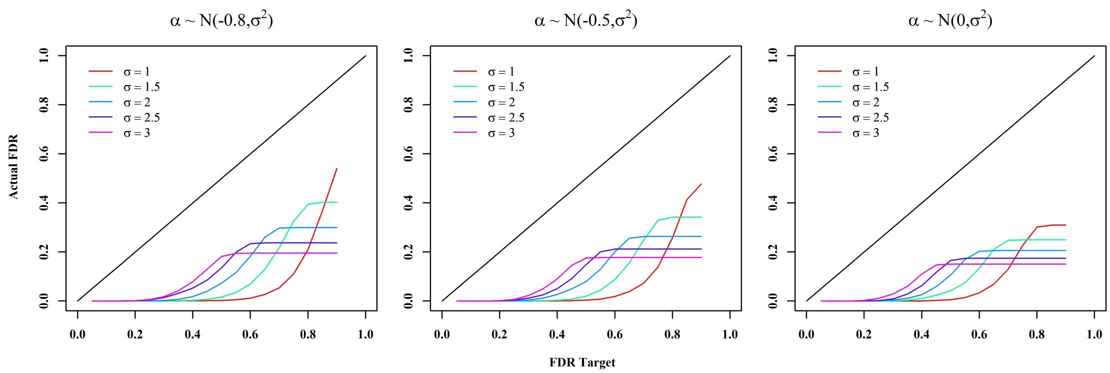


FIGURE IA13

Performance of $fFDR^{+}$ for single normal α with unbalanced panel data

The graphs show the performance of the $fFDR^{+}$ 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

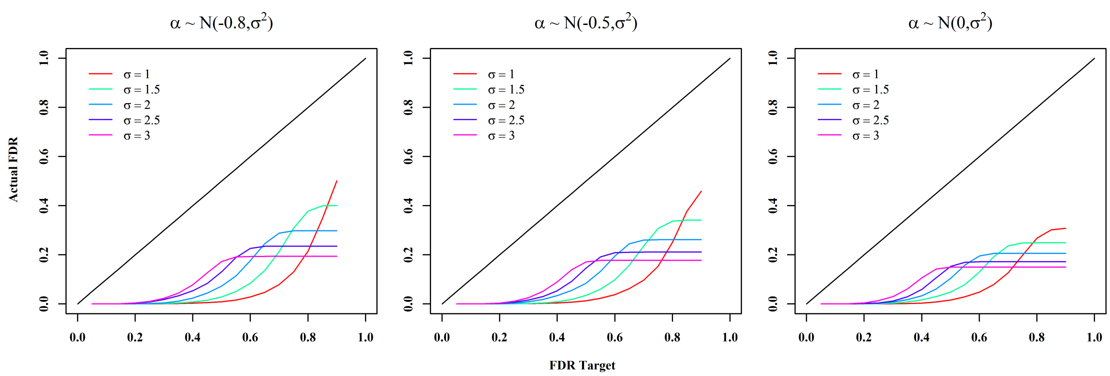


FIGURE IA14

Performance comparison between the FDR^{+} and the $fFDR^{+}$

The graphs compare the performance of the $fFDR^{+}$ in terms of FDR control and power when the generated data are based on a real data sample.

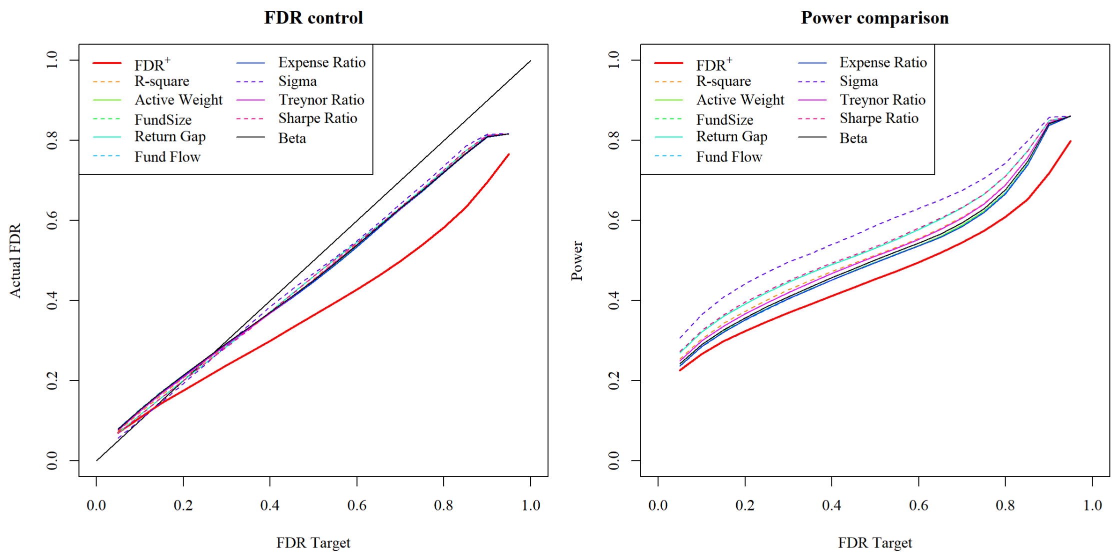


FIGURE IA15

Evolution of wealth

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

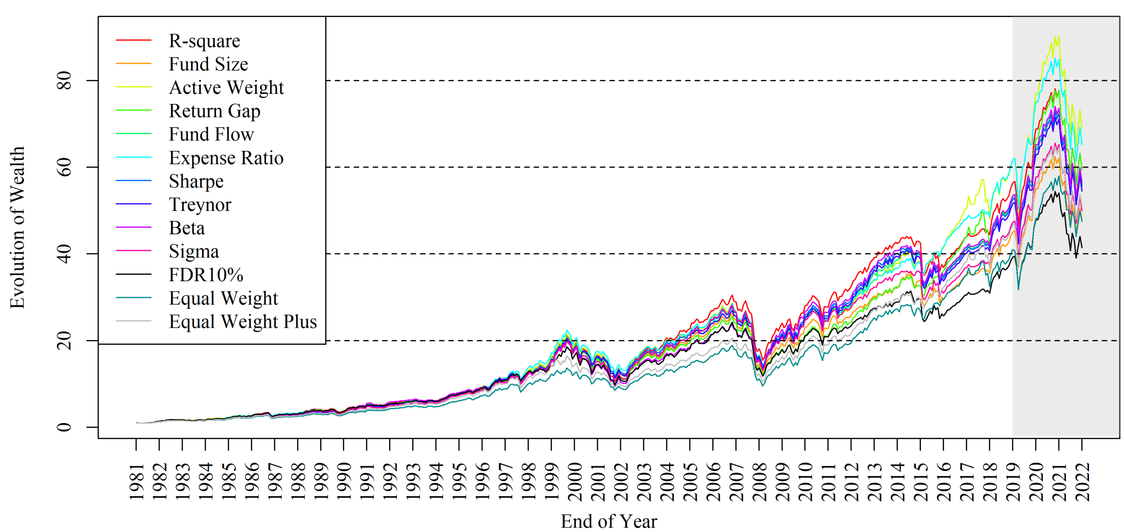


FIGURE IA16

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 ten $FDR10\%$ portfolios corresponding to the ten covariates, $fFDR10\%$, Equal Weight and Equal Weight Plus portfolios.

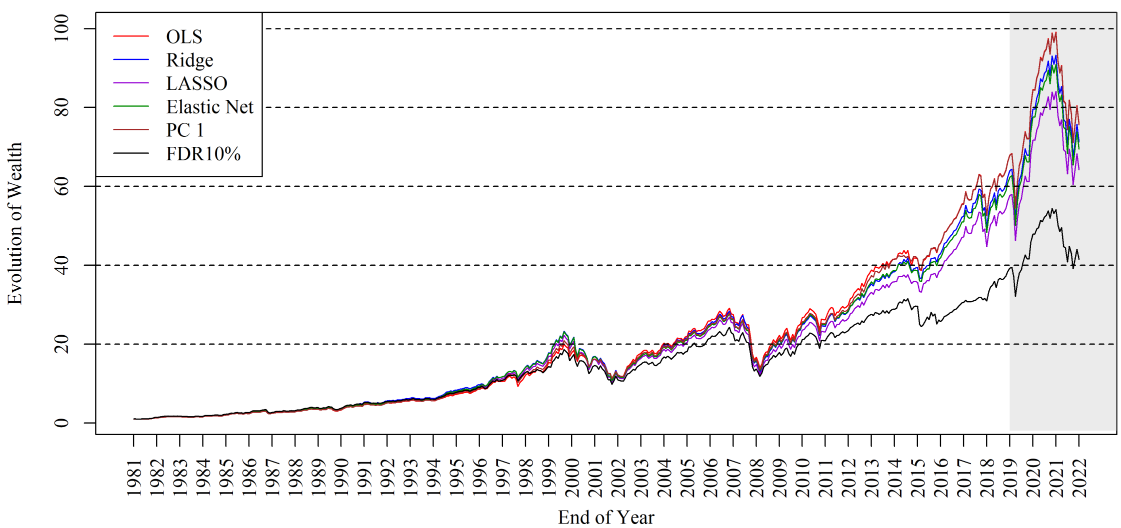


FIGURE IA17

Alpha evolution of $fFDR20\%$ and $FDR20\%$ portfolios over time

The graph presents the evolution of annualized alpha of the ten $fFDR20\%$ portfolios corresponding to the ten covariates, the $FDR20\%$ of BSW and the two equally weighted portfolios.

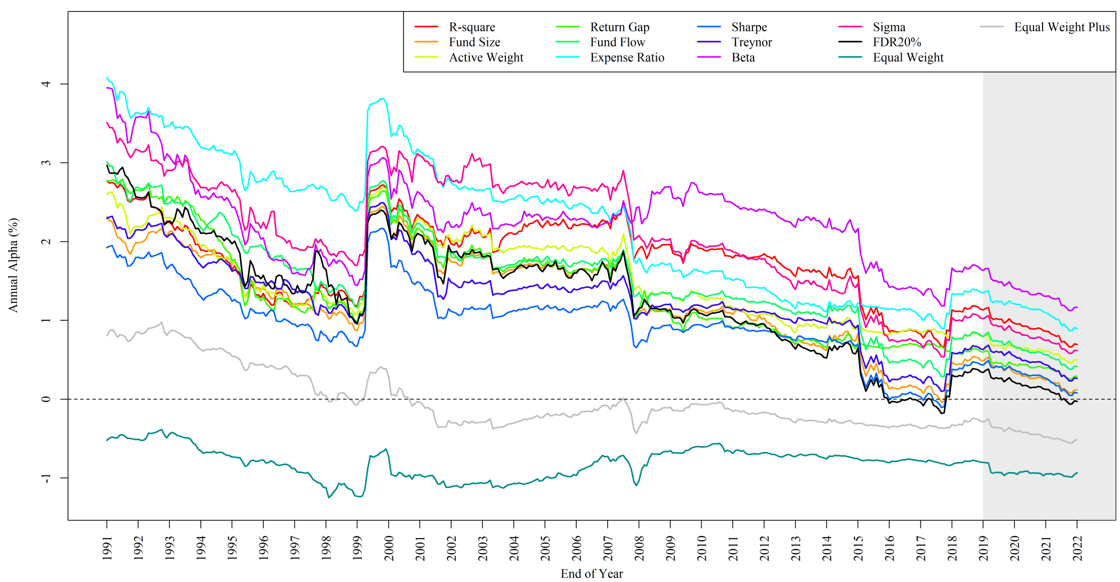


FIGURE IA18

Alpha evolution of $fFDR\tau$ portfolios under alternative coviariates' proxy.

The graph presents the evolution of annualized alpha (in %) of the ten $fFDR10\%$ portfolios (corresponding to ten covariates), the portfolio $FDR10\%$ of BSW and the two equally weighted portfolios. The proxy of a 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.

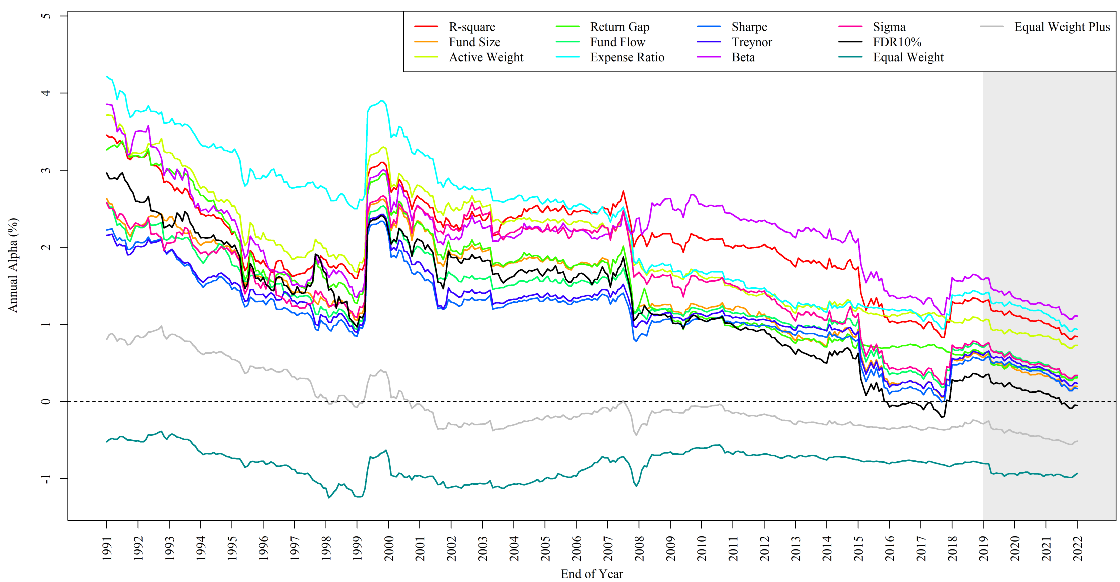


FIGURE IA19

Alpha evolution of $fFDR10\%$ and $FDR10\%$ portfolios over time

The graph presents the evolution of annualized alphas (in %) of the ten $fFDR10\%$ portfolios corresponding to the ten covariates, the portfolio $FDR10\%$ of BSW and the two equally weighted portfolios.

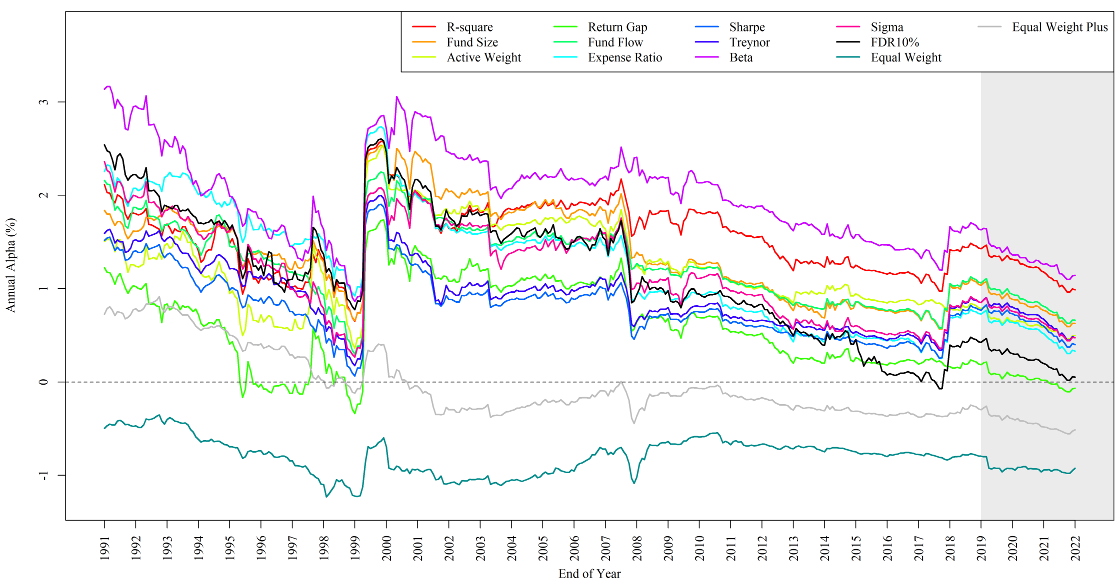
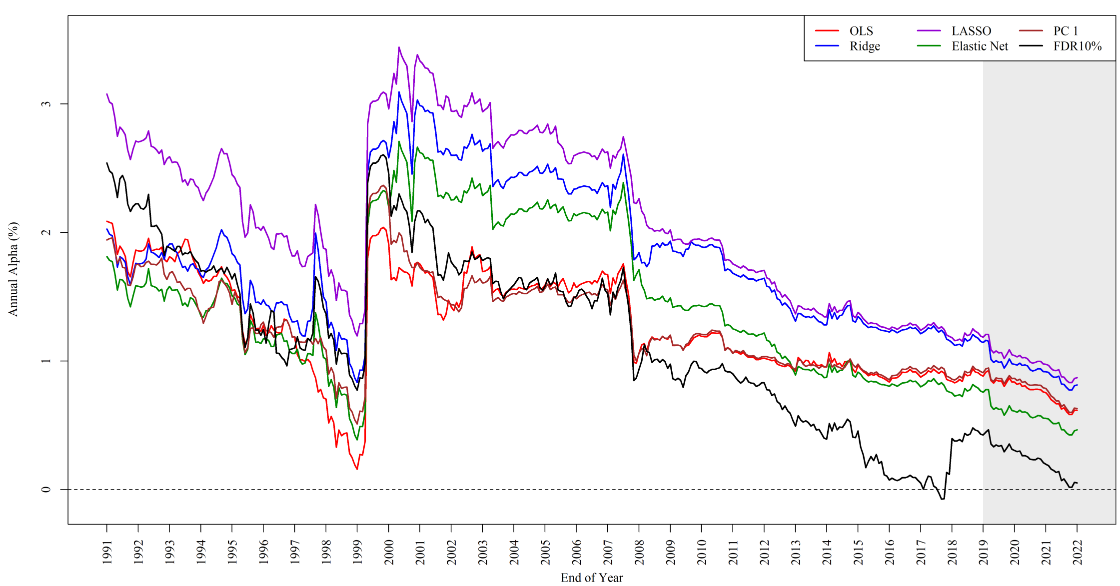


FIGURE IA20

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.



References

Amihud, Y., and R. Goyenko. "Mutual Fund's R^2 as Predictor of Performance." *The Review of Financial Studies*, 26 (2013), 667–694.

Andrikogiannopoulou, A., and F. Papakonstantinou. "Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power?" *The Journal of Finance*, 74 (2019), 2667–2688.

Barras, L.; P. Gagliardini; and O. Scaillet. "Skill, Scale, and Value Creation in the Mutual Fund Industry." *The Journal of Finance*, 77 (2022), 601–638.

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

Barras, L.; O. Scaillet; and R. Wermers. "Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? A Reply." *The Journal of Finance*. Replications and Corrigenda (web-only).

Benjamini, Y., and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)*, 57 (1995), 289–300.

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

Berk, J. B., and J. H. van Binsbergen. "Measuring skill in the mutual fund industry." *Journal of Financial Economics*, 118 (2015), 1–20.

Bessembinder, H.; S. Chen; M. J. Cooper; J. Xue; and F. Zhang. “Mutual Fund Flows and Investor Disappointment.” (2023). Working paper.

Bris, A.; H. Gulen; P. Kadiyala; and P. R. Rau. “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 (2007), 953–982.

Capponi, A.; P. Glasserman; and M. Weber. “Swing Pricing for Mutual Funds: Breaking the Feedback Loop Between Fire Sales and Fund Redemptions.” *Management Science*, 66 (2020), 3581–3602.

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

Chen, J.; H. Hong; M. Huang; and J. D. Kubik. “Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization.” *The American Economic Review*, 94 (2004), 1276–1302.

DeBruine, L. *faux: Simulation for Factorial Designs* (2021). 10.5281/zenodo.2669586. R package version 1.1.0.

Doshi, H.; R. Elkamhi; and M. Simutin. “Managerial Activeness and Mutual Fund Performance.” *The Review of Asset Pricing Studies*, 5 (2015), 156–184.

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

Harvey, C. R., and Y. Liu. “Decreasing Returns to Scale, Fund Flows, and Performance.” (2017). Working paper.

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

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

McLemore, P. “Do Mutual Funds Have Decreasing Returns to Scale? Evidence from Fund Mergers.” *Journal of Financial and Quantitative Analysis*, 54 (2019), 1683–1711.

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

Pastor, L.; R. F. Stambaugh; and L. A. Taylor. “Do Funds Make More When They Trade More?” *The Journal of Finance*, 72 (2017), 1483–1528.

Politis, D. N., and J. P. Romano. *Journal of the American Statistical Association*, 89 (1994), 1303–1313.

Sirri, E. R., and P. Tufano. “Costly Search and Mutual Fund Flows.” *The Journal of Finance*, 53 (1998), 1589–1622.

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

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

Yan, X. S. "Liquidity, Investment Style, and the Relation between Fund Size and Fund Performance." *Journal of Financial and Quantitative Analysis*, 43 (2008), 741–767.

Zheng, L. "Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability." *The Journal of Finance*, 54 (1999), 901–933.

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