Picking Hedge Funds with High Confidence

Po-Hsuan Hsu[a], Tren Ma[b], Ioannis Psaradellis[c], Georgios Sermpinis[d]

[a] *College of Technology Management, National Tsing Hua University, Taiwan*
[b] *Nottingham University Business School, University of Nottingham, Nottingham NG8 1BB, UK*
[c] *School of Economics and Finance, University of St Andrews, St Andrews KY16 9AJ, UK*
[d] *Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ, UK*

Abstract

This paper introduces a new procedure to control for family error rate (FWER) in picking out-performers. The method utilizes multiple side information to more precisely estimate the FWER and gains much higher power in detecting out-performers compared to existing ones. In assessing hedge fund performance context, the new method allows investors picking out-performing funds with high confidence, that is, with very low FWER. The yearly rebalancing portfolios of hedge funds constructed by the new method with use of available covariates beat passive benchmarks in various settings. Our further experiments show that the new method detects truly out-performing hedge fund managers who can repeat their past performance over a long horizon.

Email addresses: pohsuanhsu@mx.nthu.edu.tw (Po-Hsuan Hsu), tren.ma@nottingham.ac.uk (Tren Ma), ioannis.psaradellis@st-andrews.ac.uk (Ioannis Psaradellis), georgios.sermpinis@glasgow.ac.uk (Georgios Sermpinis)

1. Introduction

Literature on hedge fund performance has extensively focused on understanding the risk-return characteristics of hedge funds. Those studies cover the persistent performance of hedge fund managers (see, e.g., Agarwal and Naik, 2000; Baquero *et al.*, 2005; Kosowski *et al.*, 2007; Sun *et al.*, 2018), and the relationship between characteristics of the funds and their performance. Based on these findings, investors can create portfolios by sorting hedge funds based on their characteristics and past performance. However, forming portfolios in this manner has several caveats that make them impractical in practice. First, due to the large number of funds available, sorting portfolios, such as those created through quintile partition, can become large in size. Each hedge fund typically requires a significant minimum investment, which implies a huge investment requirement. Second, if only a few funds is chosen, via the rankings of the characteristics and other potential predictors, it is likely that some lucky funds are selected without proper control. Third, existing approaches in fund performance literature that controls for lucky funds focus on false discovery rate (FDR), which is expected proportion of non-outperforming funds among those selected as out-performers. In hedge fund application, controlling for this type I error is too loose as there exist many out-performing hedge funds (see, e.g., Chen *et al.*, 2017). Consequently, portfolios of hedge funds are again too large in size and more importantly, there are virtually some non-outperforming funds in portfolio at all time.

To have high confidence in hedge fund portfolio selection, investors require a tool to control for a more stringent error. For this purpose, controlling for the family wise error rate (FWER), which is probability of selecting at least one non-outperforming fund in forming the portfolio, provides a well-suited solution. To explain this, suppose the investors control for FDR at 5% when forming their portfolio. Then they would expect there are always about 5% of funds in the portfolio are non-outperforming. In contrast, if they control for FWER at 5%, the chance of having non-outperforming funds in the portfolio is 5%. Loosely speaking, if they form such portfolios yearly over 100 years, they should expect only about 5 years where their portfolio containing some non-outperforming funds. Thus, when the investors opt to control the FWER instead of the FDR, they gain a much higher level of confidence in their investment decision, especially when substantial amounts of capital is involved, as is often the case in hedge fund investments.

Literature in controlling for FWER in detecting out-performers is rich with notable contributions of White (2000), Romano and Wolf (2005), Hansen (2005) and Hsu *et al.* (2010). The main focuses are developing testing procedures that control for FWER while enhancing the performance in terms of power. All of the mentioned works rely on bootstrapping procedures and exploit only raw information such as return of funds or trading strategies. Investors' flows chase for funds that are truly out-performing, i.e., the ones that generate positive alpha - the excess return adjusted for some passive benchmark. However, hedge fund return series are usually short. Consequently, the investors assess funds based on a short periods of time, typically 24 or 36 months (see, e.g., Kosowski *et al.*, 2007; Cumming *et al.*, 2012; Chen *et al.*, 2017). Given the small number of observations, the mentioned existing FWER methods are struggling in detecting even a small number of the out-performing hedge funds. A more powerful procedure is therefore in high demand.

This study fills the gap by introducing a new approach to control for FWER. The new approach is based on statistical framework of Zhou *et al.* (2021), which estimates the FWER as a function of multiple informative covariates. The new approach deviates from the framework of Zhou *et al.* (2021) by specifically aiming to control the FWER among the discoveries in the right tail. It is well-suited to controlling for luck in the hedge fund portfolio selection. First, it has a high power in detecting out-performing hedge funds, which allows investors picking fund with a very low FWER. Secondly, while effectively controlling for such a low error rate the approach is able to select a reasonable number of funds, thereby making the investment size more feasible. Third, and more importantly, the hedge funds identified by the new approach, though based on assessment over a short period, perform persistently in some long out-of-sample (OOS) periods. This makes the method a high potential in real world practice.

Distinguished from existing methods that solely rely on funds' adjusted returns or alpha, our approach harnesses multiple sources of side information to enhance the detection power. More specifically, we assess performance of a fund via testing its alpha against zero. We implement the framework of Zhou *et al.* (2021) with use of the side information to detect all funds that having significant non-zero alpha. As we are focusing on the right tail of the alpha distribution, we select as out-performing funds the subset of those detected non-zero alpha funds having positive estimated alpha. We name this procedure the *FWER "plus"* ($fwer^+$). We show that the procedure controls well for FWER at any given targets, and when an informative covariate is available, it gains an impressive power in detecting truly positive alpha funds. Controlling for FWER at 5% target, our procedure outperforms the stepwise approaches of Romano and Wolf (2005) and Hsu *et al.* (2010), the two most powerful ones in recent finance and economics applications, with gaps that varies from 1% to 15% depending on the number of observations per fund and the informativeness level of the covariates.

In empirical experiments, we first construct yearly rolling portfolios of out-performing hedge funds with control for FWER at small targets spanning from 0.1% to 5%. We use 20 covariates that are available and easily calculated from the hedge funds return. The $fwer^+$ with use of single or multiple covariates always detects non-empty group of out-performing funds despite of the small targets of FWER and the choice of short in-sample (IS) periods such as 24, 36 or 48 months. We then invest in the selected funds in the following year as OOS and roll forward till the end of 2021. The portfolios gain statistically significant positive alphas which spans from 4% to more than 5% per annum. We see that, the portfolios controlling for smaller target of FWER tend to gain higher alphas, which transform to Sharpe ratio of more than 2. These results are robust to the choice of IS periods and asset pricing models. To further examine the persistence in the performance of the funds detected by the $fwer^+$, we consider the choices of longer OOS periods. We see that even with OOS of four years, the $fwer^+$ portfolios still gain roughly as high alpha as the choice OOS of one year. This suggests that the $fwer^+$ helps the investors selecting truly out-performing funds based on assessing funds over a relative short past performance. We then enhance the informativeness of the covariates via using machine learning techniques to forecast future funds' return and use them as new covariates. The $fwer^+$ portfolios with use of those new covariates are able to generate

Sharpe ratio of more than 2.5. Finally, we construct portfolios of only a single fund who performs best among those selected by the $fwer^+$ in in-sample period. These portfolios gain slightly lower alpha than those in previous exercise but perform impressively in terms of Sharpe ratio which can reach 5.3.

Our paper thus contributes to literature in both econometrics and economics aspects. First, it provides a powerful approach to detecting out-performers, which is in demand not only in hedge funds context but also in many other topics in economics and finance. Second, it shows that the performance of the genuine hedge funds detected by the new procedure persists. By using the new method, the investors can detect hedge fund managers who are able to repeat their performance.

The remains of the paper are organised as follows. Section 2 introduces our $fwer^+$ procedure. Section 3 describes our hedge fund dataset while Section 4 discusses on our choice of funds' performance measure. Section 5 provides simulations to show the performance of the $fwer^+$. Section 6 is devoted for empirical analyses and Section 7 concludes our paper.

2. FWER and informative covariates

Suppose that we are assessing n hedge funds based on a performance metric ϕ. We test for each fund i a hypothesis

$$H_0 : \phi_i = 0 \quad \text{against} \quad H_1 : \phi_i \neq 0. \tag{1}$$

where ϕ_i is the true but unknown value ϕ of the fund i, $i = 1, \dots, n$.

This study focuses on detecting out-performing hedge funds based on their alpha, i.e., the metric ϕ_i is alpha of the fund i, $i = 1, \dots, n$. We aim to detect the funds having positive ϕ, with controlling for the probability of selecting at least one non-outperforming fund at a predetermined target $\tau \in (0, 1)$. Formally, let R be number of funds selected as out-performing funds and among them F funds actually having $\phi \leq 0$. We are attempting to control a type I error which is defined as FWER $= \mathbb{P}(F \geq 1)$ at the target τ.

Literature in detecting out-performers with controlling for FWER focuses on one-sided test with a composite null, and numerous testing procedures have been developed. Most recent contributions are the procedures of Romano and Wolf (2005), Hansen (2005) and Hsu *et al.* (2010). The common ground of those procedures are based on bootstrapping with use of funds' returns (or more generally, some relative performance variable). These procedures suffer from the computational burden and are not using additional information, which are informative and available alongside the tests. In picking out-performing funds, which typically assesses funds' performance based on short time series of return, these approaches appear to be lack of power.

The low power issue of multiple testing procedures has been gained attention across areas of science. Recent developments in statistics attempt to incorporate side information to raise the power of detecting false nulls, such as the contributions of Ignatiadis and Huber (2021) and Zhou *et al.* (2021). Nevertheless, the implementation of those innovative approaches in the selecting out-performers has not been addressed. In this study we are proposing a simple procedure to further develop the framework of Zhou *et al.* (2021) to solving the low power problem of the existing approaches.[1] In the follows, we summarise the framework and subsequently propose our procedure.

2.1. The use of informative covariates in controlling FWER

Suppose we set a target of FWER at $\tau \in (0,1)$ and there is a set of d informative covariates $Z^1, \ldots, Z^d$ carrying the information on probability that the null of tests (1) being true. Each Z^k is a column vector $(Z_1^k, \ldots, Z_n^k)'$, $k = 1, \ldots, d$. For convenience, we denote $\boldsymbol{Z} = (Z^1, \ldots, Z^d)$ and thus $\boldsymbol{Z}_i$ means $(Z_i^1, \ldots, Z_i^d)$. For each $i = 1, \ldots, n$, let P_i be the random variable representing the $p-$value of the test (1) corresponding to the fund i and p_i be its realization. Conditional on $\boldsymbol{Z}_i = \boldsymbol{z}_i$, we denote the prior probability of the null hypothesis i being true by $\pi_0(\boldsymbol{z}_i)$. We model the distribution of P_i as a mixture of two groups in which the weights of the first and the second group is $\pi_0(\boldsymbol{z}_i)$ and $1 - \pi_0(\boldsymbol{z}_i)$, respectively. Let f_0 and f_{alt} be the density functions of the first

[1] The framework of Zhou *et al.* (2021) is more efficient in terms of computation and has been shown to be more powerful than that of Ignatiadis and Huber (2021).

and second group, respectively. Formally, we have

$$P_i|(\boldsymbol{Z}_i = \boldsymbol{z}_i) \sim \pi_0(\boldsymbol{z}_i)f_0(\cdot) + (1 - \pi_0(\boldsymbol{z}_i))f_{alt}(\cdot) \tag{2}$$

Thus, the covariates $\boldsymbol{Z}$ carry their information through the $\pi_0(\boldsymbol{z}_i)$ which takes value differently across the tests. In contrast, the f_0 and f_{alt} respectively are the density functions of those $p-$values under true nulls and false nulls, and they are in the same form for all tests, i.e., not depending on i. In this model, the two density functions do not depend on $\boldsymbol{Z}$ neither. We assume the $p-$value of a true null is uniformly distributed, i.e., $f_0(p) = 1 \; \forall p$.

In conventional approaches, the rejection region is determined by a common threshold T which is fixed for all tests, i.e, a hypothesis i is rejected if and only if $p_i \leq \Theta$. The idea now is to determine for each hypothesis i a threshold which is a function of $\boldsymbol{z}_i$ denoted by $\Theta(\boldsymbol{z}_i)$, i.e., the null of hypothesis i is rejected if and only if $p_i \leq \Theta(\boldsymbol{z}_i)$. For this purpose, we assume the $f_{alt}(p)$ to be a strictly decreasing function of p and follow the developments of Zhou *et al.* (2021), the mentioned threshold is defined as

$$\Theta(\boldsymbol{z}_i) = f_{alt}^{-1}\left(\frac{\pi_0(\boldsymbol{z}_i)\theta^*}{1 - \pi_0(\boldsymbol{z}_i)}\right) \tag{3}$$

where $\theta^* = \min\left\{ \theta > 0 : \; \sum_{i=1}^n \pi_0(\boldsymbol{z}_i)f_{alt}^{-1}\left(\frac{\pi_0(\boldsymbol{z}_i)\theta}{1-\pi_0(\boldsymbol{z}_i)}\right) \leq \tau \right\}$ and f_{alt}^{-1} is the inverse function of f_{alt}.

In practice, the $\pi_0(\boldsymbol{z})$ is modelled as a logit function which has a form of $\pi_0(\boldsymbol{z}_i) = 1/(1 + e^{-b_0 - \boldsymbol{b}'\boldsymbol{z}_i})$, where $\boldsymbol{b} = (b_1, \ldots, b_d)$ is the column vector of the coefficients of the d covariates, while $f_{alt}(p)$ is modelled as a beta distribution $f_{alt}(p) = kp^{k-1}$ for $k \in (0, 1)$. The parameters $b_0, \ldots, b_d$ and k are estimated via an expectation and maximization algorithm.[2]

[2] For the details of developments and algorithms, readers are referred to Zhou *et al.* (2021).

2.2. Application in picking outperforming funds

As we aim to picking truly positive alpha funds, we actually need to controlling for FWER in the group of the selected out-performing funds. In this section, we propose a simple procedure to applying the framework of Zhou *et al.* (2021) to the problem.

Given a target $\tau \in (0, 1)$ of FWER, our procedure, namely $fwer^+$, consists of two steps. First, we implement FWER procedure of Zhou *et al.* (2021) on the population of funds with controlling FWER at the target τ. The procedure will detect a set of abnormal funds, say A, which includes both under- and out-performing funds. Second, we pick from this set only a subset consisting of those funds having positive estimated alpha, say A^+. Since the probability of having at least one zero-alpha funds in A is less than τ, this also conservatively holds for the set A^+ as it is a subset of A. Assuming that there are no under-performing funds that are very lucky and selected in A^+, then the set A^+ consists of out-performing funds with FWER being controlled at the target τ. As controlling for FWER is stringent, this assumption is likely to be valid.

As will be shown in Section 5, the $fwer^+$ controls well for FWER at any given targets and, when informative covariates are available, it is more powerful than existing methods.

3. Data

Our hedge fund data is collected from Lipper TASS database. Following previous research, we impose screenings to deal with common sample biases (see, Fung and Hsieh (2001); Bali *et al.* (2012); Chen *et al.* (2021b)). We include only US dollar-based hedge funds in our sample to avoid duplicate funds listed in different currencies. We do not consider funds that have not reported any data during during the study period as well as we include both "live" and "graveyard" funds from January 1994 to account for survivorship bias. To address the back-fill bias issue, we exclude the first 12 months of returns for each fund. To control for multi-period sampling bias, we require all funds have at least 36 months of return history. For each fund, we consider only months where the fund's net-of-fee return and asset under management data are available. After the above restrictions, we end up with a sample of 5,314 funds covering the period January 1994 to

December 2021.[3]

4. Performance measure

Following majority of the existing literature on hedge fund performance, we use the seven-factor model alpha of Fung and Hsieh (2004) as our baseline performance measure of a fund.[4] For each fund i we regress

$$r_{i,t} = \alpha_i + \boldsymbol{F_t}\boldsymbol{\beta}_i + \varepsilon_{i,t} \tag{4}$$

where $r_{i,t}$ is the excess return, $\boldsymbol{F_t} = [F_t^1, F_t^2, \ldots, F_t^7]$ is the 1×7 matrix of the seven risk factors, $\boldsymbol{\beta}_i = [\beta_i^1, \beta_i^2, \ldots, \beta_i^7]'$ is the 7×1 matrix of coefficients, and $\varepsilon_{i,t}$ is the noise of the fund i at month t. The seven factors include an equity market factor which is the S&P500 excess return (MKT); the Wilshire small cap minus large cap return; the change in the constant maturity yield of the 10-year Treasury ($\Delta 10Y$); the change in the spread between Moody's Baa yield and the 10-year Treasury ($\Delta CredSpr$) (i.e., credit yield spread); and 3 trend-following factors for bonds ($BDTF$), currency ($FXTF$), and commodities ($CMTF$).

In Section 6.6, we additionally conduct robustness check of the fund portfolio performance of our proposed method under the use of different multi-factor models in selecting funds with significant alphas following previous and recent studies, such as those of Bali *et al.* (2012, 2014) and Chen *et al.* (2023). Those include the four-factor model of Carhart (1997), a six-factor model in which we add the two risk factors $\Delta 10Y$ and $\Delta CredSpr$ into the four-factor model of Carhart (1997), and a nine-factor model in which three more risk factors including $BDTF$, $FXTF$ and $CMTF$ are added into the six-factor model above. The four risk factors in the four-factor model consist of the market's excess return on the CRSP NYSE/Amex/NASDAQ value-weighted market portfolio, the Fama–French small minus big factor, the high minus low factor, the momentum factor.[5] We also con-

[3]Following hedge fund literature, e.g., Chen *et al.* (2023), we exclude monthly net-of-fee returns that are below -90% or excess 300%.

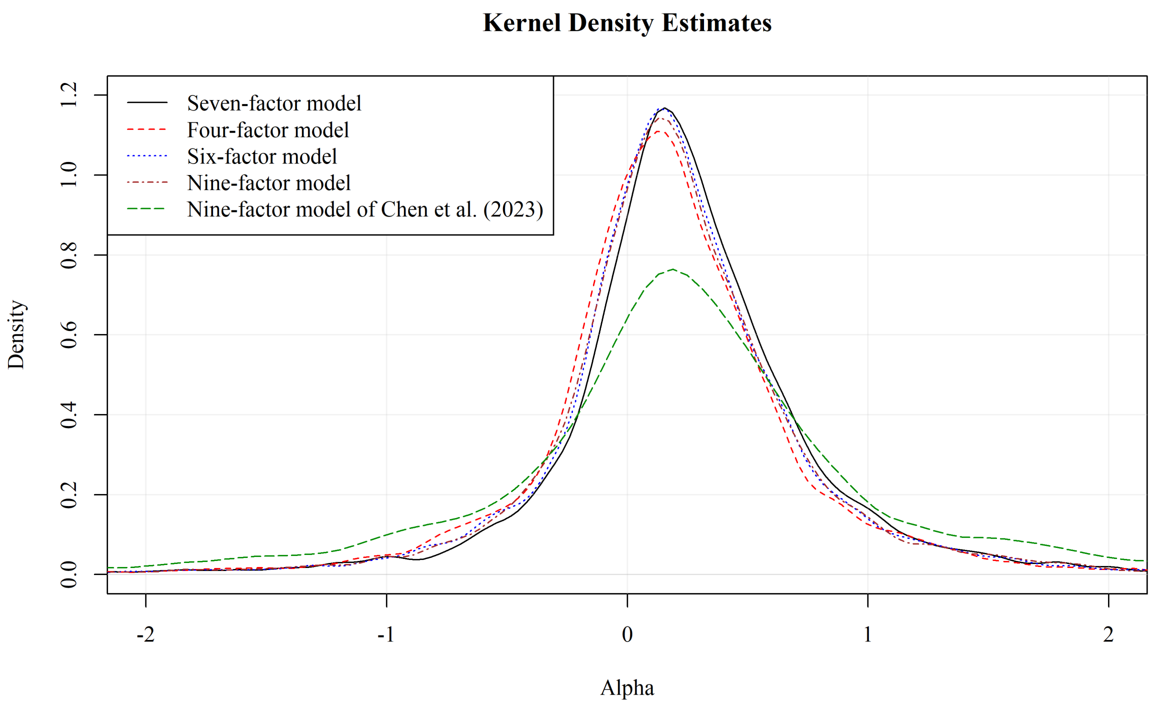
[4]See, e.g., Kosowski *et al.* (2007) and Chen *et al.* (2017).

[5]We follow David Hsieh's website to collect the seven risk factors of the Fung and Hsieh (2004) factor

sider the proposed nine-factor model of Chen *et al.* (2023), which outperforms previous multi-factor models in explaining hedge fund returns. Their multi-factor model includes the MKT, asset growth (AGR), betting-against-beta (BAB), low-risk ($LRSK$), return on assets (ROA), time-series momentum ($TSMOM$), $\Delta 10Y$, $\Delta CredSpr$ and the spread between the constant maturity yield of the 10-Year Treasury bill and constant maturity of the 3-Month Treasury bill (term spread) ($TERM$) factors. [6]

Figure 1 depicts the distributions of hedge fund alphas under the use of the mentioned four models via estimating their kernel density curves. We see that all curves have a peak at some positive alpha point. This is similar to reports in hedge fund literature such as those of Chen *et al.* (2023). It indicates the presence of a majority positive alpha hedge

Figure 1: Distribution of hedge fund alphas. The figure shows the kernel density estimates of funds' alpha with use of different factor models including four-, six-, seven-, and two nine-factor models. We require at least 36 observations per fund and for each model we regress excess return of each individual fund on the model's risk factors to obtain its alpha and then estimate the kernel density of the alpha population.



model, the one-month Treasury bill rate and four risk factors of the Carhart (1997) factor model have been obtained from the data library of Kenneth French.

[6]We collect the nine risk factors of the Chen *et al.* (2023) factor model from the the data library of Kenneth French, Andrew Chen's Open Source Asset Pricing data library, AQR's online datasets, Global Factor Data and Federal Reserve Bank of St. Louis.

funds under the considering factor models. Compared to Chen *et al.* (2023), our density curve of Chen *et al.* (2023)'s nine-factor model alpha has a peak at positive point rather than at zero due to the fact that our data includes only funds in the Lipper TASS database, a subset of the data studied in their paper.

5. Simulation studies

In this section, we conduct a set of simulations to show: i) our proposed approach works well in terms of controlling FWER and outperforms the existing methods in terms of power; ii) the excellent performance of $fwer^+$ under variants of important factors.

As presented in Section 2.1, the FWER is estimated based on a mixture model assumption in which the informativeness of the covariates is conveyed via the null proportion $\pi_0(\boldsymbol{z})$. Thus, the relationship between $\boldsymbol{Z}$ and the non-zero alpha is not be reflected in the model.[7] This implies that, in simulation design, the non-zero alpha component can be freely generated and not depending on the value $\boldsymbol{Z}$. Thus, the value of $\boldsymbol{Z}$ will be generated in the first step. Based on these values and assumption on $\pi_0(\boldsymbol{z})$ we assign the true nulls, i.e., determine which funds have zero-alpha. The remaining funds will be assigned with non-zero alpha values from a predetermined distribution. Remark that, the signals of being false null, i.e. the magnitude of the non-zero alphas, are transformed to rejection rule via the estimated f_{alt}.

To estimate necessary parameters for data generating process, we use data of all the 5,314 funds in our sample and the risk factors of our baseline model. Specifically, we utilize the seven risk factors data which start from January 1995 to December 2021 and estimate their mean and matrix of correlation coefficients. We calculate the factor loadings ($\boldsymbol{\beta}_i$) for the funds by regressing each of the 5,131 funds on the risk factors. We design simulations to cover different scenarios in applications. We generate balanced panel data of n funds with T observations per fund. As hedge fund literature focuses on constructing portfolio with use of 24-, 36- and 48- month IS periods, we first consider

[7]This differs from other models, such as Chen *et al.* (2021a), where they take into count the joint distribution of alternative hypothesis' $p-$value and covariates. This way, the link between the non-zero alpha and value of covariates is reflected in the models.

$T \in \{24, 36, 48\}$.

In each iteration, we conduct 7 steps as followings.

1. We generate two covariates $\boldsymbol{Z} = (U, V)$ from a bivariate normal distribution with mean 0 and standard deviation of 1 and with specific correlation coefficient ρ of U and V. We consider two cases of $\rho \in \{0, 0.5\}$.

2. The $\pi_0(u, v)$ has a logit form $1/(1 + e^{-b_0 - b_1 u - b_2 v})$ where the triple (b_0, b_1, b_2) is one of the three cases $(0.5, 1, 1)$, $(0.5, 1.5, 1.5)$ and $(0.5, 2, 2)$. These three cases cover a weak, moderate and strong relationship between covariates and the probability of a fund being zero alpha, respectively. The choice of $b_0 = 0.5$ is to generate a set of simulated hypotheses with a null proportion, denoted by π_0, of 60%. For non-zero alpha funds, we generate data sets such that a half of them have positive alpha.[8] Given a specific choice of $\pi_0(u, v)$, we determine funds having zero-alpha funds as follows. For each fund i, we draw a random value from Bernoulli distribution which takes value 0 with probability of $\pi_0(u_i, v_i)$. Funds with drawn values of 0 are assigned as zero-alpha funds. We assign randomly a half of the remaining funds with alpha of $\alpha > 0$ and the rests with alpha of $-\alpha$ where the monthly alpha $\alpha \in \{0.5\%, 1\%, 1.5\%\}$. The choice of 0.5% is close to the third quantile of the estimated alphas in our data sample while other values are chosen under assumption that the true α is some value in between the third quantile and the maximum of the estimated alphas.

3. We generate the risk factors, $\boldsymbol{F}^s$ and their loadings $\boldsymbol{\beta}^s$ from normal distributions such that their parameters are the same as those of the real sample counterpart in whole sample period (i.e., from January 1995 to December 2021).

4. We generate the simulated excess return of each fund via the following formula

$$R_{i,t} = \alpha_i^s + \boldsymbol{F}_t^s \boldsymbol{\beta}_i^s + \varepsilon_{i,t} \tag{5}$$

[8] By using procedures of Storey (2002) and Barras *et al.* (2010) we find the estimated proportions of zero-alpha (π_0) and out-and under-performing funds (π^+ and π^-) are 59%, 41% and 0%. To cover general scenarios in applications we consider $\pi_0 = 60\%$ and $\pi^+ = \pi^- = 20\%$. We additionally conduct simulations with $\pi^+ = 40\%, \pi^- = 0\%$ and present the results in Section IB of the Internet Appendix.

where the noise $\epsilon_{i,t}$ is drawn independently from a normal distribution $N(0, \sigma^2)$ with σ to be set at 2.2% as the median of standard deviation of error terms estimated from the real sample fund-by-fund regressions.

5. For each fund, we then regress its simulated returns on the seven simulated factors, $\boldsymbol{F}_t^s$, to obtain its estimated $\hat{\alpha}$ and the $p-$value of testing its alpha against 0.
6. We implement the StepM of Romano and Wolf (2005), Stepwise-SPA of Hsu *et al.* (2010) and $fwer^+$ procedures, controlling for FWER at predetermined targets $\tau \in (0, 1)$, to detect truly positive alpha funds with use of the $\hat{\alpha}$s, calculated $p-$values and simulated covariates.[9] We consider $\tau \in \{0.1\%, 1\%, 2\%, \ldots, 20\%\}$.
7. By comparing the simulated α^s and the selected out-performing funds, we record the family wise error (FWE) which takes value 1 if there is at least one of the funds in the negative or zero α^s groups classified as out-performers. We also calculate the detected proportion which is the ratio of truly out-performing funds detected by each procedure.

We repeat the steps 1 to 7 across 1000 iterations and calculate estimates of the actual FWER as the ratio of number of times we observe FWE = 1 over 1000, i.e. the frequency of error, and the power as average of the detected proportion recorded in the step 7 above.

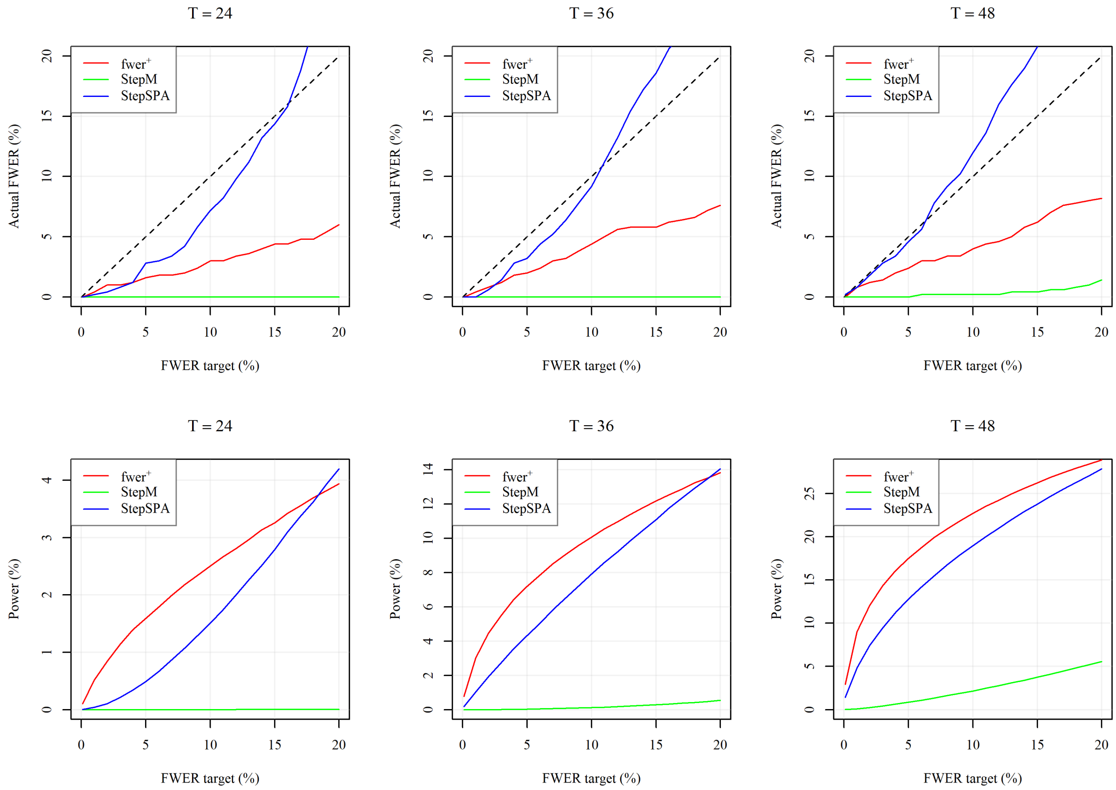
5.1. A comparison to existing methods

In order to compare the performance of the $fwer^+$ to existing procedures, the StepM of and StepSPA, we opt a specific simulated data setting with $n = 1000$ funds and alpha magnitude of non-zero alpha fund $\alpha = 1$.

The performance of the procedures are presented in Figure 2 where all numbers are in percentage. In each of the top three sub-figures, we depict the estimated actual FWER given the targets where each curve represents a procedure. At a specific target, a procedure controls well for FWER if the corresponding represented point on the curve at that target is lying below or on the dashed 45° line. We see that the $fwer^+$ and StepM procedures control well for considering FWER targets regardless the number of observations

[9] The detail of the StepM and StepSPA procedures under our framework is presented in Section IA of the Internet Appendix.

Figure 2: Performance comparison. The figure compares the $fwer^+$ and existing procedures including StepM of Romano and Wolf (2005) and StepSPA of Hsu *et al.* (2010) in terms of FWER control (top three sub-figures) and power (bottom three sub-figures). The simulated data are balanced panels of 1000 funds with T observations per fund. From the left to the right, T takes values 24, 36 and 48. The input covariates U, V of the $fwer^+$ are independent.



per fund whereas the StepSPA starts to lose its controlling of FWER when the target is higher than 15%, 10% and 5% in $T = 24, 36$ and 48 settings, respectively.

In terms of power, provided that the FWER is controlled well, the $fwer^+$ always performs better than the other two with gaps depending on the FWER target and number of observations per fund T. For instance, at target $\tau = 5\%$ and $T = 36$, the gaps in power of the $fwer^+$ compared to the StepM and StepSPA are 7.5% and 3.5%, respectively. Those numbers are larger (smaller) for $T = 48$ ($T = 24$) case which are about 15% and 5% (1.5% and 1%), respectively.

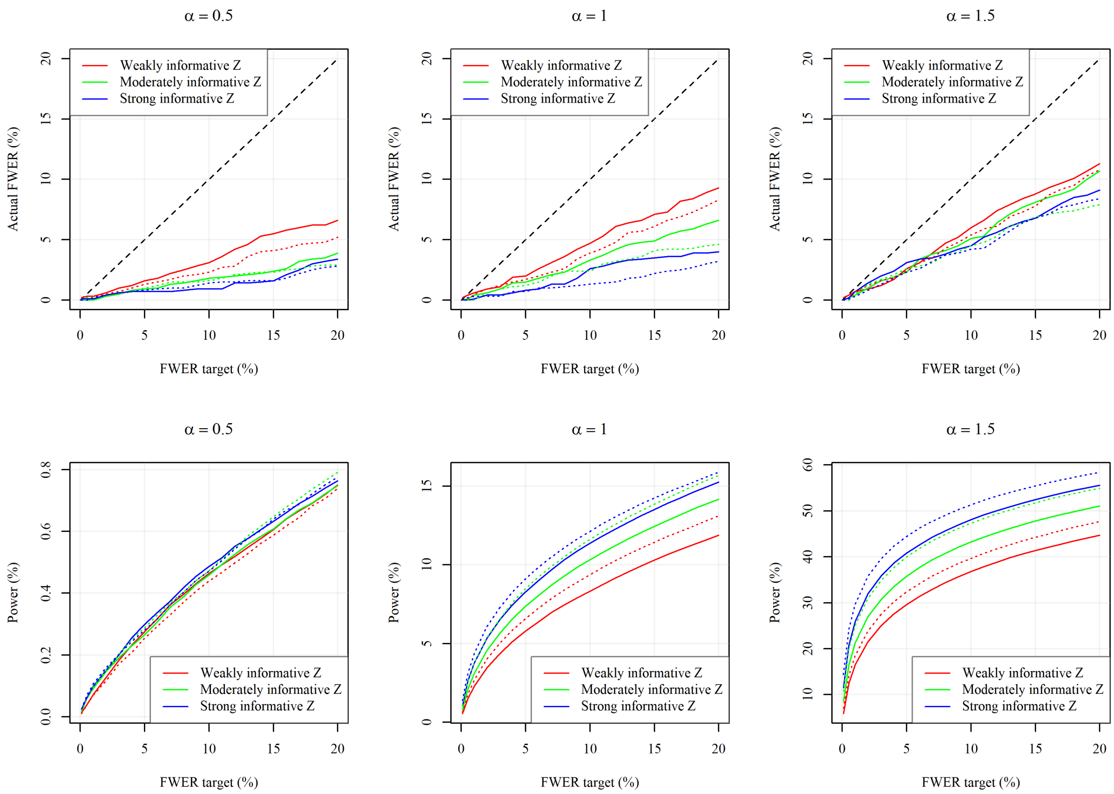
5.2. Performance of the $fwer^+$ under varying signals

In applications, the parameters of the input data are varying. For instance, in our portfolio construction, which will be presented in Section 6, we need to assess funds' performance based on a short window of 36 months. The magnitude of alpha of non-zero alpha funds and the informativeness level of covariates are varying across different periods. To study impacts of these factors on performance of the $fwer^+$, in this section we vary the alpha of the non-zero alpha funds and the informativeness of the covariate.

As such, we fix number of funds $n = 1000$ and $T = 36$, which are close to the representative IS sample in our baseline empirical experiment, while α is varying from 0.5% to 1.5% and the relationship between the covariates and the prior null are weak, moderate and strong.

We report in Figure 3 the performance of the $fwer^+$ procedure under both independent and correlated covariates settings. The top three sub-figures show the estimated actual FWER at the given targets whereas the bottom three sub-figures the power. From

Figure 3: Performance of the $fwer^+$ under varying setting of signals. The figure shows impact of signals, i.e., the magnitude of true non-zero alpha and informativeness of covariates, on the performance of the $fwer^+$ in terms of FWER control (top three sub-figuers) and power (bottom three figures). The simulated data are balanced panels of $n = 1000$ funds where each of them has $T = 36$ observations. The funds population consists of around 60%, 20% and 20% zero-alpha, under- and out-performing funds, respectively. The out-performing (under-performing) funds in population have alpha of α $(-\alpha)$ which varies in $\{0.5\%, 1.0\%, 1.5\%\}$. We consider three settings of the two covariates $\boldsymbol{Z} = (u, v)$ including weakly, moderately and strongly informative. The covariates can be independent (solid curves) or correlated with a correlation coefficient of 0.5 (dotted curves).



left to right, each sub-figure represents for a setting of non-zero alpha magnitude including 0.5%, 1% and 1.5%. In each of the top three sub-figures, the red- (green- and blue-) solid curves present for the estimated actual FWER of the $fwer^+$ under a setting of the weakly (moderately and strongly) informative and independent covariates (i.e., $\rho = 0$). The dotted curve of the same color as the solid one is the estimated actual FWER of the corresponding informativeness level under the correlated covariates setting (i.e., $\rho = 0.5$). It is clear that the $fwer^+$ controls well for FWER at any given targets, regardless the dependence of covariates, as all points of the curves are below or on the 45° line.

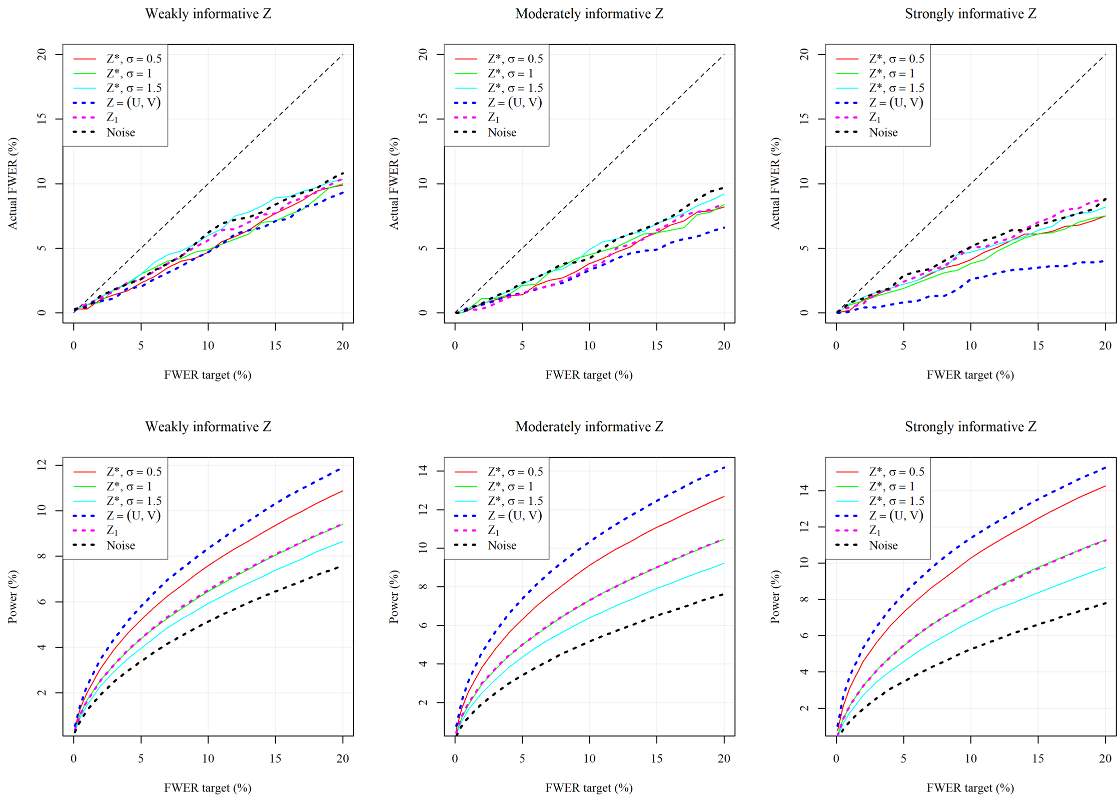
The presentations in the bottom three sub-figures are similar but representing for the power. We observe that the stronger the informativeness of the covariates the higher power the $fwer^+$ gains. Moving from the left to the right sub-figures, the alpha magnitude of the out-performing funds is increasing and the $fwer^+$ gains higher power. This is unsurprising as the out-performing funds are being easier to be detected. When the two covariates are correlated, the power might be slightly lower (e.g., $\alpha = 0.5$ case), or higher (e.g., $\alpha = 1$ and 1.5 cases). This indicates that the dependence among covariates does not significantly effect the performance of the $fwer^+$.

5.3. Performance of the $fwer^+$ under insufficient, noisy and non-informative covariates

We have investigated performance of the $fwer^+$ under the use of two covariates $\boldsymbol{Z} = (U, V)$. Often, in real applications we do not know how many covariates actually convey information on the performance of funds. Consequently, it happens the case we use less covariates than we should, or we might use the covariates that are estimated with noise or, even worse, unrelated to the funds' performance (i.e, non-informative covariates).

For the first scenario, we implement the $fwer^+$ with use of only one of the two covariates $Z_1 = U$. For the covariates estimated with noise case, we generate two new covariates $\boldsymbol{Z}^* = (U + \eta, V + \zeta)$ where η and ζ are noise drawn from normal distribution $N(0, \sigma^2)$. We investigate different levels of the noise via varying the $\sigma \in \{0.5, 1.0, 1.5\}$. Finally, for the non-informative covariate case, i.e., a covariate that is totally noise drawn from $N(0, 1)$ is used as a single input covariate. The performance of the $fwer^+$ for all mentioned scenarios are depicted in Figure 4. In this figure, we add the performance of the $fwer^+$ with use of the informative covariates $\boldsymbol{Z}$ for comparison purpose.

Figure 4: Performance of the $fwer^+$ under use of insufficient information. The figure shows impact of using insufficient covariates, covariates containing different levels of noise and non-informative covariate on the performance of the $fwer^+$ in terms of FWER control (top three sub-figuers) and power (bottom three figures). The simulated data are balanced panels of $n = 1000$ funds where each of them has $T = 36$ observations. The funds population consists of around 60%, 20% and 20% zero-alpha, under- and out-performing funds, respectively. The out-performing (under-performing) funds in population have alpha of 1%. We consider three settings of the two covariates $\boldsymbol{Z} = (U, V)$ including weakly, moderately and strongly informative. The simulated data are generated based on $\boldsymbol{Z}$ via $\pi_0(\boldsymbol{Z})$. In noisy covariates cases, instead of using $\boldsymbol{Z}$, the $fwer^+$ uses $\boldsymbol{Z}^* = (U+\eta, V+\zeta)$ where $\eta, \zeta \sim N(0, \sigma^2)$ and $\sigma \in \{0.5, 1.0, 1.5\}$. In insufficient covariates case, $fwer^+$ uses only $Z_1 = U$ while in the non-informative case it uses only one covariate which is a noise drawn from $N(0, 1)$ without any connection to $\pi_0(\boldsymbol{Z})$. We include performance of $fwer^+$ with use of $\boldsymbol{Z}$ for comparison purpose.



We see that in all scenarios, the FWER is controlled well at all considering targets. This is an excellent property of the $fwer^+$. The non-informative case implies that it is safe, in terms of controlling FWER, to implement the $fwer^+$ even if we wrongly include an unrelated covariate. Unsurprisingly, in terms of power, the $fwer^+$ performs best when we use the truly and sufficiently information while it is least powerful in case the covariate is irrelevant or non-informative. The power of the $fwer^+$ with use of the covariates estimated with noise lies in between the two extreme cases and decreases with respect to the level of the noise, i.e., the magnitude of σ. This also implies that, adding into a given informative covariates set an non-informative one might damage the power of the $fwer^+$.

5.4. *Performance of the $fwer^+$ under varying of sample size and observations*

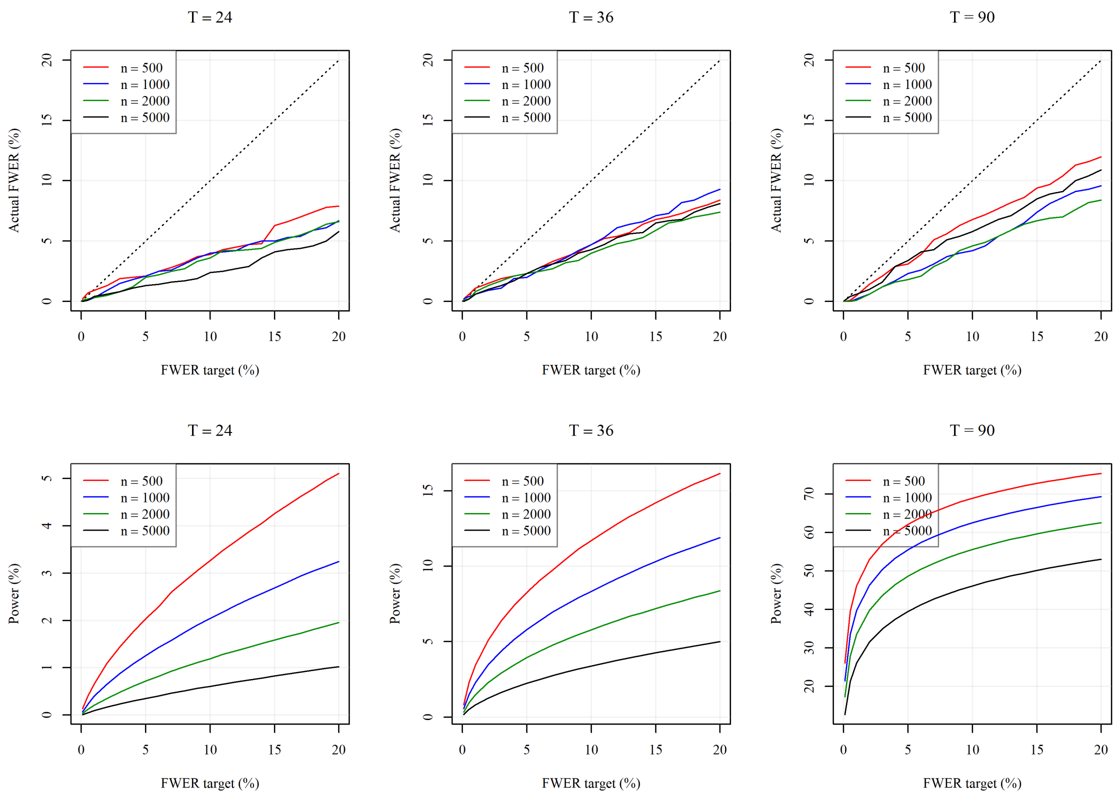
Hitherto, we have investigated the performance of the $fwer^+$ under different scenarios of the informative covariates as well as the strength of the signals of the out-performing funds. In this section, we further investigate the impact of sample size, i.e., number of funds in sample, and number of observations per fund on the performance of the procedure.

As such, we consider balanced panel data with varying the number of funds n and the number of observations per fund T. As the IS horizons for portfolio selection in hedge fund literature are typically 24 or 36 months we consider $T = 24$, and 36. We additionally experiment with much longer time series of $T = 90$, which is also the median number of observations per fund in our whole sample data. The n is also varying to cover all cases of our application in empirical experiments which spreads from around 500 to 2000. We also add a case $n = 5000$ which is close to our whole sample size. For the interest of space, in this set of simulations, we present results for data generated under the independent and weakly informative covariates and with $\alpha = 1$ setting.[10] The results are depicted in Figure 5.

In sub-figures of the Figure 5, the number of observations per fund is increasing from left to right. In each sub-figure, we present the results corresponding to different setting in number of funds, $n = 500, 1000, 2000$ and 5000. From the top three sub-figures, we again witness the excellent performance of the $fwer^+$ in terms of FWER control. It is clear from the bottom three sub-figures that, the power gains are higher for the data with longer time series (i.e., larger T). This is consistent with the fact that the out-performing funds are easier to be detected if they outperform in a longer period. Interestingly, the $fwer^+$ is more powerful when the sample size is smaller. This is a good property since the procedure can be applied in a wider problems both with small and large number of

[10] Our conclusions are robust to other settings such as dependent and moderately and strongly informative covariates and the results are available upon request.

Figure 5: Varying sample size and number of observations. The figure presents the performance of the $fwer^+$ under varying sample size (n) and number of observations per fund (T). The simulation data are balanced panels with T observations per fund under weakly informative and independent covariates.



hypotheses.

In conclusion, the simulations show the excellent performance of the $fwer^+$ in terms of controlling for the FWER in various scenarios of data. We witness the higher power of the proposed procedure when we have one of the followings: i) the stronger the relationship between the covariates and the prior null; ii) the larger magnitude of out-performing funds' alpha; iii) the more sufficient set of informative covariates; iv) the out-performing funds do well in a longer period (larger T); and v) the smaller number of funds in the population. In Section IB of the Internet Appendix, we show that our conclusions are robust to alternative setting of the out-performing funds proportion.

6. Empirical analysis

In this section, we use the $fwer^+$ procedure to detect out-performing funds based on past short IS performance and invest in those detected funds in a rolling forward fashion. We describe the covariates that we are studying, the formation of our $fwer^+$-

based portfolios and show their performance in various choice of IS horizons and models that we use to assessing the performance of funds.

6.1. Covariates

As hedge fund data reveal little information on funds' holdings, we focus on the covariates that are calculated based on excess return of the funds. Since we are assessing the performance of a fund via testing its adjusted return - the alpha, we include covariates that are potentially adding information alongside the alpha itself. There exists a number of such side information that have been shown to be linked with the performance of hedge funds.

First, Titman and Tiu (2011) regress individual hedge fund returns on a group of risk factors and find that funds with low R-squares gain higher alpha. The authors further document that those low R-square funds charge higher incentive and management fees. Thus, the R-square of the funds is not only conveying the fund's managerial skill but also some other fund's characteristics. We use the R-square of the considering factor model as a covariate.

Second, as documented in Boyson (2008), funds' performance is more consistent among the younger and smaller funds. As investors' flows chase funds outperforming in the past, funds become larger and more passive. Thus, the size, i.e., the asset under management (AUM) of funds have a link with the funds' performance, and is chosen as one of our covariates.[11]

Third, Khandani and Lo (2011) argue that fund's excess return auto-correlation can measure the illiquidity in hedge funds and find a significant link between the auto-correlation of a fund and its expected return. Thus we consider as our covariates the first, second and third degrees of auto-correlation coefficients, which are denoted by ACF1, ACF2 and ACF3 respectively, of the fund's past 12-, 24- and 36-month excess return. These make up nine covariates and constitute our "persistent covariates" group.

Last, we study the risk measures based on fund's excess return as they are potentially

[11]In our experiment, we follow literature to use the logarithm of fund's AUM instead of the AUM itself.

informative. For example, Liang and Park (2007) document that downside risk measures incorporating higher moments help explain the cross-sectional variation of hedge fund performance and have predictive power. Wu *et al.* (2021) also find that the kurtosis of the excess return is an important variable in forecasting future hedge fund return. Thereby, we study risk metrics that consist of the variance (total volatility), kurtosis and skewness the fund's excess return over the past 12-, 24- and 36-month periods. These make up nine covariates and form our "moment covariates" group.

6.2. Portfolios of out-performing funds

In this section, we use the $fwer^+$ to construct portfolios of hedge funds based on assessing their short-term performance over a past period. We investigate OOS performance of the $fwer^+$ portfolios under various FWER target $\tau \in (0, 1)$.

We first describe our baseline portfolios. At the end of each year from 1997, we use most recent three years data up to that point of time as the IS period to calculate needed information. Specifically, we assess funds based on alpha of the 7-factor model and conduct for each fund the test of its alpha against zero, calculate $p-$value and estimate the mentioned covariates with use of only the data in IS period. A fund is eligible if it has returns data for all months of the IS period and the data of all considering covariates at the portfolio constructing time.[12] We implement the $fwer^+$ to picking out-performing funds with control for FWER at the given target τ. We then invest equally weighted in those selected funds in the following year. The performance of the portfolio in this OOS year is recorded. If there are no funds selected, we invest on bond to earn a return at the interest rate. When a selected fund stops reporting its returns during the OOS year, we redistribute fund equally into the remaining funds in the portfolio.[13] Our portfolios

[12] We additionally conduct exercises where we restrict to consider only the funds that have at least 5 million USD in AUM and find that our empirical conclusions remain unchanged. For the interest of space, the results are presented in Appendix A.

[13] The IS horizon, which is used to estimate alpha (and covariates), could be 24 months as in Chen *et al.* (2017) and Kosowski *et al.* (2007) or 36 month as in Cumming *et al.* (2012). The OOS period is also varying in literature, Chen *et al.* (2017) use 3, 6, 9, 12, 24 and 36 months while Kosowski *et al.* (2007) use 12 months. Practically, hedge fund is a long term investment vehicle and there is usually a lock-up period which is varying up to one year depending on funds. Thus, in this study we use the holding period of at least one year.

are rolling forward yearly. The first OOS period is the year 1998, which we invest in the funds selected based on the data in the IS period from January 1995 to December 1997. The final OOS period is the year 2021, which we invest in the funds selected based on using data from January 2018 to December 2020. Thus, each of our portfolios has OOS returns spanning over 24 years, from January 1998 to December 2021.

To further study the empirical performance of the $fwer^+$, we construct portfolios that control for FWER at different targets $\tau \in \{0.01\%, 1\%, 5\%\}$. The input $p-$value is calculated with use of heteroskedasticity and auto-correlation consistent (HAC) correction of Newey and West (1987).[14]

As benchmarks, we conduct two equally weighted portfolios as followings. At the end of each year from 1997, the first (second) equally weighted portfolio, denoted by EW (EW^+), simply selects all funds that are eligible (eligible and having positive estimated alpha) in the IS period to invest with equal weights in the following year. We repeat yearly until the end of 2020 to have a set of funds to invest in the year 2021.

As the numbers of covariates in the persistent and moment groups are large, we construct representative covariates that are the first principal component (PC1) of each group.[15] We thus have four covariates and construct our $fwer^+$ portfolios with use each of the four. Their OOS performance metrics are reported in panels A to D of Table 1. The metrics include annualized alpha as well as its HAC correction $t-$statistics and $p-$value, annualized excess return and Sharpe ratio. As a measurement of empirical power, we report the average, minimum, maximum and standard deviation of the number of out-performing funds detected by the $fwer^+$.

[14]For the purpose of selecting out-perorming funds with low FWER targets, bootstrapped $p-$value has a limitation since the $p-$value is lower bounded at $1/(B+1)$ where B is number of bootstrapped iterations. Consequently, the highly out-performing funds with truly smaller $p-$value lose their advantage to be selected and empty portfolios are generated as a result. We therefore use $p-$value calculated from $t-$score with HAC standard error correction. In small sample size time series, the HAC correction might be biased as documented in Boudoukh *et al.* (2022) and Muller (2014). We further conduct experiments based on $p-$value calculated without using HAC correction and report the results in Section ID of the Internet Appendix. We see that the performances of portfolios are better and our main conclusions are even stronger.

[15]We report comprehensively the OOS performance of the $fwer^+$ portfolios with use of each of individual covariates in Tables I and II in Section IC of the Internet Appendix. We see that portfolios with use of individual covariates in the same group perform similarly. This suggests the use of PC1s as the representative covariates.

As our first observation, the $fwer^+$ detects non-empty group of out-performing funds in all 24 times of portfolio constructions even when controlling for FWER at 0.01%, which is very low target. This reflects the superior power of the $fwer^+$ procedure and thus it allows investors picking funds with high confidence, i.e., with very low error. More importantly, all portfolios with use one of the considering covariates gain psotive abnormal alpha from around 4.2% to 5.3% in OOS period which are statistically significant with $t-$statistics varying from around 6 to 8. Of the four considering covariates, the PC1 of the persistent group performs best followed by others which are somewhat similar. It is clear that portfolios which control for a lower FWER target tend to perform better and can gain a Sharpe ratio of more than 2.

Next, we construct our $fwer^+$ portfolios with use of multiple covariates. As such,

Table 1: OOS performance of $fwer^+$ portfolios. Panels A to D of the table report OOS performance metrics of the $fwer^+$ portfolios with use each of R-square, AUM, and PC1s of moment and persistent group as the sole input covariate. The performance metrics include annualized alpha as well as its $t-$statistic and $p-$value, excess return and Sharpe ratio and summary on number of out-performing funds detected by $fwer^+$. Panel E reports these metrics of the $fwer^+$ portfolio with use of all four mentioned covariates whereas panel F the performance metrics of the equally weighted (EW) and equally weighted plus (EW^+) portfolios.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Number of detected funds			
						Average	Min	Max	Std
Panel A: $fwer^+$ with use of R-square as the covariate									
0.01	4.98	7.5	0.00	5.42	2.17	16	4	25	6
1.00	4.44	7.2	0.00	4.94	1.59	33	10	55	13
5.00	4.45	6.8	0.00	4.95	1.58	47	14	88	20
Panel B: $fwer^+$ with AUM as the covariate									
0.01	4.87	7.2	0.00	5.33	2.09	16	4	25	7
1.00	4.33	6.9	0.00	4.82	1.55	35	8	64	15
5.00	4.25	6.6	0.00	4.80	1.48	49	14	95	23
Panel C: $fwer^+$ with use of PC1 of moment group as the covariate									
0.01	4.85	7.0	0.00	5.26	2.03	16	4	29	7
1.00	4.48	7.4	0.00	4.99	1.63	34	10	65	15
5.00	4.21	6.5	0.00	4.75	1.53	49	14	97	22
Panel D: $fwer^+$ with use of PC1 of persistent group as the covariate									
0.01	5.27	8.4	0.00	5.64	2.38	15	4	24	6
1.00	4.60	7.5	0.00	5.10	1.69	33	10	56	13
5.00	4.41	6.9	0.00	4.88	1.59	46	15	84	19
Panel E: $fwer^+$ with use of the R-square, AUM and PC1s of the two groups as the covariates									
0.01	5.12	8.2	0.00	5.47	2.32	17	4	31	7
1.00	4.47	7.2	0.00	4.97	1.62	37	10	69	16
5.00	4.10	6.1	0.00	4.66	1.44	52	14	104	25
Panel F: equally weighted portfolios									
EW	2.58	2.9	0.00	4.65	0.72	1067	350	1570	361
EW^+	3.00	3.7	0.00	4.77	0.80	761	273	1418	324

we use the R-square, AUM and the two PC1s of the persistent and moment groups as the four input covariates. As shown in Panel E of Table 1, those portfolios gain higher power than those with use of a sole underlying covariate. This is consistent with the fact shown in our simulation, that is, the more input informative covariates we use the higher power the $fwer^+$. The OOS performance of these portfolios is roughly at the average performance of the portfolios based on each of the four underlying covariates. The results suggest that using more covariates does not necessarily imply a higher alpha. This is not implausible because the $fwer^+$ is more powerful with more covariates and it might select some more smaller truly positive alpha funds.

As benchmarks, we report the performance of the equally weighted portfolios in panel F of the same table. We see that all the considering $fwer^+$ portfolios outperform the equally weighted ones. It is also noted that, the equally weighted portfolio does not select all funds but ones that pass the screening based on the number of observations, and the equally weighted plus one further requires funds having positive estimated alpha in the IS period. Thus it is not surprised that those portfolios also gain significantly positive alpha.

Overall, the $fwer^+$ portfolios perform well with all of the considering covariates. The $fwer^+$ shows its power in detecting outperforming funds even when we control for a very small error. The selected funds perform persistently in the OOS period and those selected with lower FWER targets tend to perform better on average.

6.3. Persistent analysis

As documented in Section 6.2, the performance of the funds selected by $fwer^+$ is persistent at least over the rolling OOS of one year. In this section, we provide further evidence on this advantage of the $fwer^+$ portfolios. Thereby, we examine the performance of those funds selected by the $fwer^+$ over longer OOS horizons. As such, we implement the $fwer^+$ every m years and we hold the detected funds over m years where $m = 2, 3$ and 4. For the interest of space, we report in Table 2 the performance of only the $fwer^+$ portfolios with use of R-square, AUM, and PC1s of moment and persistent group as the four input covariates. In long horizons, the attrition rate becomes important since the selected funds might not survive throughout the holding periods, leading to potential empty portfolios. We thus report the summary of monthly portfolio size rather than that of the number of funds selected by the $fwer^+$ as in previous discussions. We see that all considering portfolios are non-empty throughout the holding periods even with the holding horizon of four years.

Table 2: Performance of $fwer^+$ in various OOS horizons. The table reports performance metrics of the $fwer^+$ portfolios with use of R-square, AUM, and PC1s of moment and persistent group as the four input covariate with different OOS holding horizons. In OOS horizon of 2 (3 and 4) years, outperforming funds are selected by $fwer^+$ every 2 (3 and 4) years and invested in the following 2 (3 and 4) years. The performance metrics include annualized alpha as well as its $t-$statistic and $p-$value, excess return, Sharpe ratio and summary on monthly portfolio size. Panels A, B, and C report these metrics for portfolios with holding horizons of 2, 3, and 4 years, respectively.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Portfolio Size			
						Average	Min	Max	Std
Panel A: 2-year OOS horizon									
0.01	5.36	8.0	0.00	5.71	2.13	17	3	31	7
1.00	4.60	6.9	0.00	5.22	1.66	36	9	69	16
5.00	4.55	6.8	0.00	5.32	1.52	52	17	104	24
Panel B: 3-year OOS horizon									
0.01	4.95	7.8	0.00	5.40	2.03	16	2	31	7
1.00	3.55	5.3	0.00	4.30	1.24	39	7	60	17
5.00	3.38	5.4	0.00	4.16	1.19	55	15	98	25
Panel C: 4-year OOS horizon									
0.01	4.99	8.2	0.00	5.27	2.07	17	1	31	9
1.00	4.31	7.2	0.00	4.85	1.63	33	3	69	19
5.00	4.27	6.9	0.00	4.96	1.50	45	9	104	28

In terms of alpha and Sharpe ratio, we see that the $fwer^+$ portfolios with 2- and 4-year holding horizons perform as well as those with one year holding whereas those of 3-year holding horizon are slightly worse. Given the use of only 3-year IS periods, the persistence in performance of the 4-year holding horizon portfolios is impressive.

Holding for a longer period also implies a less re-balanced cost. However, investors also face a risk of funds' attrition which might lead to a low diverse portfolio. As shown in summary of portfolio size columns, the minimum portfolio size is reducing with respect to the holding horizon. Nevertheless, in this particular case, the investors will not face any diversification problem if they set a target of FWER at 5%.

6.4. Sub-sample analysis

In this section, we further investigate the performance of the $fwer^+$ portfolios in sub-periods. We partition the whole OOS period, which spans from 1998 to 2021, into five non-overlapping sub-periods: 1998–2001, 2002–2006, 2007–2011, 2012–2016, and 2017–2021. Of those sub-periods, only the first one lasts for four years, others are five-year periods. We calculate the performance metrics of the portfolios in each sub-period and report them in each panel of Table 3.

Table 3: OOS performance of $fwer^+$ portfolios in sub-samples. Table report OOS performance metrics of the $fwer^+$ portfolios with use of R-square, AUM, and PC1s of moment and persistent group as the four input covariate in five non-overlapping sub-periods. For each sub-period we construct the equally weighted (EW) and equally weighted plus (EW^+) portfolios as benchmarks. The performance metrics include annualized alpha as well as its $t-$statistic and $p-$value, excess return, Sharpe ratio and summary on monthly portfolio size. We report in each panel the performance of the portfolios in the sub-period shown in its title.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Portfolio size			
						Average	Min	Max	Std
Panel A: Period 1998–2001									
0.01	4.96	3.9	0.00	3.75	0.97	22	18	30	4
1.00	5.01	4.1	0.00	3.81	1.00	50	40	60	8
5.00	5.10	4.3	0.00	3.95	0.96	67	55	84	11
EW	6.01	2.9	0.01	4.63	0.59	526	333	724	133
EW^+	5.54	3.3	0.00	4.24	0.61	401	258	571	111
Panel B: Period 2002–2006									
0.01	4.83	5.6	0.00	6.59	3.65	18	8	27	7
1.00	4.81	6.6	0.00	6.73	3.99	48	31	69	14
5.00	5.17	6.7	0.00	6.96	4.07	72	46	106	24
EW	3.90	3.2	0.00	7.25	1.49	1056	766	1364	188
EW^+	3.99	3.0	0.00	7.57	1.56	824	676	977	110
Panel C: Period 2007–2011									
0.01	2.86	1.9	0.06	2.70	1.10	10	2	22	7
1.00	2.63	1.7	0.10	1.87	0.44	26	7	61	18
5.00	2.44	1.6	0.12	1.60	0.35	41	10	87	27
EW	1.82	1.0	0.32	0.39	0.09	1393	1191	1544	78
EW^+	2.20	1.2	0.24	0.72	0.13	1125	927	1399	115
Panel D: Period 2012–2016									
0.01	7.07	6.3	0.00	8.18	4.06	14	7	20	4
1.00	6.58	7.7	0.00	7.99	4.86	35	27	41	4
5.00	6.81	7.9	0.00	8.40	4.95	51	35	61	9
EW	0.46	0.4	0.67	5.18	1.23	1135	924	1422	126
EW^+	1.84	1.7	0.09	5.49	1.56	774	616	1028	117
Panel E: Period 2017–2021									
0.01	4.69	6.0	0.00	5.17	3.30	18	7	27	6
1.00	3.29	3.3	0.00	4.12	1.33	24	10	36	6
5.00	3.08	2.9	0.01	3.88	1.35	28	12	42	8
EW	-0.52	-0.4	0.70	5.95	0.82	763	540	962	120
EW^+	0.83	0.7	0.48	5.87	1.04	410	267	561	96

The table shows that the $fwer^+$ portfolios gain positive alpha and Sharpe ratio in all sub-periods. Except the period 2007–2011, which covers the global financial crisis 2007–2008, the portfolios' alphas are statistically significant at all considering FWER targets. Compared to equally weighted portfolios, the $fwer^+$ portfolios gain higher alpha for four over five sub-periods. In only the first sub-period, the $fwer^+$ portfolios gain lower alpha but their $t-$statistic are higher.[16] Also, the $fwer^+$ portfolios always have a higher Sharpe ratio than the equally weighted portfolios regardless the considered FWER targets and sub-periods.

The $fwer^+$ portfolios perform best during the period 2012–2016 with alphas roughly 7% and Sharpe ratios spanning from 4 to roughly 5. The most recent sub-period of our sample, the $fwer^+$ portfolios perform as well as the average of the whole sample reported in previous section and the Sharpe ratio can reach 3.3.

6.5. Boosting the informativeness of covariates

Hitherto, we have utilized only the informativeness of the covariates' variation. Yet, the covariates are likely containing noise and thus their informativeness is affected. In this section, we show that the performance of $fwer^+$ portfolios can be improved via boosting the informativeness of the covariates. The main idea is to generating new covariates that target for future funds' expected returns. Thereby, we first use machine learning models to predict future return of funds and then we use the predicted returns as covariates. More specifically, we are considering four well-known machine learning models including the least absolute shrinkage and selection operator (LASSO, see Tibshirani 1996), random forest (RF, see Breiman 2001), stochastic gradient boosting (GB, see Friedman 2002) and deep neural network (DNN, see LeCun *et al.* 2015).

Formally, the relationship of the funds' cumulative future return during period $t + 1, \ldots, t + h$ and the information of covariates at the end of month t can be modelled as

$$\tilde{R}_{i,t\rightarrow t+h} = f_t(X_{i,t}) + \tilde{\epsilon}_{i,t} \tag{6}$$

where $\tilde{R}_{i,t\rightarrow t+h}$ is cumulative return of fund i from month $t+1$ to $t+h$, $X_{i,t}$ is the realized covariates of the fund i measured at the end of month t, the function f_t describes the relationship of the $X_{i,t}$ and the future accumulated return over h months $\tilde{R}_{i,t\rightarrow t+h}$ whereas

[16] It is worth to note that, the $fwer^+$ aims to select highly significant alpha funds, which are reflected via the significant of the tests, i.e., the $t-$statistics.

$\tilde{\epsilon}_{i,t}$ is the noise.

Consistent with the choice of our IS horizon and rolling window, to predict the return of year corresponding to period from month $t+1$ to $t+12$ for some t, we use data at the end of each previous three years until the end of month t to train the model (6) and use it to predict future return. That is, we train the model by using target variable $\tilde{R}_{i,k\to k+12}$ and features $X_{i,k}$ with $k = t-36, t-24$ and $t-12$ across funds considered in the IS period. We fit into training model the data of features $X_{i,t}$ at the end of month t to acquire the predicting accumulated future return for period $t+1$ to $t+12$.[17] As our AUM is available from December 1997, our first predicted returns is for the year 1999. This predicted return is calculated from data up to December 1998 and used as the input covariates of the $fwer^+$ to select funds invested in the year 1999. We rolling forward and re-balance the portfolios yearly in the same fashion as the $fwer^+$ portfolios described in previous section.

Table 4 reports the OOS performance of the $fwer^+$ portfolios with use of the predicted return of each considering machine learning model as a covariate. We see that the

Table 4: OOS performance of $fwer^+$ portfolios with use of new covariates. Panel A (B, C and D) reports OOS annualized alpha as well as its $t-$statistic and $p-$value, excess return, Sharpe ratios and summary on number of out-performing funds selected by the $fwer^+$ with use of funds' future return predicted by LASSO (GB, RF and DNN) model at given FWER targets τ.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Number of detected funds			
						Average	Min	Max	Std
Panel A: $fwer^+$ with use of future return predicted by the LASSO model as the sole covariate									
0.01	5.42	9.3	0.00	5.83	2.79	15	4	25	7
1.00	4.50	7.6	0.00	5.04	1.73	33	10	59	14
5.00	4.40	7.0	0.00	4.96	1.69	46	14	86	20
Panel B: $fwer^+$ with use of future return predicted by the GB model as the sole covariate									
0.01	5.18	8.6	0.00	5.56	2.64	14	0	25	7
1.00	4.68	8.4	0.00	5.21	1.85	30	1	56	15
5.00	4.35	7.1	0.00	4.88	1.70	42	1	87	23
Panel C: $fwer^+$ with use of future return predicted by the RF model as the sole covariate									
0.01	5.42	9.4	0.00	5.87	2.52	15	4	25	7
1.00	4.87	7.0	0.00	5.66	1.38	30	5	56	14
5.00	4.52	6.2	0.00	5.28	1.31	43	7	89	21
Panel D: $fwer^+$ with use of future return predicted by the DNN model as the sole covariate									
0.01	5.37	9.3	0.00	5.78	2.72	15	4	24	6
1.00	4.84	9.2	0.00	5.39	1.96	32	10	55	14
5.00	4.54	7.4	0.00	5.12	1.76	45	14	87	20

[17] We follow Wu *et al.* (2021) in tuning the hyperparameters of the models.

performances of the portfolios are generally better than the portfolios with four covariates presented in previous section. The portfolios' alpha range from 4.4% to 5.4% and Sharpe ratio from 1.31 to 2.79. On average across the considering FWER targets, the DNN model seems to be the best with an annualized alpha varying from 4.54% to 5.37% and an annualized Sharpe ratio that can reach 2.7. These numbers are generally higher than those of the $fwer^+$ portfolio with use of the four covariates reported in Table 1. This supports for benefit of using advanced machine learning techniques in forecasting hedge funds' return.

6.6. Alternative choices of benchmarks

In this section, we assess whether the performance of the $fwer^+$ portfolios is robust to alternative benchmark multi-factor models used in fund performance literature (see, e.g., Bali *et al.*, 2012, 2014; Chen *et al.*, 2023). More specifically, we are considering three alternative factor models including the four-factor model of Carhart (1997), a six-factor and a nine-factor model, as well as the nine-factor model of (Chen *et al.*, 2023) as described in Section 4.

Table 5: Performance under alternative benchmarks. The table reports the OOS performance of the $fwer^+$ portfolios constructed by selecting truly positive alpha under alternative benchmarks. The $fwer^+$ uses all of the considering four covariates (the R-square, AUM, and two PC1s of the persistent and moment groups) as inputs. Panel A (B, C and D) presents annualized alpha of corresponding factor model as well as its $t-$statistic and $p-$value, excess return, Sharpe ratios and summary on the number of funds selected by the $fwer^+$ under the use of the four- (six- and traditional nine- and Chen *et al.* (2023)'s nine-) factor model as the benchmark.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Number of selected funds				
						Average	Min	Max	Std	#empty
Panel A: four-factor model										
0.01	5.26	7.7	0.00	5.75	2.61	14	4	28	7	
1.00	5.16	7.1	0.00	6.07	2.26	28	7	60	13	
5.00	4.61	5.4	0.00	5.58	1.81	37	12	81	18	
Panel B: six-factor model										
0.01	5.43	8.5	0.00	5.85	2.62	16	4	32	7	
1.00	4.73	6.4	0.00	5.42	1.84	31	9	57	13	
5.00	4.62	6.5	0.00	5.34	1.73	41	13	75	17	
Panel C: traditional nine-factor model										
0.01	5.26	7.9	0.00	5.63	2.23	14	3	27	7	
1.00	4.88	8.2	0.00	5.42	2.01	30	8	61	13	
5.00	4.15	6.4	0.00	4.83	1.54	42	12	79	18	
Panel D: nine-factor model of Chen *et al.* (2023)										
0.01	4.01	3.6	0.00	4.77	1.78	5	0	19	5	3
1.00	3.18	2.8	0.01	5.29	1.29	10	0	40	11	1
5.00	3.15	2.8	0.01	4.17	1.11	16	0	60	17	1

For each of the alternative benchmarks, we repeat the exercises presented in previous sections. For the interest of space, we present in Table 5 the performance of the $fwer^+$ portfolios with use of the R-square of the considering model, AUM, and two PC1s of the moment and consistent groups. Overall, the OOS alphas of the $fwer^+$ portfolios are varying across the benchmark but all are statistically significantly positive. We see that, the $fwer^+$ portfolios under the four- and six-factor models gain highest annualized alpha which varie from 4.61% to 5.43%. However, as the alphas are of different factor models, it is not appropriate to compare models based on this metric. Interestingly, comparing all considered models, including the seven-factor presented in Panel E of the Table 1, the highest Sharpe ratio is gained under the use of the four-factor model. Overall, all of our conclusions on the power of $fwer^+$ as well as the ability in detecting truly out-performing hedge funds remain. As the $fwer^+$ does not detect any truly positive alpha funds for some certain years when the Chen *et al.* (2023)'s nine-factor model is used, we report number of times (over 24 times) when this happens in the "#empty" column. For instance, there are one time the $fwer^+$ does not detect any positive alpha funds when controlling for FWER at 1%. Generally, under the use of the Chen *et al.* (2023)'s nine-factor model the $fwer^+$ finds only a third of positive alpha funds compared to when use other models. This supports for the that the model explains better the cross-sectional return of hedge funds, thus there are fewer funds that having abnormal alpha.

We have assessed the performance of hedge funds based on past 36-month IS periods. As robustness checks of for this choice of the IS horizon, we additionally conduct experiments with use of 24- and 48-month IS periods. For the interest of space, the results are presented in Appendix B. Generally, with use of 24-month IS periods, the alphas of the $fwer^+$ portfolios are slightly higher than those reported for 36-month IS case though the Sharpe ratios are slightly lower. In contrast, the Sharpe ratios of the 48-month case are comparable to the baseline while the alphas are slightly lower.

6.7. Portfolios of the best out-performing hedge fund

We have conducted portfolios of hedge funds with control for FWER at certain targets under consideration of various performance assessment settings. We have witnessed that the $fwer^+$ is being able to detect out-performing funds based on utilizing short IS data windows. In this section, we further construct the portfolios consisting of only a single fund selected by the $fwer^+$. Given a FWER target τ, as the FWER of the group of funds detected by the $fwer^+$ is controlled at the target τ, it holds so for any subgroup of the detected funds. Instead of investing on all funds selected by the $fwer^+$, our new single-fund portfolio is established by investing only the fund selected by the $fwer^+$ that performs best in the IS period, i.e., one that has highest $t-$score among those selected by the $fwer^+$.

As our $fwer^+$ portfolios under the use of traditional factor models are non-empty all the time regardless the considering FWER targets, and as the portfolio with higher FWER target contains portfolios with lower targets, the choices of the considering FWER targets (0.1%,1% and 5%) will not effect on the best fund. We see that, the best funds are also unchanged under the choices of the considering covariates. The choice of the factor model affects the best funds, however. We thus report in Panel A of Table 6 the performance of the those portfolios without showing the FWER target and covariates. We report results for all considering factor models. We see that all portfolios performs impressive, especially in terms of Sharpe ratio with the best reaching 5.3. In terms of alpha, the best fund portfolios perform as good as those $fwer^+$ with use of all four covariates at FWER target 0.1% reported in 1 and 5. On the downside, the portfolios are empty for 4 to 11 months over 288 months of the investing period.

In contrast, the $fwer^+$ portfolios under the use of Chen *et al.* (2023)'s nine-factor model are empty for number of years depending on the FWER targets. We thus report similar metrics for those portfolios in Panel B of the same table. At target of 1%, the portfolio gain higher return and alpha compared to the use of the traditional factor models. However, the Sharpe ratio is lower and the portfolio is empty during 18 over 288 months.

Table 6: Performance under alternative benchmarks. The table reports the OOS performance of the portfolio that consists of the fund performed best in IS period among those selected by the $fwer^+$. Panel A presents annualized alpha of corresponding factor model as well as its $t-$statistic and $p-$value, excess return, Sharpe ratios and empty rate of the portfolios under the use of traditional factor models. Panel B presents similar metrics under the use of the Chen *et al.* (2023)'s nine-factor model.

Panel A: traditional factor models						
Model	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Empty rate (%)
4 factors	5.16	14.8	0.00	5.36	4.86	11/288
6 factors	5.08	15.2	0.00	5.27	5.33	7/288
7 factors	5.01	10.6	0.00	5.32	3.50	4/288
9 factors	5.16	15.0	0.00	5.37	5.31	7/288
Panel B: nine-factor model of Chen *et al.* (2023)						
τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Empty rate (%)
0.01	4.70	4.0	0.00	5.10	2.30	42/288
1.00	5.56	4.5	0.00	5.85	1.84	18/288
5.00	5.56	4.5	0.00	5.85	1.84	18/288

7. Concluding discussion

We have introduced the $fwer^+$ to control FWER in picking out-performers. The procedure utilizes multiple side information in estimating the FWER. Via simulations we show that when informative covariates are available the method gains significant higher power than existing methods which are not using the covariates.

Empirical experiments in hedge funds context show that the method is so powerful that it can detect out-performing funds even with a very low target of FWER. The portfolios of the detected funds are able to generate statistically significantly positive alpha and the performance of those funds are persistent for a long period. This is robust to various choices of IS horizons and asset pricing models. All experiments suggest a powerful and promising tool for investors who desire to picking hedge funds with high confidence.

The new method has a high potential in applications, especially for ones that require a low level of error. In similar applications to our study, the method can be used in picking out-performing mutual fund, bond fund and trading strategies. It can be also used to guard the data snooping in predictive model and factor selection.

References

Agarwal, V. and Naik, N. Y. (2000) Multi-period performance persistence analysis of hedge funds. *The Journal of Financial and Quantitative Analysis*, **35**, 327–342.

Bali, T. G., Brown, S. J. and Caglayan, M. O. (2012) Systematic risk and the cross section of hedge fund returns. *Journal of Financial Economics*, **106**, 114–131.

Bali, T. G., Brown, S. J. and Caglayan, M. O. (2014) Macroeconomic risk and hedge fund returns. *Journal of Financial Economics*, **114**, 1–19.

Baquero, G., ter Horst, J. and Verbeek, M. (2005) Survival, look-ahead bias, and persistence in hedge fund performance. *The Journal of Financial and Quantitative Analysis*, **40**, 493–517.

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

Boudoukh, J., Israel, R. and Richardson, M. (2022) Biases in long-horizon predictive regressions. *Journal of Financial Economics*, **145**, 937–969.

Boyson, N. M. (2008) Hedge fund performance persistence: A new approach. *Financial Analysts Journal*, **64**, 27–44.

Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.

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

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

Chen, Y., Cliff, M. and Zhao, H. (2017) Hedge funds: The good, the bad, and the lucky. *The Journal of Financial and Quantitative Analysis*, **52**, 1081–1109.

Chen, Y., Han, B. and Pan, J. (2021b) Sentiment trading and hedge fund returns. *Journal of Finance*, **76**, 2001–2033.

Chen, Y., Li, S. Z., Tang, Y. and Zhou, G. (2023) Anomalies as new hedge fund factors: A machine learning approach. *Working paper*.

Cumming, D., Dai, N., Haß, L. H. and Schweizer, D. (2012) Regulatory induced performance persistence: Evidence from hedge funds. *Journal of Corporate Finance*, **18**, 1005–1022.

Friedman, J. H. (2002) Stochastic gradient boosting. *Computational Statistics & Data Analysis*, **38**, 367–378.

Fung, W. and Hsieh, D. A. (2001) Empirical characteristics of dynamic trading strategies: the case of hedge funds. *Review of Financial Studies*, **10**, 275–302.

Fung, W. and Hsieh, D. A. (2004) Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal*, **60**, 65–80.

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

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

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

Khandani, A. E. and Lo, A. W. (2011) Illiquidity premia in asset returns: An empirical analysis of hedge funds, mutual funds, and us equity portfolios. *The Quarterly Journal of Finance*, **01**, 205–264.

Kosowski, R., Naik, N. Y. and Teo, M. (2007) Do hedge funds deliver alpha? a bayesian and bootstrap analysis. *Journal of Financial Economics*, **84**, 229–264.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep learning. *Nature*, **521**, 436–444.

Liang, B. and Park, H. (2007) Risk measures for hedge funds: a cross-sectional approach. *European Financial Management*, **13**, 333–370.

Muller, U. K. (2014) Hac corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics*, **32**, 311–322.

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

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

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

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

Sun, Z., Wang, A. W. and Zheng, L. (2018) Only winners in tough times repeat: Hedge fund performance persistence over different market conditions. *Journal of Financial and Quantitative Analysis*, **53**, 2199–2225.

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

Titman, S. and Tiu, C. (2011) Do the Best Hedge Funds Hedge? *The Review of Financial Studies*, **24**, 123–168.

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

Wu, W., Chen, J., Yang, Z. B. and Tindall, M. L. (2021) A cross-sectional machine learning approach for hedge fund return prediction and selection. *Management Science*, **67**, 4577–4601.

Zhou, H., Zhang, X. and Chen, J. (2021) Covariate adaptive familywise error rate control for genome-wide association studies. *Biometrika*, **108**, 915–931.

Appendix A. Empirical results with restriction on AUM

In this section we present the performance of the $fwer^+$ portfolios when we restrict to considering only funds having at least 5 millions of AUM at the time we construct the portfolios. Table A reports OOS performance of the $fwer^+$ portfolios with use of the R-square, AUM, and two PC1s of the moment and persistent groups as solely input covariate (Panels A to D) and four input covariates (Panel E). As benchmarks, we also report in Panel F the performance of the equally weighted portfolios which simply select all eligible funds in IS period (EW) and subset of those additionally having positive estimated IS alpha (EW^+). Generally, the performances of the portfolios remain similar to those with only requirement on the availability of the AUM reported in the Table 1 of the main manuscript.

Table A: OOS performance of $fwer^+$ portfolios under restrictions on AUM. Panels A to D of the table report OOS performance metrics of the $fwer^+$ portfolios with use each of R-square, AUM, and PC1s of moment and persistent group as the sole input covariate. The performance metrics include annualized alpha as well as its $t-$statistic and $p-$value, excess return, Sharpe ratio and summary on number of out-performing funds detected by the $fwer^+$ procedure. Panel E reports these metrics of the $fwer^+$ portfolio with use of all four mentioned covariates whereas panel F the performance metrics of the equally weighted (EW) and equally weighted plus (EW^+) portfolios.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Number of detected funds			
						Average	Min	Max	Std
Panel A: $fwer^+$ with use of R-square as the covariate									
0.01	4.94	7.2	0.00	5.35	2.09	16	4	26	7
1.00	4.39	7.3	0.00	4.88	1.58	34	9	54	13
5.00	4.38	6.8	0.00	4.90	1.56	47	15	85	19
Panel B: $fwer^+$ with AUM as the covariate									
0.01	4.80	6.8	0.00	5.23	2.00	16	4	26	7
1.00	4.37	7.1	0.00	4.86	1.55	35	9	61	15
5.00	4.19	6.5	0.00	4.75	1.46	49	15	94	23
Panel C: $fwer^+$ with use of PC1 of moment group as the covariate									
0.01	4.78	6.7	0.00	5.15	1.96	17	4	30	7
1.00	4.54	7.7	0.00	5.03	1.66	35	9	66	15
5.00	4.30	6.7	0.00	4.86	1.53	49	15	98	22
Panel D: $fwer^+$ with use of PC1 of persistent group as the covariate									
0.01	5.21	8.1	0.00	5.56	2.29	15	4	25	6
1.00	4.54	7.3	0.00	5.03	1.65	33	9	53	13
5.00	4.37	6.9	0.00	4.83	1.57	46	14	82	19
Panel E: $fwer^+$ with use of the R-square, AUM and PC1s of the two groups as the covariates									
0.01	5.00	7.6	0.00	5.32	2.16	17	4	30	7
1.00	4.43	7.0	0.00	4.93	1.55	38	10	69	16
5.00	4.38	6.5	0.00	4.97	1.48	53	14	106	25
Panel F: equally weighted portfolios									
EW	2.50	2.8	0.01	4.58	0.70	1018	336	1533	353
EW^+	2.91	3.5	0.00	4.68	0.78	739	266	1389	319

Appendix B. Alternative choices of in-sample horizons

Literature in hedge fund performance construct portfolios based on assessing funds' performance over a short past performance, i.e, a short IS horizon. The most common choices are 24 and 36 months. Beside, a choice of 48 month is also considered. As robustness checks, we repeat the discussed experiments with the choices of 24- and 48-month IS horizons and present the OOS performance in Tables B and C, respectively.

In both cases, our conclusions on both power and performance remain as in the 36-month IS case. On average, the $fwer^+$ gain higher power for a longer IS period. We also observe that the $fwer^+$ portfolios with a longer IS period tend to gain higher Sharpe ratio but lower alpha. Nevertheless, the differences are small.

Table B: OOS performance of $fwer^+$ portfolios with use of 24-month IS periods. Panels A to D of the table report OOS performance metrics of the $fwer^+$ portfolios with use each of R-square, AUM, and PC1s of moment and persistent group as the sole input covariate. The performance metrics include annualized alpha as well as its $t-$statistic and $p-$value, excess return and Sharpe ratio of the portfolios and summary on number of funds selected by $fwer^+$. Panel E reports these metrics of the $fwer^+$ portfolio with use of all four mentioned covariates as inputs whereas panel F the performance metrics of the equally weighted (EW) and equally weighted plus (EW^+) portfolios.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Number of selected funds			
						Average	Min	Max	Std
Panel A: $fwer^+$ with use of R-square as the covariate									
0.01	5.15	6.4	0.00	5.40	1.70	9	1	18	5
1.00	4.67	5.7	0.00	5.12	1.56	21	7	47	10
5.00	3.80	4.8	0.00	4.32	1.24	30	11	72	15
Panel B: $fwer^+$ with AUM as the covariate									
0.01	5.67	6.1	0.00	5.99	1.68	10	1	20	5
1.00	4.40	5.1	0.00	4.85	1.44	22	7	52	11
5.00	3.78	4.5	0.00	4.31	1.22	31	12	79	17
Panel C: $fwer^+$ with use of PC1 of moment group as the covariate									
0.01	6.03	7.0	0.00	6.39	1.91	10	1	18	5
1.00	4.82	6.1	0.00	5.34	1.68	22	7	52	11
5.00	3.92	4.9	0.00	4.46	1.33	31	10	75	17
Panel D: $fwer^+$ with use of PC1 of persistent group as the covariate									
0.01	5.71	6.2	0.00	6.04	1.66	9	1	18	5
1.00	4.68	5.6	0.00	5.10	1.54	21	7	47	10
5.00	4.06	5.1	0.00	4.53	1.39	30	11	72	15
Panel E: $fwer^+$ with use of the R-square, AUM and PC1s of the two groups as the covariates									
0.01	5.57	7.2	0.00	5.84	1.95	10	2	20	5
1.00	4.59	5.6	0.00	5.08	1.56	24	7	56	12
5.00	3.85	4.9	0.00	4.40	1.28	34	12	79	17
Panel F: equally weighted portfolios									
EW	2.85	3.2	0.00	5.00	0.78	1098	500	1570	335
EW^+	3.24	4.0	0.00	5.13	0.87	783	323	1418	313

Table C: OOS performance of $fwer^+$ portfolios with use of 48-month IS periods. Panels A to D of the table report OOS performance metrics of the $fwer^+$ portfolios with use each of R-square, AUM, and PC1s of moment and persistent group as the sole covariate. The performance metrics include annualized alpha as well as its $t-$statistic and $p-$value, excess return and Sharpe ratio of the portfolios and summary on number of funds selected by $fwer^+$. Panel E reports these metrics of the $fwer^+$ portfolio with use of all four mentioned covariates as inputs whereas panel F the performance metrics of the equally weighted (EW) and equally weighted plus (EW^+) portfolios.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Number of selected funds			
						Average	Min	Max	Std
Panel A: $fwer^+$ with use of R-square as the covariate									
0.01	4.82	7.3	0.00	5.18	2.34	18	3	33	8
1.00	4.19	6.2	0.00	4.69	1.72	38	13	74	19
5.00	3.85	4.9	0.00	4.44	1.46	52	18	100	27
Panel B: $fwer^+$ with AUM as the covariate									
0.01	4.76	8.2	0.00	5.13	2.48	19	3	35	8
1.00	4.18	6.2	0.00	4.70	1.69	41	12	80	21
5.00	3.84	5.0	0.00	4.45	1.40	56	17	111	31
Panel C: $fwer^+$ with use of PC1 of moment group as the covariate									
0.01	4.73	7.0	0.00	5.09	2.12	19	2	34	8
1.00	4.34	6.0	0.00	4.86	1.68	38	4	77	20
5.00	4.11	5.4	0.00	4.73	1.56	54	8	105	30
Panel D: $fwer^+$ with use of PC1 of persistent group as the covariate									
0.01	4.92	7.6	0.00	5.29	2.38	18	3	34	8
1.00	4.20	6.1	0.00	4.71	1.70	37	13	70	19
5.00	3.99	5.1	0.00	4.59	1.50	52	15	100	27
Panel E: $fwer^+$ with use of the R-square, AUM and PC1s of the two groups as the covariates									
0.01	4.55	7.3	0.00	4.86	2.10	20	1	37	9
1.00	4.43	6.7	0.00	4.98	1.81	43	2	82	23
5.00	4.04	5.0	0.00	4.70	1.46	59	2	118	34
Panel F: equally weighted portfolios									
EW	2.77	3.1	0.00	4.94	0.76	898	337	1271	283
EW^+	3.09	3.7	0.00	5.00	0.84	672	263	1153	263

Internet Appendix

In section IA we present the StepM and StepSPA procedures in our alpha testing framework. Section IB presents additional simulation results in which we consider an alternative setting on the proportion of out-performing funds. Section IC reports the performance of the $fwer^+$ portfolios under the use of individual covariates. Finally, Section ID presents empirical results when the simple $p-$value is used instead of the one without HAC correction.

IA. The implementation of the StepM and StepSPA procedures

In this section we present the StepM and StepSPA procedures for multiple tests where the testing metric is the alpha of an asset pricing model. In line with frameworks of Romano and Wolf (2005) and Hansen (2005), Hsu *et al.* (2010) we first consider performance of n funds and conduct for each fund i a hypothesis test:

$$H_0 : \mu_i \leq 0 \qquad H_1 : \mu_i > 0 \tag{A.1}$$

where μ_i is the expectation of a time-varying metric $d_{i,t}$ which represents for the performance of the fund i relative to a benchmark at time t, $i = 1, \ldots, n$. The relative performance can be expressed in a form of $d_{i,t} = L_{0,t} - L_{i,t}$ where $L_{0,t}$ and $L_{i,t}$ are values of a loss function measured at time t of the benchmark and fund i, respectively. The choice of the loss function is flexible and depends on the goal of researchers.

For instance, in the framework of Hsu *et al.* (2010), where they assess the performance of trading rules, the $L_{i,t}$ is set to be -1 multiplied by the return of a trading rule i in excess of interest rate in day t. The benchmark strategy is one that earns the interest rate, whose $L_{0,t} = 0$ which is -1 multiplied by 0 (the benchmark return excess of the interest rate). The $d_{i,t}$ turns out to be the excess return of the strategy i and μ_i is its expected return.

In our framework, the testing performance is the alpha of a fund, the choice of the loss function will be different. Suppose we are testing the alpha of the model (4). We consider funds surviving through periods t from 1 to T and assess their performances in this period. As the adjusted return of a fund i is $r_{i,t} - \boldsymbol{F}_t \hat{\boldsymbol{\beta}}_i$ where $\hat{\boldsymbol{\beta}}_i$ is the estimate of $\boldsymbol{\beta}_i$, we define a loss function as $L_{i,t} = -[r_{i,t} - \boldsymbol{F_t} \hat{\boldsymbol{\beta}}_i]$. This is a natural setting as the smaller the loss $L_{i,t}$, the better the performance of the fund. The benchmark is the portfolio that invests on the considering risk factors and thus the adjusted return is 0 and its loss is $L_{0,t} = 0$. We have that the expectation of $L_{0,t} - L_{i,t} = r_{i,t} - \boldsymbol{F_t} \hat{\boldsymbol{\beta}}_i$ is the alpha of the fund i. Thus, in our framework

$$d_{i,t} = r_{i,t} - \boldsymbol{F_t} \hat{\boldsymbol{\beta}}_i = \hat{\alpha}_i + \hat{\varepsilon}_{i,t} \tag{A.2}$$

The StepM and StepSPA procedures rely on a bootstrapped resampling where the stationary bootstrap procedure of Politis and Romano (1994), with average length $1/q$ where $q \in (0, 1)$, is adopted.

First we estimate the variance $\hat{\omega}_i^2$ of $d_{i,t}$ as in Hansen (2005). More specifically, let $\bar{d}_i$ be the average of $d_{i,1}, \ldots, d_{i,T}$. Then,

$$\hat{\omega}_i^2 = \hat{\gamma}_{i,0} + 2 \sum_{t=1}^{T-1} \kappa(T, t) \hat{\gamma}_{i,t} \tag{A.3}$$

where $\hat{\gamma}_{i,t} = 1/T. \sum_{k=1}^{T-t} (d_{i,k} - \bar{d}_i)(d_{i,k+t} - \bar{d}_i)$, $t = 0, \ldots, T-1$ and $\kappa(T, t) = \frac{T-t}{T}(1-q)^t + \frac{t}{T}(1-q)^{T-t}$.

For each bootstrapped iteration b, a cross-sectionally and jointly bootstrapped return of each fund i and risk factors are generated. We calculate $\bar{d}_{i,b} = \sum_{t=1}^{T} d_{i,t,b}/T$ where $d_{i,t,b}$ is the relative performance obtained by implementing (A.2) on the bootstrapped return of the fund i.[18]

For the (studentized) StepM procedure, we calculate the variance $\hat{\omega}_{i,b}^2$ for the bootstrapped differential $d_{i,t,b}$ via using (A.3) where $d_{i,t}$ and $\bar{d}_i$ are replaced by $d_{i,t,b}$ and $\bar{d}_{i,b}$, respectively.

After B iterations, we establish a bootstrapped critical point for StepM, $c^*_{\tau,StepM}$, as the $(1-\tau)^{th}$ quantile of the bootstrapped population $\max_i[(\bar{d}_{i,b}-\bar{d}_i)/\hat{\omega}_{i,b}]$, $b = 1,\ldots,B$.[19]

In the (studentized) StepSPA, we define $\hat{\mu}_i = \bar{d}_i.\mathbf{1}_{\{\sqrt{T}\bar{d}_i \le -\hat{\omega}_i\sqrt{2\log\log T}\}}$ where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function, $i = 1,\ldots,n$. and the bootstrapped critical point is defined as $c^*_\tau = \max\{c_\tau, 0\}$ where c_τ is the $(1-\tau)^{th}$ quantile of the bootstrapped population $\sqrt{T}\max_i[(\bar{d}_{i,b}-\bar{d}_i+\hat{\mu}_i)/\hat{\omega}_i]$, $b = 1,\ldots,B$.

Both the StepM and StepSPA are processed with the same steps but different in the statistics and bootstrapped critical point. In particular, the StepSPA procedure is as followings:

- Sort $\sqrt{T}\bar{d}_i/\hat{\omega}_i$ in a descending order.
- Select the top k funds if $\sqrt{T}\bar{d}_k/\hat{\omega}_k > c^*_\tau$. If there is no hypothesis rejected then stop the procedure. Otherwise,
    1. Remove the selected funds to obtain a sub-sample. Recalculate the critical c^*_τ with use of the sub-sample, denoted by c^s_τ.
    2. The top k' funds in the sub-sample with $\sqrt{T}\bar{d}_{k'}/\hat{\omega}_{k'} > c^s_\tau$ are selected. If there is no hypothesis rejected then stop the procedure. Otherwise go to step 3.
    3. Repeat the steps 1 and 2 above until there is no hypothesis that can be rejected.

In the StepM procedure, the statistics $\sqrt{T}\bar{d}_k/\hat{\omega}_k$ and bootstrapped critical point c^*_τ are replaced by $\bar{d}_k/\hat{\omega}_k$ and $c^*_{\tau,StepM}$, both in the whole sample and sub-sample, respectively.

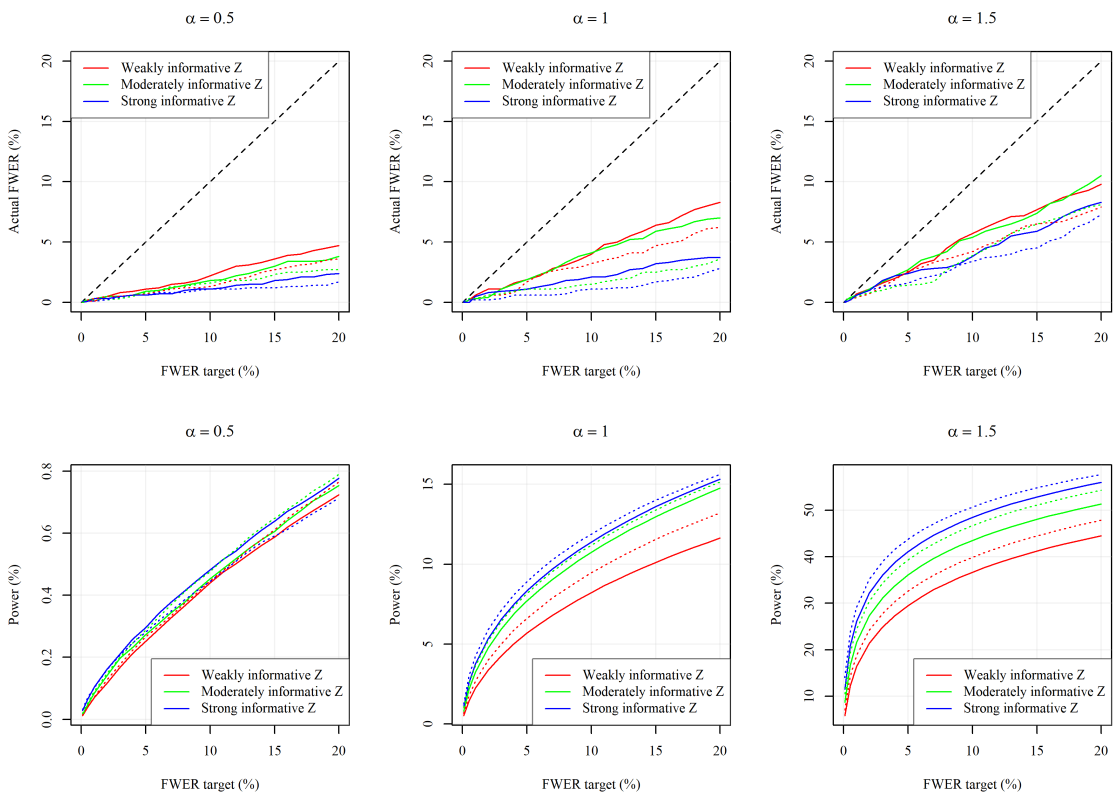
[18] We bootstrap from fund returns and risk factors as adjusted return of a fund i is changing via both its return and the estimated $\boldsymbol{\beta}_i$ calculated with use of the return.

[19] Our implementation of the StepM procedure is similar to Hsu *et al.* (2010) which differs from the original procedure of Romano and Wolf (2005) in three aspects. First, they use a circular block bootstrap while we use the stationary bootstrap. Second, they adopt data-driven algorithm in determining the block size of bootstrap, while we use a fixed value $q = 0.9$ following literature. Third, they use bootstrapped standard errors whereas we adopt HAC estimators as described in A.3. These differences might affect the finite sample performance of the StepM performance reported in our simulations results.

IB. Additional simulation studies

To complement our simulation studies presented in Section 5 of our main manuscript, this section presents simulation results for a different setting of the triple (π^+, π_0, π^-). More specifically, we additionally consider the case $\pi^+ = 40\%, \pi_0 = 60\%$ and $\pi^- = 0\%$. Generally, the results are very similar to the case $\pi^+ = 20\%, \pi_0 = 60\%$ and $\pi^- = 20\%$ and for the interest of space, we present the main results in Figure I where 18 variant settings are considered. These variants cover various signal settings in terms of α magnitude, the covariate signal strength and the correlation among the covariates.

Figure I: Performance of the $fwer^+$ under varying setting of signals under the alternative setting. The figure shows impact of signals, i.e., the magnitude of true non-zero alpha and informativeness of covariates, on the performance of the $fwer^+$ in terms of FWER control (top three sub-figuers) and power (bottom three figures). The simulated data are balanced panel of $n = 1000$ funds where each of them has $T = 36$ observations. The funds population consists of around 60%, 0% and 40% zero-alpha, under- and out-performing funds, respectively. The out-performing (under-performing) funds in population have alpha of α $(-\alpha)$ which varies in $\{0.5\%, 1.0\%, 1.5\%\}$. We consider three settings of the two covariates $\boldsymbol{Z} = (u, v)$ including weakly, moderately and strongly informative. The covariates can be independent (solid curves) or correlated with a coefficient of 0.5 (dotted curves).



IC. Performance of $fwer^+$ portfolios with use of individual covariates

In this section, we present the OOS performance metrics of the $fwer^+$ portfolios with use of underlying individual covariates. Particularly, we present in Table I and II the metrics corresponding to individual covariates in the persistent and moment groups, respectively.

Table I: OOS performance of $fwer^+$ portfolios with use of persistent covariates. The table reports the OOS performance of $fwer^+$ portfolios with use of individual covariates in persistent group. Panels A to I report OOS annualized alpha as well as its Newey and West (1987) $t-$statistic and $p-$value, excess return and Sharpe ratios and summary on size of $fwer^+$ portfolios with use of the sole covariate at various given FWER target τ.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Portfolio Size			
						Average	Min	Max	Std
Panel A: $fwer^+$ with use as the covariate the ACF1 of past 12-month excess returns									
0.01	5.38	8.1	0.00	5.81	2.31	15	4	25	6
1.00	4.77	7.6	0.00	5.34	1.73	33	10	56	13
5.00	4.54	6.9	0.00	5.09	1.63	46	11	84	19
Panel B: $fwer^+$ with use as the covariate the ACF1 of past 24-month excess returns									
0.01	5.18	8.0	0.00	5.59	2.26	15	4	25	6
1.00	4.33	6.9	0.00	4.84	1.55	33	10	56	14
5.00	4.37	6.9	0.00	4.83	1.60	46	15	82	19
Panel C: $fwer^+$ with use as the covariate the ACF1 of past 36-month excess returns									
0.01	5.24	8.1	0.00	5.64	2.26	15	4	25	6
1.00	4.44	7.2	0.00	4.94	1.59	34	10	56	13
5.00	4.45	7.0	0.00	4.91	1.61	46	15	87	19
Panel D: $fwer^+$ with use as the covariate the ACF2 of past 12-month excess returns									
0.01	4.89	7.0	0.00	5.36	2.07	15	4	26	6
1.00	4.40	7.3	0.00	4.90	1.60	33	8	54	14
5.00	4.43	6.9	0.00	4.94	1.60	46	14	88	19
Panel E: $fwer^+$ with use as the covariate the ACF2 of past 24-month excess returns									
0.01	4.91	7.2	0.00	5.38	2.11	15	4	25	6
1.00	4.48	7.3	0.00	5.01	1.62	33	10	55	14
5.00	4.41	6.8	0.00	4.94	1.58	46	14	87	20
Panel F: $fwer^+$ with use as the covariate the ACF2 of past 36-month excess returns									
0.01	4.93	7.2	0.00	5.37	2.06	15	3	25	6
1.00	4.33	6.6	0.00	4.87	1.53	34	10	55	13
5.00	4.22	6.4	0.00	4.73	1.50	47	14	88	19
Panel G: $fwer^+$ with use as the covariate the ACF3 of past 24-month excess returns									
0.01	5.26	8.3	0.00	5.64	2.35	15	4	23	6
1.00	4.59	7.4	0.00	5.09	1.66	33	10	54	13
5.00	4.28	6.7	0.00	4.76	1.52	46	15	87	20
Panel H: $fwer^+$ with use as the covariate the ACF3 of past 12-month excess returns									
0.01	5.24	8.4	0.00	5.62	2.39	15	4	25	6
1.00	4.42	7.1	0.00	4.93	1.61	33	10	55	13
5.00	4.15	6.6	0.00	4.60	1.47	46	15	83	19
Panel I: $fwer^+$ with use as the covariate the ACF3 of past 36-month excess returns									
0.01	4.96	7.4	0.00	5.42	2.15	15	4	25	7
1.00	4.56	7.5	0.00	5.07	1.65	33	10	55	13
5.00	4.22	6.6	0.00	4.68	1.51	46	14	87	20

Table II: OOS performance of $fwer^+$ portfolios with use of moment covariates. The table reports the OOS performance of $fwer^+$ portfolios with use of individual covariates in moment group. Panels A to I report OOS annualized alpha as well as its $t-$statistic and $p-$value, excess return, Sharpe ratio and a summary on the size of the $fwer^+$ portfolios with use of the each covariate in moment group.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Portfolio Size			
						Average	Min	Max	Std
Panel A: $fwer^+$ with use as the covariate the kurtosis of past 12-month excess returns									
0.01	4.94	7.2	0.00	5.38	2.11	16	4	25	6
1.00	4.37	7.2	0.00	4.87	1.63	32	9	55	14
5.00	4.39	7.0	0.00	4.88	1.58	45	13	86	20
Panel B: $fwer^+$ with use as the covariate the kurtosis of past 24-month excess returns									
0.01	4.89	7.0	0.00	5.35	2.06	15	4	26	6
1.00	4.50	7.8	0.00	5.00	1.68	33	9	55	13
5.00	4.18	6.4	0.00	4.70	1.48	45	14	86	20
Panel C: $fwer^+$ with use as the covariate the kurtosis of past 36-month excess returns									
0.01	4.94	7.2	0.00	5.40	2.10	15	4	25	6
1.00	4.35	7.3	0.00	4.85	1.60	33	10	55	13
5.00	4.17	6.5	0.00	4.67	1.48	45	13	87	20
Panel D: $fwer^+$ with use as the covariate the skewness of past 12-month excess returns									
0.01	4.84	6.9	0.00	5.26	2.05	15	4	25	7
1.00	4.38	7.2	0.00	4.87	1.58	33	9	57	13
5.00	4.30	6.8	0.00	4.78	1.56	45	13	86	20
Panel E: $fwer^+$ with use as the covariate the skewness of past 24-month excess returns									
0.01	5.17	7.6	0.00	5.57	2.26	15	4	26	7
1.00	4.25	6.9	0.00	4.69	1.56	34	10	56	13
5.00	4.33	6.7	0.00	4.83	1.56	46	14	88	20
Panel F: $fwer^+$ with use as the covariate the skewness of past 36-month excess returns									
0.01	5.43	7.9	0.00	5.85	2.35	15	3	26	7
1.00	4.29	7.1	0.00	4.75	1.57	32	10	55	14
5.00	4.25	6.3	0.00	4.75	1.49	46	14	87	21
Panel G: $fwer^+$ with use as the covariate the variance of past 12-month excess returns									
0.01	4.90	7.0	0.00	5.31	2.03	16	4	29	7
1.00	4.51	7.5	0.00	5.01	1.65	34	10	64	15
5.00	4.15	6.2	0.00	4.68	1.49	48	13	90	22
Panel H: $fwer^+$ with use as the covariate the variance of past 24-month excess returns									
0.01	4.86	7.1	0.00	5.26	2.04	16	4	30	7
1.00	4.49	7.4	0.00	4.99	1.63	34	10	65	15
5.00	4.24	6.5	0.00	4.79	1.51	49	14	92	22
Panel I: $fwer^+$ with use as the covariate the variance of past 36-month excess returns									
0.01	4.91	7.3	0.00	5.34	2.12	16	4	27	7
1.00	4.45	7.2	0.00	4.96	1.60	35	10	65	15
5.00	4.36	6.8	0.00	4.90	1.58	49	14	97	23

Generally, we see that the performances of the $fwer^+$ portfolios with use of different individual covariates of the same sub-groups but differing in estimation windows tend to be similar. This supports the use of a representative covariate such as PC1 for each group as presented in our main manuscript.

ID. Performance of $fwer^{+}$ portfolios with use of simple $p-$values

As mentioned in our main manuscript, there might be concern in use of $p-$values with HAC correction given the short IS time series. In this section, we present OOS performance of the $fwer^{+}$ portfolios with use of simple $p-$values, i.e., the p-values calculated without using the HAC correction. The results are shown in Table III. The performance metrics gained by the portfolios are uniformly higher than those gained by the portfolios constructed with use of HAC correction. As also shown in Table IV, when the new covariates obtained by the four famous machine learning techniques are used, the $fwer^{+}$ portfolios perform impressively with Sharpe ratio of more than 2 at all considering FWER targets and more than 3 at the lowest considering FWER target.

Table III: OOS performance of $fwer^{+}$ portfolios with use of simple $p-$values. Panels A to D of the table report OOS performance metrics of the $fwer^{+}$ portfolios with use each of R-square, AUM, and PC1s of moment and persistent group as the sole covariate. The performance metrics include annualized alpha as well as its $t-$statistic and $p-$value, excess return and Sharpe ratio and summary on number of out-performing funds detected by the $fwer^{+}$. Panel E reports these metrics of the $fwer^{+}$ portfolio with use of all four mentioned covariates as inputs whereas panel F the performance metrics of the equally weighted (EW) and equally weighted plus (EW^{+}) portfolios.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Number of detected funds			
						Average	Min	Max	Std
Panel A: $fwer^{+}$ with use of R-square as the covariate									
0.01	5.45	9.2	0.00	5.71	2.60	13	2	24	6
1.00	4.90	7.7	0.00	5.40	2.09	28	9	57	11
5.00	4.56	7.0	0.00	5.11	1.74	39	13	84	18
Panel B: $fwer^{+}$ with AUM as the covariate									
0.01	5.32	8.3	0.00	5.56	2.29	14	2	26	7
1.00	4.70	7.5	0.00	5.20	1.98	29	9	60	13
5.00	4.40	6.2	0.00	5.03	1.59	41	10	97	22
Panel C: $fwer^{+}$ with use of PC1 of moment group as the covariate									
0.01	5.37	8.7	0.00	5.62	2.45	14	2	30	7
1.00	4.51	6.6	0.00	4.99	1.91	29	9	69	13
5.00	4.81	6.8	0.00	5.45	1.65	41	11	108	22
Panel D: $fwer^{+}$ with use of PC1 of persistent group as the covariate									
0.01	5.39	8.6	0.00	5.63	2.42	13	2	23	6
1.00	4.78	7.2	0.00	5.28	1.96	27	9	59	12
5.00	4.69	7.0	0.00	5.26	1.83	37	10	83	17
Panel E: $fwer^{+}$ with use of the R-square, AUM and PC1s of the two groups as the covariates									
0.01	5.55	8.7	0.00	5.84	2.48	15	3	30	7
1.00	5.10	7.6	0.00	5.70	1.81	32	10	72	15
5.00	4.99	8.0	0.00	5.69	1.62	47	13	115	25
Panel F: equally weighted portfolios									
EW	2.58	2.9	0.00	4.65	0.72	1067	350	1570	361
EW^{+}	3.00	3.7	0.00	4.77	0.80	761	273	1418	324

Table IV: OOS performance of $fwer^+$ portfolios with use of new covariates and simple $p-$values. Panel A (B, C and D) reports OOS annualized alpha as well as its $t-$statistic and $p-$value, excess return and Sharpe ratios and summary on the size of the $fwer^+$ portfolios with use of funds' future return predicted by LASSO (GB, RF and DNN) model at given FWER targets τ.

τ (%)	Alpha (%)	$t-$statistic	$p-$value	Return (%)	Sharpe Ratio	Number of detected funds			
						Average	Min	Max	Std
Panel A: $fwer^+$ with use of future return predicted by the LASSO model as the sole covariate									
0.01	5.96	11.8	0.00	6.18	3.33	13	2	22	6
1.00	5.49	9.8	0.00	5.98	2.58	25	8	49	10
5.00	5.16	8.6	0.00	5.76	2.24	35	10	81	16
Panel B: $fwer^+$ with use of future return predicted by the GB model as the sole covariate									
0.01	5.75	11.4	0.00	6.01	3.22	13	2	23	7
1.00	5.01	7.9	0.00	5.54	2.39	26	9	56	12
5.00	4.93	7.6	0.00	5.55	2.13	36	10	80	19
Panel C: $fwer^+$ with use of future return predicted by the RF model as the sole covariate									
0.01	6.61	9.0	0.00	6.81	2.46	12	2	22	6
1.00	5.23	8.2	0.00	5.75	2.38	25	9	50	10
5.00	5.12	7.8	0.00	5.82	2.01	35	10	81	17
Panel D: $fwer^+$ with use of future return predicted by the DNN model as the sole covariate									
0.01	5.91	11.7	0.00	6.20	3.28	13	2	23	7
1.00	5.08	8.3	0.00	5.61	2.39	26	5	56	12
5.00	5.16	8.2	0.00	5.80	2.20	36	9	77	18