

Assessing Hedge Fund Performance with an Information-based Multiple Test

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*University of Edinburgh Business School, University of Edinburgh, Edinburgh, EH8 9JS UK*

^d*Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ, UK*

Abstract

We develop $fwer^+$, an informative covariate-based testing procedure for selecting outperforming hedge funds while controlling the family-wise error rate (FWER). The method conditions rejection thresholds on informative covariates (such as fund characteristics and macroeconomic variables) and outperforms prior unconditional tests in simulations. Applied to U.S. hedge fund data, $fwer^+$ constructs portfolios that outperform passive benchmarks and generate significant out-of-sample alphas across FWER targets and factor models. Contribution analyses indicate that return-based covariates account for the largest standalone share of alpha gains, while macroeconomic exposures appear to complement other information sources in identifying persistent outperformance.

Keywords: data snooping, multiple testing, multivariate family-wise error rate, informative covariates, hedge fund.

JEL Classification: C12, C15, G11, G12, G23.

*We thank participants at the 2024 European Financial Management Association Annual Meeting, the 2024 Spanish Finance Association Annual Meeting, the 2025 International Finance and Banking Society Conference, 2026 Asia Meeting of the Econometric Society, East & Southeast Asia, and the Durham University and University of Nottingham seminars for helpful comments. Replication code and sample data are available at: <https://github.com/TrenMA/fwer-plus-in-assessing-hedge-fund-performance>. Hsu acknowledges research support from the Mack Institute for Innovation Management, The Wharton School, University of Pennsylvania. We acknowledge access to the University of Nottingham's Ada HPC service.

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

1. Introduction

Hedge funds offer several advantages to investors: they are free to change strategies, can apply high leverage, and operate in a relaxed regulatory framework. This has led to an increase in the size of the hedge fund industry over the past few decades, reaching \$4.51 trillion in assets under management in 2024 (a 9.75 % increase from the previous year, [Mandle, 2024](#)). Paradoxically, the actual skill of hedge fund managers is not directly observable by investors ([Wu et al., 2021](#)). These facts raise the following long-standing question in the hedge fund literature: Is there a method to detect skilled hedge fund managers? Is there any information input that can assist investors in fund selection? To answer all these questions, researchers need a statistical methodology that can handle the following four issues simultaneously: (i) appropriately controlling downside risk; (ii) updating information in a timely way by incorporating various fund features in the selection process; (iii) controlling for *luck*; and (iv) being feasible for a large panel. In this paper, we propose the family-wise error rate plus ($fwer^+$) method, which helps us assess hedge fund performance and examine the covariate information associated with persistent hedge fund outperformance.

Prior hedge fund studies differ in their approaches to identifying hedge fund managers who can produce significant alphas: one strand of the literature focuses on fund-level performance covariates, while the other uses multiple-hypothesis testing to control for luck. The first strand of the literature examines the persistence of hedge fund performance and the relationships between funds' characteristics, macroeconomic variables, and performance (see, e.g., [Agarwal and Naik, 2000](#); [Avramov et al., 2011](#); [Bali et al., 2014](#); [Bali and Weigert, 2024](#); [Sun et al., 2018](#); [Chen et al., 2021](#)). The majority of those studies create hedge fund portfolios by sorting funds based on specific fund characteristics and past performance.

In the second strand, the existing literature on hedge fund performance accounting for *lucky* funds is limited. It focuses on controlling either the family-wise error rate (FWER) or

the false discovery rate (FDR).¹ The FDR is the expected proportion of non-outperforming funds among those selected as outperformers. Suppose that investors control FDR at the 5% level when forming their portfolio. Then they would expect about 5% of the portfolio of funds to be non-outperforming. In contrast, if they control for FWER at 5%, the chance of having *any* non-outperforming funds in the portfolio is only 5%. Roughly speaking, if they form such portfolios annually over 100 years, they should expect only about 5 years in which any selected funds are non-outperforming.

It is also worth noting that existing multiple-testing methods in this literature are unconditional; that is, they rely only on funds' returns and do not account for other relevant information affecting hedge fund performance (e.g., fund characteristics). In practice, it appears unrealistic to evaluate hedge funds without accounting for all relevant information available to investors. On the one side, hedge fund investments are characterised by illiquidity, correlation risk, and tail risk.² Moreover, hedge funds trade in illiquid assets and tend to smooth returns when pricing them. Smoothing returns can lead to inflated risk-adjusted performance (Getmansky et al., 2004). All these elements lead them to report negative skewness and fat tails relative to normal distribution, and their population to be littered with *lucky* funds, or else hedge funds that report a positive performance that is not statistically significant (see, Giglio et al., 2020; Ferson and Chen, 2021; Chen et al., 2017). On the other

¹In particular, Romano and Wolf (2005) and Romano and Shaikh (2006) have developed FWER control methods, apply them to hedge fund performance and identified a handful of hedge funds as outperformers. Ferson and Chen (2021) build upon the FDR method of Barras et al. (2010) and probability theory to infer the performance of a cross-section of hedge funds. Similarly, Giglio et al. (2020) develop a machine learning FDR procedure that accounts for omitted factors and missing data, and conduct multiple hypothesis testing for hedge funds. In contrast with the findings of Romano and Wolf (2005) and Romano and Shaikh (2006), both FDR studies argue that hundreds of hedge funds outperform after controlling for luck. This finding is expected given FDR's lenient nature.

²Hedge funds often employ market-neutral arbitrage strategies to immunise their portfolios. Those strategies are built on unrealistic assumptions about correlations (e.g., time-varying market betas), which can lead to exposure to correlation risk and result in large losses (see Buraschi et al., 2013). Also, hedge funds extensively trade derivatives, which exposes them to correlation and volatility shocks, as well as tail risk (see Agarwal et al., 2017).

hand, hedge funds require a significant minimum investment and have lock-up periods. This leads to a high investment threshold to form portfolios, which, in simple terms, makes the risk from choosing the wrong hedge fund considerably higher. All of these issues demand a strict method to control for luck and to account for all available information in a dynamic manner.

This study bridges the aforementioned two strands of literature and fills the gap in hedge funds' selection by introducing the $fwer^+$ procedure, an approach based on the statistical framework of [Zhou et al. \(2021\)](#) that estimates FWER as a function of multiple informative covariates. The idea is to allow the test's rejection threshold for different hedge funds to vary by fund-specific information. We deviate from the framework of [Zhou et al. \(2021\)](#) by (i) specifically aiming to control the FWER among the discoveries in the right tail (i.e., detecting hedge funds with positive abnormal returns); and (ii) accommodating correlated covariates that are common in financial data. $fwer^+$ is well-suited to hedge fund evaluation for three reasons. First, the FWER control protects against the downside risk of selecting the wrong funds. Second, its multiple testing framework controls for luck in fund performance evaluation. Finally, its use of informative covariates to differentiate rejection thresholds allows testing the *conditional* performance of hedge funds. Given that investors continually update their information, additional information captured by those covariates allows us to evaluate investors' rational expectations. In fact, using those covariates gives us greater power to detect additional outperforming hedge funds, making it feasible to form a hedge fund portfolio.

We first compare the finite-sample performance of $fwer^+$ with two established FWER-controlling procedures: the stepwise multiple testing procedure of [Romano and Wolf \(2005\)](#) (StepM) and the stepwise superior predictive ability test of [Hsu et al. \(2010\)](#) (StepSPA). The simulations show that $fwer^+$ controls FWER well across target levels while retaining meaningful power compared to its benchmarks. In contrast, StepM is highly conservative in our hedge fund setting, with power close to zero across all FWER targets considered. StepSPA

is more powerful than StepM, but its empirical FWER exceeds the nominal target across stationary-bootstrap average block lengths ranging from 2 to 12 months. These results suggest that, in our short-window, large-cross-section setting, the two unconditional procedures are useful for assessing finite-sample trade-offs but are not reliable tools for empirical fund selection. Accordingly, we report StepM and StepSPA as simulation benchmarks, while the empirical portfolio analysis focuses on $fwcr^+$.

We then consider an extensive list of simulations, which shows that the $fwcr^+$ procedure achieves higher power when (i) covariates are highly informative about the prior null probability; (ii) outperforming funds' alphas are higher; (iii) informative covariates are more comprehensive; (iv) the noise-to-signal ratio is lower in covariates; (v) a longer period is considered; and (vi) a smaller number of funds are tested. Moreover, the correlation among covariates does not materially reduce the power of $fwcr^+$ and its control at all FWER targets.

In our empirical analysis, we implement $fwcr^+$ to identify outperforming hedge funds using a comprehensive set of 24 informative covariates grouped into four categories: return characteristics, fund characteristics, managerial skill, and macroeconomic variables. Beginning in late 1999, we estimate fund alphas, compute covariates, and apply $fwcr^+$ using a rolling 36-month in-sample (IS) window, with FWER target levels ranging from 0.1% to 5%. For each selection set, we construct both equal-weighted and alpha-weighted portfolios ("FWER portfolios" hereafter), with the latter assigning larger weights to funds with stronger estimated IS alphas. The selected portfolios are then held out-of-sample (OOS) for one year, after which the entire procedure is rolled forward annually through the end of 2023. To assess whether the selection step adds value beyond simple rules based on the same investment universe, we compare these portfolios with three passive benchmarks: EW , an equal-weighted portfolio of all eligible funds; EW^+ , an equal-weighted portfolio of eligible funds with positive IS alpha; and αW^+ , an alpha-weighted version of EW^+ .

We conduct several experiments using this rolling design. First, we construct portfolios

using each of the 24 covariates individually. We find that all portfolios outperform the passive benchmarks and generate significantly positive alphas in the seven-factor model of [Fung and Hsieh \(2004\)](#) (up to 4.2% per annum). Second, we form portfolios using the full set of covariates simultaneously. We find that these portfolios achieve significant seven-factor alphas of up to 6.4% per annum for the equal-weighted FWER portfolio and up to 9.0% per annum for the alpha-weighted FWER portfolio, respectively. The $fwer^+$ procedure using multiple covariates is sufficiently powerful to form non-empty portfolios even when controlling the FWER at 0.1%, with more than 70% of selected funds producing positive alpha in the OOS. The economic gains are also substantial. Over the full OOS period, the $fwer^+$ portfolios strongly outperform the passive EW , EW^+ , and αW^+ . A \$1 investment in the alpha-weighted (equal-weighted) $fwer^+$ portfolios grows to approximately \$11.0–\$13.5 (\$6.8–\$7.5), compared with roughly \$3.0–\$4.0 for the passive EW , EW^+ , and αW^+ benchmarks. This corresponds to cumulative returns of almost 1,200% over the last two decades. Notably, the selected funds display performance patterns consistent with managerial skill that not only help them avoid losses during downturns but also generate substantial gains during the 2008–2009 financial crisis, in contrast to the large losses observed in the broader hedge fund industry. Although performance moderates in parts of the later sample, both the equal-weighted and alpha-weighted $fwer^+$ portfolios retain statistically significant full-sample alphas, unlike their passive benchmarks. These findings remain when the holding periods of the $fwer^+$ portfolios are extended up to four years and when alternative factor models are considered.

Third, we decompose the $fwer^+$ -based portfolio into sub-portfolios by fund style and show that these sub-portfolios outperform their respective industry counterparts. The $fwer^+$ -based portfolio does not include funds from all styles at all times; rather, it selectively identifies truly outperforming funds from certain styles when such funds exist. Relative Value funds appear most frequently in the portfolio and have been dominant in recent decades. Nevertheless, the full set of Relative Value style funds, as obtained by our dataset, do not

outperform other styles, but those funds selected by $fwer^+$ do so.

Fourth, to better understand the economic information underlying the superior performance of the proposed procedure, we investigate how informative covariates contribute to the OOS alpha of the equal-weighted FWER portfolios. We focus on equal-weighted portfolios to isolate the contribution of covariate information to fund selection from the additional effect of alpha-based portfolio weighting. To this end, we conduct two complementary analyses. First, a leave-one-out ablation analysis quantifies the marginal contribution of individual covariates and covariate groups. Second, we adapt the Shapley value framework (Shapley, 1953) to decompose the portfolio alpha gain across covariate groups while accounting for overlapping information and to examine whether different sources of information reinforce one another. The results suggest distinct roles for return-based and macroeconomic covariates. Return-based covariates provide the largest standalone contribution to portfolio alpha, whereas macroeconomic covariates complement return-based covariates, managerial skill proxies, and fund characteristics. Together with additional robustness analyses, the findings suggest that $fwer^+$ benefits from integrating both common and covariate-specific information contained in the original covariates.

Our paper contributes to the literature in three ways. First, we contribute methodologically by adapting the informative-covariate FWER framework of Zhou et al. (2021) to the problem of selecting outperforming investment opportunities. The resulting $fwer^+$ procedure is designed for one-sided fund-selection problems and accommodates correlated covariates, which are common in financial applications. This provides a conditional multiple-testing framework that uses economically motivated information to improve power while maintaining stringent control of false discoveries.

Second, we contribute to the hedge fund performance literature by showing that conditioning multiple testing on informative covariates identifies portfolios with economically large and statistically significant OOS performance. The selected portfolios outperform passive benchmarks formed from the same eligible fund universe, remain significant across FWER

targets, weighting schemes, sub-periods, holding horizons, and factor models, and perform particularly well during market stress periods. These findings suggest that informative-covariate multiple testing can make stringent FWER control practically useful in large-scale fund-selection problems.

Third, we provide new evidence on the sources of information that drive the performance of the selected portfolios. Using the equal-weighted FWER portfolios to isolate the selection effect of covariates from the additional effect of alpha-based weighting, we develop leave-one-out and Shapley analyses to quantify the contribution of covariate information to portfolio alpha. These analyses show that return-based covariates provide the largest standalone contribution, whereas macroeconomic covariates reinforce the information contained in return-based covariates, managerial-skill proxies, and fund characteristics. These results highlight the importance of combining historical fund performance with information about funds' sensitivity to changing economic conditions.

The remainder of the paper is organized as follows. Section 2 introduces the *fwcr*⁺ procedure. Section 3 describes the hedge fund dataset, while Section 4 discusses the performance measures used. Section 5 presents the Monte Carlo simulation design and results. Section 6 reports the empirical analyses, Section 7 analyzes the contribution of covariate information to portfolio performance, and Section 8 concludes the paper.

2. Methodology

Suppose that we are assessing n hedge funds with respect to the Fung and Hsieh (2004) seven-factor model alpha (α). We test for each fund i a hypothesis

$$H_0 : \alpha_i = 0 \quad \text{against} \quad H_1 : \alpha_i \neq 0, \quad (1)$$

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

The literature on detecting outperformers while controlling for FWER has focused primarily on one-sided tests with a composite null, leading to the development of numerous

testing procedures.³ We aim to detect the funds having positive α , while controlling for the probability of selecting at least one non-outperforming fund at a predetermined target $\tau \in (0, 1)$. Among the funds selected as outperformers, let F denote the number of funds that in fact have $\alpha \leq 0$. We are attempting to control a type I error which is defined as $\text{FWER} = \mathbb{P}(F \geq 1)$ at the target τ .

The procedures proposed in the literature share a common foundation: they rely on bootstrap methods applied to funds' returns (or, more generally, some relative performance variable). However, these methods do not incorporate additional information – often available and informative – alongside the tests. Given that fund performance is typically assessed over short time series of returns, these limitations reduce the power of such approaches in detecting truly outperforming funds or managers that are expected to continue their performance in the (short) future.

The issue of low power in multiple testing procedures has attracted growing attention across many areas of science. Recent advances in statistics have sought to address this by incorporating additional information to improve the detection of false nulls, as in [Ignatiadis and Huber \(2021\)](#) and [Zhou et al. \(2021\)](#). However, these innovative approaches have not yet been applied to the problem of selecting outperformers. In this study, we propose a simple procedure that builds on the framework of [Zhou et al. \(2021\)](#) to overcome the low-power limitations of existing methods.⁴ In the following sections, we first summarize the framework and then introduce our $fwer^+$ procedure.

2.1. Application in Picking Outperforming Funds

At a given time t , we assess hedge funds' performance using their return series over the past 36-month period. We conduct tests (1) on these return series using the [Fung and Hsieh](#)

³This literature includes [Romano and Wolf \(2005\)](#), [Hansen \(2005\)](#), and [Hsu et al. \(2010\)](#).

⁴The framework of [Zhou et al. \(2021\)](#) is computationally more efficient and has been shown to be more powerful than that of [Ignatiadis and Huber \(2021\)](#).

(2004) seven-factor model as our benchmark.⁵ Our first step is to detect truly non-zero alpha funds while controlling FWER at a target $\tau \in (0, 1)$. Suppose there is a set of d informative covariates $Z^1, \dots, Z^d$ carrying the information on the probability that the null of tests (1) being true. Each Z^k is a column vector $(Z_1^k, \dots, Z_n^k)'$, $k = 1, \dots, d$. These covariates may be time-varying, such as the exposure of fund returns to macroeconomic conditions over the past 36 months, or time invariant, such as the characteristics of the funds.

For convenience, we denote $\mathbf{Z} = (Z^1, \dots, Z^d)$; thus, $\mathbf{Z}_i$ means $(Z_i^1, \dots, Z_i^d)$. For each $i = 1, \dots, n$, let P_i be the random variable representing the p -value of the test (1) corresponding to fund i and p_i be its realization. Conditional on $\mathbf{Z}_i = \mathbf{z}_i$, we denote the prior probability of the null hypothesis i being true by $\pi_0(\mathbf{z}_i)$. We model the distribution of P_i as a mixture of two groups in which the weights of the first and the second group are $\pi_0(\mathbf{z}_i)$ and $1 - \pi_0(\mathbf{z}_i)$, respectively. Let f_0 and f_{alt} be the density functions of the first and second group, respectively. Formally, we have

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

Thus, the covariates $\mathbf{Z}$ carry their information through the $\pi_0(\mathbf{z}_i)$, which takes a different value 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 . That is, the two density functions do not depend on $\mathbf{Z}$. Following common assumptions in the statistics literature, we assume the p -value of a true null is uniformly distributed, i.e., $f_0(p) = 1$ for all p and $f_{alt}(p)$ is a strictly decreasing function of p .⁶

In conventional approaches such as those of Bonferroni (1936) and Holm (1979), the rejection region is determined either by a common threshold Θ that is fixed across all tests

⁵We also consider other factor models in our robustness checks.

⁶Under the framework of Zhou et al. (2021), the tests are assumed to be cross-sectionally independent. For each test, the covariates are independent. However, in Section 5 we show that these assumptions can be substantially relaxed.

(i.e., hypothesis i is rejected if and only if $p_i \leq \Theta$) or by the rank of the test's p -value within the set of hypotheses being tested. Our idea is to determine for each hypothesis i a threshold which is a function of $\mathbf{z}_i$ denoted by $\Theta(\mathbf{z}_i)$ (i.e., the null hypothesis i is rejected if and only if $p_i \leq \Theta(\mathbf{z}_i)$). With this covariate-dependent threshold, FWER can be conservatively estimated by ⁷

$$\widehat{\text{FWER}} = \sum_{i=1}^n \pi_0(\mathbf{z}_i) \Theta(\mathbf{z}_i). \quad (3)$$

For each fund i , the prior odds of the null versus the alternative conditional on covariates is

$$\frac{\mathbb{P}(\alpha_i = 0 \mid \mathbf{Z}_i = \mathbf{z}_i)}{\mathbb{P}(\alpha_i \neq 0 \mid \mathbf{Z}_i = \mathbf{z}_i)} = \frac{\pi_0(\mathbf{z}_i)}{1 - \pi_0(\mathbf{z}_i)}.$$

When comparing multiple funds, those with higher prior odds are more likely to be zero-alpha funds. It is therefore natural to impose more stringent thresholds $\Theta(\mathbf{z}_i)$ for such funds, while allowing more lenient thresholds for those with lower prior odds. Given a fixed FWER budget, this reallocation allows more non-zero-alpha funds to be identified. This can be achieved by setting

$$\Theta(\mathbf{z}_i) = f_{alt}^{-1} \left(\frac{\pi_0(\mathbf{z}_i) \theta^*}{1 - \pi_0(\mathbf{z}_i)} \right), \quad (4)$$

where f_{alt}^{-1} denotes the inverse of f_{alt} , and $\theta^* > 0$ is a common scaling parameter. Since f_{alt}^{-1} is a decreasing function, higher prior odds translate into lower thresholds $\Theta(\mathbf{z}_i)$, as desired.

Similarly, a smaller value of θ^* leads to higher thresholds $\Theta(\mathbf{z}_i)$, thereby increasing the number of non-zero-alpha funds being detected. Hence, θ^* is chosen to be as small as possible while ensuring that FWER is controlled at the target level τ . Using (3) and (4), we have

$$\theta^* = \min \left\{ \theta > 0 : \sum_{i=1}^n \pi_0(\mathbf{z}_i) f_{alt}^{-1} \left(\frac{\pi_0(\mathbf{z}_i) \theta}{1 - \pi_0(\mathbf{z}_i)} \right) \leq \tau \right\}. \quad (5)$$

⁷Specifically, conditioning on $\mathbf{Z}_i = \mathbf{z}_i$, we have

$$\begin{aligned} \text{FWER} &= \mathbb{P}(p_i \leq \Theta(\mathbf{z}_i) \text{ for some zero-alpha fund } i) \leq \sum_{i=1}^n \mathbb{P}(p_i \leq \Theta(\mathbf{z}_i), \alpha_i = 0) \\ &= \sum_{i=1}^n \mathbb{P}(p_i \leq \Theta(\mathbf{z}_i) | \alpha_i = 0) \mathbb{P}(\alpha_i = 0 | \mathbf{Z}_i = \mathbf{z}_i) = \sum_{i=1}^n \Theta(\mathbf{z}_i) \pi_0(\mathbf{z}_i). \end{aligned}$$

In practice, to derive the asymptotic control of FWER, [Zhou et al. \(2021\)](#) replace the $\pi_0(\mathbf{z})$ outside f_{alt}^{-1} in (5) with a conservative quantity $\mathbf{1}_{p_i > \lambda}/(1 - \lambda)$ for some $\lambda \in (0, 1)$, where $\mathbf{1}_{p_i > \lambda}$ takes the value 1 if $p_i > \lambda$ and 0 otherwise.⁸ The $\pi_0(\mathbf{z})$ inside (4) and (5) as a logit function which has a form of $\pi_0(\mathbf{z}_i) = 1/(1 + e^{-b_0 - \mathbf{b}'\mathbf{z}_i})$, where $\mathbf{b} = (b_1, \dots, b_d)$ is the column vector of the coefficients of the d covariates. The alternative density function $f_{alt}(p)$ is modelled as a beta distribution $f_{alt}(p) = kp^{k-1}$ for $k \in (0, 1)$. The parameters $b_0, \dots, b_d$ and k are estimated via an expectation–maximization algorithm.

Of the mentioned parameters, $\mathbf{b} = (b_1, \dots, b_d)$ reflects the strength of the informativeness of the covariates, i.e., the link between the covariates and the probability of a particular hypothesis being null. Covariates with higher absolute coefficients are more informative. When all covariates are non-informative, one should expect these coefficients to be close to zero. The remaining coefficient b_0 reflects the proportion of the nulls, that is, the proportion of zero-alpha funds. These zero-alpha funds either randomly select stocks or replicate passive benchmarks or factors. All else equal, the higher b_0 , the higher the proportion of these funds.

As we aim to select truly positive alpha funds, we need to control FWER within the group of selected outperforming funds. In this Section, we propose a simple procedure that applies the framework of [Zhou et al. \(2021\)](#) to this problem.

Given a target FWER level $\tau \in (0, 1)$, our procedure, denoted $fwer^+$, consists of two steps. First, we implement the FWER procedure of [Zhou et al. \(2021\)](#) on the entire population of funds, controlling FWER at the target level τ . This procedure identifies a set of abnormal funds, say A , which includes both under-performing and outperforming funds.

⁸Indeed, for any $\lambda \in (0, 1)$, conditional on $\mathbf{Z}_i = \mathbf{z}_i$, we have that

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{p_i > \lambda} | \mathbf{z}_i) &= 0 \cdot \mathbb{P}(p_i \leq \lambda | \mathbf{z}_i) + 1 \cdot \mathbb{P}(p_i > \lambda | \mathbf{z}_i) = \mathbb{P}(p_i > \lambda, \alpha_i = 0 | \mathbf{z}_i) + \mathbb{P}(p_i > \lambda, \alpha_i \neq 0 | \mathbf{z}_i) \\ &= \mathbb{P}(p_i > \lambda | \alpha_i = 0, \mathbf{z}_i) \cdot \mathbb{P}(\alpha_i = 0 | \mathbf{z}_i) + \mathbb{P}(p_i > \lambda | \alpha_i \neq 0, \mathbf{z}_i) \cdot \mathbb{P}(\alpha_i \neq 0 | \mathbf{z}_i) \\ &= (1 - \lambda) \cdot \pi_0(\mathbf{z}_i) + \mathbb{P}(p_i > \lambda | \alpha_i \neq 0, \mathbf{z}_i) \cdot \mathbb{P}(\alpha_i \neq 0 | \mathbf{z}_i) \\ &\geq (1 - \lambda) \cdot \pi_0(\mathbf{z}_i) \end{aligned}$$

Hence, $\mathbb{E}(\mathbf{1}_{p_i > \lambda}/(1 - \lambda) | \mathbf{z}_i) \geq \pi_0(\mathbf{z}_i)$, i.e., $\mathbf{1}_{p_i > \lambda}/(1 - \lambda)$ is a conservative estimate of $\pi_0(\mathbf{z}_i)$.

Second, we select from this set only those funds with positive estimated alphas, denoted A^+ . Since the probability of having at least one zero-alpha fund in A is less than τ , the same bound conservatively holds for the subset $A^+ \subseteq A$. Assuming that no under-performing funds appear in A^+ by chance (i.e., no very lucky under-performers are selected), the set A^+ consists of truly outperforming funds with FWER controlled at the target level τ . As FWER control is stringent, this assumption is likely to hold in practice.

As we will show in Section 5, the $fwer^+$ procedure controls FWER effectively at any specified target level and, when informative covariates are available, is more powerful than existing methods.

2.2. Implementation Steps

We now describe how the $fwer^+$ procedure is implemented in practice. The goal is to identify hedge funds with positive alpha while controlling the probability of selecting any non-outperforming fund at a predetermined FWER target $\tau \in (0, 1)$. Suppose we have n hedge funds, with each of these having a return series for $t = 1, \dots, T$. For simplicity, we illustrate the procedure using two informative covariates (U, V) . The same implementation naturally extends to any number of covariates. The procedure consists of four steps.

1. *Calculate inputs.* For each fund i , we conduct test (1), collect the estimated alpha $\hat{\alpha}_i$ and calculate its p -value p_i using the heteroskedasticity and autocorrelation consistent (HAC) standard error of [Newey and West \(1987\)](#). We also record the realized values of the informative covariates $(U = u_i, V = v_i)$, $i = 1, \dots, n$. These quantities serve as inputs to the multiple testing procedure. The p -values summarize the statistical evidence for non-zero alpha, while the covariates capture additional information that may help distinguish skilled funds from zero-alpha funds.
2. *Estimate the relationship between covariates and the null probability.* The intuition behind $fwer^+$ is that funds with different characteristics have different probabilities of being zero-alpha funds. To incorporate this information, we estimate the probability

that fund i has zero alpha as a function of its covariates. We first determine a $\lambda \in (0, 1)$ such that p -values above λ are predominantly generated under the null (i.e., correspond to zero-alpha funds).⁹ To estimate the functions $\pi_0(u_i, v_i)$ and f_{alt} , the modelling of the p -value distribution of funds in model (2) is reformulated to the modelling of whether a fund has zero-alpha. Specifically, by defining $Y_i = \mathbf{1}_{p_i > \lambda}$, model (2) is equivalent to

$$Y_i | (u_i, v_i) \sim \pi_0(u_i, v_i) \text{Ber}(1 - \lambda) + [1 - \pi_0(u_i, v_i)] \text{Ber}(1 - F_{alt}(\lambda)) \quad (6)$$

where $\text{Ber}(1 - \lambda)$ is the Bernoulli distribution with success probability $1 - \lambda$ and F_{alt} is the cumulative distribution function corresponds to f_{alt} . Assume that $\pi_0(u_i, v_i)$ is a logistic function of the covariates (u_i, v_i)

$$\pi_0(u_i, v_i) = \frac{1}{1 + e^{-b_0 - b_1 u_i - b_2 v_i}}$$

and the density function of the truly non-zero-alpha funds' p -value as

$$f_{alt}(p_i) = k p_i^{k-1}.$$

The parameters b_0, b_1, b_2 and $k \in (0, 1)$ are estimated using an expectation–maximization algorithm that maximizes the likelihood of model (6).¹⁰ This step estimates the association between the covariates and the probability that a fund has zero alpha. This step determines which fund characteristics are informative about the likelihood of a fund having zero alpha. The coefficients describe how the covariates are associated with the probability that a fund belongs to the zero-alpha group.

3. *Determine covariate-dependent rejection thresholds.* For each fund i , given the estimated functions $\pi_0(u_i, v_i)$ and $f_{alt}(p_i)$, the rejection threshold for the fund is calculated

⁹To determine the value of λ , the approach of Storey (2002) and Storey et al. (2017) in estimating the proportion of zero-alpha funds π_0 , is adopted. Specifically, for a given λ , π_0 is estimated by $\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda\}}{(1-\lambda)n}$, where $\#\{p_i > \lambda\}$ denotes the number of funds with p -values exceeding λ . λ is chosen over the set $\{0.05, 0.10, \dots, 0.95\}$ to minimize the estimated mean squared error of $\hat{\pi}_0(\lambda)$.

¹⁰Specifically, the algorithm initializes values for (b_0, b_1, b_2) and k . At each iteration, using the parameter values from the previous iteration, the expectation step computes, for each fund, the posterior probability that the fund has zero alpha, given its Y_i and covariates. The maximization step updates (b_0, b_1, b_2) by maximizing the expected quasi-log-likelihood (White, 1982) of the Bernoulli mixture model (6), and updates k through the corresponding likelihood of the alternative distribution. These steps are repeated until convergence.

as

$$\Theta(u_i, v_i) = \left[\frac{\pi_0(u_i, v_i)\theta^*}{(1 - \pi_0(u_i, v_i))k} \right]^{1/(k-1)},$$

where

$$\theta^* = k \left[\frac{1}{\tau(1-\lambda)} \sum_{i=1}^n Y_i \left(\frac{\pi_0(u_i, v_i)}{1 - \pi_0(u_i, v_i)} \right)^{1/(k-1)} \right]^{1-k}.$$

These results are driven by, first, imposing the FWER constraint (5) to obtain θ^* , then, applying model (4) to achieve $\Theta(u_i, v_i)$ with a note that $f_{alt}^{-1}(x) = (x/k)^{1/(k-1)}$.¹¹ Funds that are more likely to be zero-alpha funds (i.e., have larger $\pi_0(u_i, v_i)$) receive more stringent rejection thresholds, whereas funds with more favourable characteristics receive more lenient thresholds. This reallocation improves the statistical power of the procedure to identify genuinely non-zero-alpha funds.

4. *Apply the rejection rule.* Finally, we apply the rejection rule to each fund. A fund i is deemed to have an alpha significantly different from zero if $p_i \leq \min\{\Theta(u_i, v_i), \lambda\}$.¹² Among these significant non-zero alpha funds, select all those with $\hat{\alpha}_i > 0$ as the final set of outperforming funds. The additional condition $p_i \leq \lambda$ ensures that funds with large p -values, those likely to be generated under the null, are excluded from the selection.

3. Data

Our hedge fund data is collected from the Lipper TASS and Hedge Fund Research (HFR) databases. We follow previous studies to merge both datasets and remove duplicate funds

¹¹Specifically, by replacing $\pi_0(u_i, v_i)$ outside the f_{alt}^{-1} with $\mathbf{1}_{p_i > \lambda}/(1-\lambda) = Y_i/(1-\lambda)$ and using $f_{alt}^{-1}(x) = (x/k)^{1/(k-1)}$, the constraint in (4) becomes $\sum_{i=1}^n \frac{Y_i}{1-\lambda} \left(\frac{\pi_0(u_i, v_i)\theta}{[1-\pi_0(u_i, v_i)]k} \right)^{1/(k-1)} \leq \tau$. By isolating common terms and rearranging, we obtain $\theta \geq k \left[\frac{1}{\tau(1-\lambda)} \sum_{i=1}^n Y_i \left(\frac{\pi_0(u_i, v_i)}{1-\pi_0(u_i, v_i)} \right)^{1/(k-1)} \right]^{1-k}$. This yields the expression of θ^* .

¹²The condition $p_i \leq \min\{\Theta(u_i, v_i), \lambda\}$ is used instead of $p_i \leq \Theta(u_i, v_i)$ because funds with $p_i > \lambda$ are classified as having zero alpha.

existing in both data providers (see Joenväärä et al., 2021; Coutts et al., 2024).¹³ Following previous research, we impose several screenings to deal with common sample biases (see, Fung and Hsieh, 2001; Bali et al., 2012; Chen et al., 2021; Coutts et al., 2024; Chen et al., 2025). We include only US-dollar-based hedge funds in our sample to avoid duplicate funds listed in different currencies. We include both *live* and *graveyard* funds from January 1994 to account for survivorship bias. To address the backfill bias, for each fund, we retain only observations starting from the month the fund is added to the database. For each fund, we consider only months where the fund’s net-of-fee return and asset under management (AUM) data are available. Following the literature, we consider only funds’ returns once their AUM first reaches \$5 million in the sample and require all funds to have at least 36 months of return history (see, e.g., Coutts et al., 2024). Finally, we consider only those funds whose category is defined. These categories include: Convertible Arbitrage, Dedicated Short Bias, Emerging Markets, Equity Market Neutral, Event Driven, Fixed Income Arbitrage, Global Macro, Long/Short Equity Hedge, Managed Futures, Multi-Strategy, Fund of Funds, and Relative Value. After applying the above restrictions, we obtain a sample of 7,241 funds spanning January 1997 to December 2023.

4. Performance Metrics

Following the majority of the existing literature on hedge fund performance, we use the Fung and Hsieh (2004) seven-factor model alpha as our baseline performance measure (see, e.g., Kosowski et al. (2007) and Chen et al. (2017)). For each fund i we regress

$$r_{i,t} = \alpha_i + \mathbf{F}_t \boldsymbol{\beta}_i + \varepsilon_{i,t} \quad (7)$$

where $r_{i,t}$ is the excess return in month t (i.e., simple return minus risk-free rate). Given that hedge funds often invest in illiquid assets, their reported returns tend to be highly serially

¹³We apply fuzzy-matching to fund names across the two databases, and we declare a pair of funds as duplicates if their correlation of returns exceeds 99%. Finally, we retain only the fund with the longest series of returns and assets under management (AUM) for each duplicate pair.

correlated.¹⁴ Therefore, we calculate and test funds' alpha using their unsmoothed returns, obtained using the procedure proposed by [Getmansky et al. \(2004\)](#).¹⁵ This standard practice is widely adopted in the relevant literature.¹⁶

$\mathbf{F}_t = [F_t^1, F_t^2, \dots, F_t^7]$ is the 1×7 matrix of the seven risk factors, $\boldsymbol{\beta}_i = [\beta_i^1, \beta_i^2, \dots, \beta_i^7]'$ 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 6.7, we additionally conduct robustness checks 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. \(2025\)](#). Those include the four-factor model of [Carhart](#)

¹⁴Other sources can also lead to smoothed returns, such as misreporting ([Bollen and Pool, 2008, 2009](#)) or the intention to earn higher fees ([Agarwal et al., 2011](#)).

¹⁵[Getmansky et al. \(2004\)](#) assume that the observed return of fund i in period t , denoted by $r_{i,t}^*$, is a weighted average of the fund's unobserved true economic returns over the most recent $k + 1$ periods:

$$r_{i,t}^* = w_0 R_{i,t} + w_1 R_{i,t-1} + \dots + w_k R_{i,t-k}$$

where $R_{i,t}$ denotes the true economic return in period t , the weights satisfy $w_j \in [0, 1]$ for all $j = 0, \dots, k$, and $\sum_{j=0}^k w_j = 1$. [Getmansky et al. \(2004\)](#) model the economic return as $R_{i,t} = \mu_i + \eta_{i,t}$ where μ_i is the average return of the fund and $\eta_{i,t} \stackrel{\text{iid}}{\sim} N(0, \sigma_\eta^2)$. This implies

$$r_{i,t}^* = \mu_i + w_0 \eta_{i,t} + w_1 \eta_{i,t-1} + \dots + w_k \eta_{i,t-k} = \mu_i + w'_1 \epsilon_{i,t-1} + \dots + w'_k \epsilon_{i,t-k} + \epsilon_t$$

which is a moving average process of order k where $\epsilon_{i,t'} = w_0 \eta_{i,t'}$ for $t' = t, \dots, t - k$, and $w'_j = w_j / w_0$ for $j = 1, \dots, k$. Following [Couts et al. \(2024\)](#), we conservatively set $k = 3$ and estimate the moving average process to obtain estimates of the mean return μ_i , weights $w'_j, j = 1, \dots, k$, and residual $\epsilon_{i,t}$. We recover w_0 by using $\sum_{j=1}^k w'_j = (1 - w_0) / w_0$ and then extract the residual $\eta_{i,t} = \epsilon_{i,t} / w_0$. Finally, the unsmoothed (economic) return is then recovered as $R_{i,t} = \mu_i + \eta_{i,t}$. In our regression (7), $r_{i,t} = R_{i,t} - r_{f,t}$ where $r_{f,t}$ is the one-month Treasury bill rate in month t .

¹⁶See, e.g., [Kosowski et al. \(2007\)](#), [Fung et al. \(2008\)](#), [Li et al. \(2016\)](#), [Titman and Tiu \(2011\)](#), [Agarwal et al. \(2018\)](#), and [Wu et al. \(2021\)](#), among others.

(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, and the momentum factor.¹⁷ We also consider the nine-factor model of Chen et al. (2025), which outperforms previous multi-factor models in explaining hedge fund returns. Their multi-factor model includes the market factor (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) ($TERM$).¹⁸

5. Simulations

In this Section, we show that: i) our proposed approach works well in terms of controlling FWER and outperforms the existing methods in terms of power; ii) the performance of $fwer^+$ is robust across variations in key simulation parameters.

5.1. Simulation Settings

As presented in Section 2.1, FWER is estimated using a mixture model in which the informativeness of the covariates is captured via only the null proportion function $\pi_0(\mathbf{z})$. Therefore, the simulation design will be simple as follows. The value of $\mathbf{Z}$ will be generated

¹⁷We follow David Hsieh’s website to collect the seven risk factors of the Fung and Hsieh (2004) factor 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.

¹⁸We collect the nine risk factors of the Chen et al. (2025) factor model from 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.

in the first step. Based on these values and the assumption on the form of $\pi_0(\mathbf{z})$, we assign the true nulls (i.e., determine which funds have zero-alpha). The remaining funds will be assigned non-zero alpha values. Under our framework, the signals of being false null (i.e., the magnitude of the non-zero alphas) are transformed to an estimation of FWER via the density function of the non-zero alpha funds' p -value, f_{alt} .

To determine realistic values for the parameters, we use data from all 7,241 funds in our sample and utilize the seven risk factors data, which start from January 1997 to December 2023. We calculate the factor loadings (β_i) for the funds by regressing each of the funds on the risk factors. We estimate their mean and correlation coefficient matrix. We generate balanced panel data of n funds with T observations per fund.

In each iteration, we conduct seven steps as follows.

1. We generate two covariates $\mathbf{Z} = (U, V)$ from standard normal distributions. We investigate two cases: independent covariates and correlated covariates with a correlation coefficient $\rho = 0.5$.
2. Consistent with the $fwer^+$ framework, we assume $\pi_0(u, v)$ having a logit form $\frac{1}{1+e^{-b_0-b_1u-b_2v}}$ where the triplet (b_0, b_1, b_2) is one of the three cases $(2, 1, 1)$, $(2.4, 1.5, 1.5)$ and $(2.8, 2, 2)$. These three cases of (b_1, b_2) cover a weak, moderate and strong relationship between covariates and the probability of a fund having a zero-alpha, respectively. Given a specific choice of $\pi_0(u, v)$, for each fund i , we draw a random value from a Bernoulli distribution which takes the value 0 with a probability of $\pi_0(u_i, v_i)$. Funds with drawn values of 0 are assigned as zero-alpha funds, which is around 81% of the population. We randomly assign the remaining funds with alpha of $\alpha > 0$ and $-\alpha$ with proportion of 13% and 6%, respectively. We consider the monthly alpha $\alpha \in \{0.5, 1.0, 1.5\}\%$.¹⁹

¹⁹To set these values, we first estimate the proportions of zero (π_0), positive (π^+), and negative (π^-) alpha funds using the approach of [Barras et al. \(2010\)](#). The method gives us $\pi_0 \approx 81\%$, $\pi^+ \approx 13\%$, and $\pi^- \approx 6\%$. Depending on b_1 and b_2 , the choice of b_0 aims to generate $\pi_0 \approx 81\%$. Our conclusions remain unchanged for different choices of π_0 , π^+ , and π^- . As the empirical monthly alpha distribution (in %) has

3. We generate the risk factors, $\mathbf{F}^s$ and their loadings $\boldsymbol{\beta}^s$ from normal distributions such that their parameters are the same as those of the real sample counterpart in the whole sample period (i.e., from January 1997 to December 2023).
4. We generate the simulated excess return of each fund via the following formula

$$r_{i,t} = \alpha_i^s + \mathbf{F}_t^s \boldsymbol{\beta}_i^s + \varepsilon_{i,t} \quad (8)$$

where the noise $\varepsilon_{i,t}$ is drawn independently from a normal distribution $N(0, \sigma^2)$ with σ to be set at 2.6% as the median of the 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, $\mathbf{F}_t^s$, to obtain its estimated $\hat{\alpha}$ and the p -value.
6. Using the estimated $\hat{\alpha}$ s, the corresponding p -values, and simulated covariates, we implement the $fwer^+$ procedure to detect truly positive alpha funds while controlling FWER at predetermined targets $\tau \in (0, 1)$. We benchmark our approach against the StepM procedure of [Romano and Wolf \(2005\)](#) and the StepSPA procedure of [Hsu et al. \(2010\)](#).²⁰ As our focus is on small FWER targets, we consider $\tau \in \{0.1\%, 1\%, 2\%, \dots, 10\%\}$.
7. By comparing the simulated α^s and the selected outperforming funds, we record the family-wise error (FWE), which takes the value 1 if there is at least one of the funds in the negative or zero α^s groups classified as outperformers. We also calculate the detected proportion, which is the ratio of truly outperforming funds detected by each procedure.

We repeat the steps 1 to 7 across 1000 iterations and calculate the actual FWER as the ratio

an 87th percentile of 0.5, and the average alpha for the top 13% of funds is 0.9, it is realistic to simulate with truly positive alphas of 0.5, 1.0, and 1.5.

²⁰Details of the StepM and StepSPA procedures within our framework are provided in Section IA of the Internet Appendix. [Romano and Wolf \(2005\)](#) and [Romano and Shaikh \(2006\)](#) apply the StepM procedure to relatively small sets of hedge funds (105 and 210, respectively) in earlier periods. [Hsu et al. \(2010\)](#) propose the StepSPA procedure, which extends the superior predictive ability (SPA) test of [Hansen \(2005\)](#) in an analogous manner that the StepM procedure extends the reality check of [White \(2000\)](#). [Hsu et al. \(2010\)](#) apply StepSPA to evaluate the predictive performance of technical trading rules.

of the number of times we observe $\text{FWE} = 1$ over 1000 and the power as the average of the detected proportion recorded in the step 7 above.

5.2. A Comparison to Existing Methods

To compare the performance of the $fwer^+$ procedure with existing methods, including the StepM and StepSPA tests, we consider a specific simulated data setting with $n = 1000$ funds and moderately informative covariates.²¹ We first present simulation results for $T = 36$, with α varying in $\{0.5, 1.0, 1.5\}\%$.

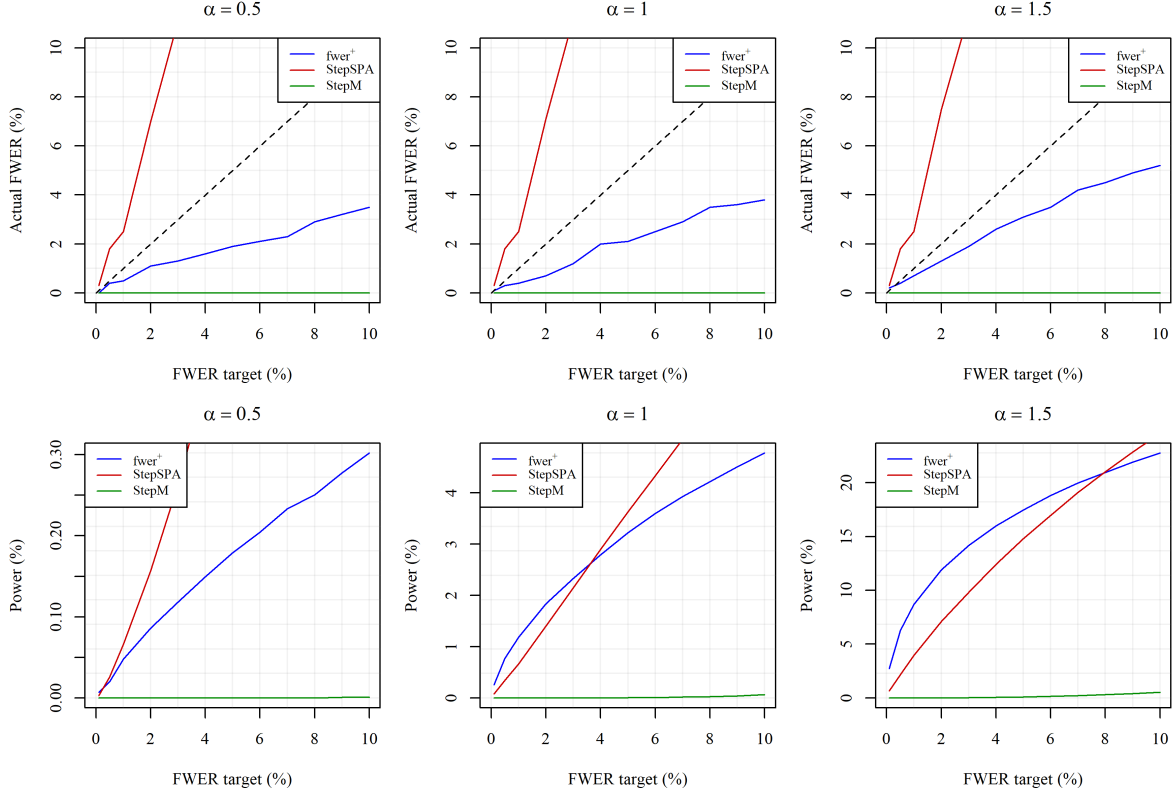
The performance of the procedures is shown in Figure 1, where all numbers are reported in percentage terms. In each of the top three sub-figures, we display the actual FWER for a range of targets, with each curve representing a different procedure. At a given target level, a procedure is said to control FWER well if its corresponding point lies on or below the dashed 45° line. We observe that both the $fwer^+$ and StepM procedures consistently control FWER across different alpha magnitudes, whereas the StepSPA test loses control for all considered FWER targets.

We present the power in the bottom three sub-figures. Provided that FWER is well controlled, the $fwer^+$ procedure outperforms the other two methods in all cases, with the performance gap depending on the FWER target. For instance, at target $\tau = 5\%$ and $\alpha = 1.5\%$, the power advantage of $fwer^+$ over StepM is 16%. We note that StepSPA does not control FWER, and StepM has power close to 0. Our findings explain the absence of prior studies in hedge fund or mutual fund selection based on FWER.²²

²¹Our main conclusions remain unchanged when using weakly or strongly informative covariates. In the StepM and StepSPA implementations, we adopt the stationary bootstrap procedure of Politis and Romano (1994), with an average block length of 6 months to preserve the realistic dependence structure in hedge fund returns. In the Internet Appendix IB, we report results for average block lengths ranging from 2 to 12 months. Our conclusions remain unchanged.

²²Unlike FDR methods, which aim for a looser Type I control, the power of FWER methods tends to be lower, especially when the number of funds is large and the truly positive alpha is small. Note that with $n = 1000$ funds in our simulation setting, power gaps of 6% to 16% translate into 60 to 160 more truly positive alpha hedge funds being detected.

Figure 1: Performance comparison. The figure compares the $fwer^+$ procedure with existing methods, including the StepM of [Romano and Wolf \(2005\)](#) and the StepSPA of [Hsu et al. \(2010\)](#), in terms of FWER control (top three sub-figures) and power (bottom three sub-figures). The simulated data consist of balanced panels of 1000 funds, each with 36 monthly observations. StepM and StepSPA are implemented using the stationary bootstrap with an average block length of six months. All values are reported in percentage terms. From left to right, α takes values 0.5, 1.0, and 1.5 (in %). The input covariates U and V used in the $fwer^+$ procedure are moderately informative and independent.



5.3. Performance of $fwer^+$ under Different Signal Strengths

In empirical experiments, the characteristics of the input data can vary over time. For instance, in constructing our portfolio presented in Section 6, we assess fund performance using a rolling window of 36 months. Both the magnitude of non-zero alphas and the informativeness of covariates can change across different periods. To understand the impact of these factors on the performance of the $fwer^+$ procedure, this Section varies the alpha level of the non-zero-alpha funds and the informativeness of the covariates.

Specifically, we fix the number of funds at $n = 1000$ and the time window at $T = 36$, which reflects the representative IS configuration in our baseline empirical analysis. We then vary α from 0.5% to 1.5%, and consider covariates with weak, moderate, and strong relationships to the proportion of zero-alpha funds.

Figure 2 reports the performance of the $fwer^+$ procedure under both independent and correlated covariate settings. The top three sub-figures show the actual FWER at the given targets, while the bottom three depict power. From left to right, each subfigure corresponds to a different non-zero alpha magnitude: 0.5%, 1%, and 1.5%.

In each of the top three sub-figures, the red (green and blue) solid curves represent the actual FWER of the $fwer^+$ procedure under weakly (moderately and strongly) informative and independent covariates. The dotted curve of the same color corresponds to the same informativeness level under the correlated covariate setting (i.e., the two covariates are correlated with a correlation coefficient $\rho = 0.5$). It is evident that the $fwer^+$ procedure effectively controls FWER at all target levels, regardless of covariate dependence, as all curve points lie on or below the 45° line.

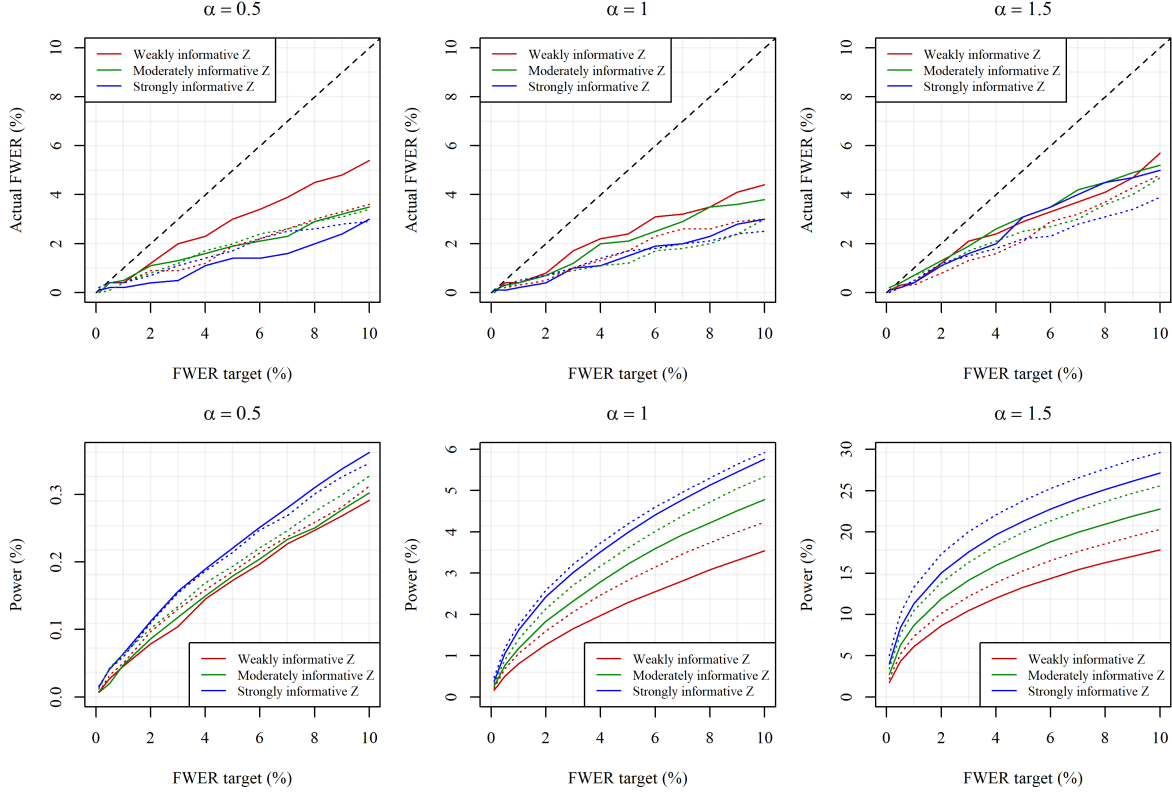
The bottom three sub-figures display the power. We observe that the greater the informativeness of the covariates, the higher the power of the $fwer^+$ procedure. Moving from the left to the right sub-figures, the magnitude of alpha among the outperforming funds increases, and accordingly, the power of $fwer^+$ also improves. This is not surprising, as outperforming funds become easier to detect when the signal is stronger.²³

5.4. Performance of $fwer^+$ under Varying Covariates Strength

We have investigated the performance of $fwer^+$ using two covariates, $\mathbf{Z} = (U, V)$. In practice, however, the number of truly informative covariates is often unknown. As a result,

²³In our simulation setting, the true null probability $\pi_0(u, v)$ (and thus the p -value) is generated based on the values of u and v . Consequently, we cannot compare the power of the $fwer^+$ procedure across different (u, v) settings. Therefore, the powers of $fwer^+$ under independent and dependent covariate structures are not directly comparable.

Figure 2: Performance of $fwer^+$ under varying signal settings. The figure shows the impact of signals (i.e., the magnitude of true non-zero alphas and the informativeness of covariates) on the performance of $fwer^+$ in terms of FWER control in the top three sub-figures and power in the bottom three sub-figures. The simulated data consist of balanced panels of $n = 1000$ funds, each with $T = 36$ observations. The outperforming (underperforming) funds in the population have alphas of α ($-\alpha$), where α varies in $\{0.5, 1.0, 1.5\}\%$. We consider three settings of the covariates $\mathbf{Z} = (u, v)$ including weakly, moderately, and strongly informative. Covariates may be independent (solid curves) or correlated with a correlation coefficient of 0.5 (dotted curves).



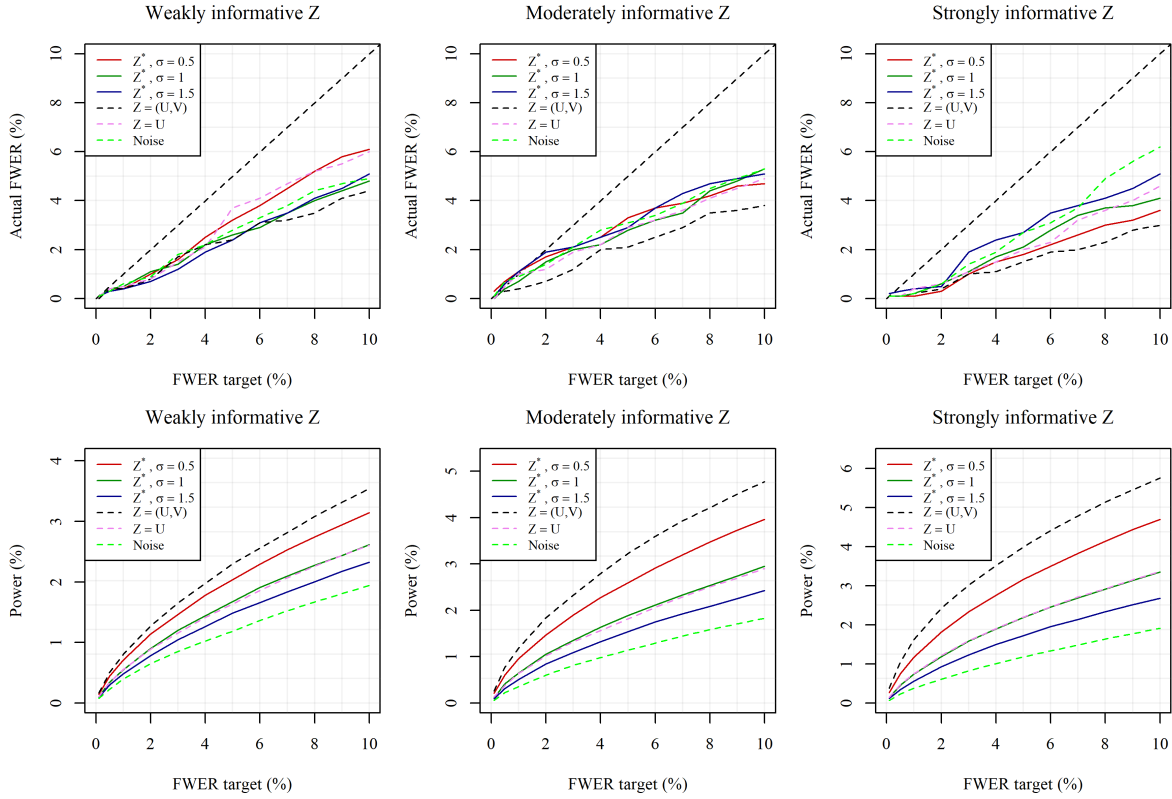
practitioners may use too few covariates, or include variables that are estimated with noise or, worse, completely unrelated to fund performance (i.e., non-informative covariates). In this Section, we assess the robustness of $fwer^+$ under these scenarios.

First, to model the case of insufficient covariate information, we implement the $fwer^+$ procedure using only one covariate, $Z_1 = U$. Second, to evaluate the impact of noisy covariates, we generate perturbed versions $\mathbf{Z}^* = (U + \eta, V + \zeta)$, where η and ζ are noise terms drawn from a normal distribution $N(0, \sigma^2)$. We vary the noise level by setting $\sigma \in \{0.5, 1.0, 1.5\}\%$.

Finally, for the non-informative covariate case, we use a single covariate that is pure noise, drawn independently from $N(0, 1)$.

Figure 3 presents the performance of $fwer^+$ under all these scenarios. For comparison, we also include the benchmark case using the original informative covariates $\mathbf{Z}$.

Figure 3: Performance of $fwer^+$ under use of insufficient information. The figure shows the impact of using insufficient covariates, covariates containing different levels of noise, and non-informative covariates on the performance of $fwer^+$ in terms of FWER control (top three sub-figures) and power (bottom three sub-figures). The simulated data are balanced panels of $n = 1000$ funds, each with $T = 36$ observations and a non-zero alpha of 1%. We consider three settings of the two independent covariates $\mathbf{Z} = (U, V)$, including weakly, moderately, and strongly informative. The simulated data are generated based on $\mathbf{Z}$ via $\pi_0(\mathbf{Z})$. In the noisy covariates cases, instead of using $\mathbf{Z}$, $fwer^+$ uses $\mathbf{Z}^* = (U + \eta, V + \zeta)$ where $\eta, \zeta \sim N(0, \sigma^2)$ and $\sigma \in \{0.5, 1.0, 1.5\}$. In the insufficient covariates case, the $fwer^+$ uses only $Z_1 = U$, while in the non-informative case it uses a single covariate drawn from $N(0, 1)$, unrelated to $\pi_0(\mathbf{Z})$. For comparison purposes, we include the performance of the $fwer^+$ using full information $\mathbf{Z}$.



We observe that in all scenarios, FWER is well controlled at all considered targets. The non-informative case suggests that it is safe, in terms of FWER control, to implement the

$fwer^+$ procedure even when an unrelated covariate is mistakenly included. Unsurprisingly, in terms of power, $fwer^+$ performs best when truly informative covariates are used, and worst when the covariate is irrelevant or non-informative. The power of $fwer^+$ with covariates estimated with noise lies between these two extremes and decreases as the noise level increases, i.e., with the magnitude of σ . This also implies that adding a non-informative covariate to an otherwise informative set may reduce the power of $fwer^+$.

5.5. Performance of $fwer^+$ under Varying Sample Size

Hitherto, we have investigated the performance of $fwer^+$ across scenarios regarding the informativeness of covariates and the strength of signals from outperforming funds. In this Section, we further examine the impact of sample size, i.e., the number of funds in the sample (cross-section) and the number of observations per fund (time-series), on the procedure's performance.

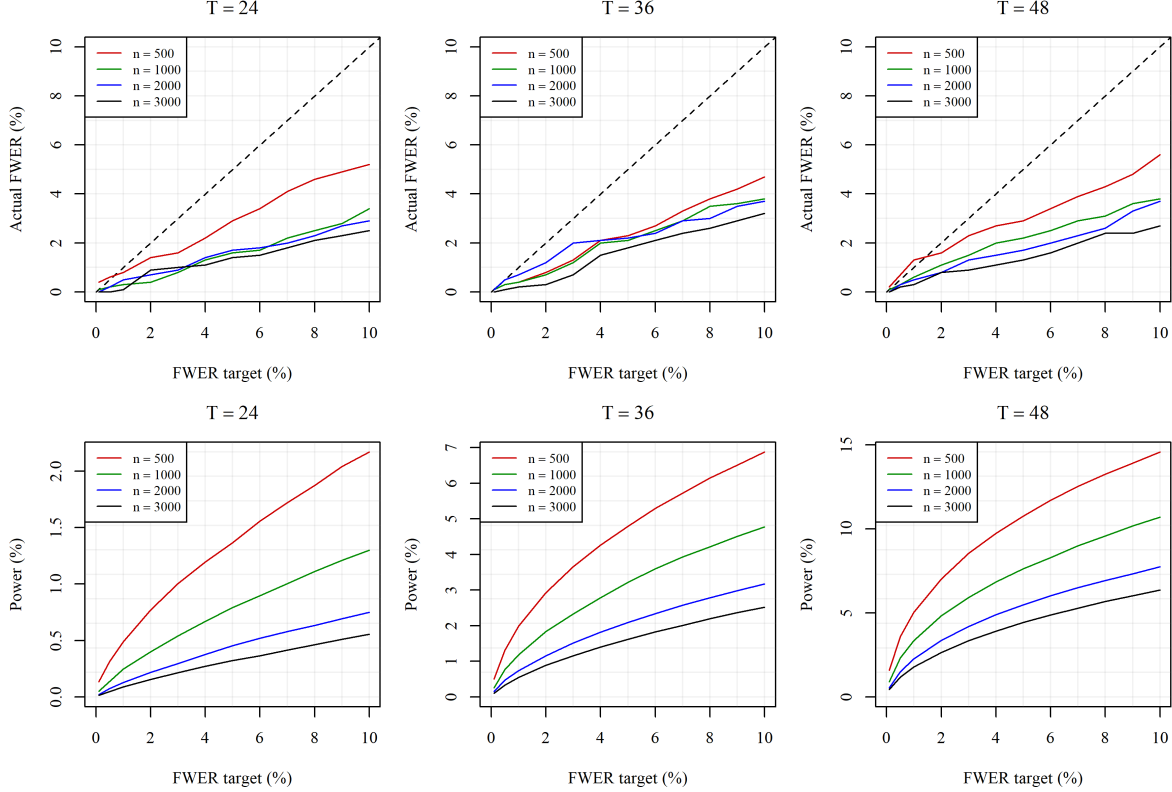
To that end, we consider balanced panel data by varying both the number of funds, n , and the number of observations per fund, T . As the IS horizons for portfolio selection in the hedge fund literature are typically 24, 36 or 48 months, we consider $T = 24, 36$, and 48. The value of n is also varied to cover the range of cases encountered in our empirical experiments, spanning from 500 to 3000.

For the sake of space, in this set of simulations, we report results for data generated under the setting of moderately informative and independent covariates, with $\alpha = 1\%$.²⁴ The results are depicted in Figure 4.

In the sub-figures of Figure 4, the number of observations per fund increases from left to right. In each sub-figure, we present the results in different colours corresponding to the number of funds: $n = 500, 1000, 2000$, and 3000. From the top three sub-figures, we again observe the excellent performance of $fwer^+$ in terms of FWER control. It is clear from

²⁴Our conclusions are robust to other settings, including dependent covariates and moderately or strongly informative covariates; the corresponding results are available upon request.

Figure 4: Varying sample size and number of observations. The figure shows the performance of $fwer^+$ under varying sample sizes (n) and numbers of observations per fund (T). The simulated data are balanced panels with T observations per fund, generated under moderately informative and independent covariates, and with a non-zero alpha of 1%. From left to right, the top three sub-figures show performance with respect to FWER control as T varies from 24 to 48. The bottom three sub-figures present the corresponding power.



the bottom three sub-figures that power gains are higher when the time series are longer (i.e., T increases from left to right). This is consistent with the fact that outperforming funds are easier to detect when they outperform over a longer period. Interestingly, $fwer^+$ is more powerful when the number of funds is smaller. This is a desirable property, as the procedure can be applied to a wide range of problems involving both small and large numbers of hypotheses.

5.6. Performance of $fwer^+$ under Cross-sectional Dependence in Covariates

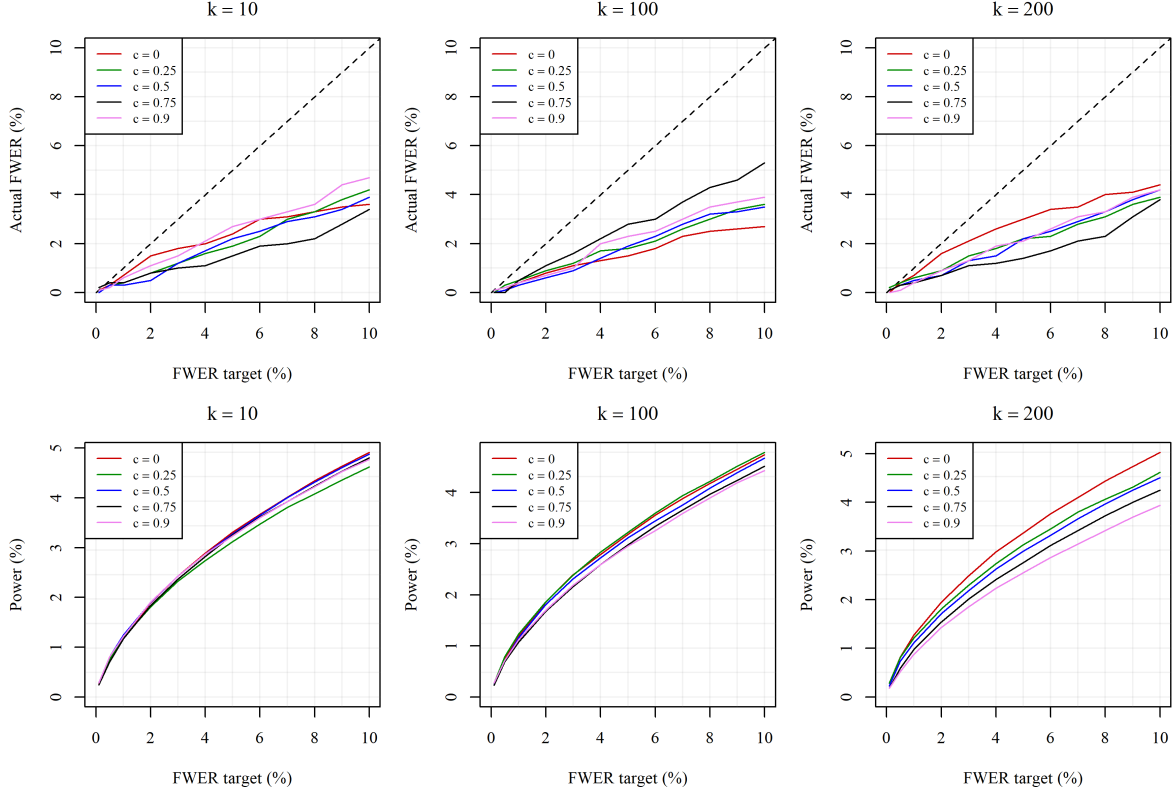
The FWER framework of Zhou et al. (2021) assumes that p -values and covariates are independently and identically distributed across tests. This assumption is strong and often violated in finance. Hedge funds may be exposed to common shocks, leading to correlated returns and covariates, particularly among funds with the same investment style. In this Section, we evaluate the performance of $fwer^+$ under correlated covariates.

More specifically, in step 1 of the simulation design, we partition 1,000 funds into groups of size k . Each covariate is drawn from a normal distribution such that: (i) within each group, the covariate values are correlated with a correlation coefficient c , and (ii) across groups, the covariate values are independent. We consider three cases for group size, $k \in \{10, 100, 200\}$, and five cases for correlation, $c \in \{0, 0.25, 0.5, 0.75, 0.9\}$. Larger values of k or c imply stronger dependence among tests. The remaining steps, from steps 2 to 7, follow the procedure described in Section 5.1. Figure 5 reports the results (FWER in the upper panels and power in the lower panels). We find that $fwer^+$ maintains control for all FWER targets and that dependence does not materially reduce the procedure's power.

5.7. Summary of Simulations

All our simulations demonstrate the excellent performance of $fwer^+$ in controlling FWER across various data scenarios. We observe that the procedure achieves higher power when any of the following conditions hold: (i) a stronger relationship between the covariates and the prior null probability; (ii) a larger magnitude of the outperforming funds' alpha; (iii) using a more comprehensive set of informative covariates; (iv) lower noise levels in the covariates; (v) the outperforming funds exhibit superior performance over a longer period (i.e., larger T); and (vi) a smaller number of funds in the population. Moreover, the correlation among covariates does not materially reduce the power of the $fwer^+$ procedure and its control at all FWER targets.

Figure 5: Cross-sectional dependence in covariates. The figure shows the performance of $fwer^+$ under cross-sectional dependence in covariates. Funds are partitioned into independent groups of size k . Within each group, the covariate is generated so that the values for the k funds are correlated with a coefficient c , while values across groups are independent. The simulated data are balanced panels with 36 observations per fund, generated under moderately informative covariates with cross-sectional dependence as described above, and with a non-zero alpha of 1%. From left to right, the top three sub-figures report FWER control as k varies from 10 to 200. In each sub-figure, each curve corresponds to a value of $c \in \{0, 0.25, 0.5, 0.75, 0.9\}$. The bottom three sub-figures show the corresponding power.



6. Empirical Analysis

In this Section, we apply the $fwer^+$ procedure to detect outperforming funds in the IS period and invest in those identified funds in OOS. We first describe the covariates under consideration and outline the construction of our $fwer^+$ -based portfolios. We then examine portfolio performance using individual and multiple covariates, compare it with passive benchmarks, and assess whether selected funds continue to perform well over longer holding

horizons. Finally, we decompose performance by fund style and assess robustness under alternative performance evaluation models.

6.1. *Covariates*

We create a comprehensive set of 24 hedge fund covariates documented in the literature as predictors of hedge fund performance. In particular, we focus on studies that provide statistically significant evidence that the selected covariates explain or predict hedge fund alphas.²⁵ We categorise our covariates into four groups, namely return-based and moments, fund characteristics, managerial skill, and macroeconomic covariates. Table 1 presents the exact covariates we consider.

Previous studies have shown the importance of trend movements such as cumulative returns, after unsmoothing returns, and higher moments and downside risk measures in predicting hedge fund returns due to the non-normality of hedge fund returns (see [Liang and Park, 2007](#); [Bali et al., 2012](#); [Wu et al., 2021](#)). For example, [Liang and Park \(2007\)](#) document that downside risk measures reveal information on the cross-sectional variation of hedge fund returns. [Wu et al. \(2021\)](#) also find that the hedge fund returns' kurtosis is an important moment in predicting hedge fund returns. Panel A of Table 1 includes covariates calculated from funds' past returns, such as cumulative return, standard deviation (second moment), skewness (third moment), kurtosis (fourth moment), and the Sharpe ratio, all based on sample funds' past 12-month returns.

Panel B of the same table contains covariates based on fund-specific characteristics, such as the logarithm of assets under management, age, incentive fee, and management fee. Earlier studies have highlighted the importance of hedge fund characteristics in explaining cross-sectional variation in fund returns and, by extension, fund performance (see [Ackermann](#)

²⁵We incorporate covariates regardless of the approach used to establish their significance (e.g., time series, Fama and MacBeth regression, etc.), as well as whether that significance was established in raw returns, factor model alphas, or returns controlled for characteristic exposure.

Table 1: Hedge fund return covariates. The table reports the 24 covariates used by $fwer^+$ to select outperforming hedge funds. We categorise them into four different groups, namely, *return-based* (Panel A), *fund characteristics* (Panel B), *managerial skill* (Panel C), and *macroeconomic covariates* (Panel D).

Covariate	Note
<u>Panel A: Return-based Covariates</u>	
Cumulative return	Cumulative return over the past 12 months
Return Volatility	Standard deviation of returns over the past 12 months
Skewness	Skewness of returns over the past 12 months
Kurtosis	Kurtosis of returns over the past 12 months
Sharpe ratio	Sharpe ratio over the past 12 months
<u>Panel B: Fund Characteristics</u>	
Fund size	Logarithm of the fund’s assets under management
Fund Age	Age of fund in months
Incentive fee	Fund’s incentive fee
Management fee	Fund’s management fee
Minimum investment	Fund’s minimum investment
Lockup period	Fund’s lockup period
High-water mark	Fund’s high-water mark
Redemption notice period	Fund’s redemption notice period
<u>Panel C: Managerial Skill Covariates</u>	
Rsquare	Rsquare of the testing model over the past 36 months
Idiosyncratic risk	Idiosyncratic risk (IVOL) based on the testing model
Six-factor Idiosyncratic risk	Equity six-factor model’s IVOL of Bali and Weigert (2024)
<u>Panel D: Beta Coefficient on Macroeconomic Covariates</u>	
Sentiment Change	Investor sentiment change of Baker and Wurgler (2006) (beta coeff.)
Macroeconomic risk	Macroeconomic risk factor of Bali et al. (2014) (beta coeff.)
Liquidity timing	Market liquidity timing (loading coeff. of squared liquidity)
Inflation	Inflation (beta coeff.)
Term spread	Term spread (beta coeff.)
Default yield spread	Default yield spread (beta coeff.)
Equity Market Uncertainty	Equity Market Uncertainty index (beta coeff.)
Economic Policy Uncertainty	Economic Policy Uncertainty index (beta coeff.)

[et al., 1999](#); [Aragon, 2007](#); [Boyson, 2008](#); [Agarwal et al., 2009](#)). Such explanatory power remains robust even after controlling for other variables (see [Bali et al., 2014, 2021](#); [Bali and Weigert, 2024](#)). For instance, [Boyson \(2008\)](#) reports that hedge fund performance is more persistent 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) of funds is directly linked to their performance.

Covariates based on factor models are also informative on managerial skill. For example, [Titman and Tiu \(2011\)](#) show that the R^2 of a factor model measures managerial skill because less skilled managers will choose greater exposure to systematic factors (i.e., higher R^2) to generate alphas, whereas managers with high skill (i.e., lower R^2) do not. Similarly, highly skilled fund managers can successfully trade idiosyncratic volatility stocks and profit from the positive relation between idiosyncratic volatility and futures returns (see [Bali and Weigert, 2024](#)). For that reason, we employ the R^2 and the idiosyncratic risk of the factor model used as a benchmark for measuring alpha, which is [Fung and Hsieh \(2004\)](#) in our baseline analysis, and the idiosyncratic risk based on the [Fama and French \(2015\)](#) five-factor model plus the momentum factor of [Carhart \(1997\)](#), as reported in Panel C. The idiosyncratic risk is the unbiased estimate of the variance of the model’s error terms.²⁶

Finally, we consider hedge fund loadings on macroeconomic variables as covariates. Hedge funds aggressively seek abnormal returns on risk assets driven by changes in economic conditions. Hence, fund performance is affected by the extent to which funds vary their exposure to economic indicators such as inflation, interest rates, and economic uncertainty. Those macroeconomic variables and/or funds’ loadings on them have been extensively reported to explain hedge fund returns and alphas over the past years (see [Bali et al., 2011](#); [Avramov et al., 2013](#); [Cao et al., 2013](#); [Bali et al., 2014](#); [Chen et al., 2021](#); [Wu et al., 2021](#); [Chen et al., 2025](#)).²⁷ Panel D of Table 1 reports the macroeconomic indicators and their correspond-

²⁶We employ the [Fung and Hsieh \(2004\)](#) seven-factor model for our main analysis. For robustness, we also consider alternative models (see Section 6.7), including the [Carhart \(1997\)](#) four-factor, six-factor, and traditional nine-factor models (as used in [Bali et al., 2012](#)), as well as the recent nine-factor model proposed by [Chen et al. \(2025\)](#). For each model, we regress excess return on the factors and obtain R^2 of the regression.

²⁷For example, [Bali et al. \(2014\)](#) examine whether macroeconomic risk predicts the cross-sectional dispersion of hedge fund returns. They find that time-varying uncertainty betas of fund returns on different measures of macroeconomic risk can predict the cross-section of fund returns. Similarly, [Chen et al. \(2021\)](#) document a significant cross-sectional relation between hedge fund sentiment beta and fund future perfor-

ing fund loadings used as covariates for our study. We consider loadings on the sentiment changes index of [Baker and Wurgler \(2006\)](#) and the [Pástor and Stambaugh \(2003\)](#) squared liquidity measure (timing), and the beta coefficients of the following: the macroeconomic risk factor of [Bali et al. \(2014\)](#), the inflation, the term spread, the default yield spread, the Equity Market Uncertainty, and the Economic Policy Uncertainty index.²⁸

6.2. Portfolios of Outperforming Funds

We focus on the performance of a portfolio of selected hedge funds rather than on a single fund. At the end of each year from 1999 onward, we use the return data over the past 36 months for eligible funds to calculate their unsmoothed returns using the model of [Getmansky et al. \(2004\)](#). These unsmoothed returns over the past 36 months are used to compute the [Fung and Hsieh \(2004\)](#) seven-factor model alpha and its p -value using HAC standard errors of [Newey and West \(1987\)](#).²⁹ For covariates, except the R^2 obtained along with the alpha, we use the original return and other data fields to be consistent with the findings in the literature. All covariates are standardized. A fund is eligible if it has return data for all months of the IS period and complete data for all covariates considered at the time of portfolio formation.

mance using the loading coefficient of the sentiment changes index in the regression of fund returns on the benchmark model augmented by the sentiment changes index of [Baker and Wurgler \(2006\)](#).

²⁸We calculate the sentiment changes index following [Baker and Wurgler \(2006\)](#) using data available on Jeffrey Wurgler’s webpage (<https://pages.stern.nyu.edu/~jwurgler/>). We obtain the timing liquidity measure from Robert F. Stambaugh’s webpage (<https://finance.wharton.upenn.edu/~stambaugh/>). Other macro indicators are obtained from the Federal Reserve Bank of St. Louis (<https://fred.stlouisfed.org/>), Turan Bali’s webpage (<https://sites.google.com/a/georgetown.edu/turan-bali/>) and the economic policy uncertainty webpage (<https://www.policyuncertainty.com/>). The beta coefficients for macroeconomic covariates were estimated using 36-month rolling-window regressions of hedge fund returns on each macroeconomic covariate. Using the same estimation window, the loading on squared liquidity was estimated by regressing hedge fund returns on the liquidity measure and its square.

²⁹If we extend the data back to 1995, our sample includes only a few funds with all covariates at the end of 1997 and 1998. Therefore, we restrict our sample to start from the beginning of 1997 and begin constructing portfolios from the end of 1999. Following most studies in the hedge fund literature, we use the seven-factor model of [Fung and Hsieh \(2004\)](#) and adopt a 36-month IS horizon (see, e.g., [Cumming et al., 2012](#); [Chen et al., 2017](#); and [Chen et al., 2021](#)). We later conduct robustness checks using alternative factor models in Section 6.7.

We apply the $fwer^+$ procedure to detect outperforming funds while controlling FWER at a given target level τ . Following the notation in Section 2, let A_t^+ denote the set of funds selected by $fwer^+$ as outperformers at the end of year t . We construct both an equal-weighted and an alpha-weighted portfolio from this same selection set. In the equal-weighted portfolio, each selected fund receives weight $1/|A_t^+|$. In the alpha-weighted portfolio, capital is allocated in proportion to each selected fund's estimated IS alpha over the IS period of 36 months. To be specific, the weight for fund i is determined as

$$w_{i,t}^\alpha = \frac{\hat{\alpha}_{i,t}^{IS}}{\sum_{j \in A_t^+} \hat{\alpha}_{j,t}^{IS}}, \quad i \in A_t^+$$

where $\hat{\alpha}_{i,t}^{IS}$ denotes the IS estimated alpha for fund i . Note that, by the construction $\hat{\alpha}_{i,t}^{IS} > 0$ for all funds $i \in A_t^+$. Thus, the equal-weighted and alpha-weighted portfolios differ only in their allocation rule, not in the $fwer^+$ selection step.

The returns of the resulting portfolios in the following year (OOS) are recorded. If no funds are selected, we invest in the one-month Treasury bill to earn the risk-free rate. If a selected fund stops reporting during the OOS year, its weight is redistributed among the remaining funds in the portfolio according to the corresponding portfolio-weighting rule.³⁰

Our portfolios are formed on a rolling yearly basis. The first OOS period is the year 2000, during which we invest in the funds selected based on IS data from January 1997 to December 1999. The final OOS period is 2023, using fund selections based on IS data from January 2020 to December 2022. We construct the $FWER\tau$ portfolios using the $fwer^+$ procedure with FWER controlled at the target level τ . We consider a range of target levels, including two conventional choices $\tau \in \{1\%, 5\%\}$ as well as two more stringent levels, $\tau = 0.5\%$ and $\tau = 0.1\%$.

³⁰The OOS holding period varies in the literature: Kosowski et al. (2007) use 12 months while Chen et al. (2017) consider 3, 6, 9, 12, 24, and 36 months. In practice, a hedge fund is a long-term investment vehicle and typically imposes a lock-up period. In our data, the lock-up period spans from 0 to several years with an average of 3.5 months. Accordingly, we first adopt a one-year holding period and later extend it to 2, 3, and 4 years (see Section 6.5) to illustrate the practical relevance of our portfolios.

We evaluate portfolio performance using unsmoothed portfolio returns. We first compute the original portfolio return as the weighted average of constituent fund returns, using either equal weights or the past alpha weights defined above. We then apply the unsmoothing procedure of [Getmansky et al. \(2004\)](#) to the resulting portfolio return series.³¹

6.3. $FWER\tau$ Portfolios and Covariates

This Section presents our experiments in which we construct a set of $FWER\tau$ portfolios using the mentioned covariates. We first construct portfolios using only a single covariate. As such, we have 24 portfolios for each target τ . For each portfolio, we calculate the annualized OOS seven-factor model alpha, the related t-statistic, portfolio size, and the proportion of months in which the portfolio is empty. In the interest of space, we report in Panel A of Table 2 a summary of these metrics across each of the FWER targets. The $FWER\tau$ portfolios yield alphas ranging from 2.4% to 6.8%, with mean values ranging from 4.0% to 4.2%, depending on the FWER target.

Thus far, we have implemented $fwer^+$ using only a single covariate. Different covariates might capture various aspects of fund managers' skills. Utilizing multiple covariates can help $fwer^+$ perform better as the information set is more comprehensive. In Panel B of Table 2, we present the performance of the equal-weighted $FWER\tau$ portfolio using all 24 considered covariates as inputs. We find that the new portfolio contains roughly twice as many funds compared to the $FWER\tau$ portfolios with individual covariates. As a result, it is a non-empty portfolio in every year, even with control at a very low error level. This allows investors to pick funds with high confidence. More importantly, its alpha is substantially higher than the average alpha of the $FWER\tau$ portfolios based on individual covariates, ranging from 5.9% to 6.4% with larger gains for portfolios controlling for more stringent FWER levels.

³¹The results using the original returns of the selected funds show similar portfolio alpha gains and are available upon request.

Table 2: OOS performance for the $FWER\tau$ portfolios compared to benchmarks. This table reports the OOS performance of portfolios constructed from funds selected by the $fwer^+$ procedure. Panel A reports the minimum, average, and maximum values across the 24 individual-covariate $FWER\tau$ portfolios for alpha and its Newey–West t -statistics, average excess return, and portfolio sizes across FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$. Panels B and C report portfolio performance for equal-weighted and alpha-weighted $FWER\tau$ portfolios constructed using all 24 covariates simultaneously. As benchmarks, Panel D reports the same performance metrics for the passive equal-weighted portfolio of all eligible funds (EW), the equal-weighted portfolio of funds with positive IS alpha (EW^+), and the alpha-weighted version of EW^+ (αW^+). Portfolio alphas are estimated using the [Fung and Hsieh \(2004\)](#) seven-factor model. Average excess returns and alphas are annualized.

τ (%)	FH Alpha (%)			t-statistic			Excess Return (%)			Portfolio Size			
	Average	Min	Max	Average	Min	Max	Average	Min	Max	Average	Min	Max	Empty Rate (%)
Panel A: Summary performance of $FWER\tau$ portfolios with individual covariates													
0.1	4.0	2.4	6.7	5.1	1.8	6.4	4.5	3.4	7	13	0	51	2.2
0.5	4.2	3.2	6.8	4.9	1.2	6.1	4.7	3.9	7	17	0	63	1.7
1.0	4.2	3.3	6.8	4.8	1.3	5.9	4.7	3.9	7	19	0	67	1.6
5.0	4.0	3.0	6.7	4.4	1.2	5.2	4.6	3.8	6	26	0	84	1.4
τ (%)	FH Alpha (%)			t-statistic			Excess Return (%)			Portfolio size			
	Average	Min	Max	Average	Min	Max	Average	Min	Max	Average	Min	Max	Empty Rate (%)
Panel B: Equal-weighted $FWER\tau$ portfolios using all covariates													
0.1	6.4			6.7			6.9			27	2	71	0
0.5	6.3			6.9			6.8			36	2	95	0
1.0	6.1			6.8			6.7			41	2	117	0
5.0	5.9			6.8			6.5			56	2	172	0
Panel C: Alpha-weighted $FWER\tau$ portfolios													
0.1	9.0			5.0			9.7			27	2	71	0
0.5	8.7			5.2			9.4			36	2	95	0
1.0	8.6			5.1			9.3			41	2	117	0
5.0	8.3			5.5			9.1			56	2	172	0
Panel D: Benchmark portfolios													
EW	0.2			0.2			3.4			1331	398	1961	0
EW^+	0.8			0.9			3.4			832	224	1487	0
αW^+	1.6			1.1			4.5			832	224	1487	0

As mentioned, we additionally construct an alpha-weighted portfolio based on the in-sample alpha of the selected funds. The performance of these portfolios is reported in Panel C of the same table. Relative to the equal-weighted $FWER\tau$ portfolios presented in Panel B, the alpha-weighted portfolios exhibit even stronger performance, with alpha improvements ranging from 8.3% to 9.0%.

As the existing FWER procedures either fail to adequately control the FWER or suffer from low statistical power, we benchmark the empirical performance of the $fwcr^+$ portfolios against equal- or alpha- weighted portfolios of all eligible funds, i.e., those having a complete 36-month return history over the past 3 years at the time of portfolio formation and all considered covariates. These portfolios are constructed as follows. At the end of each year from 1999, the first (second) equally weighted portfolio, denoted by EW (EW^+), simply selects all those eligible funds (eligible funds having a positive alpha in the IS period) to invest in with equal weights in the following year. We repeat the procedure yearly until the end of 2022 to have a set of funds invested in 2023. We also construct an αW^+ portfolio that weights funds in EW^+ by their IS alpha. We report the performance of those portfolios (EW , EW^+ and αW^+) in Panel D of the same table. We find that all the $FWER\tau$ portfolios outperform EW , EW^+ and αW^+ . It should also be noted that the EW , EW^+ and αW^+ portfolios gain alphas that are not significantly different from zero.

We further investigate the performance of the $FWER\tau$ portfolios in sub-periods. We partition the entire OOS period, spanning 2000 to 2023, into three non-overlapping sub-periods, each lasting 8 years: 2000–2007, 2008–2015, and 2016–2023. We calculate the performance metrics for the portfolios in each sub-period and report them in Panels A-C of Table 3.

The table shows that the equal-weighted $FWER\tau$ portfolios generate positive alphas and positive excess returns across all sub-periods. The portfolio alphas are statistically significant for all FWER targets and sub-periods considered. In terms of alphas, the $FWER\tau$ portfolios perform best during the first sub-period, with annualized alphas ranging from 6.8% to 7.0%.

Table 3: OOS performance of the $FWER\tau$ portfolios using all considered covariates. This table reports the OOS performance of portfolios constructed from funds selected by $fwer^+$ using all 24 covariates across three non-overlapping sub-periods: 2000–2007 (Panel A), 2008–2015 (Panel B), and 2016–2023 (Panel C). For each weighting scheme, the reported statistics are the annualized Fung and Hsieh (2004) seven-factor alpha, the corresponding Newey–West t-statistic, and the annualized excess return. The final three columns report the average, minimum, and maximum number of funds held in the portfolio. Results are reported for FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$.

τ (%)	Equal-weighted $FWER\tau$ portfolio			Alpha-weighted $FWER\tau$ portfolio			Portfolio size		
	FH Alpha (%)	Excess Return (%)	t-stat	FH Alpha (%)	Excess Return (%)	t-stat	Average	Min	Max
<u>Panel A: Period 2000–2007</u>									
0.1	6.9	5.5	8.0	9.5	4.4	11.0	32	6	71
0.5	6.8	5.3	8.0	9.5	4.3	11.2	44	6	91
1.0	6.8	5.4	8.1	9.6	4.2	11.5	49	6	103
5.0	7.0	5.7	8.2	9.7	4.5	11.4	68	6	144
<u>Panel B: Period 2008–2015</u>									
0.1	5.4	4.3	5.5	5.8	2.3	5.5	27	2	69
0.5	5.5	4.4	5.6	5.7	2.5	5.2	37	2	95
1.0	5.5	4.0	5.7	5.7	2.4	5.3	43	2	117
5.0	5.3	3.8	5.5	5.7	2.7	5.4	61	2	172
<u>Panel C: Period 2016–2023</u>									
0.1	4.7	2.1	7.2	10.3	2.0	12.6	22	6	43
0.5	4.4	2.3	6.8	9.3	2.0	11.7	28	7	56
1.0	4.0	2.1	6.3	8.7	1.9	11.3	31	7	68
5.0	3.4	2.1	5.8	7.5	1.9	10.4	40	8	93

Although the performance of the $FWER\tau$ portfolios is uniformly deteriorating over time, in the most recent sub-period, they continue to deliver statistically significant alphas ranging from 3.4% to 4.7%, with annualized excess returns reaching up to 7.2%. When it comes to the alpha-weighted portfolio, the performance is even better across all sub-periods, with the seven-factor alpha generated being up to 10.3% per annum for the most recent period (i.e., 2016–2023). But most importantly, both portfolio formations show that the $FWER\tau$ selected funds maintain a positive performance even during the second sub-period (i.e., an annual seven-factor alpha of almost 6%), which includes the Global Financial Crisis period.

6.4. Alpha and Wealth Evolution

To examine how portfolio performance evolves over time, we compute expanding-window alpha estimates for each portfolio. Specifically, for each portfolio and each month from December 2004 onward, we estimate the seven-factor alpha using portfolio returns from January 2000 through that month. Each portfolio therefore generates a time series of estimated alphas, which we plot in Panel A of Figure 6. For comparison, we also report the alpha evolution of the passive EW , EW^+ , and αW^+ portfolios. The alphas of the $FWER\tau$ portfolios remain consistently above those of the passive benchmarks over most of the OOS period.

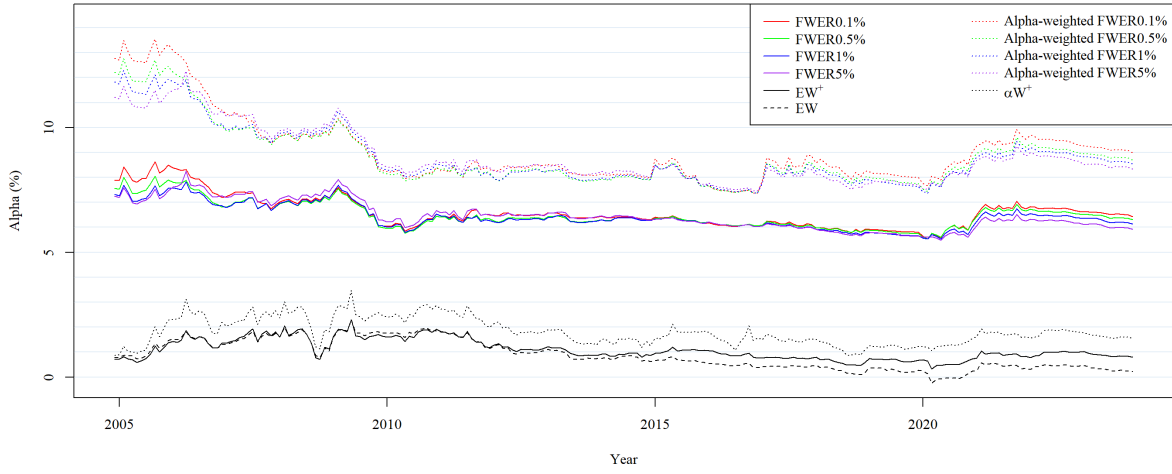
Panel B of Figure 6 shows the corresponding accumulated wealth from investing \$1 in each portfolio from January 2000 to December 2023. The equal-weighted $FWER\tau$ portfolios grow to approximately \$6.8–\$7.5, compared with roughly \$3 for the passive equal-weighted benchmarks. The alpha-weighted $FWER\tau$ portfolios perform even better, with accumulated wealth reaching approximately \$11.0–\$13.5, compared with about \$4 for the passive αW^+ benchmark. Notably, while the broader hedge fund industry suffered large losses during the 2008 financial crisis, the funds selected by $fwer^+$ avoided the downturn and generated substantial gains.³²

Regarding the hedge funds' lock-up periods, a couple of portfolios contain funds with lock-up periods of up to 3 years. More specifically, the proportions of the selected funds with lock-up periods greater than 1 year and 2 years are 5% and 2%, respectively. For this reason, to test the practical application of the method, we present in Section 6.5 the performance of the portfolio with holding horizons of up to 3 and 4 years.

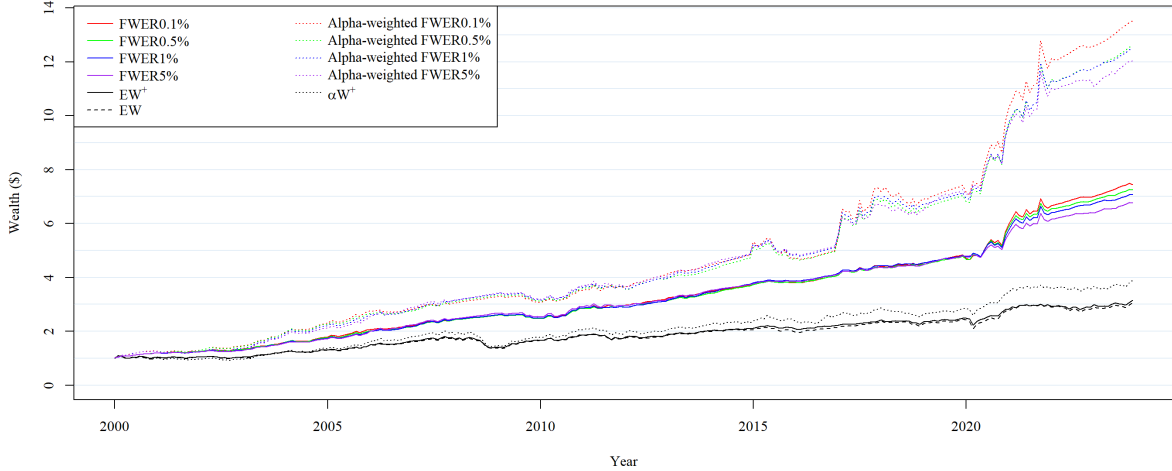
³²Specifically, while the equally weighted portfolio incurred a loss of -21% in 2008, the $FWER\tau$ portfolio with $\tau = 5\%$ achieved a gain of 8% . Over the entire crisis period, spanning July 2007 to June 2009, the corresponding figures are -11% and 11.5% , respectively.

Figure 6: Alpha and wealth evolution. This figure reports the evolution of OOS portfolio alphas and accumulated wealth for the portfolios reported in Table 2. The $FWER\tau$ portfolios are constructed using all 24 covariates and are reported for FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$. The figure includes both equal-weighted and alpha-weighted $FWER\tau$ portfolios, together with the passive benchmark portfolios: the equal-weighted portfolio of all eligible funds (EW), the equal-weighted portfolio of funds with positive IS alpha (EW^+), and the alpha-weighted version of EW^+ (αW^+). Panel A shows the evolution of portfolio alphas. For each portfolio and each month starting at the end of 2004, we estimate the seven-factor alpha using all available portfolio returns from January 2000 through that month. The first observation therefore corresponds to January 2000–December 2004, while the final observation uses data through December 2023. Panel B shows the corresponding accumulated wealth of a \$1 investment over the full OOS period from January 2000 to December 2023. Portfolio alphas are annualized.

Panel A: Alpha evolution



Panel B: Wealth evolution



6.5. Retention, Outperforming Rates and Longer Horizons

Each year, a large number of hedge funds are delisted from commercial databases (see [Van der Sluis and Posthuma, 2004](#)). [Hodder et al. \(2014\)](#) reveals that the delisted funds with positive prior performance have an insignificant difference in average returns from the listed funds over the next year. In contrast, delisted funds with negative prior performance have a mean delisting return of -5.67% in the following year. Delisted funds may affect the performance of the $fwer^+$ procedure presented in the previous Sections.

For this purpose, we consider the $FWER\tau$ portfolios using all covariates and compute the retention rate of the selected funds (or, otherwise, the funds that remained listed in our database) over the entire OOS holding period. We first calculate, for each of the 24 exercise years, the proportion of portfolio funds that report returns over the entire 12-month OOS holding period. We then average these 24 retention proportions to obtain the overall retention rate. We estimate the seven-factor model alpha using 12-month returns for each fully listed fund, then calculate the proportion of fully listed funds with positive alpha. Averaging proportions yields the OOS outperforming rate in the retained funds. We also calculate the OOS outperforming rate that accounts for all selected funds (regardless of listing status). The related findings are presented in [Table 4](#).

Table 4: OOS outperformance rate of selected funds. This table reports the OOS retention and outperforming rates of funds selected by the $fwer^+$ procedure using all 24 covariates. For each of the 24 portfolio formation years, the retention rate is defined as the proportion of selected funds that continue to report returns throughout the full 12-month OOS holding period. The full-sample retention rate is the average of these annual retention rates. For funds that remain listed throughout the OOS year, we estimate the seven-factor alpha using their 12 monthly OOS returns and compute the proportion with positive alpha. The OOS outperforming rate is the average of these annual proportions. The table also reports the OOS outperforming rate relative to the initial set of selected funds, which treats funds that stop reporting during the OOS year as non-outperformers.

τ (%)	Average number of selected funds	OOS retention rate (%)	OOS outperforming rate in retained funds (%)	OOS outperforming rate in selected funds (%)
0.1	28	92	84	79
0.5	37	93	81	76
1.0	43	92	80	75
5.0	58	93	78	73

We note that the proportion of delisted funds in our portfolios is lower than that reported in the literature, averaging only 8% (Hodder et al., 2014 report a 15.14% delisting rate of hedge funds between 1994 and 2009). This means that, on average, only 1 or 2 selected funds are delisted during our OOS periods. In addition, we note that the OOS outperformance rates for the retained and selected funds are close and sufficiently high. These observations, along with the reported small or insignificant penalty for delisted funds (see Hodder et al., 2014), lead us to argue that the effect of delisted funds, positive or negative, is negligible for our $fwer^+$ portfolios.

We now study whether the funds selected by $fwer^+$ can sustain their past performance over horizons longer than one year. Hedge funds are characterized by lock-up periods that restrict investors from redeeming capital over short horizons. To account for this, we extend the rebalancing interval to every two, three, or four years instead of forming a new set of outperformers annually. For example, in the 2-year holding case, we implement the $fwer^+$ procedure at the end of 1999 using data over the previous three years, 1997–1999, select significant outperforming funds, and hold them through 2000 and 2001. Similarly, we form a new set of funds to hold through 2002 and 2003 at the end of 2001, and so on, until reaching the end of 2021, when we select funds for 2022–2023. The procedure is analogous for the 3- and 4-year holding horizon cases.

The performance of $FWER\tau$ portfolios using all considered covariates is presented in Table 5. We note that all $FWER\tau$ portfolios generate positive alphas, and most of them are statistically significant, even when the holding horizon is extended to four years, which is longer than the IS period (i.e., three years) used to assess and select the outperforming funds. This illustrates $fwer^+$'s ability to identify funds whose performance persists over longer horizons, highlighting its practical relevance to investors.

Table 5: OOS performance of the $FWER\tau$ portfolios using all considered covariates with varying holding horizons. This table reports the OOS performance of portfolios constructed from funds selected by $fwer^+$ using all 24 covariates under longer holding horizons. Panels A, B and C report results for two-, three-, and four-year holding horizons, respectively. For each weighting scheme, the reported statistics are the annualized [Fung and Hsieh \(2004\)](#) seven-factor alpha, the corresponding Newey–West t-statistic, and the annualized excess return. The final three columns report the average, minimum, and maximum number of funds held in the portfolio. Results are reported for FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$.

τ (%)	Equal-weighted $FWER\tau$ portfolio			Alpha-weighted $FWER\tau$ portfolio			Portfolio size		
	FH Alpha (%)	Excess Return (%)	t-stat	FH Alpha (%)	Excess Return (%)	t-stat	Average	Min	Max
Panel A: 2-year holding period									
0.1	5.5	5.9	5.9	7.7	8.2	4.1	23	1	69
0.5	5.3	5.7	5.8	6.9	7.5	3.6	30	1	95
1.0	5.3	5.8	6.2	7.2	7.8	4.1	35	1	117
5.0	5.0	5.5	5.8	6.8	7.4	4.4	49	2	172
Panel B: 3-year holding period									
0.1	3.4	4.3	3.0	4.0	6.2	1.0	19	2	53
0.5	2.9	3.9	2.6	2.9	5.0	0.8	25	2	68
1.0	2.5	3.9	2.0	2.7	5.0	0.8	29	2	74
5.0	1.9	3.4	1.4	1.7	4.2	0.6	39	2	95
Panel C: 4-year holding period									
0.1	4.9	5.5	4.3	7.0	7.9	3.8	14	1	38
0.5	4.7	5.3	4.0	6.2	7.2	2.9	18	1	48
1.0	4.7	5.2	4.4	6.3	7.2	3.1	21	1	51
5.0	4.1	4.5	4.0	5.4	6.3	3.0	28	2	60

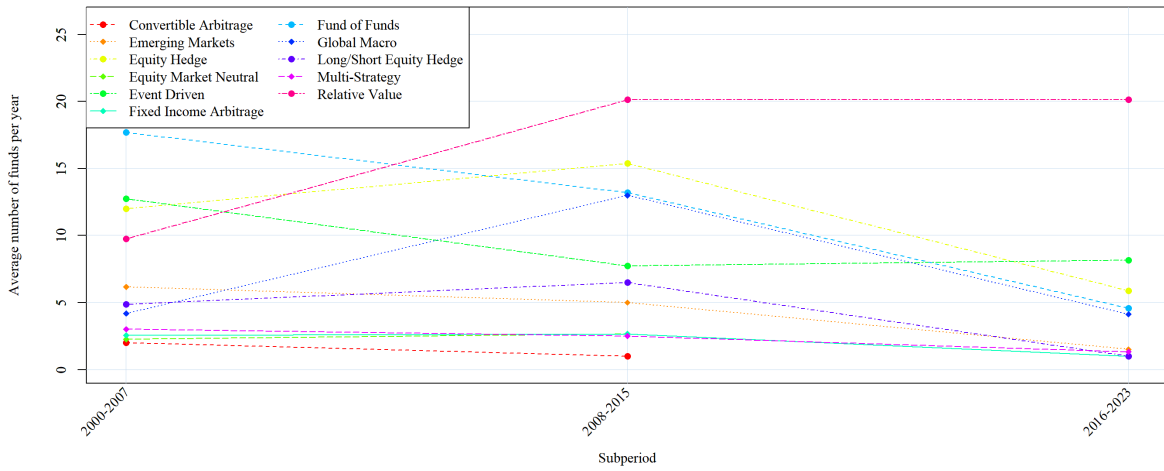
6.6. Portfolio Performance Decomposition by Fund Style

Next, we examine which fund styles are most frequently selected. For this purpose, we consider the $FWER5\%$ portfolio using all covariates. We calculate the average number of funds selected by $fwer^+$ per year for each style across different sub-periods. Panel A of [Figure 7](#) shows that Relative Value becomes the dominant selected style after 2009, despite ranking below several other styles in the first sub-period. Equity Hedge is also consistently selected across all sub-periods, but it is not selected as intensively as Relative Value in the latter part of the sample. Other fund styles frequently selected during our sample period are

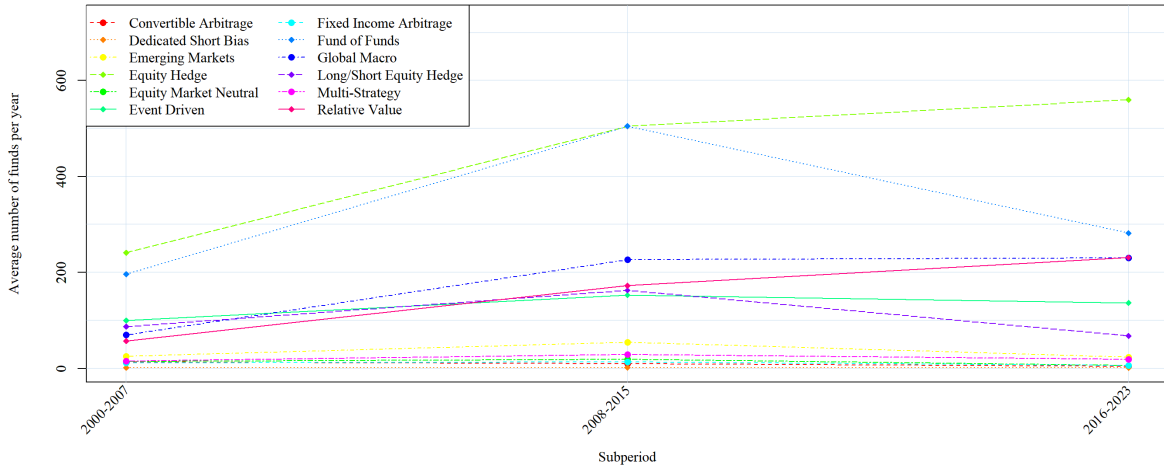
those of Global Macro and Fund of Funds.³³

Figure 7: Fund style analysis. The figure shows the style composition of the $FWER5\%$ portfolio using all considered covariates (Panel A) and of all eligible funds at the corresponding portfolio formation dates (Panel B). For each sub-period, we count the number of funds in each style and then take the average to obtain the number of funds per year.

Panel A: Funds selected in the $FWER5\%$ portfolio by style



Panel B: All eligible funds by style



³³Results for the $FWER\tau$ portfolios with $\tau \in \{0.1, 0.5, 1\}\%$ are similar, and our conclusions remain unchanged.

To investigate whether this pattern simply reflects the composition of the hedge fund universe used, Panel B of Figure 7 reports the style composition of all eligible funds. The high selection rate of Relative Value funds cannot be explained by the prevalence of this style in the overall sample. In particular, Equity Hedge and Fund of Funds account for much larger shares of eligible funds, whereas Relative Value accounts for a substantially smaller share. This suggests that the prominence of Relative Value in the *FWER5%* portfolio reflects selective identification by $fwer^+$, rather than the unconditional size of the style category.³⁴

Panel A of Table 6 sheds more light on the contribution of each fund style to the OOS performance of the *FWER5%* portfolio. We partition the selected funds into style-based groups and treat each group as a new portfolio. We track their OOS returns over the holding periods and calculate their corresponding annualized seven-factor model alpha and excess returns. As not all fund styles are selected every year, we count the number of months during which funds from each style are present in the portfolio, along with the average number of funds belonging to that style in those months, in the last column. We report the results for the entire OOS period (288 months) in Panel A. We find that the Relative Value style appears in the portfolio for the longest period (281 out of 288 months), followed by the Event-driven (264 out of 288 months) and Equity Hedge (252 out of 288 months) styles. On average, the Relative Value style also contributes the most funds to the portfolio, more than twice that of the second group (Event-driven).

In terms of alpha, the contribution of each style is more difficult to evaluate because it depends on both the number of selected funds in the style and the number of months the style is present in the portfolio. For example, the Emerging Markets funds yield the highest alpha. For example, Emerging Markets funds generate the highest alpha. However, this style appears in fewer than half of the OOS months (137 out of 288 months), and its portfolio size

³⁴Relative Value funds seek to exploit mispricings between assets within the same or similar markets. Their objective is to capitalise on short-term pricing inefficiencies across comparable securities and generate returns uncorrelated with the broader market.

Table 6: OOS performance by style. The table decomposes the OOS performance of *FWER5%* portfolios (Panel A) and all eligible funds in our sample (Panel B) into groups by fund style. For each portfolio, constituent funds are partitioned by style, and each style group is treated as a separate sub-portfolio. We track the monthly returns of each style sub-portfolio during the OOS holding periods and report performance over the full OOS sample of 288 months, from January 2000 to December 2023. Portfolio alphas are estimated using the seven-factor model and are annualized. The table reports the annualized alpha, the corresponding Newey–West HAC-adjusted t-statistic, and the annualized excess return. For each style sub-portfolio, we also report the number of months in which the style is present in the corresponding main portfolio and the average number of funds in that style during those months.

Fund Style	FH Alpha (%)	t-statistic	Excess	Portfolio Representation	
			Return (%)	Presence months	Avg. #funds
<u>Panel A: Selected Funds</u>					
Convertible Arbitrage	5.2	4.4	6.0	92	1.8
Emerging Markets	10.8	3.0	17.0	137	4.5
Equity Hedge	7.4	3.6	9.3	252	10.2
Equity Market Neutral	8.5	4.7	9.0	129	2.4
Event Driven	4.5	3.1	5.2	264	9.4
Fixed Income Arbitrage	3.2	1.6	4.3	192	2.2
Fund of Funds	4.3	4.7	4.7	228	11.6
Global Macro	8.0	4.2	7.6	216	6.5
Long/Short Equity Hedge	8.3	2.0	10.2	167	4.6
Multi-Strategy	8.2	2.5	7.3	120	2.3
Relative Value	4.7	6.4	4.5	281	16.3
<u>Panel B: All eligible funds</u>					
Convertible Arbitrage	2.3	1.4	5.0	288	8.6
Dedicated Short Bias	1.5	0.3	-1.5	174	1.5
Emerging Markets	-0.1	-0.0	6.0	288	31.3
Equity Hedge	-0.6	-0.6	4.0	288	413.6
Equity Market Neutral	3.9	2.0	5.0	288	11.5
Event Driven	1.0	1.1	4.2	288	122.6
Fixed Income Arbitrage	1.6	0.9	3.7	264	9.4
Fund of Funds	-0.3	-0.4	1.9	288	306.3
Global Macro	3.1	2.7	3.6	288	165.7
Long/Short Equity Hedge	-1.4	-1.0	3.4	288	97.7
Multi-Strategy	1.7	1.3	4.0	288	19.4
Relative Value	0.7	0.6	3.1	288	144.8

is relatively small (4.5 funds on average). The Equity Market Neutral, Long/Short Equity Hedge, Multi-Strategy and Global Macro funds follow, generating an annualized alpha of around 8%. Nevertheless, all selected style portfolios generate positive alphas over the full OOS period.

To assess the practical value of $fwer^+$, we construct style-based portfolios consisting of

all eligible funds in the IS of 36 months (i.e., those with a complete 36-month return history and all considered covariates). We exhibit the performance of these portfolios in Panel B of the same table. Note that two groups, including Dedicated Short Bias and Fixed Income Arbitrage, do not span all 288 months because they either lack eligible funds in certain years or all eligible funds are delisted during the OOS holding period.

Our first observation is that the entire Relative Value group does not outperform the other styles. In fact, it significantly underperforms the Global Macro and Equity Market Neutral groups. This implies that the prominence of Relative Value funds in the $FWER5\%$ portfolio reflects selective identification by $fwer^+$, rather than superior unconditional performance of the style as a whole. Second, when comparing Panels A and B on a style-by-style basis, we find that every style-specific $FWER5\%$ portfolio outperforms its equally weighted (EW) counterpart. The higher alpha gains and shorter presence periods suggest that $fwer^+$ effectively identifies outperforming funds within each style and does so during periods when such funds exist.

6.7. Alternative Factor Models

The ability of factor models to estimate the entire distribution, as well as the alpha and beta components, can vary due to misspecification. Using different models, potentially less susceptible to misspecification, can yield better insights into the true alphas and betas (see [Ardia et al., 2024](#)). In this Section, we assess whether the performance of the $FWER\tau$ portfolios is robust to alternative benchmark multi-factor models used in fund performance literature (see, e.g., [Bali et al., 2012, 2014](#)). More specifically, we consider four alternative factor models including the four-factor model of [Carhart \(1997\)](#), a six factor model, including the four-factor model and $\Delta 10Y$ and $\Delta CredSpr$ risk factors, a nine-factor model including $BDTF$, $FXTF$ and $CMTF$ factors on top of the six-factor model above, as well as the recent nine-factor model of [Chen et al. \(2025\)](#) that is suitable for explaining hedge fund performance as described in Section 4.

We repeat the exercises presented in the previous sections for each alternative factor model. We present in Table 7 the performance of the $FWER\tau$ portfolios using all informative covariates. Overall, the OOS alphas of the $FWER\tau$ portfolios vary by factor models, but all are positive and statistically significant at the 5% level.

Table 7: Performance of the $FWER\tau$ portfolios using all covariates under alternative factor models. This table reports the OOS performance of $FWER\tau$ portfolios constructed from funds selected by $fwer^+$ using all 24 covariates under alternative performance evaluation models. Panel A reports results using the four-factor model of Carhart (1997); Panel B reports results using the six-factor model which augments the Carhart four-factor model with $\Delta 10Y$ and $\Delta CredSpr$; Panel C reports results using the nine-factor model which further includes bond, currency, and commodity trend-following factors; and, Panel D reports results using the nine-factor model of Chen et al. (2025). For each weighting scheme, the reported statistics are the annualized portfolio alpha, the corresponding Newey–West t-statistic, and the annualized excess return. The final four columns report the average, minimum, and maximum number of funds held in the portfolio and the percentage of OOS months in which the portfolio is empty. Results are reported for $FWER$ target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$.

τ (%)	Equal-weighted $FWER\tau$ portfolio			Alpha-weighted $FWER\tau$ portfolio			Portfolio size			
	Alpha (%)	Excess Return (%)	t-stat	Alpha (%)	Excess Return (%)	t-stat	Average	Min	Max	Empty rate (%)
Panel A: Four-factor model of Carhart (1997)										
0.1	5.4	6.3	4.8	6.5	7.9	3.1	39	3	140	0
0.5	5.0	6.0	4.4	6.2	7.7	3.1	48	4	161	0
1.0	5.0	6.1	4.6	6.4	7.9	3.4	54	5	175	0
5.0	4.4	5.7	3.6	5.6	7.3	3.0	69	6	221	0
Panel B: Six-factor model										
0.1	4.7	5.6	4.1	8.4	9.1	2.7	30	4	102	0
0.5	4.6	5.5	4.0	7.7	8.3	2.8	38	6	125	0
1.0	4.5	5.6	3.7	7.5	8.4	2.7	44	10	139	0
5.0	4.0	5.1	3.4	6.0	7.1	2.2	58	13	179	0
Panel C: Nine-factor model										
0.1	6.5	7.2	3.8	9.5	10.3	2.1	27	1	91	0
0.5	6.0	6.8	4.0	8.6	9.7	2.2	33	1	110	0
1.0	5.9	6.7	3.9	8.5	9.6	2.3	37	1	116	0
5.0	5.4	6.2	4.2	7.6	8.9	2.3	50	1	135	0
Panel D: Nine-factor model of Chen et al. (2025)										
0.1	6.2	6.2	3.0	8.8	7.9	2.4	17	0	55	9.7
0.5	6.8	6.9	3.7	9.2	8.7	2.7	23	0	97	4.2
1.0	6.7	6.9	3.7	8.8	8.4	2.8	27	0	129	4.2
5.0	6.6	6.9	4.1	9.0	8.6	3.3	38	0	208	4.2

7. The Contribution of Covariate Information

Before assessing the contribution of covariates to OOS portfolio performance, we first verify that the estimated prior null probabilities behave consistently with the selection mechanism of $fwer^+$. As discussed in Section 2.1, the procedure assigns fund-specific rejection thresholds according to the estimated prior odds of the null hypothesis. Because the prior odds are a strictly increasing transformation of $\pi_0(\mathbf{z})$, funds with lower estimated $\pi_0(\mathbf{z})$ receive more lenient rejection thresholds. Section IC of the Internet Appendix shows that the estimated prior null probabilities exhibit substantial dispersion across covariate quartiles. This diagnostic confirms that the covariates are informative for the threshold-allocation mechanism of $fwer^+$, although their informativeness varies considerably across covariates. While this exercise examines the informativeness of individual covariates in the fund selection process, it does not quantify their contribution to OOS portfolio performance. We therefore turn to leave-one-out and Shapley analyses to quantify the performance contribution of individual covariates and broader covariate groups.

The two analyses examine the contribution of covariate information from complementary perspectives. In the leave-one-out analysis, we take the $FWER\tau$ portfolio constructed using all covariates as the benchmark and measure the alpha loss resulting from excluding either a covariate group or an individual covariate. In the Shapley analysis, we instead take the $FWER\tau$ portfolio constructed without informative covariates as the benchmark and decompose the total alpha gain from incorporating all covariates across the four covariate groups while accounting for their overlapping information. We further compute pairwise Shapley interactions to examine whether different covariate groups reinforce or substitute one another. All analyses in this section use the equal-weighted $FWER\tau$ portfolios to isolate the role of covariate information in fund selection from the additional effect of alpha-based portfolio weighting.

7.1. Marginal Contribution of Covariates to Portfolio Performance

To quantify the contribution of the covariates to portfolio performance, we first conduct a leave-one-out ablation analysis. Starting from the full specification with all 24 covariates, we reconstruct the OOS portfolio after excluding one covariate group or one individual covariate at a time. Alpha loss is then estimated from the monthly return of the full-covariate portfolio minus that of the reduced-covariate portfolio. A positive (negative) alpha loss means that excluding the group or the individual covariate harms (improves) the performance of the resulting portfolio.

In the interest of space, Table 8 reports the results for the 5% FWER target.³⁵ Panel A shows that excluding any covariate group results in a positive alpha loss, indicating that each group contributes to portfolio performance. The return-based covariate group has the largest marginal contribution, with an annualized alpha loss of 2.37%, substantially larger than those associated with macroeconomic covariates (0.96%), managerial skill proxies (0.80%), and fund characteristics (0.66%). Although none of the group-level alpha losses is statistically significant at conventional levels, their magnitudes suggest that return-based covariates contribute the most to the OOS performance of the portfolio.

Panel B reports the corresponding leave-one-covariate-out analysis. Default yield spread, Sharpe ratio, cumulative return, macro risk, and lockup period generate the largest alpha losses when excluded, whereas removing return volatility or kurtosis slightly improves portfolio performance. Incentive fee is the only individual covariate associated with a statistically significant alpha loss at the 5% level, while the contribution of liquidity timing is marginally significant at the 10% level. The results reveal substantial heterogeneity in the marginal contribution of individual covariates to the OOS performance of $fwer^+$. Importantly, each covariate group contains variables whose exclusion reduces portfolio alpha, indicating that all four groups provide economically valuable information for fund selection.

³⁵Results for other FWER targets are qualitatively similar.

Table 8: Marginal contribution of covariate information to portfolio alpha. The table reports the marginal contribution of different covariate groups and individual covariates to the performance of $fwer^+$ using a leave-one-out ablation analysis. For each analysis, the corresponding covariate group or individual covariate is excluded from the input of $fwer^+$, while all remaining covariates are retained. Portfolio returns are then reconstructed under the reduced specification. We report alpha loss, estimated as the intercept from a regression of the monthly return difference between the full-covariate and reduced-covariate portfolios on the [Fung and Hsieh \(2004\)](#) seven-factor model, together with the corresponding Newey–West t-statistics. Results are reported for the 5% FWER target.

Ablation set	Alpha loss	t-statistic
<u>Panel A: Group of covariates</u>		
Return-based covariates	2.37	1.86
Macroeconomic covariates	0.96	1.47
Managerial skill proxies	0.80	1.22
Fund characteristics	0.66	1.62
<u>Panel B: Individual covariates</u>		
Default Yield Spread	1.91	1.55
Sharpe Ratio	1.34	1.08
Cumulative Return	0.86	1.73
Macro Risk	0.83	1.79
Lockup Period	0.72	1.73
Idiosyncratic Risk	0.70	1.65
Incentive Fee	0.63	2.23
Inflation	0.60	1.73
R-squared	0.52	1.05
Fund Age	0.45	1.31
Six-factor Model Idiosyncratic Risk	0.43	1.24
Skewness	0.41	1.12
Equity Market Uncertainty	0.41	1.04
Liquidity Timing	0.40	1.93
Management Fee	0.39	1.11
Minimum Investment	0.38	0.89
Term Spread	0.37	1.02
Redemption Notice Period	0.30	0.68
High-Water Mark	0.24	0.64
Investor Sentiment Change	0.16	0.44
Economic Policy Uncertainty	0.08	0.20
Fund Size	0.06	0.29
Kurtosis	-0.08	-0.50
Return Volatility	-0.26	-0.92

Because the covariates contain overlapping information, the alpha losses from the leave-one-out analysis cannot be interpreted as additive measures of importance.³⁶ We therefore

³⁶For example, summing the alpha losses from excluding the individual covariates in Panel B yields

next perform a Shapley decomposition of the portfolio alpha generated by $fwer^+$ to attribute the overall performance gain across different sources of covariate information while accounting for their overlap.

7.2. Shapley Decomposition of Alpha Gains

To quantify the contribution of different sources of covariate information to the portfolio alpha generated by $fwer^+$, we adapt the Shapley value framework originally proposed by Shapley (1953) and subsequently applied to attribution problems in economics (Israeli, 2007) and machine learning (Lundberg and Lee, 2017), including recent applications to mutual fund selection (DeMiguel et al., 2023). Whereas previous applications focus on attributing *model performance*, we use the Shapley framework to attribute the *economic value* generated by incorporating informative covariates, measured by the incremental OOS portfolio alpha. Thereby, we first construct an $fwer^+$ -based portfolio using a constant variable as the sole covariate input. Because the constant variable contains no cross-sectional information, this specification is equivalent to applying $fwer^+$ without informative covariates. Using this portfolio as the benchmark, the alpha gain from incorporating all covariates is calculated. The Shapley decomposition then measures how much each covariate group contributes to this gain.³⁷

Table 9 reports the results. The first three columns report the annualized alpha obtained

11.85%, which exceeds the total alpha gain attributable to incorporating informative covariates.

³⁷Let $\mathcal{G}$ denote the set of four covariate groups. For any subset (coalition) $S \subseteq \mathcal{G}$, let $v(S)$ denote the annualized alpha of the $fwer^+$ portfolio constructed using only the covariates in the groups in S . In particular, $v(\emptyset)$ denotes the annualized alpha obtained by $fwer^+$ without informative covariates (i.e., implementing $fwer^+$ using a constant variable as the sole covariate input), while $v(\mathcal{G})$ denotes the annualized alpha obtained using all covariates. The Shapley value of group $g \in \mathcal{G}$ is defined as

$$\phi_g = \sum_{S \subseteq \mathcal{G} \setminus \{g\}} \frac{\#S! (\#\mathcal{G} - \#S - 1)!}{\#\mathcal{G}!} [v(S \cup \{g\}) - v(S)].$$

where $\#S$ denotes the number of groups in subset S and “!” denotes the factorial operator, and $v(S \cup \{g\}) - v(S)$ is the marginal increase in annualized portfolio alpha obtained by adding covariate group g to subset

with no covariates, the annualized alpha obtained using all covariates, and the corresponding alpha gain, respectively. Compared to the no-covariate portfolios, incorporating covariate information increases annualized portfolio alpha by between 1.77% and 2.63% across the four FWER targets. The remaining columns decompose this alpha gain into the contributions of the four covariate groups. Notably, return-based covariate group accounts for the largest contribution at every FWER target, contributing between 1.18% and 1.49% of annualized alpha, corresponding to approximately 48% to 84% of the total alpha gain in relative terms. This finding indicates that recent return characteristics constitute the primary source of information driving the superior OOS performance of the $FWER\tau$ portfolios.

Table 9: Shapley decomposition of alpha gains by covariate group. The table reports the exact Shapley decomposition of the improvement in portfolio alpha from incorporating covariate information into the $fwcr^+$ procedure. “No covariates” and “All covariates” report the annualized alpha of the portfolios formed by $fwcr^+$ with a constant covariate and with all covariates, respectively, while “Alpha gain” (in %) denotes the difference between the two. The remaining columns report the Shapley contribution of each covariate group to the total alpha gain. Results are reported for the four FWER targets $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$.

τ (%)	No covariates	All covariates	Alpha gain	Shapley decomposition of alpha gains			
				Return-based	Macro	Managerial skill	Characteristics
0.1	3.80	6.43	2.63	1.27	0.70	0.36	0.30
0.5	4.25	6.30	2.05	1.18	0.28	0.18	0.41
1.0	4.14	6.13	1.99	1.36	0.20	0.14	0.28
5.0	4.15	5.93	1.77	1.49	0.05	0.11	0.13

The remaining alpha gain is distributed across macroeconomic covariates, managerial skill proxies, and fund characteristics. At the most stringent FWER target ($\tau = 0.1\%$), macroeconomic covariates provide the second-largest contribution, accounting for 0.70% of

S. The Shapley values satisfy

$$\sum_{g \in \mathcal{G}} \phi_g = v(\mathcal{G}) - v(\emptyset),$$

so that they exactly decompose the improvement in portfolio alpha attributable to incorporating covariate information. Since there are four covariate groups, the decomposition requires forming and evaluating only $2^4 = 16$ portfolios. Applying the same decomposition to the 24 individual covariates would instead require forming and evaluating $2^{24} = 16,777,216$ portfolios, making an exact Shapley decomposition computationally prohibitive.

annualized alpha (27% of the total gain). As the FWER target becomes less stringent, however, the contribution of macroeconomic covariates declines substantially, whereas those of fund characteristics and managerial skill proxies remain positive but relatively modest throughout. Overall, the Shapley decomposition indicates that recent return characteristics account for the largest standalone contribution to the incremental alpha generated by incorporating covariate information. This finding is consistent with the leave-one-out evidence in Section 7.1 and with the broader empirical result that past-performance information plays an important role in fund selection. As we show below, however, the contribution of macroeconomic covariates is more apparent through their complementarity with the remaining sources of information.

While the Shapley decomposition attributes the overall portfolio alpha gain to each covariate group, it does not reveal whether different sources of information complement or substitute one another. To address this question, we further compute the pairwise Shapley interactions between covariate groups. Intuitively, a pairwise Shapley interaction measures whether two covariate groups jointly generate more (or less) portfolio alpha than would be expected from the sum of their individual contributions after averaging over all possible combinations of the remaining covariate groups.³⁸

³⁸Using the notation in Footnote 37, for two covariate groups $g, h \in \mathcal{G}$ and a background coalition $S \subseteq \mathcal{G} \setminus \{g, h\}$, define the second-order marginal contribution as

$$\Delta_{g,h}(S) = v(S \cup \{g, h\}) - v(S \cup \{g\}) - v(S \cup \{h\}) + v(S),$$

which measures the additional annualized alpha generated by including g and h jointly beyond the sum of their individual contributions under the same background coalition. The pairwise Shapley interaction between g and h is then defined as

$$I_{g,h} = \sum_{S \subseteq \mathcal{G} \setminus \{g,h\}} \frac{\#S!(\#\mathcal{G} - \#S - 2)!}{(\#\mathcal{G} - 1)!} \Delta_{g,h}(S),$$

which averages the second-order marginal contribution over all possible background coalitions using the standard Shapley weights. Positive values indicate complementarity between the two covariate groups, whereas negative values indicate substitutable or overlapping information.

Table 10 reports the pairwise Shapley interactions between the four covariate groups. Two findings emerge. First, the strongest interactions on average involve macroeconomic covariates. Across all FWER target levels, the interactions between macroeconomic variables and return-based covariates, as well as between macroeconomic variables and managerial skill proxies, are positive and among the largest observed. Averaged across the four FWER targets, these interactions equal 0.90% and 0.88%, respectively. These results indicate that macroeconomic covariates complement both return-based information and managerial skill proxies in generating OOS portfolio alpha, suggesting that their contribution extends beyond their individual effects.

Table 10: Pairwise Shapley interactions between covariate groups. The table reports the pairwise Shapley interactions between the four covariate groups. Positive values indicate that the two covariate groups jointly contribute more to OOS portfolio alpha than would be expected from their individual contributions, whereas negative values indicate overlapping or substitutable information. Results are reported across the four FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$ together with their average.

τ (%)	Pairwise Shapley interaction of covariate groups					
	Return × Macro	Return × Skill	Return × Characteristics	Macro × Skill	Macro × Characteristics	Skill × Characteristics
0.1	0.48	-0.11	0.57	0.40	0.22	-0.06
0.5	0.85	-0.08	-0.16	1.16	0.71	-0.12
1.0	1.22	0.15	-0.15	0.84	0.28	-0.08
5.0	1.05	0.35	0.55	1.14	0.41	0.22
Average	0.90	0.08	0.20	0.88	0.40	-0.01

Second, interactions not involving macroeconomic covariates are generally smaller and less stable across FWER targets. In particular, the average interactions between return-based variables and managerial skill proxies (0.08%), return-based variables and fund characteristics (0.20%), and managerial skill proxies and fund characteristics (-0.01%) are close to zero. Taken together, the results suggest that the superior OOS performance of $FWER\tau$ portfolios arises not only from the individual contributions of informative covariates but also from the complementarity between macroeconomic covariates and other sources of economically motivated information.

Taken together, the analyses provide complementary perspectives on the contribution of covariate information to the OOS performance of the $FWER\tau$ portfolios. The leave-one-out analysis quantifies the marginal contribution of individual covariates and covariate groups, the Shapley decomposition attributes the overall alpha gain across the four covariate groups while accounting for overlapping information, and the pairwise Shapley interactions examine whether different sources of information reinforce one another. Although these approaches measure different aspects of covariate importance, they deliver a consistent overall message. In particular, the results suggest distinct roles for the return-based and macroeconomic covariate groups. Return-based covariates provide the dominant standalone contribution to portfolio alpha, whereas macroeconomic covariates appear to serve as contextual information that consistently reinforces the information contained in return-based covariates, managerial skill proxies, and fund characteristics.

These findings are further supported by the supplementary analyses reported in the Internet Appendix. Section ID shows that macroeconomic and risk-exposure covariates account for a substantial share of the common latent structure underlying the covariates. Nevertheless, replacing the original covariates with latent factors or principal components leads to inferior portfolio performance, indicating that $fwer^+$ benefits from exploiting both common and covariate-specific information. Moreover, the alternative approaches for enhancing covariate informativeness presented in Sections IE and IF, including statistical screening and machine-learning-based transformations, do not improve upon the original 24-covariate specification. Overall, the evidence suggests that preserving the original covariates allows $fwer^+$ to exploit both their individual predictive content and the complementarities among different sources of economically motivated information.

8. Conclusion

This paper introduces the $fwer^+$ procedure as a powerful and flexible method for controlling FWER when selecting outperforming investment opportunities. By incorporating

multiple informative covariates, $fwer^+$ estimates FWER conditionally on the available information, allowing rejection thresholds to vary across tests. Through simulations, we demonstrate that when informative covariates are available, $fwer^+$ effectively controls FWER at any target level and achieves considerably higher power than existing methods that ignore such information.

Empirically, applying $fwer^+$ to U.S. hedge funds identifies portfolios that deliver economically large and statistically significant OOS performance. The selected portfolios outperform passive benchmarks formed from the same eligible fund universe, including the equal-weighted portfolio of all funds, the equal-weighted portfolio of funds with positive in-sample alpha, and its alpha-weighted counterpart. These results hold across FWER target levels, weighting schemes, sub-periods, longer holding horizons, and alternative factor models. The evidence therefore suggests that incorporating informative covariates into multiple testing improves the practical usefulness of FWER control in fund selection, where unconditional procedures are often too conservative for short rolling windows and large cross-sections. Finally, the leave-one-out ablation, Shapley decomposition, and pairwise Shapley interaction analyses provide new insights into the contribution of informative covariates to portfolio performance. Return-based covariates provide the largest standalone contribution to the superior performance of $fwer^+$, whereas macroeconomic covariates consistently complement return-based covariates, managerial skill proxies, and fund characteristics. These findings highlight the importance of integrating multiple, complementary sources of covariate information in hedge fund selection.

Overall, our evidence suggests that the $fwer^+$ procedure is an effective tool for investors seeking to identify outperformers with high confidence and limited false discoveries. Its flexibility allows users to specify any desired FWER threshold, making it particularly suitable for applications in which downside risk is severe. Beyond hedge funds, the method can be applied to other large-scale selection problems in economics and finance where false discoveries are costly and informative covariates are available, such as selecting outperforming

mutual funds (see, e.g., [Barras et al., 2010](#); [Hsu et al., 2025](#)), factor selection (see, e.g., [Harvey and Liu, 2020](#)), or algorithmic trading strategies (see, e.g., [Bajgrowicz and Scaillet, 2012](#)). The $fwer^+$ framework thus provides a general-purpose approach for high-confidence decision-making under covariate information.

References

- Ackermann, C., McEnally, R., Ravenscraft, D., 1999. The performance of hedge funds: Risk, return, and incentives. *The Journal of Finance* 54, 833–874.
- Agarwal, V., Daniel, N.D., Naik, N.Y., 2009. Role of managerial incentives and discretion in hedge fund performance. *The Journal of Finance* 64, 2221–2256.
- Agarwal, V., Daniel, N.D., Naik, N.Y., 2011. Do hedge funds manage their reported returns? *The Review of Financial Studies* 24, 3281–3320.
- Agarwal, V., Green, T.C., Ren, H., 2018. Alpha or beta in the eye of the beholder: What drives hedge fund flows? *Journal of Financial Economics* 127, 417–434.
- Agarwal, V., Naik, N.Y., 2000. Multi-period performance persistence analysis of hedge funds. *The Journal of Financial and Quantitative Analysis* 35, 327–342.
- Agarwal, V., Ruenzi, S., Weigert, F., 2017. Tail risk in hedge funds: A unique view from portfolio holdings. *Journal of Financial Economics* 125, 610–636.
- Aragon, G.O., 2007. Share restrictions and asset pricing: Evidence from the hedge fund industry. *Journal of Financial Economics* 83, 33–58.
- Ardia, D., Barras, L., Gagliardini, P., Scaillet, O., 2024. Is it alpha or beta? decomposing hedge fund returns when models are misspecified. *Journal of Financial Economics* 154, 103805.

- Avramov, D., Barras, L., Kosowski, R., 2013. Hedge fund return predictability under the magnifying glass. *The Journal of Financial and Quantitative Analysis* 48, 1057–1083.
- Avramov, D., Kosowski, R., Naik, N.Y., Teo, M., 2011. Hedge funds, managerial skill, and macroeconomic variables. *Journal of Financial Economics* 99, 672–692.
- Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics* 106, 473–491.
- Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance* 61, 1645–1680.
- Bali, T.G., Brown, S.J., Caglayan, M.O., 2011. Do hedge funds’ exposures to risk factors predict their future returns? *Journal of Financial Economics* 101, 36–68.
- Bali, T.G., Brown, S.J., 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., Caglayan, M.O., 2014. Macroeconomic risk and hedge fund returns. *Journal of Financial Economics* 114, 1–19.
- Bali, T.G., Brown, S.J., Caglayan, M.O., Celiker, U., 2021. Does industry timing ability of hedge funds predict their future performance, survival, and fund flows? *Journal of Financial and Quantitative Analysis* 56, 2136–2169.
- Bali, T.T., Weigert, F., 2024. Hedge funds and the positive idiosyncratic volatility effect. *Review of Finance* 28, 1611–1661.
- Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance* 65, 179–216.
- Bollen, N.P.B., Pool, V.K., 2008. Conditional return smoothing in the hedge fund industry. *Journal of Financial and Quantitative Analysis* 43, 267–298.

- Bollen, N.P.B., Pool, V.K., 2009. Do hedge fund managers misreport returns? evidence from the pooled distribution. *The Journal of Finance* 64, 2257–2288.
- Bonferroni, C.E., 1936. *Teoria statistica delle classi e calcolo delle probabilita*. Bagni di Lucca, Italy: Libreria Internazionale Seeber .
- Boyson, N.M., 2008. Hedge fund performance persistence: A new approach. *Financial Analysts Journal* 64, 27–44.
- Buraschi, A., Kosowski, R., Trojani, F., 2013. When there is no place to hide: Correlation risk and the cross-section of hedge fund returns. *The Review of Financial Studies* 27, 581–616.
- Cao, C., Chen, Y., Liang, B., Lo, A.W., 2013. Can hedge funds time market liquidity? *Journal of Financial Economics* 109, 493–516.
- Carhart, M.M., 1997. On persistence in mutual fund performance. *The Journal of Finance* 52, 57–82.
- Chen, Y., Cliff, M., 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., Pan, J., 2021. Sentiment trading and hedge fund returns. *The Journal of Finance* 76, 2001–2033.
- Chen, Y., Li, S.Z., Tang, Y., Zhou, G., 2025. Anomalies as new hedge fund factors Forthcoming in *Journal of Financial and Quantitative Analysis*.
- Couts, S.J., Gonçalves, A.S., Rossi, A., 2024. Unsmoothing returns of illiquid funds. *The Review of Financial Studies* 37, 2110–2155.
- Cumming, D., Dai, N., Haß, L.H., Schweizer, D., 2012. Regulatory induced performance persistence: Evidence from hedge funds. *Journal of Corporate Finance* 18, 1005–1022.

- DeMiguel, V., Gil-Bazo, J., Nogales, F.J., Santos, A.A.P., 2023. Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics* 150, 103737.
- Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116, 1–22.
- Ferson, W., Chen, Y., 2021. How many good and bad funds are there, really?, in: *Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning*. World Scientific Publishing Co.. chapter 108, pp. 3753–3827.
- Fung, W., Hsieh, D.A., 2001. Empirical characteristics of dynamic trading strategies: the case of hedge funds. *Review of Financial Studies* 10, 275–302.
- Fung, W., Hsieh, D.A., 2004. Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal* 60, 65–80.
- Fung, W., Hsieh, D.A., Naik, N.Y., Ramadorai, T., 2008. Hedge funds: Performance, risk, and capital formation. *The Journal of Finance* 63, 1777–1803.
- Getmansky, M., Lo, A.W., Makarov, I., 2004. An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics* 74, 529–609.
- Giglio, S., Liao, Y., Xiu, D., 2020. Thousands of alpha tests. *The Review of Financial Studies* 34, 3456–3496.
- Hansen, P.R., 2005. A test for superior predictive ability. *Journal of Business & Economic Statistics* 23, 365–380.
- Harvey, C.R., Liu, Y., 2020. False (and missed) discoveries in financial economics. *The Journal of Finance* 75, 2503–2553.

- Hodder, J.E., Jackwerth, J.C., Kolokolova, O., 2014. Recovering delisting returns of hedge funds. *Journal of Financial and Quantitative Analysis* 49, 797–815.
- Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.
- Hsu, P.H., Hsu, Y.C., 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.
- Hsu, P.H., Kyriakou, I., Ma, T., Sermpinis, G., 2025. Mutual funds’ conditional performance free of data snooping bias. *Journal of Financial and Quantitative Analysis* 60, 1373–1400.
- Ignatiadis, N., Huber, W., 2021. Covariate powered cross-weighted multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 83, 720–751.
- Israeli, O., 2007. A shapley-based decomposition of the R-square of a linear regression. *The Journal of Economic Inequality* 5, 199–212.
- Joenväärä, J., Kauppila, M., Kosowski, R., Tolonen, P., 2021. Hedge fund performance: Are stylized facts sensitive to which database one uses? *Critical Finance Review* 10, 271–327.
- Kosowski, R., Naik, N.Y., Teo, M., 2007. Do hedge funds deliver alpha? a bayesian and bootstrap analysis. *Journal of Financial Economics* 84, 229–264.
- Li, H., Xu, Y., Zhang, X., 2016. Hedge fund performance evaluation under the stochastic discount factor framework. *Journal of Financial and Quantitative Analysis* 51, 231–257.
- Liang, B., Park, H., 2007. Risk measures for hedge funds: a cross-sectional approach. *European Financial Management* 13, 333–370.

- Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc.
- Mandle, C., 2024. Hedge fund industry reaches \$4.5 trillion in 2024. Reuters July 2025.
- Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Pástor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111, 642–685.
- Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313.
- Romano, J.P., Shaikh, A.M., 2006. Stepup procedures for control of generalizations of the familywise error rate. *The Annals of Statistics* 34, 1850 – 1873.
- Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73, 1237–1282.
- Shapley, L.S., 1953. A value for n-person games, in: Kuhn, H.W., Tucker, A.W. (Eds.), *Contributions to the Theory of Games II*. Princeton University Press, pp. 307–317.
- Storey, J.D., 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64, 479–498.
- Storey, J.D., Bass, A.J., Dabney, A., Robinson, D., 2017. qvalue: Q-value estimation for false discovery rate control. URL: <http://github.com/StoreyLab/qvalue>. r package version 2.15.0.

- Sun, Z., Wang, A.W., 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.
- Titman, S., Tiu, C., 2011. Do the Best Hedge Funds Hedge? *The Review of Financial Studies* 24, 123–168.
- Van der Sluis, P., Posthuma, N., 2004. A critical examination of historical hedge fund returns. *RISK Books*. pp. 365–386.
- White, H., 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.
- White, H., 2000. A reality check for data snooping. *Econometrica* 68, 1097–1126.
- Wu, W., Chen, J., Yang, Z.B., 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., Chen, J., 2021. Covariate adaptive familywise error rate control for genome-wide association studies. *Biometrika* 108, 915–931.

Internet Appendix:

Assessing Hedge Fund Performance with An Information-based Multiple Test

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:

$$H_0 : \mu_i \leq 0 \quad H_1 : \mu_i > 0 \quad (9)$$

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 , $i = 1, \dots, n$. 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 (7). 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 $r_{i,t} - \mathbf{F}_t \hat{\beta}_i$ where $\hat{\beta}_i$ is the estimate of β_i , we define a

loss function as $L_{i,t} = -[r_{i,t} - \mathbf{F}_t \hat{\boldsymbol{\beta}}_i]$. 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 $L_{0,t} - L_{i,t} = r_{i,t} - \mathbf{F}_t \hat{\boldsymbol{\beta}}_i$ is the alpha of the fund i . Thus, in our framework

$$d_{i,t} = r_{i,t} - \mathbf{F}_t \hat{\boldsymbol{\beta}}_i = \hat{\alpha}_i + \hat{\varepsilon}_{i,t} \quad (10)$$

The StepM and StepSPA procedures rely on a bootstrapped resampling where the stationary bootstrap procedure of [Politis and Romano \(1994a\)](#), 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 $d_{i,1}, \dots, d_{i,T}$. Then,

$$\hat{\omega}_i^2 = \hat{\gamma}_{i,0} + 2 \sum_{t=1}^{T-1} \kappa(T, t) \hat{\gamma}_{i,t} \quad (11)$$

where $\hat{\gamma}_{i,t} = 1/T \cdot \sum_{k=1}^{T-t} (d_{i,k} - \bar{d}_i)(d_{i,k+t} - \bar{d}_i)$, $t = 0, \dots, T-1$ 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 (10) on the bootstrapped return of the fund i .³⁹

For the (studentized) StepM procedure, we calculate the variance $\hat{\omega}_{i,b}^2$ for the bootstrapped differential $d_{i,t,b}$ via using (11) 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, $c_{\tau, \text{StepM}}^*$, as the $(1 - \tau)^{th}$ quantile of the bootstrapped population $\max_i [(\bar{d}_{i,b} - \bar{d}_i)/\hat{\omega}_{i,b}]$, $b = 1, \dots, B$.⁴⁰

³⁹We 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.

⁴⁰Our implementation of the StepM procedure is similar to [Hsu et al. \(2010\)](#) which differs from the original

In the (studentized) StepSPA, we define $\hat{\mu}_i = \bar{d}_i \cdot \mathbf{1}_{\{\sqrt{T}\bar{d}_i \leq -\hat{\omega}_i \sqrt{2 \log \log T}\}}$ where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function, $i = 1, \dots, n$. 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 $\sqrt{T} \max_i [(\bar{d}_{i,b} - \bar{d}_i + \hat{\mu}_i)/\hat{\omega}_i]$, $b = 1, \dots, B$.

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_τ^s .
 2. The top k' funds in the sub-sample with $\sqrt{T}\bar{d}_{k'}/\hat{\omega}_{k'} > c_\tau^s$ 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.

procedure of [Romano and Wolf \(2005\)](#) in three aspects. First, they use a circular block bootstrap while [Hsu et al. \(2010\)](#) and us use the stationary bootstrap. Second, they adopt data-driven algorithm in determining the block size of bootstrap, while we use a fixed values of q that correspond to average bootstrap lengths from 2 to 12 months. Third, they use bootstrapped standard errors whereas [Hsu et al. \(2010\)](#) and us adopt HAC estimators as described in [11](#). In unreported simulations, these differences do not affect the finite sample performance of the StepM performance reported in our simulations results. Finally it should be noted that both [Hsu et al. \(2010\)](#) and [Romano and Wolf \(2005\)](#) apply their methods to technical rules and daily data.

IB. Performance comparison of $fwer^+$ against StepM and StepSPA with varying average block lengths

As discussed in Section IA, we adopt the stationary bootstrap procedure of Politis and Romano (1994b) for the StepM and StepSPA methods. This procedure relies on a parameter q , which corresponds to an average block length of $1/q$ in the data-generating process. Section 5.2 presents comparisons between $fwer^+$ and the StepM and StepSPA procedures when $q = 1/6$, corresponding to an average block length of 6 months. In this section, we further conduct simulations to demonstrate that our conclusions remain unchanged across plausible values of $q \in \{1/12, 1/3, 1/2\}$ which corresponds to average block lengths of 12, 3 and 2 months, respectively. These values are realistic for preserving the dependence structure in hedge fund returns.

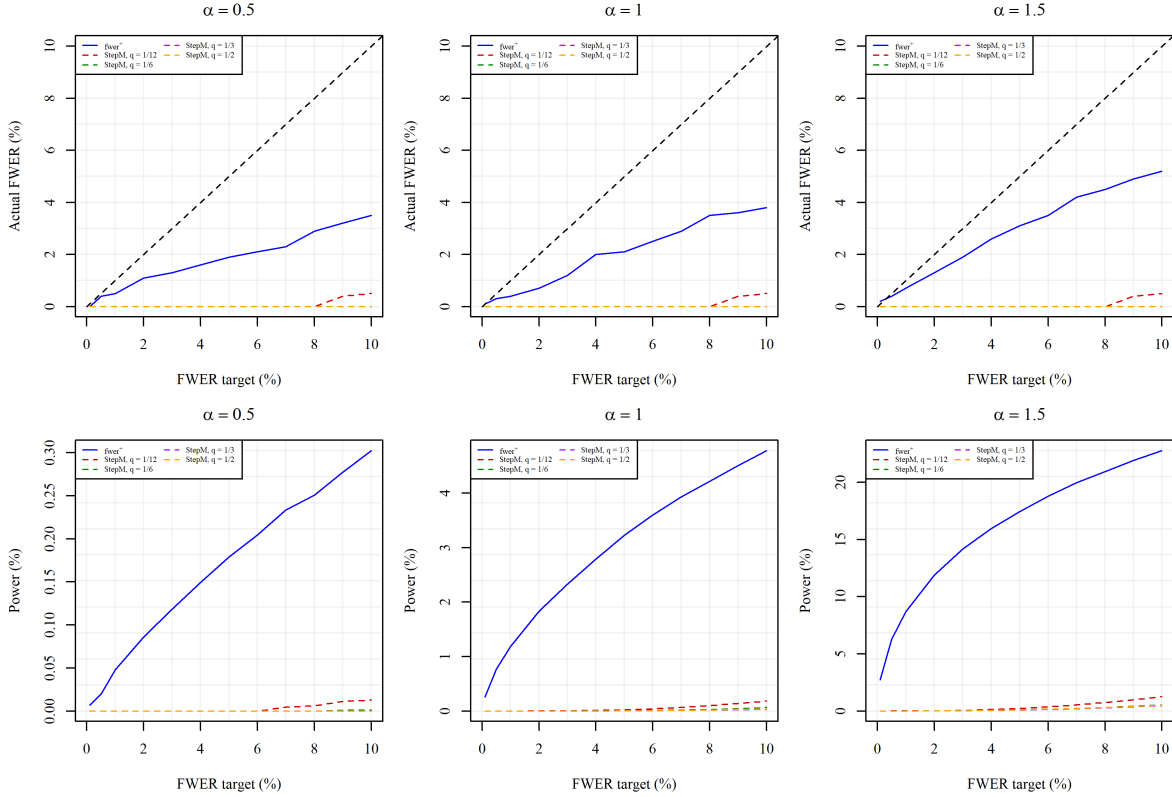
Figures I and II present the results comparing $fwer^+$ with the StepM and StepSPA procedures, respectively. We find that StepM provides satisfactory FWER control across all considered values of q , whereas StepSPA fails to do so. In addition, the $fwer^+$ procedure consistently achieves higher power than StepM.

IC. Informativeness of Covariates for Fund Selection

As discussed in Section 2.1, the $fwer^+$ procedure assigns fund-specific rejection thresholds according to the prior odds of the null hypothesis (see Equation (4)). Since the prior odds are a strictly increasing transformation of the estimated prior null probability, $\pi_0(\mathbf{z})$, the two measures induce identical rankings of funds. Consequently, lower values of $\pi_0(\mathbf{z})$ imply a lower estimated probability that a fund is a zero-alpha fund and hence more lenient rejection thresholds under $fwer^+$.

Using the estimated prior null probabilities, we next assess the informativeness of the covariates for fund selection under $fwer^+$. For each covariate, funds are sorted into quartiles within each estimation period according to the covariate values. The average estimated prior

Figure I: Performance comparison: $fwer^+$ versus StepM under varying average block lengths. The figure compares the performance of the $fwer^+$ procedure and the StepM procedure of Romano and Wolf (2005) in terms of FWER control (top row) and statistical power (bottom row). The simulated data consist of balanced panels with 1000 funds, each observed over 36 monthly periods. The simulated data consist of balanced panels of 1000 funds, each with 36 monthly observations. From left to right, α takes values 0.5, 1.0, and 1.5 (in %). The input covariates U and V used in the $fwer^+$ procedure are moderately informative and independent. The StepM procedure employs the stationary bootstrap with average block length parameter q . We consider five block-length specifications $q \in \{1/12, 1/6, 1/3, 1/2\}$.



null probability is first computed within each quartile and estimation period, and Table I reports the time-series average of these quartile means together with the difference between the fourth and first quartiles (Q4–Q1). Covariates are ranked by the absolute value of this difference. We interpret this absolute difference as a descriptive measure of the covariate’s discriminatory ability. Larger absolute differences indicate that the covariate more effectively separates funds assigned high and low estimated prior null probabilities and, consequently, more differentiated rejection thresholds under the proposed $fwer^+$ procedure.

Figure II: Performance comparison: $fwer^+$ versus StepSPA under varying average block lengths. The figure compares the performance of the $fwer^+$ procedure and the StepM procedure of Hsu et al. (2010) in terms of FWER control (top row) and statistical power (bottom row). The simulated data consist of balanced panels with 1000 funds, each observed over 36 monthly periods. The simulated data consist of balanced panels of 1000 funds, each with 36 monthly observations. From left to right, α takes values 0.5, 1.0, and 1.5 (in %). The input covariates U and V used in the $fwer^+$ procedure are moderately informative and independent. The StepSPA procedure employs the stationary bootstrap with average block length parameter q . We consider five block-length specifications $q \in \{1/12, 1/6, 1/3, 1/2\}$.

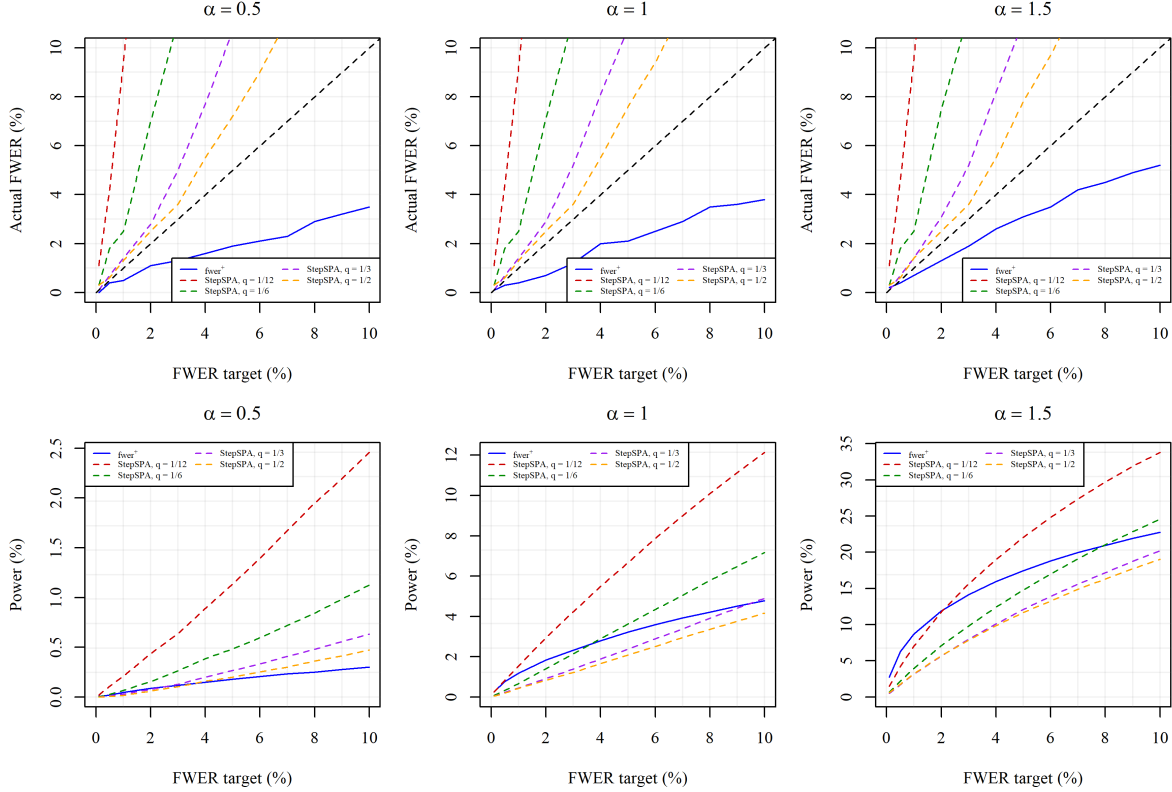


Table I reveals substantial heterogeneity in the discriminatory ability of the covariates. The Sharpe ratio exhibits the strongest pattern, with the average estimated prior null probability declining from 0.79 in the lowest quartile to 0.45 in the highest quartile. Similar monotonic decreasing patterns are observed for fund size, cumulative return, equity market uncertainty, and redemption notice period, whereas return volatility, idiosyncratic risk, and six-factor model idiosyncratic risk display increasing patterns. These results suggest that higher values of the former covariates are associated with lower estimated prior null prob-

abilities, while higher values of the latter are associated with higher estimated prior null probabilities.

Table I: Covariate information and estimated prior null probabilities. For each covariate, funds are sorted into quartiles within each estimation period according to the covariate values. The average estimated prior null probability, $\pi_0(\mathbf{z})$, is first computed within each quartile and estimation period, and the table reports the time-series average of these quartile means. The final two columns report the difference between the fourth and first quartiles (Q4–Q1) and its absolute value.

Covariate	Mean $\pi_0(\mathbf{z})$				Differences in mean $\pi_0(\mathbf{z})$	
	Q1	Q2	Q3	Q4	(Q4–Q1)	Q4–Q1
Sharpe Ratio	0.79	0.74	0.65	0.45	-0.34	0.34
Fund Size	0.75	0.69	0.63	0.54	-0.21	0.21
Cumulative Return	0.78	0.69	0.59	0.57	-0.20	0.20
Return Volatility	0.53	0.66	0.71	0.72	0.19	0.19
Idiosyncratic Risk	0.55	0.66	0.70	0.71	0.16	0.16
Six-factor Model Idiosyncratic Risk	0.55	0.67	0.70	0.70	0.15	0.15
Redemption Notice Period	0.71	0.69	0.65	0.58	-0.13	0.13
Equity Market Uncertainty	0.72	0.68	0.62	0.60	-0.12	0.12
R-squared	0.61	0.65	0.67	0.69	0.08	0.08
Minimum Investment	0.69	0.68	0.64	0.61	-0.08	0.08
Economic Policy Uncertainty	0.70	0.66	0.62	0.64	-0.06	0.06
Liquidity Timing	0.69	0.66	0.64	0.64	-0.05	0.05
Lockup Period	0.67	0.66	0.66	0.63	-0.05	0.05
Management Fee	0.67	0.68	0.64	0.63	-0.04	0.04
Inflation	0.70	0.62	0.63	0.66	-0.04	0.04
Fund Age	0.64	0.65	0.66	0.68	0.04	0.04
Kurtosis	0.67	0.66	0.66	0.63	-0.03	0.03
Incentive Fee	0.66	0.68	0.65	0.63	-0.03	0.03
Term Spread	0.69	0.64	0.64	0.66	-0.03	0.03
Default Yield Spread	0.69	0.64	0.63	0.66	-0.03	0.03
Investor Sentiment Change	0.70	0.63	0.62	0.68	-0.03	0.03
Macro Risk	0.68	0.64	0.64	0.67	-0.01	0.01
Skewness	0.66	0.66	0.66	0.64	-0.01	0.01
High-Water Mark	0.66	0.66	0.65	0.65	-0.01	0.01

Several other covariates, including inflation, investor sentiment change, high-water mark, macro risk, and management fee, exhibit relatively flat or non-monotonic patterns, indicating limited discriminatory ability when considered individually. This does not imply that they are uninformative in $fwer^+$, however, since $\pi_0(\mathbf{z})$ is estimated from a multivariate model in which covariates may contribute jointly with other variables. In addition, the quartile-based analysis is less informative for binary or highly discrete covariates, such as High-Water Mark,

where the limited number of distinct values restricts variation across quartiles.

ID. Performance of dimensionality reduction methods

Section shows that $fwer^+$ identifies hedge funds that generate significant OOS alphas when supplied with a rich set of informative covariates. However, the 24 covariates considered in this study are unlikely to provide entirely distinct information. Instead, many are expected to capture overlapping aspects of fund performance, risk, managerial behaviour, and macroeconomic conditions. This raises the question of whether the information relevant for $fwer^+$ can be summarized by a smaller number of latent factors without materially reducing portfolio performance. To investigate this question, we replace the original covariates with latent factors extracted using factor analysis and principal component analysis (PCA), and compare the resulting portfolio performance with that obtained using the full 24-covariate specification.

This exercise serves two purposes. First, it provides an interpretability check by identifying which covariate groups are most strongly represented in the common information structure. Second, it evaluates whether a small number of latent factors can preserve the information content of the full set of covariates when used as inputs to $fwer^+$. The aim is to understand the economic content of the information that $fwer^+$ uses when selecting outperforming funds.

We start with factor analysis method which models the observed informative covariates $\mathbf{Z} = (Z^1, \dots, Z^d)$ (standardized) as a linear combination of the unobserved m latent factors $f_i, i = 1, \dots, m$, plus noise:

$$\begin{aligned} Z^1 &= l_{11}f_1 + \dots + l_{1m}f_m + \epsilon_1 \\ Z^2 &= l_{21}f_1 + \dots + l_{2m}f_m + \epsilon_2 \\ &\vdots \\ Z^d &= l_{d1}f_1 + \dots + l_{dm}f_m + \epsilon_d \end{aligned} \tag{12}$$

where ϵ_i is independent noise with zero mean and finite variance, and $\mathbf{f} \sim N(0, \mathbf{I}_m)$ where $\mathbf{f} = (f_1, f_2, \dots, f_m)'$ is the column vector of the latent factors.

The number of latent factors is determined by the Kaiser–Guttman rule, which retains only those factors associated with eigenvalues of the covariance matrix of the standardized covariates greater than one (Kaiser, 1960).⁴¹ Across 24 years, the number of latent factors varies from six to nine, with an average of eight, and the model (12) shows a good fit with a Root Mean Square residual (RMS) from 0.02 to 0.04.⁴² Given the good fit of the factor analysis model, it is of interest to investigate how the information in the covariates considered is aggregated by the latent factors. For this purpose, we calculate the *communality* of each standardized covariate Z^j , which represents the proportion of its variance explained by the common factors.

Specifically, model (12) implies that the part of Z^j not explained by the factors is

$$\text{Var}(\epsilon_j) = \text{Var}(Z^j) - \text{Var}\left(\sum_{i=1}^m l_{ji} f_i\right) = 1 - \sum_{i=1}^m l_{ji}^2. \quad (13)$$

Consequently, Z^j is better explained by the factors when the quantity $\sum_{i=1}^m l_{ji}^2$, called the communality, is higher. This communality quantifies the extent to which a covariate shares information with the extracted latent factors. From a variance perspective, it represents the proportion of the covariate’s variance attributable to a linear combination of the latent factors.

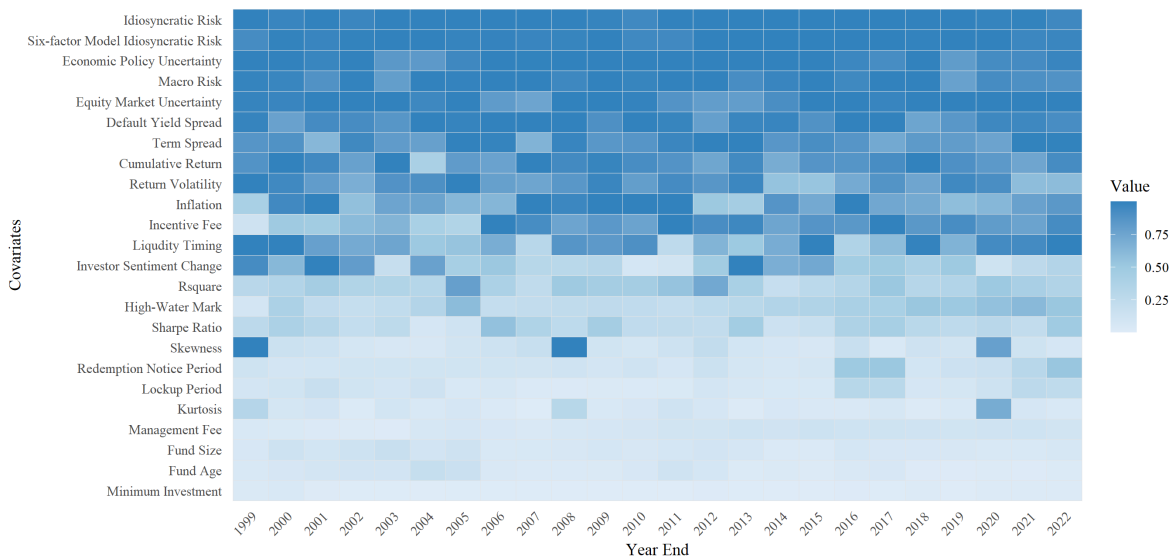
Figure III shows the communality of each covariate over time. To construct this figure, we calculate each covariate’s communality each time we conduct the factor analysis, from

⁴¹This quantitative approach is among the most commonly used in the literature (Goretzko et al., 2021). The rationale for this criterion is that an eigenvalue represents the variance explained by a factor. Since each covariate has a maximum variance of one, selecting factors with eigenvalues above one ensures that each retained factor explains at least as much variance as a single covariate on average.

⁴² $\text{RMS} = \sqrt{\frac{\sum_{i=1}^{d-1} \sum_{j=i+1}^d (r_{ij} - \hat{r}_{ij})^2}{d(d-1)/2}}$ where r_{ij} and $\hat{r}_{ij}$ are the observed correlation and the correlation reproduced by the factor model of covariates Z^i and Z^j , respectively. As a rule of thumb, the model is a good fit if the RMS is less than 0.05.

the first portfolio formation at the end of 1999 to the final formation at the end of 2022. We then rank the covariates by their average communality across years, from highest to lowest, and display the results in a heat map.

Figure III: Covariate communalities over time. This figure reports the communality of each covariate from the factor analysis conducted at each portfolio formation date, from 1999 to 2022. The communality of a covariate is the proportion of its variance explained by the common latent factors, with values closer to one indicating that the covariate is well explained by the common factor structure and values closer to zero indicating that it contains more covariate-specific information.



Overall, 12 covariates have an average communality above 0.7. They range from the idiosyncratic volatilities of the seven-factor and equity six-factor models of [Bali and Weigert \(2024\)](#) to liquidity timing. These covariates include representatives from all four groups of covariates listed in Table 1, with more than half coming from macroeconomic covariates (7), followed by managerial skill (2), past return-based covariates (2), and fund characteristics (1).⁴³ In contrast, the bottom panel of the figure shows covariates that are least explained

⁴³The macroeconomic covariates include the Economic Policy Uncertainty Index, Macro Risk, Equity Market Uncertainty, Default Yield Spread, Term Spread, Inflation and Liquidity Timing. The managerial skill covariates include the Idiosyncratic Risks of the seven-factor model of [Fung and Hsieh \(2004\)](#) and the six-factor model of [Bali and Weigert \(2024\)](#). The past return-based covariates include the Cumulative Return and Return Volatility, while the fund characteristics include the Incentive Fee.

by the latent factors. In our context, these covariates are less represented in the latent-factor specification, indicating that their information is more idiosyncratic and not fully captured by the common factors. Notably, most of them are fund characteristics (7), followed by past return-based (3), managerial skill (1), and macroeconomic covariates (1).⁴⁴

Our findings suggest that the ability to identify future outperformers is not driven only by past returns or static fund characteristics. Instead, it is closely connected to how funds are exposed to, and potentially respond to, changing economic conditions. At the same time, the fact that the full 24-covariate specification performs better than the latent-factor specification suggests (to be discussed below) that $fwer^+$ benefits from both common information across covariates and idiosyncratic information contained in individual variables.

Table II presents the performance of the $FWER\tau$ portfolios constructed from funds selected by the $fwer^+$ procedure using latent factors extracted from the full set of 24 covariates across different FWER target levels. Thereby, when forming portfolios at the end of each year from 1999 onward, we compute latent factors from 24 covariates and use these factors as new covariates. The number of latent factors is determined by the Kaiser–Guttman rule, which retains only those factors associated with eigenvalues of the covariance matrix of the standardized covariates greater than one (Kaiser, 1960).⁴⁵

Compared with the results obtained using all 24 covariates directly (reported in Panel B of Table 2), the portfolios based on the latent factors generate lower OOS alphas and excess returns. These results suggest that a small number of latent factors cannot fully capture the information contained in the original covariates, implying that covariate-specific information

⁴⁴The fund characteristics include the Redemption Notice Period, Lockup Period, Management Fee, Fund Size, Fund Age, High-Water Mark and Minimum Investment. The past return-based covariates include the Sharpe Ratio, Skewness and Kurtosis. The managerial skill covariate includes the R-square of the seven-factor model, while the macroeconomic covariate includes the Investor Sentiment Change.

⁴⁵This quantitative approach is among the most commonly used in the literature (Goretzko et al., 2021). The rationale for this criterion is that an eigenvalue represents the variance explained by a factor. Since each covariate has a maximum variance of one, selecting factors with eigenvalues above one ensures that each retained factor explains at least as much variance as a single covariate on average.

Table II: Performance of the $FWER\tau$ portfolios using factor analysis. This table reports the OOS performance of $FWER\tau$ portfolios constructed from funds selected by the $fwer^+$ procedure using latent factors extracted from the full set of 24 covariates. At each portfolio formation date, latent factors are obtained via factor analysis and used as new covariates in the $fwer^+$ procedure. The number of latent factors is selected using the Kaiser–Guttman rule. Portfolio performance is reported across FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$. The table reports the annualized Fung and Hsieh (2004) seven-factor alpha, the corresponding Newey–West t -statistic, the annualized excess return, and portfolio-size summary statistics. The portfolio-size summary reports the average, minimum, and maximum number of funds held in the portfolio, as well as the percentage of OOS years in which the portfolio is empty.

τ (%)	FH Alpha (%)	t-statistic	Excess Return (%)	Portfolio size summary			
				Average	Min	Max	Empty rate (%)
0.1	4.7	4.6	5.2	20	0	47	4
0.5	3.8	3.3	4.7	28	0	68	4
1.0	3.6	2.9	4.6	32	0	76	4
5.0	3.2	2.7	4.2	46	0	112	4

also contributes to the performance of $fwer^+$.

Similarly, instead of the latent factors from factor analysis, we use principal components extracted from principal component analysis as the input covariates. Table III reports the OOS performance of the resulting $FWER\tau$ portfolios. Although the principal-component specification performs slightly better than the latent-factor specification, it still underperforms the full 24-covariate specification. Taken together, these findings suggest that $fwer^+$ benefits from exploiting both the common information shared across covariates and their covariate-specific information.

IE. Covariates Screening using Statistical Test of Informativeness

Including uninformative covariates also adds noise to the estimation process. It might therefore be useful to test whether a covariate provides meaningful information. In this section, we introduce the omnibus test developed by Huang et al. (2020). This test assesses the dependence between a covariate and the p -value through a permutation-based approach. Specifically, the test computes a sample statistic by categorizing both the p -values and the covariate into groups to measure their dependence. To obtain the sample statistic, a simple test such as the Chi-square test for dependence between the two newly categorized variables

Table III: Performance of the $FWER\tau$ portfolios using principal components as covariates. This table reports the OOS performance of $FWER\tau$ portfolios constructed from funds selected by the $fwer^+$ procedure using principal components extracted from the full set of 24 covariates. At each portfolio formation date, principal components are obtained from the standardized covariates and used as new covariates in the $fwer^+$ procedure. The number of retained principal components is selected using the Kaiser–Guttman rule. Portfolio performance is reported across FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$. The table reports the annualized Fung and Hsieh (2004) seven-factor alpha, the corresponding Newey–West t -statistic, the annualized excess return, and portfolio-size summary statistics. The portfolio-size summary reports the average, minimum, and maximum number of funds held in the portfolio.

τ (%)	FH Alpha (%)	t-statistic	Excess Return (%)	Portfolio size summary		
				Average	Min	Max
0.1	5.5	6.2	5.6	17	1	53
0.5	5.3	6.3	5.4	21	1	65
1.0	5.0	5.4	5.2	25	1	77
5.0	5.0	5.4	5.3	33	1	100

is used. It then resamples the hypotheses without replacement to generate a distribution of test statistics under the null. The p -value of the omnibus test is obtained by comparing the observed statistic with this resampling distribution.

At the end of each year from 1999 to 2022, we construct portfolios to invest in the following year using a three-step testing procedure. First, we test the informativeness of the twenty-four considered covariates. Second, since we conduct multiple omnibus tests, one for each covariate, we need to correct for multiple testing when selecting the significant informative covariates. Given the small number of tests (a maximum of 24), we apply the procedure of Holm (1979), known as the Holm procedure.⁴⁶ Third, we perform the $fwer^+$ procedure using the informative covariate(s) selected by the Holm procedure to detect truly outperforming funds. We invest equally in this set of selected funds in the following year and record the return. We roll forward the three-step procedure until 2022 to select the funds

⁴⁶Given the small number of tests, two appropriate procedures can be adopted, including those of Bonferroni (1936) and Holm (1979). To control FWER at a target τ , the Bonferroni (1936) procedure simply uses a fixed p -value threshold, τ/M , for all tests, where M is the number of tests. The Holm procedure, which first sorts the p -values in increasing order, $p_1 \leq p_2 \leq \dots \leq p_M$, then rejects a null hypothesis i if $p_i \leq \tau/(M - i + 1)$. The latter is more powerful and is thus used. Results using the procedure of Bonferroni (1936) are similar and are available upon request.

invested in 2023.

We exhibit the OOS performance of the new $FWER\tau$ portfolios in Table IV. The table present the results for portfolios where informative covariates are selected using the Holm procedure with FWER controlled at 1%. We consider a range of τ values from 0.1% to 5%. We find that all resulting portfolios deliver positive and statistically significant alphas, ranging from 5.3% to 5.5%. Compared to Panels A and B of Table 2, the gains are higher than those of the average of the portfolios using individual covariates but lower than those of the one using all covariates.

Table IV: OOS performance of $FWER\tau$ portfolios using significantly informative covariates. The table reports OOS performance metrics of the $FWER\tau$ portfolios constructed using significantly informative covariates identified by the omnibus test. Informative covariates are selected using the Holm procedure with FWER controlled at 1%. Portfolio performance is reported across FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$. The table reports the annualized Fung and Hsieh (2004) seven-factor alpha, the corresponding Newey–West t -statistic, the annualized excess return, and portfolio-size summary statistics. The portfolio-size summary reports the average, minimum, and maximum number of funds held in the portfolio.

τ (%)	FH Alpha (%)	t-statistic	Excess Return (%)	Portfolio size summary		
				Average	Min	Max
0.1	5.5	6.6	6.0	24	2	60
0.5	5.4	5.7	6.0	32	2	78
1.0	5.4	5.3	6.0	36	2	92
5.0	5.3	5.4	5.9	49	2	134

IF. Boosting the Informativeness of Covariates via Machine Learning

In this section, we investigate whether the informativeness of covariates can be enhanced using machine learning techniques. The main idea is to generate new covariates that better target the future expected returns of funds. Specifically, we first employ machine learning models to predict the future alphas of funds and then use these predicted alphas as new covariates. We consider four well-known machine learning models: the least absolute shrinkage and selection operator (LASSO; Tibshirani 1996), random forest (RF; Breiman

2001), stochastic gradient boosting (GB; Friedman 2002), and deep neural network (DNN; LeCun et al. 2015).

Formally, the relationship between the fund’s future alpha over the period $t + 1, \dots, t + h$ and the covariate information available at the end of month t can be modelled as

$$\tilde{\alpha}_{i,t \rightarrow t+h} = f_t(X_{i,t}) + \tilde{\epsilon}_{i,t}, \quad (14)$$

where $\tilde{\alpha}_{i,t \rightarrow t+h}$ is the seven-factor alpha of fund i calculated using returns from month $t + 1$ to $t + h$, $X_{i,t}$ denotes the realized covariates of fund i measured at the end of month t , f_t is an unknown function capturing the relationship between $X_{i,t}$ and the future alpha $\tilde{\alpha}_{i,t \rightarrow t+h}$, and $\tilde{\epsilon}_{i,t}$ is the noise term.

Consistent with the choice of our IS horizon and rolling window, to predict the alpha over the period from month $t + 1$ to $t + 12$ for some t , we use data from the previous three years ending at month t to train the model (14) and generate predictions. Specifically, we train the model using the target variable $\tilde{\alpha}_{i,k \rightarrow k+12}$ and features $X_{i,k}$ for $k = t - 36, t - 24$, and $t - 12$, all calculated within the IS period. Then, we apply the trained model to the features $X_{i,t}$ observed at the end of month t to obtain the predicted alpha over the period $t + 1$ to $t + 12$.⁴⁷

Aligned with our baseline portfolios, the first predicted alpha corresponds to the year 1999. This predicted alpha, computed using data up to December 1999, is used as an input covariate in the $fwer^+$ procedure to select funds for investment in the year 2000.⁴⁸ We then roll the sample forward and rebalance the portfolios annually, following the same procedure as for the $FWER\tau$ portfolios described in Section 6 of the main manuscript.

Table V reports the OOS performance of the $FWER\tau$ portfolios using the predicted alphas from each machine learning model as a covariate. We find that these portfolios

⁴⁷We follow Wu et al. (2021) in tuning the hyperparameters of the models.

⁴⁸To obtain the first predicted alpha at the end of 1999, we extend the data back to January 1995 to calculate the input covariates required for training the machine learning model.

Table V: OOS performance of $FWER\tau$ portfolios using machine learning-based covariates. This table reports the OOS performance of $fwcr^+$ portfolios constructed using machine-learning-based covariates across FWER target levels $\tau \in \{0.1\%, 0.5\%, 1\%, 5\%\}$. The covariates are one-year-ahead predicted alphas generated from LASSO (Panel A), gradient boosting (GB; Panel B), random forest (RF; Panel C), and deep neural network (DNN; Panel D) models. Panel E reports the performance of $fwcr^+$ portfolios using all four machine-learning-based covariates jointly. Portfolio alphas are estimated using the [Fung and Hsieh \(2004\)](#) seven-factor model. Monthly returns and alphas are annualized. Newey–West t -statistics are reported in the “ t -statistic” column. The portfolio-size summary reports the average, minimum, and maximum number of funds held in the portfolio, together with the empty rate, defined as the percentage of OOS years in which the portfolio is empty.

τ (%)	FH Alpha (%)	t-statistic	Excess Return (%)	Portfolio size summary			
				Average	Min	Max	Empty rate (%)
Panel A: $FWER\tau$ with the covariate using LASSO							
0.1	3.9	5.5	4.3	13	2	34	0
0.5	4.4	6.3	4.8	17	3	41	0
1.0	4.3	5.8	4.7	20	5	46	0
5.0	3.8	5.0	4.3	26	7	61	0
Panel B: $FWER\tau$ with the covariate using GB model							
0.1	4.0	3.7	4.8	12	0	34	8
0.5	4.4	4.1	5.2	15	0	43	8
1.0	4.6	4.3	5.4	18	0	46	8
5.0	4.6	4.6	5.3	23	0	61	8
Panel C: $FWER\tau$ with the covariate using RF model							
0.1	5.4	4.6	5.9	13	0	34	4
0.5	4.9	3.9	5.5	16	0	41	2
1.0	5.0	3.8	5.6	18	0	46	2
5.0	4.4	3.4	5.1	24	0	61	2
Panel D: $FWER\tau$ with the covariate using DNN model							
0.1	4.1	4.8	4.3	12	1	34	0
0.5	3.7	3.7	4.1	16	1	41	0
1.0	3.8	3.9	4.1	18	1	45	0
5.0	3.6	3.7	4.1	25	1	62	0
Panel E: $FWER\tau$ with use of all covariates generated by the four models							
0.1	5.4	6.6	5.5	12	0	36	8
0.5	5.3	6.6	5.7	15	0	44	4
1.0	5.2	6.4	5.5	17	0	46	4
5.0	5.0	6.1	5.4	22	0	61	4

perform generally similarly to the average of the underlying individual ones. When all four machine learning-based covariates are used simultaneously, the performance is improved but underperforms the one obtained from using all 24 underlying covariates directly.

References

- Bali, T.T., Weigert, F., 2024. Hedge funds and the positive idiosyncratic volatility effect. *Review of Finance* 28, 1611–1661.
- Bonferroni, C.E., 1936. *Teoria statistica delle classi e calcolo delle probabilita*. Bagni di Lucca, Italy: Libreria Internazionale Seeber .
- Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- Friedman, J.H., 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38, 367–378.
- Fung, W., Hsieh, D.A., 2004. Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal* 60, 65–80.
- Goretzko, D., Pham, T.T.H., Bühner, M., 2021. Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology* 40, 3510–3521.
- Hansen, P.R., 2005. A test for superior predictive ability. *Journal of Business & Economic Statistics* 23, 365–380.
- Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.
- Hsu, P.H., Hsu, Y.C., 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.
- Huang, J., Bai, L., Cui, B., Wu, L., Wang, L., An, Z., Ruan, S., Yu, Y., Zhang, X., Chen, J., 2020. Leveraging biological and statistical covariates improves the detection power in epigenome-wide association testing. *Genome Biology* 21, 1–19.

- Kaiser, H.F., 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20, 141–151.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Politis, D.N., Romano, J.P., 1994a. *Journal of the American Statistical Association* 89, 1303–1313.
- Politis, D.N., Romano, J.P., 1994b. The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313.
- Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73, 1237–1282.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267–288.
- Wu, W., Chen, J., Yang, Z.B., Tindall, M.L., 2021. A cross-sectional machine learning approach for hedge fund return prediction and selection. *Management Science* 67, 4577–4601.