



# City Research Online

## City St George's, University of London

**Citation:** Chen, T., Xing, S., Yang, C. & Pilbeam, K. (2026). Do active Chinese equity fund managers produce positive alpha? A comprehensive performance evaluation. *Journal of International Financial Markets, Institutions and Money*, 109, 102312. doi: 10.1016/j.intfin.2026.102312

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/37057/>

**Link to published version:** <https://doi.org/10.1016/j.intfin.2026.102312>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Journal of International Financial Markets, Institutions & Money

journal homepage: [www.elsevier.com/locate/intfin](http://www.elsevier.com/locate/intfin)

## Do active Chinese equity fund managers produce positive alpha? A comprehensive performance evaluation<sup>☆</sup>

Tingting Cheng<sup>a</sup>, Shuo Xing<sup>b,c</sup>, Cheng Yan<sup>d</sup>, Keith Pilbeam<sup>e</sup> <sup>\*</sup><sup>a</sup> School of Finance, Nankai University, Tianjin, 300350, PR China<sup>b</sup> Hua Xia Bank Co., Limited, 100005, PR China<sup>c</sup> PBC School of Finance, Tsinghua University, Beijing, 100083, PR China<sup>d</sup> School of Management, Huazhong University of Science and Technology, Wuhan, 430074, PR China<sup>e</sup> Department of Economics, City St George's, University of London, EC1V 0HB, UK

### ARTICLE INFO

#### JEL classification:

C10  
G11  
G23

#### Keywords:

Mutual fund performance  
Skill vs. luck  
False discovery  
Bootstrap  
Simulation

### ABSTRACT

This paper evaluates the performance of mutual funds in China with the bootstrap-based false discovery rate (FDR) method based on a battery of factor models. We find robust evidence of a significantly higher proportion of skilled funds in China (19.25 percent) than is found for developed countries in the existing literature. We also examine the heterogeneity across sub-samples of different fund styles and find positive alphas for 27 percent of growth funds, 14.56 percent for balance-oriented funds and 11.6 percent for value-oriented funds. We complement the FDR accuracy assessment literature by validating the applicability of the FDR method through elaborate simulations.

### 1. Introduction

Identifying skilled funds has received substantial attention (e.g., [Cheng and Yan, 2017](#); [Zhang and Yan, 2018](#); [Yan and Cheng, 2019](#)) and can be seen as an indirect test of the Efficient Market Hypothesis (EMH) (e.g., [Glode and Green, 2011](#); [Pástor and Stambaugh, 2002](#); [Cai et al., 2018](#)). Mutual funds usually promote themselves as generating risk-adjusted returns through their active trading ([Huang et al., 2020, 2021](#)). However, the majority of existing studies claim that most fund managers could not outperform the market in a persistent way in the United States (U.S.) or other developed markets (e.g., [Baks et al., 2001](#); [Cuthbertson et al., 2012](#); [Blake et al., 2017](#); [Cai et al., 2018](#); [Cheng et al., 2021](#)). However, the conclusions in relation to developed markets do not necessarily apply to developing countries such as China which have a different mix of retail and institutional ownership and may be less informationally efficient (e.g., [Dyck et al., 2013](#); [Jacobs, 2016](#); [Yan and Zhang, 2017](#); [Yang et al., 2019](#)).

The Chinese economy is the largest in terms of purchasing power parity (PPP), accompanied by the second largest stock market globally ([Allen et al., 2022](#)). As such, the Chinese stock market has emerged as a highly prominent research area in recent studies ([Allen et al., 2019, 2023](#); [Li et al., 2024](#)). The Chinese equity market is a particularly interesting one to study mutual fund managers' ability owing to its rapid growth and the dominance of retail investors. As of the end of 2020, according to the Investment Company Institute, the total value of assets under management of open ended mutual funds in China had a capitalization of \$2.81

<sup>☆</sup> We are grateful to the financial support from the National Natural Science Foundation of China (72173068, 71803091), Social Science Development Fund of Nankai University (ZB22BZ0321) and the Fundamental Research Funds for the Central Universities, China. Tingting Cheng and Shuo Xing are the co-first authors.

\* Corresponding author.

E-mail address: [k.s.pilbeam@city.ac.uk](mailto:k.s.pilbeam@city.ac.uk) (K. Pilbeam).

<https://doi.org/10.1016/j.intfin.2026.102312>

Received 14 March 2025; Accepted 25 February 2026

1042-4431/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

trillion making it the largest in the Asia-Pacific region. The Chinese equity market has a predominance of relatively unsophisticated retail investors and this means that there is the potential for more sophisticated institutional investors to have a persistent edge and the potential to generate a positive alpha. Prior studies find that in the young and still-developing Chinese A-share market, mutual funds can add value for investors and help move stock prices closer to underlying fundamentals (Chi et al., 2022; Ammer et al., 2022).

An important and recurrent issue in the literature on the performance of active mutual funds is distinguishing skill from luck (Kosowski et al., 2006; Hsu et al., 2025). Existing studies on the performance evaluation of mutual funds in China fail to comprehensively account for biases arising from nonnormal distributions and false discoveries. To rectify this deficiency, we utilize a mutual fund performance evaluation framework. Inspired by the recent study of Liu et al. (2019), we adopt their China-four-factor (hereafter LSY-CH-4) model as the benchmark model. Furthermore, we argue that employing a bootstrap-based False Discovery Rate (FDR) method can effectively address biases stemming from both non-normal distributions and false discoveries, thereby mitigating the aforementioned critiques.

Our approach employs an intra-fund bootstrap method from Cai et al. (2018) (hereafter CCY) and the FDR method from Barras et al. (2010) (hereafter BSW) to effectively distinguish skill from luck bias caused by the non-normal distribution and the multiple hypothesis testing problem. The bootstrap method is widely used in fund performance evaluation (e.g., Kosowski et al., 2006; Cuthbertson et al., 2008; Fama and French, 2010; Blake et al., 2017; Cai et al., 2018). Considering the low cross-sectional dependencies of fund residuals argued by BSW, this paper constructs a simple intra-fund bootstrap method following CCY which can remedy the luck bias caused by non-normal distribution for each fund in isolation to avoid complications. Based on the results obtained by the intra-fund bootstrap method, we use the FDR method proposed by BSW to distinguish the skill from luck bias caused by multiple hypothesis testing problem (Type I error).<sup>1</sup> If a mutual fund with a true zero-alpha is shown to be significant in the test, it will be assigned to be skilled/bad-skilled (good/bad luck), leading to a misleading conclusion. The false discovery rate is defined as the proportion of funds having significant alphas simply due to luck. As fund performance evaluation is essentially a multiple hypothesis testing problem, the false discovery rate grows with the number of hypotheses being tested. The FDR method has been frequently adopted in fund performance evaluation and other related fields (e.g., Bajgrowicz and Scaillet, 2012; Kim and In, 2012; Cuthbertson and Nitzsche, 2013).

We obtained the data of fund returns from WIND and China's risk factors from Liu et al. (2019), respectively. Specifically, we collect an unbalanced panel of monthly returns of 844 open-ended equity mutual funds (net of the trading cost and expenses) over the period January 2002 to June 2020, free of survivorship bias. As the first open-ended equity fund in China was launched in 2001, we restrict the beginning of our sample from 2002. We investigate the Chinese mutual fund performance and its heterogeneity across different fund styles. Over the full sample, we estimate that there exists a relatively larger amount of skilled funds than in developed countries with 19.25% of skilled mutual funds from 2002 to 2020. Specifically, we find that the growth-oriented mutual funds have the highest proportion (27.00%) of positive alpha, which is followed by balance-oriented mutual funds (14.56%) and value-oriented mutual funds (11.60%). These findings survive various robustness tests.

Our results reveal a better performance of the Chinese mutual funds than those in developed markets. This finding is consistent with existent literature (Chi et al., 2022; Chi, 2013; Koutmos et al., 2020). This paper provides deeper insights and methodological advancements, thereby significantly differentiating and complementing the existing literature. First, this study innovatively integrates the bootstrap method and the FDR approach within a unified analytical framework, substantially enhancing the robustness of our findings. Second, compared to commonly used international models (e.g., the Fama–French model), the LSY-CH-4 factor model we used is in characteristics more reflective of China's stock market (such as its policy-driven nature, retail investor dominance, and unique risk factors), yielding a more reliable assessment of fund manager skill in China. Third, we provide a dedicated analysis of the underlying reasons contributing to this higher observed proportion of skilled funds in the Chinese market. Finally, we address existing critiques (Andrikogiannopoulou and Papakonstantinou, 2019) regarding the applicability of the FDR method. Through simulation analysis, we systematically demonstrates that the FDR method serves as an effective and robust tool for identifying skilled funds within the specific data structure and characteristics of China's fund market.

We confirm that the unique nature of the Chinese capital market may provide more opportunities for mutual funds to beat the market from the perspective of equilibrium accounting following Fama and French (2010). Unlike in the U.S. equity market, the aggregate of active equity mutual funds in China is not a zero-sum game (aggregate  $\alpha$  is zero) for gross returns and is not a negative-sum game (aggregate  $\alpha$  is negative) for net returns, which does not echo the equilibrium accounting. Unlike in the U.S., retail investors, instead of institutional investors, dominate the turnover of the Chinese equity market. As Tang et al. (2012) argue, if a mutual fund wants to beat the market in China, it only needs to defeat a few rivals, which is a relatively easy task. Therefore, it is possible for mutual funds in China to generate positive alphas at the expense of other active investors such as retail investors.

In their work on mutual fund performance, Andrikogiannopoulou and Papakonstantinou (2019) question the accuracy of the FDR method. To address this concern, we conduct simulations where we can explicitly control the specific characteristics to investigate the accuracy of bootstrap-based FDR method: (i) We find that the bias of the FDR method does not affect the accuracy of our results. (ii) We provide a more reasonable assessment by comparing the result of the FDR method with the conventional single hypothesis t-test method, and confirm the applicability of the FDR method to the Chinese fund market. (iii) By comparing the

---

<sup>1</sup> In manager selection (or equivalently the selection of factors or trading strategies), investors can make two types of mistakes. The first involves selecting a manager who turns out to be unskilled: this is a Type I error, or a false positive. The second error is not selecting or missing a manager the investor thought was unskilled, but was not: this is a Type II error, or a false negative (Harvey et al., 2020).

simulation results for the U.S. and mutual funds in China, we find support for the applicability of the FDR method for mutual funds in China. In addition, we also provide an explanation for the technically inevitable over-modification of the FDR method, which is often overlooked by previous studies. We add to the existing literature by providing a comprehensive analysis of the performance of Chinese mutual funds, in which there is a much higher percentage of retail investors, who are less financially sophisticated and hence provide more opportunities for institutional investors such as mutual funds to make profits. The contribution of our work is to document that the Chinese mutual funds perform better than funds in developed markets.

We add to the literature by examining the accuracy of the FDR method using elaborate simulations. [Andrikogiannopoulou and Papakonstantinou \(2019\)](#) use simulations to examine the accuracy of the FDR estimator in the context of fund performance and find that the result is more biased as the proportion of zero alpha and the location of nonzero alpha decrease. Extending this line of studies, we demonstrate that the FDR method exhibits greater reliability compared to conventional single hypothesis t-test outcomes, particularly within the context of China's mutual fund market.

We also contribute to the literature on efficiency in the Chinese market. We find evidence that rejects evidence of market efficiency because the unique nature of the Chinese capital market, provides more opportunities for mutual funds to produce positive alpha. Our evidence suggests that the market efficiency of China is lower than in the case of developed markets. The existing literature (e.g., [Dyck et al., 2013](#); [Jacobs, 2016](#); [Yan and Zhang, 2017](#); [Yang et al., 2019](#)) mainly provides evidence from direct tests. We complement this strand of literature by conducting an indirect test of market efficiency by showing the existence of a positive alpha for a significant number of active mutual fund managers in the Chinese equity market.

The rest of this paper is organized as follows. Section 2 describes the bootstrap-based false discovery rate methods. Section 3 introduces our data and empirical results. Section 4 reports our simulations. Section 5 performs some robustness checks. Section 6 concludes.

## 2. Methodology

This section describes our baseline asset pricing model and then presents our proposed bootstrap-based FDR procedure.

### 2.1. Baseline asset pricing model

According to the recent literature (e.g., [Cheng and Yan, 2017](#); [Zhang and Yan, 2018](#); [Cai et al., 2018](#); [Yan and Cheng, 2019](#)), a fund with positive (negative) alpha performs well (poorly) on a risk-return basis, otherwise the fund has zero alpha. The estimated alpha is obtained by factor models. In this study, we employ the LSY-CH-4 model proposed by [Liu et al. \(2019\)](#) as our baseline asset pricing model:

$$r_{it} = \alpha_i + \beta_{1i} \cdot \text{MKT}_t + \beta_{2i} \cdot \text{SMB}_t + \beta_{3i} \cdot \text{VMG}_t + \beta_{4i} \cdot \text{PMO}_t + \epsilon_{it}, \quad (1)$$

where  $r_{it}$  denotes the excess return of fund  $i$  over the risk-free rate in month  $t$ .  $\text{MKT}_t$ ,  $\text{SMB}_t$ ,  $\text{VMG}_t$  and  $\text{PMO}_t$  denote the LSY-CH-4 factors, which are the Market excess return (MKT) factor, the Small-Minus-Big (SMB) size factor, the Value-Minus-Growth (VMG) value factor and the Pessimistic-Minus-Optimistic (PMO) sentiment factor at time  $t$ , respectively. For the size factor, the smallest 30% of firms are excluded to avoid shell-value contamination.<sup>2</sup> Comparing four candidate valuation ratios such as earnings-price (EP), book-to-market (BM), asset-to-market, and cash-flow-to-price ratios, [Liu et al. \(2019\)](#) construct the value factor based on the EP ratio, which conforms within China most. Given this sentiment-based motivation, the fourth factor is constructed using abnormal turnover, which is the past month's share turnover divided by the past year's turnover. Alpha ( $\alpha_i$ ) is the fund skill indicator in the fund performance evaluation.

### 2.2. The bootstrap-based FDR method

The bootstrap-based FDR method involves two steps. In the first step, we construct a simple intra-fund bootstrap method following CCY to generate simulated samples to approximate the true distribution of alphas, and thus remedy the luck bias caused by non-normal distribution for each fund in isolation. In the second step, we apply the false discovery rate method to remedy the luck bias caused by multiple hypothesis testing problem based on the results of step 1. In what follows, we expand on each of these two steps.

#### 2.2.1. Step 1: Intra-fund bootstrap

The normality assumption used in the conventional single hypothesis t-test of mutual fund performance evaluation may be too strict to hold for real data. [Kosowski et al. \(2006\)](#) demonstrate that the cross-section of mutual fund alphas follows a complex nonnormal distribution, and they suggest a bootstrap statistical technique to separate luck from skill using a sample of U.S. open-ended domestic equity mutual funds from 1975 to 2002. Following [Kosowski et al. \(2006\)](#), some studies (e.g., [Cuthbertson et al., 2008](#); [Fama and French, 2010](#)) also use the bootstrap method for mutual fund performance evaluation for alternative samples of funds.

<sup>2</sup> According to [Liu et al. \(2019\)](#), the smallest listed firms are the most likely shells in China. They find that when compared to other firms, the smallest 30% of firms have returns less related to operating fundamentals, proxied by earnings surprises, but more related to IPO activity.

It is necessary to select a suitable bootstrap method given that a variety of alternative bootstrap methods could be used which may lead to different conclusions. For example, [Harvey and Liu \(2022\)](#) argue that [Kosowski et al. \(2006\)](#) and [Fama and French \(2010\)](#) both evaluate the mutual fund performance in the U.S. but disagree with each other. Given the low cross-sectional dependencies of fund residuals as argued by BSW, we follow CCY and construct a simple intra-fund bootstrap method to avoid complications. The specific procedure employed is as follows:

- (1) For fund  $i, i = 1, 2, \dots, M$ , estimate the LSY-CH-4 model (1) and obtain residuals,  $\{\hat{e}_{it}\}_{t=1}^{T_i}$ .

$$r_{it} = \hat{\alpha}_i + \hat{\beta}'_i X_t + \hat{e}_{it}, \tag{2}$$

where  $\hat{\alpha}_i$  denotes the estimated alpha of fund  $i, t = 1, 2, \dots, T_i, T_i$  is the total number of observations for fund  $i, r_{it}$  is the return of fund  $i, X_t$  is a vector of risk factors (MKT, SMB, VMG, and PMO).

- (2) Draw the empirical distribution of residuals randomly with replacement to generate the bootstrap residuals  $\{\hat{e}^*_{it}\}_{t=1}^{T_i}$ . Use the new pseudo-times series of re-sampled residuals  $\{\hat{e}^*_{it}\}_{t=1}^{T_i}$  and  $\hat{\beta}_i$  to generate a time series of pseudo-monthly excess returns, denoted by  $r^*_{it}$ . The alphas are imposed to be 0.

$$r^*_{it} = \hat{\beta}'_i X_t + \hat{e}^*_{it}. \tag{3}$$

- (3) Re-estimate model (1) using  $\{r^*_{it}\}$  and obtain simulated alphas for each fund, denoted by  $\hat{\alpha}^*_i$ .
- (4) Replicate Steps (2)–(3)  $B$  times, and denote the estimated alpha for the  $b$ th bootstrap sample as  $\hat{\alpha}^*_b, b = 1, 2, \dots, B$ . Then we will get the simulated distribution of  $\alpha$  for the true zero alpha funds.
- (5) Compare the estimated alpha of fund  $i, \hat{\alpha}_i$ , and distribution of  $\hat{\alpha}^*$  to observe the fund's skill level. Let  $\alpha_q$  denote the quantile according to a specific significance level. If  $|\hat{\alpha}_i| > \alpha_q$ , we consider that fund  $i$  is skilled or bad-skilled. If  $|\hat{\alpha}_i| < \alpha_q$ , fund  $i$  is considered to be zero alpha fund. Alternatively, we can compute the simulated  $p$ -value below and use the  $p$ -value to observe fund skill.

$$\hat{p}_i = 2 \cdot \min \left( \frac{1}{B} \sum_{b=1}^B I\{\hat{\alpha}^*_b \geq \hat{\alpha}_i\}, \frac{1}{B} \sum_{b=1}^B I\{\hat{\alpha}^*_b \leq \hat{\alpha}_i\} \right).$$

For instance, at significance level 0.05, when  $p$ -value  $\leq 0.05$  and  $\hat{\alpha}_i > 0$ , we regard the sample as the skilled fund. When  $p$ -value  $\leq 0.05$  and  $\hat{\alpha}_i < 0$ , we regard the sample as the bad-skilled fund. If the  $p$ -value is larger than 0.05, we do not reject the null and deem the fund to be unskilled.

### 2.2.2. Step 2: FDR

In step 1, the bootstrap method provides a modification of luck bias caused by the non-normal distribution for each individual fund; next, in step 2 the false discovery rate method modifies the luck bias caused by multiple hypothesis testing problem based on the results of step 1.

Traditionally, for single tests, we deem funds with significantly positive (negative) alpha to be skilled (bad-skilled) funds by choosing a significance level  $\gamma$ . If a mutual fund with true zero alpha is significant in the test, it will be deemed as either skilled (good luck) or bad-skilled (bad luck) fund, which is called Type I error and the false discovery error rate is  $\gamma$ . It means that for some funds, we may erroneously reject the null hypothesis of zero alpha at the significance level  $\gamma$ .

However, as explained in [Harvey et al. \(2020\)](#), for multiple tests, the false discovery rate will be much higher than  $\gamma$ . Separating skill from luck to estimate the proportion of truly positive alpha funds merits more attention in multiple tests. To control for luck in multiple tests, [Benjamini and Hochberg \(1995\)](#), [Storey \(2002\)](#) proposed the concept of false discovery rate (FDR), which is defined as the expected proportion of funds having significant alphas (positive/negative) simply due to luck (good/bad). BSW apply the FDR method to evaluate fund performance. Here we provide the computation procedures of the FDR method briefly, and full details are in BSW ([Barras et al., 2010](#)).

To start with, we divide actively managed mutual funds into three distinct groups: bad-skilled funds ( $\alpha < 0$ ), zero alpha funds ( $\alpha = 0$ ), and skilled funds ( $\alpha > 0$ ). Among a group of  $M$  mutual funds, the conventional single hypothesis t-test method for fund  $i, (i = 1, 2, \dots, M)$  is:

$$H_0 : \alpha_i = 0 \text{ versus } H_A : \alpha_i > 0 \text{ or } \alpha_i < 0$$

At a given significance level  $\gamma$ , the probability of a zero alpha fund exhibiting luck (good/bad) equals  $\gamma/2$  since it is a two-sided test with an equal-tailed significance level. Let  $\pi_0, \pi_A^+$ , and  $\pi_A^-$  denote the proportion of true zero alpha, skilled and bad-skilled funds in the population, respectively, and the corresponding estimates are denoted as  $\hat{\pi}_0, \hat{\pi}_A^+$  and  $\hat{\pi}_A^-$ . Let  $F_\gamma^+$  ( $F_\gamma^-$ ) denote the proportion of true zero alpha funds but discovered with significantly positive (negative) alpha, then  $F_\gamma^+ = F_\gamma^- = \pi_0 \cdot (\gamma/2)$ . Then the expected proportion of “lucky funds” and “unlucky funds” denoted by  $E(F_\gamma^+)$  and  $E(F_\gamma^-)$  equals:

$$E(F_\gamma^+) = \hat{\pi}_0 \cdot (\gamma/2), \quad E(F_\gamma^-) = \hat{\pi}_0 \cdot (\gamma/2). \tag{4}$$

Let  $T_\gamma^+$  ( $T_\gamma^-$ ) denote the proportion of funds with truly positive (negative) alphas,  $E(S_\gamma^+)$  ( $E(S_\gamma^-)$ ) represent the proportion of significantly positive (negative) alpha funds obtained by the bootstrap method in Section 2.2.1.<sup>3</sup> Then the expected proportion of

<sup>3</sup> [Barras et al. \(2010\)](#) use a bootstrap approach to more accurately estimate the distribution of t-statistics for each fund and their associated p-values. In this paper, the  $E(S_\gamma^+)$  and  $E(S_\gamma^-)$  are calculated based on the alphas we obtained by the intra-fund bootstrap method in Step 1, which can eliminate the bias caused by non-normal distribution.

truly skilled (bad-skilled) funds, denoted by  $E(T_\gamma^+)$  ( $E(T_\gamma^-)$ ), is:

$$E(T_\gamma^+) = E(S_\gamma^+) - E(F_\gamma^+) = E(S_\gamma^+) - \hat{\pi}_0 \cdot (\gamma/2). \tag{5}$$

$$E(T_\gamma^-) = E(S_\gamma^-) - E(F_\gamma^-) = E(S_\gamma^-) - \hat{\pi}_0 \cdot (\gamma/2). \tag{6}$$

Next, one need to choose  $\gamma$  at different levels to make  $E(T_\gamma^+)$  and  $E(T_\gamma^-)$  converge to  $\pi_A^+$  and  $\pi_A^-$ . The estimators for the proportion of funds with positive and negative alphas are:

$$\hat{\pi}_A^+ = E(T_{\gamma^*}^+), \quad \hat{\pi}_A^- = E(T_{\gamma^*}^-), \tag{7}$$

where  $\gamma^*$  is the optimal significance level which can be determined using a mean squared error(MSE) criterion.

Let  $FDR_\gamma^+$  ( $FDR_\gamma^-$ ) denote the proportion of falsely discovered skilled (bad-skilled) funds. If the  $FDR_\gamma^+$  is 100%, then all the observed skilled funds can beat the market because of luck instead of the outperformed ability. Likewise, if the  $FDR_\gamma^-$  is 0%, then none of the funds are misjudged and the outstanding performance is all due to the managers' ability. The estimators of  $FDR_\gamma^+$  and  $FDR_\gamma^-$  are:

$$\widehat{FDR}_\gamma^+ = E(F_\gamma^+)/E(S_\gamma^+) = \hat{\pi}_0 \cdot (\gamma/2)/E(S_\gamma^+). \tag{8}$$

$$\widehat{FDR}_\gamma^- = E(F_\gamma^-)/E(S_\gamma^-) = \hat{\pi}_0 \cdot (\gamma/2)/E(S_\gamma^-). \tag{9}$$

From Eqs. (8) and (9), we see that to compute the false discovery rate, the first step requires calculating  $\hat{\pi}_0$ . Since p-values under the alternative hypothesis cluster around zero, while p-values of truly null are uniformly distributed over the interval [0,1], the calculation of  $\hat{\pi}_0$  is straightforward. Most of the p-values larger than a sufficiently high threshold, denoted by  $\lambda$ , come from zero alpha funds, the simple method to estimate  $\pi_0$  is to choose an optimal value of  $\lambda$ , denoted by  $\lambda^*$ , beyond which the histogram of the p-value is nearly flat (e.g.,  $\lambda^* = 0.6$ , as shown in BSW). The estimate of  $\pi_0$  at  $\lambda^*$ ,  $\hat{\pi}_0(\lambda^*)$ , is given by :

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

where  $\widehat{W}(\lambda^*)$  equals the number of funds of which p-values exceed  $\lambda^*$ .  $M$  denotes the total number of funds. If we choose an optimal  $\lambda$ , above which the histogram of p-values is perfectly flat, we can get an unbiased estimate of  $\hat{\pi}_0$  by counting the p-values of true null. Bias decreases as  $\lambda$  increases because of the larger proportion of non-zero alpha above  $\lambda$ . However, the variance also increases with  $\lambda$  because fewer p-values are included in our estimation. Therefore, to select  $\lambda^*$ , BSW discussed that the simplest way to determine  $\lambda^*$  is to eyeball the flat portion of the histogram of p-values. However, to be more prudent, they apply a simple MSE-bootstrap procedure introduced by Storey (2002), which minimizes the estimated MSE of  $\hat{\pi}_0$ . More details on the selection of optimal values of  $\lambda$  and  $\gamma$  are presented in Appendix A.

### 3. Empirical results

This section first introduces our fund data sets, then presents respectively the Chinese mutual fund performance evaluation results using: (i) the conventional single hypothesis t-test method without considering the luck (good/bad) bias; and (ii) the bootstrap-based FDR method with eliminating two layers of luck. We also provide some discussion on the results in the last subsection.

#### 3.1. Data and descriptive statistics

We start by identifying the equity mutual funds in mainland China from January 2002 to June 2020, free of survivorship bias, and obtain monthly returns from the WIND database. China's first open-ended equity fund was not launched until 2001, with only three funds established that year. This limited number provides an insufficient sample for robust analysis; therefore, our sample period begins in 2002. The index mutual funds are excluded because we focus on the ability of active management. Besides, for the sake of accuracy, we only keep mutual funds that live at least 30 months.<sup>4</sup> This screening process leaves us with a sample of 844 mutual funds with at least 30 observations for each. The data for risk-free rate ( $R_{ft}$ ) and risk factors (MKT, SMB, VMG, and PMO) are obtained from Liu et al. (2019). We exclude QDII funds, a fund type holding non-RMB priced assets. To ensure the accuracy of outcome, we also drop some funds which are not categorized explicitly according to their investment styles by WIND. All funds are priced in the Chinese currency RMB.

We further use a rolling window scheme to increase the sample size and observe the short-term fund performance. The primary reasons are as follows. On the one hand, the enlarged sample facilitates more direct comparisons with findings from developed markets. Our methodology aligns with established approaches (Barras et al., 2010), which employs 36-month overlapping windows to account for survivorship bias and timeliness concerns. On the other hand, unlike developed markets (e.g., the U.S.), where fund managers typically exhibit longer tenures and more stable performance, the Chinese mutual fund market is characterized by higher

<sup>4</sup> We follow Yan and Cheng (2019) and Hunter et al. (2014) by using 30 months as our threshold as it adds robustness to our results.

**Table 1**

Summary statistics of the risk-free rate and risk factors in China.  $R_f$ , MKT, SMB, VMG, and PMO denote the risk-free rate, the market factor, the size factor, the value factor, and the sentiment factor, respectively. This table reports the mean, minimum, maximum, and standard deviation of each variable from January 2002 to June 2020.

Variable	Mean	Min.	Max.	Std
$R_f$	0.0019	0.0011	0.0034	0.0006
MKT	0.0063	-0.2547	0.2458	0.0778
SMB	0.0046	-0.1720	0.1841	0.0466
VMG	0.0117	-0.1028	0.1517	0.0382
PMO	0.0085	-0.2019	0.1275	0.0362

**Table 2**

Summary statistics of fund returns in China. Monthly returns are summarized for the original sample (left panel) and enlarged sample (right panel), measured in excess of the one-month return of a risk-free rate. The values are at the cutoff points for various quantiles. Nobs is the number of observations, where a minimum of 30 is required for the original sample. For the enlarged sample, the observation is always 36, and hence we omit it for brevity. Mean is the sample mean return, and Std is the sample standard deviation of fund return. The alpha estimates are based on OLS regressions using the LSY-CH-4 factor model for mutual funds.

Pct	Original sample				Enlarged sample		
	Nobs	Mean	Std	$\hat{\alpha}_{OLS}$	Mean	Std	$\hat{\alpha}_{OLS}$
Min.	30	-0.0074	0.0268	-0.0127	-0.0259	0.0042	-0.0251
1%	30	-0.0042	0.0323	-0.0044	-0.0134	0.0347	-0.0146
5%	33	0.0001	0.0448	-0.0006	-0.0084	0.0416	-0.0083
10%	36	0.0031	0.0490	0.0012	-0.0059	0.0454	-0.0055
20%	47	0.0058	0.0560	0.0028	-0.0021	0.0507	-0.0025
30%	59	0.0074	0.0623	0.0043	0.0008	0.0549	-0.0004
Median	85	0.0099	0.0711	0.0071	0.0063	0.0661	0.0031
70%	126	0.0130	0.0772	0.0112	0.0132	0.0830	0.0068
80%	151	0.0147	0.0805	0.0141	0.0172	0.0918	0.0095
90%	169	0.0170	0.0854	0.0175	0.0226	0.1024	0.0134
95%	186	0.0191	0.0906	0.0214	0.0276	0.1105	0.0167
99%	203	0.0228	0.1008	0.0273	0.0389	0.1243	0.0246
Max.	220	0.0346	0.1209	0.0398	0.0599	0.1472	0.0437

manager turnover and more frequent strategy shifts due to regulatory changes and market volatility. The rolling-window approach helps address these unique dynamics by capturing managerial skill over shorter, more relevant periods, thereby mitigating potential biases arising from manager turnover or regime shifts that could distort long-term, fund-level assessments.

The existing literature typically uses five years as the rolling window period for the U.S. market. For example, BSW choose non-overlapping 5-year periods between 1977 and 2006 to create short-term funds. Given our sample interval in the Chinese market is much shorter than the U.S. market, we choose overlapping 3-year periods here.<sup>5</sup> Specifically, we regard 36 months as a cycle and roll over every 12 months to divide the time series into many subintervals. That is, if the beginning month of fund  $i$  is  $a_1$ , then for  $n$ th window ( $n = 1, 2, \dots, N_i$ ,  $N_i$  is the total number of subgroups of fund  $i$ ), it covers period from  $[a_1 + 12(n - 1)]$  to  $[a_1 + 12(n - 1) + 35]$ .<sup>6</sup> Thus, we get 4712 small funds in total. For clarity, we refer to those funds as the enlarged sample.

In addition, we separate the mutual funds into three categories according to their investment styles published by WIND to investigate whether there exists heterogeneity across fund styles.

1. Growth-oriented funds — are funds that aim to achieve long-term appreciation of the fund's assets; typically invest fund assets in the stocks of trustworthy companies with long-term growth prospects.
2. Value-oriented funds — are funds that aim to pursue stable recurring income; mainly invest in blue-chip stocks, bonds and other securities with stable returns.
3. Balance-oriented funds — are funds that aim to obtain current income, but also to pursue long-term appreciation.

The original sample contains 331 growth-oriented funds, 98 value-oriented funds, and 368 balance-oriented funds, while the enlarged sample contains 844 growth-oriented funds, 403 value-oriented funds, and 2241 balance-oriented funds.<sup>7</sup>

<sup>5</sup> We also use the 5-year non-overlapping rolling window and 5-year overlapping window to perform the robustness check. More detailed results can be found in Section 5.

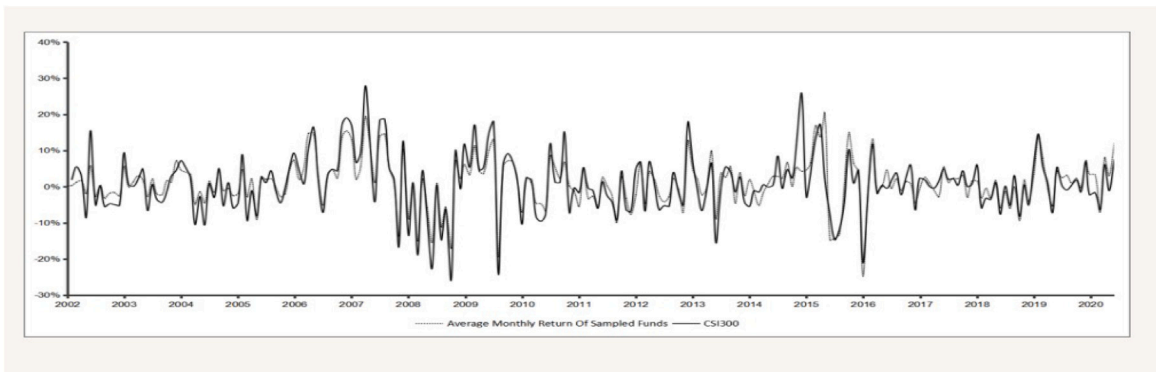
<sup>6</sup> Because the number of observations is not evenly divisible for most of the funds, we roll over every 12 months from the first available observation and abandon the redundant data.

<sup>7</sup> To avoid potential bias, we only retain those funds with clear statements of investment styles. Therefore, the total number of funds from those three investment styles is 787 (=331 + 98 + 368), which is smaller than 844. Similarly, the total number of enlarged samples from those three groups is smaller than 4712 as well.

**Table 3**

Market-level regressions. This table reports the result of regression on equally weighted portfolios in Panel A and the mean value of the estimates obtained by regression for each fund in Panel B.  $\alpha$  is the constant, which is an indicator of skill,  $\beta_1$  is the coefficient of MKT,  $\beta_2$  is the coefficient of SMB,  $\beta_3$  is the coefficient of VMG and  $\beta_4$  is the coefficient of PMO. We show the results for the full sample (844), growth-oriented (331), balance-oriented (368), and value-oriented mutual funds (98). The parameter estimates of all variables in Panel A are significant at the 5% significance level.

Panel A: Regression result of equally weighted portfolio						
Type	$\alpha$ (%)	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$R^2$
Full sample	0.68 (3.518)	0.7811 (33.612)	0.0434 (-1.393)	-0.0858 (0.881)	0.0742 (1.534)	0.86
Growth	0.64 (3.411)	0.7949 (35.436)	0.0999 (-0.898)	-0.0534 (2.101)	0.0888 (1.902)	0.88
Balance	0.62 (3.309)	0.7888 (35.347)	0.0433 (-0.931)	-0.055 (0.914)	0.0684 (1.472)	0.87
Value	-0.06 (3.309)	0.7948 (35.347)	-0.3111 (-0.931)	0.1139 (0.914)	-0.0061 (1.472)	0.72
Panel B: Mean values of estimates from regression for each fund						
Type	$\bar{\alpha}$ (%)	$\bar{\beta}_1$	$\bar{\beta}_2$	$\bar{\beta}_3$	$\bar{\beta}_4$	$R^2$
Full sample	0.85	0.8432	0.0642	-0.2365	-0.0035	0.77
Growth	0.99	0.8674	0.0128	-0.2244	-0.0452	0.76
Balance	0.73	0.8483	0.0733	-0.1826	0.0005	0.77
Value	0.28	0.857	0.1066	0.021	0.0318	0.80



**Fig. 1.** Time series plot of average monthly returns for sampled mutual funds in China. This figure plots the average monthly return of sampled mutual funds and the monthly return of CSI300 from January 2002 to June 2020.

Table 1 provides summary statistics of risk factors. We observe that the mean values of the size, value, and sentiment factors are all positive, indicating that the small-cap, value, and pessimistic stocks can earn relatively higher returns. Table 2 presents summary statistics of fund returns for the original sample (844 funds) and the enlarged sample (4712 funds). The mean values of excess returns range from  $-0.0074$  to  $0.0346$  for the original sample, and from  $-0.0259$  to  $0.0599$  for the enlarged sample. Table 2 also shows the estimated alphas based on the LSY-CH-4 factor model, ranging from  $-0.0127$  to  $0.0398$  for the original sample and from  $-0.0251$  to  $0.0437$  for the enlarged sample, respectively.

In addition, Fig. 1 displays the time series plot of the average monthly return of mutual funds and the CSI300<sup>8</sup> from January 2002 to June 2020. It shows that the trend of the Chinese mutual fund returns is consistent with the trend of the Chinese stock market.

### 3.2. Fund performance evaluation before luck adjustment

In this section, we first look at the fund performance evaluation results without considering the “luck”, that is, using the conventional single hypothesis t-test method.

<sup>8</sup> CSI300 is compiled by China Securities Index Co., Ltd, which is jointly issued by Shanghai and Shenzhen stock Exchanges on April 8, 2005, to reflect the overall trend of the Chinese A-share market. It is composed of 300 securities with large scale and good liquidity in the Shanghai and Shenzhen stock markets.

**Table 4**

Fund-level regressions. This table reports the result of regression for each fund in the original sample (Panel A) and the enlarged sample (Panel B).  $\alpha$  is the constant that indicates skill,  $\beta_1$  is the coefficient of market factor (MKT),  $\beta_2$  is the coefficient of size factor (SMB),  $\beta_3$  is the coefficient of value factor (HML), and  $\beta_4$  is the coefficient of momentum factor (MOM). In each panel, we report the number and proportion of positive and negative values for each coefficient. We also show the corresponding situation for significantly positive and significantly negative coefficients at the 5% significance level.

	Positive		Sig.		Negative		Sig.	
	No.	Pro.	No.	Pro.	No.	Pro.	No.	Pro.
Panel A: Regression result for the original sample								
$\alpha$	785	93.01%	418	49.53%	59	6.99%	1	0.12%
$\beta_1$	844	100.00%	844	100.00%	0	0.00%	0	0.00%
$\beta_2$	528	62.56%	166	19.67%	316	37.44%	36	4.27%
$\beta_3$	199	23.58%	32	3.79%	645	76.42%	311	36.85%
$\beta_4$	460	54.50%	87	10.31%	384	45.50%	384	45.50%
Panel B: Regression result for the enlarged sample								
$\alpha$	3202	67.95%	560	11.88%	1510	32.05%	78	1.66%
$\beta_1$	4712	100.00%	4704	99.83%	0	0.00%	0	0.00%
$\beta_2$	3378	71.69%	1019	21.63%	1334	28.31%	102	2.16%
$\beta_3$	1734	36.80%	243	5.16%	2978	63.20%	747	15.85%
$\beta_4$	2463	52.27%	459	9.74%	2249	47.73%	2249	47.73%

**3.2.1. Market-level regressions**

We start with a comprehensive evaluation of the performance of the whole Chinese mutual fund market. Specifically, we first perform the LSY-CH-4 factor model regression on the return of an equally weighted portfolio through Eq. (11) for the full sample, growth-oriented, balance-oriented, and value-oriented funds. The results are shown in panel A of Table 3.

$$\bar{r}_t = \alpha + \beta_1 \cdot \text{MKT}_t + \beta_2 \cdot \text{SMB}_t + \beta_3 \cdot \text{VMG}_t + \beta_4 \cdot \text{PMO}_t + \epsilon_t, \tag{11}$$

where  $\bar{r}_t = \frac{1}{M} \sum_{i=1}^M r_{it}$ .

Then we perform the regression on the return of each fund through Eq. (12)

$$r_{it} = \alpha_i + \beta_{1i} \cdot \text{MKT}_t + \beta_{2i} \cdot \text{SMB}_t + \beta_{3i} \cdot \text{VMG}_t + \beta_{4i} \cdot \text{PMO}_t + \epsilon_{it}, \tag{12}$$

and define

$$\bar{\alpha} = \frac{1}{M} \sum_{i=1}^M \hat{\alpha}_i, \quad \bar{\beta}_1 = \frac{1}{M} \sum_{i=1}^M \hat{\beta}_{1i}, \quad \bar{\beta}_2 = \frac{1}{M} \sum_{i=1}^M \hat{\beta}_{2i}, \quad \bar{\beta}_3 = \frac{1}{M} \sum_{i=1}^M \hat{\beta}_{3i}, \quad \bar{\beta}_4 = \frac{1}{M} \sum_{i=1}^M \hat{\beta}_{4i}.$$

The average values of all coefficients ( $\bar{\alpha}$ ,  $\bar{\beta}_1$ ,  $\bar{\beta}_2$ ,  $\bar{\beta}_3$ , and  $\bar{\beta}_4$ ) are presented in panel B of Table 3.

Overall, the LSY-CH-4 factor model can explain mutual fund returns well since the value of  $R^2$  reaches 0.86 in Panel A and 0.77 in Panel B. Meanwhile, the average values of estimated alphas from regression for each fund for the full sample and subgroups are all positive, with the highest for the growth-oriented funds (0.85) and the lowest for the value-oriented funds (0.28). This result indicates that mutual funds in China can earn an excess return over the benchmark, with the growth-oriented funds performing the best while the value-oriented funds performing the worst.

In addition, several interesting findings emerge. First, the estimated coefficients of market factors are all positive and less than 1. This suggests that the correlation between the return of mutual funds and the stock market is positive, and the performance of mutual funds is more conservative than the market performance. Second, the estimated coefficients of the size factor in Panel A are positive with the highest (0.0999) for the growth-oriented funds. It shows that growth-oriented funds prefer small-cap stocks mostly. Third, the estimated coefficients of the value factor for the full sample, growth-oriented, and balance-oriented funds are all negative, indicating that the mutual funds in China prefer stocks with a high P/E ratio overall, except for the value-oriented funds. The estimate for the growth-oriented funds is the smallest while the estimate for the value-oriented funds is the largest. This finding shows that growth-oriented funds tend to invest in high P/E stocks.

The findings presented in Table 3 provide a comprehensive overview of the performance of Chinese mutual funds, both from a market-level perspective and an average individual fund perspective. To provide an in-depth evaluation of the Chinese mutual fund performance, we conduct further analysis for each fund in Section 3.2.2.

**3.2.2. Fund-level regressions**

In this section, we perform OLS regressions on the excess return of each individual fund through Eq. (1) and report the results of the original sample and enlarged sample in Panels A and B of Table 4, respectively. For all coefficients, each panel tabulates the amount and proportion of positive, negative, significantly positive, and significantly negative estimates (all at the 5% significance level).

**Table 5**

The estimated alpha results of the bootstrap method. This table reports the distribution of estimated alpha for the full sample, growth-oriented, value-oriented, and balance-oriented funds of the enlarged sample (4712 small simulated mutual funds). It presents the amount and proportion of positive, negative, and zero alpha funds.

Type	$\alpha > 0$	$\alpha < 0$	$\alpha = 0$
Full sample	602	77	4033
	12.78%	1.63%	85.59%
Growth	175	13	656
	20.73%	1.54%	77.73%
Balance	243	33	1965
	10.84%	1.47%	87.68%
Value	33	15	355
	8.19%	3.72%	88.09%

First, the proportion of significantly positive alpha is 49.53% for the original sample, but only 11.88% for the enlarged sample. This suggests that nearly half of the mutual funds in China can beat the market based on the conventional single hypothesis t-test method in the long term.<sup>9</sup> In addition, the proportion of significantly negative (positive) funds for the enlarged sample is larger (lower) than that for the original sample, 1.66% versus 0.12% (11.88% versus 49.53%), which indicates that the short-term performance is worse than the long-term for the mutual funds in China.

Second, the main findings for the four factor loadings are as follows.<sup>10</sup> (i) Almost all of the estimated coefficients of the market factor are significantly positive, which characterizes that most funds have the same trend as the market. (ii) About 21.63% (2.16%) of estimated coefficients of the size factor are significantly positive (negative), which suggests that small-cap stocks affect the return of mutual funds more than large-cap stocks, but the explanatory power of the size factor is limited for most mutual funds due to the insignificance of  $\beta_2$  for most funds. (iii) The estimated coefficients of the value factor for most funds are significantly negative (-5.16%), which indicates that most funds prefer growth-oriented stocks (i.e., stocks with a low book-to-market ratio). (iv) About 9.74% (47.73%) of the mutual funds have significantly positive (negative) estimated coefficients of sentiment factor, which confirms that more mutual funds are likely to show great optimism.

To sum up, about one-fifth (a tenth) of mutual funds of the original (enlarged) sample is considered to have a positive alpha once the luck element has been accounted for. Consistent with the results in Section 3.2.1, mutual funds prefer small-cap stocks, growth stocks, and blue-chip stocks in general.

### 3.3. Fund performance evaluation after luck adjustment

Instead of simply using the conventional single hypothesis t-test method, we now apply our bootstrap-based FDR method to take "luck" into account when evaluating the mutual fund performance.

#### 3.3.1. Performance evaluation applying the intra-fund bootstrap in step 1

We first apply the intra-fund bootstrap statistical technique discussed in Section 2.2.1 on the mutual funds in China to eliminate the bias caused by the nonnormal distribution for each individual fund.

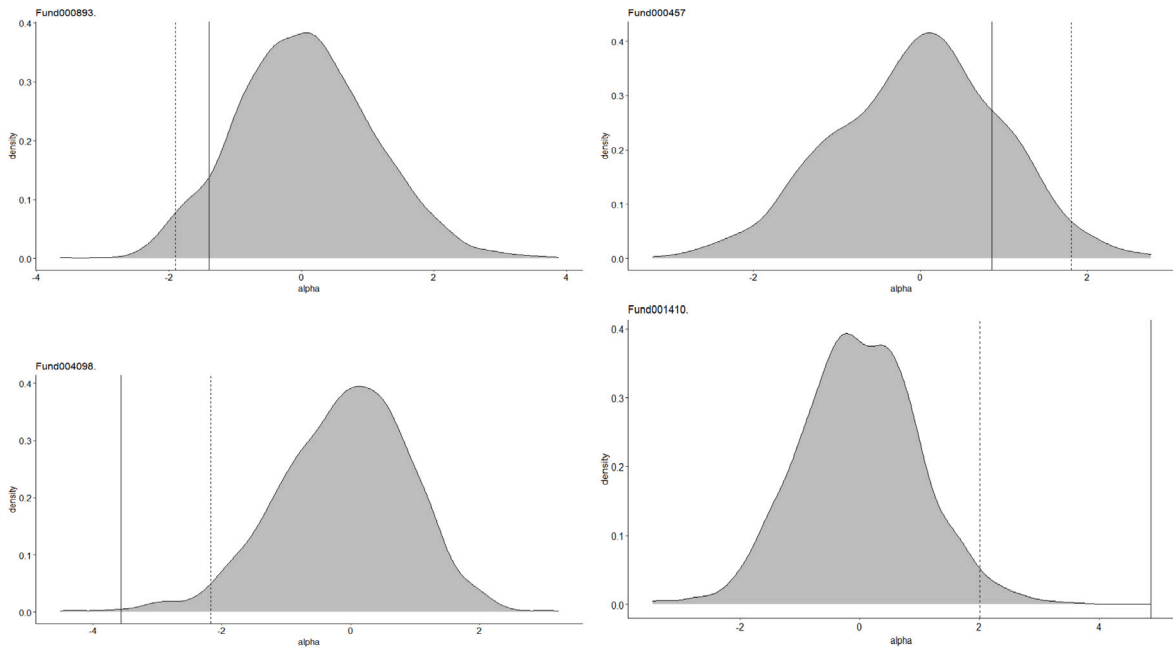
After drawing from the regression estimated residuals with replacement to make a pseudo-times series in which the alpha value is set to zero and performing the regression again, we can get a simulated probability density distribution of estimated alpha for each fund. Specifically, we decide whether the significantly positive alpha is from skill or luck by observing the simulated probability density distribution of the estimated alpha. For example, Fig. 2 shows the simulated probability density distribution of estimated alpha for four mutual funds (code 000893, 000457, 004098, and 001410). In each sub-figure, we mark the OLS estimated alpha ( $\hat{\alpha}_i$ ) with the solid line and the critical value at 5% significance level ( $\hat{\alpha}_q$ ) with the dashed line. We can conclude that Fund 001410 is skilled while Fund 004098 is bad-skilled because the estimates are significant ( $|\hat{\alpha}_i| > |\alpha_q|$ ). Conversely, Fund 000893 and Fund 000457 are zero alpha funds because the estimates are not significant ( $|\hat{\alpha}_i| < |\alpha_q|$ ).

Likewise, we apply the bootstrap method to the enlarged sample of 4712 small mutual funds obtained by using the rolling window. Table 5 shows the amount and proportion of positive and negative estimated alpha at the 5% significance level obtained from the bootstrap method for each subgroup. The proportion of positive, negative, and zero estimated alpha for the full sample is 12.78%, 1.63%, and 85.59%, respectively. From the perspective of investment style, growth-oriented mutual fund managers get the highest proportion (20.73%) of positive estimates, which is followed by balance-oriented mutual funds (10.84%) and value-oriented mutual funds (8.19%).

Table 5 shows more nonzero alpha (602 positive alpha funds and 77 negative alpha funds) than the results by the conventional single t-test in Panel B of Table 4 (560 positive alpha funds and 78 negative alpha funds). After checking the codes of selected funds,

<sup>9</sup> Conventional single hypothesis t-test simply counts the number of funds with significantly positive estimated alphas to estimate the prevalence of skilled fund managers. This is different from the joint test of the null hypothesis that all fund alphas are equal to zero in a sample as argued by Barras et al. (2010).

<sup>10</sup> Since the results in Panel A and Panel B are quite similar, here we only analyze the results of the enlarged sample for brevity.



**Fig. 2.** Distribution of simulated alpha by bootstrap. This figure plots the probability density distribution of simulated alpha for four representative individual funds: Fund 000893, Fund 000457, Fund 004098, and Fund 001410. Each sub-figure marks the estimated alpha ( $\hat{\alpha}_i$ , denoted by the solid line) and the critical value at 5% significance level ( $\alpha_c$ , denoted by the dashed line). By comparing the distribution of the simulated alpha with the actual OLS  $\hat{\alpha}$ , we can identify the fund alpha types.

we find that the non-overlapping proportion of the significantly positive and negative funds between these two methods are 9.14% (55 in 602) and 9.09% (7 in 77). This implies that the nonnormal distribution causes a certain percentage of significant alpha funds to be missed.

**3.3.2. Performance evaluation applying FDR in step 2**

Based on the results in Section 3.3.1, we further eliminate the luck bias by taking into account possible false discoveries. Specifically, we apply the false discovery rate method to the enlarged sample to examine the performance of mutual funds in China from January 2002 to June 2020. Table 6 reports the results of fund performance evaluation after remedying the luck bias for the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panels A–D, respectively. We select the optimal threshold  $\lambda^*$  as 0.4 with the MSE-bootstrap selection procedure and set significance level  $\gamma^*$  to be 0.05.<sup>11</sup> Each panel of Table 6 first shows the estimated percentage of true zero alpha funds, true positive alpha (skilled) funds, and true negative alpha (bad-skilled) funds and then shows the specific results at various significance levels from 0.05 to 0.25.

We now discuss the main results presented in Table 6. For the full sample group, the skilled mutual funds account for 19.25%, while zero alpha and bad-skilled funds account for 81.14% and -0.39%, respectively. The false discovery rate of the superior funds is much lower than the inferior funds at any level of  $\gamma$ . The results illustrate that most funds cannot beat the market. However, there exists a certain number of fund managers who possess the ability to generate excess returns over the passive benchmarks. We find robust evidence of a larger proportion of skilled funds in China, while other studies find a larger proportion of unskilled funds. For example, Cuthbertson et al. (2012) find 23.3% unskilled funds in the UK and Kim et al. (2014) find 27.5% unskilled funds in Australia. Both of them find few skilled funds in the UK (3.7%) and in Australia (1.8%). Conversely, we find 19.25% skilled mutual funds and few unskilled funds. On the other hand, the FDR range varies a lot with the sample size, sample period, alpha location, significance level, etc. For example, the positive FDR range in Cuthbertson et al. (2012) is from 30.4% to 45.4%, and the negative FDR range is from 5% to 17.7%. Kim et al. (2014) find that the positive (negative) FDR range is from 4.9% to 18.5% (27.3% to 74.8%). The positive FDR range in this paper is lower than Cuthbertson et al. (2012) and higher than Kim et al. (2014).

The proportion of bad-skilled funds is negative (-0.39%) because the left-hand side is slightly over-modified. It is caused by some natural characteristics of the FDR method. Since some p-values under the alternative hypothesis may mix in the interval above  $\lambda^*$ , the proportion of zero alpha funds will be slightly overestimated inevitably, and thus lead to the over-modification problem.<sup>12</sup>

<sup>11</sup> Note that here we can also simply set  $\lambda^*$  equal to 0.4 by observing the flat portion of the histogram of p-values shown in Fig. 3, which shows the distribution histogram of the p-value obtained from the bootstrap method for each subgroup.

<sup>12</sup> In addition to the negative proportion of bad-skilled funds, the over-modified problem also leads to the  $\widehat{FDR}$  larger than 100%. The  $\widehat{FDR}$  larger than 100% is essentially the same as the negative proportion of bad-skilled funds.

**Table 6**

Luck versus the skill of the mutual funds in China. This table reports the result of fund performance evaluation after correcting the luck bias with the bootstrap-based FDR method for the enlarged sample. We exhibit the results of the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panel A–D, respectively. In each panel, we first report the estimates of  $\pi_0$ ,  $\pi^+$ , and  $\pi^-$ . Then we show the result at various significance levels from 0.05 to 0.25. At selected  $\lambda^*$ , the variables are defined as follows:  $E(S_{\lambda^*}^+)$  ( $E(S_{\lambda^*}^-)$ ) denotes the expected proportion of significantly positive (negative) alpha funds obtained by the bootstrap method.  $E(F_{\lambda^*}^+)$  ( $E(F_{\lambda^*}^-)$ ) denotes the proportion of funds that are mistakenly classified into skilled (bad-skilled) funds.  $E(T_{\lambda^*}^+)$  ( $E(T_{\lambda^*}^-)$ ) denotes the expected proportion of truly skilled (bad-skilled) funds.  $\widehat{FDR}^+$  ( $\widehat{FDR}^-$ ) denotes the estimated false discovery rate which is calculated by  $E(F_{\lambda^*}^+)/E(S_{\lambda^*}^+)$  ( $E(F_{\lambda^*}^-)/E(S_{\lambda^*}^-)$ ).

Panel A: Full sample funds						Panel B: Growth-oriented funds					
	$\pi_0$	$\pi_+$	$\pi_-$				$\pi_0$	$\pi_+$	$\pi_-$		
%	81.14%	19.25%	-0.39%			%	73.29%	27.00%	-0.29%		
$\gamma$	At different significant level					$\gamma$	At different significant level				
	0.05	0.1	0.15	0.2	0.25		0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda^*}^+)$	602	858	1084	1259	1425	$E(S_{\lambda^*}^+)$	175	223	257	283	312
	12.78%	18.21%	23.01%	26.72%	30.24%		20.73%	26.42%	30.45%	33.53%	36.97%
$E(S_{\lambda^*}^-)$	77	142	211	275	335	$E(S_{\lambda^*}^-)$	13	24	34	49	57
	1.63%	3.01%	4.48%	5.84%	7.11%		1.54%	2.84%	4.03%	5.81%	6.75%
$F(S_{\lambda^*}^+)$	2.03%	4.06%	6.09%	8.11%	10.14%	$F(S_{\lambda^*}^+)$	1.83%	3.66%	5.50%	7.33%	9.16%
$F(S_{\lambda^*}^-)$	2.03%	4.06%	6.09%	8.11%	10.14%	$F(S_{\lambda^*}^-)$	1.83%	3.66%	5.50%	7.33%	9.16%
$F(T_{\lambda^*}^+)$	10.75%	14.15%	16.92%	18.60%	20.10%	$F(T_{\lambda^*}^+)$	18.90%	22.76%	24.95%	26.20%	27.81%
$F(T_{\lambda^*}^-)$	-0.39%	-1.04%	-1.61%	-2.28%	-3.03%	$F(T_{\lambda^*}^-)$	-0.29%	-0.82%	-1.47%	-1.52%	-2.41%
$\widehat{FDR}^+$	15.88%	22.28%	26.45%	30.37%	33.54%	$\widehat{FDR}^+$	8.84%	13.87%	18.05%	21.86%	24.78%
$\widehat{FDR}^-$	124.13%	134.62%	135.90%	139.03%	142.66%	$\widehat{FDR}^-$	118.96%	128.87%	136.45%	126.24%	135.65%
Panel C: Value-oriented funds						Panel D: Balance-oriented funds					
	$\pi_0$	$\pi_+$	$\pi_-$				$\pi_0$	$\pi_+$	$\pi_-$		
%	86.85%	11.60%	1.55%			%	86.12%	14.56%	-0.68%		
$\gamma$	At different significant level					$\gamma$	At different significant level				
	0.05	0.1	0.15	0.2	0.25		0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda^*}^+)$	33	45	64	74	90	$E(S_{\lambda^*}^+)$	234	346	456	530	617
	8.19%	11.17%	15.88%	18.36%	22.33%		10.44%	15.44%	20.35%	23.65%	27.53%
$E(S_{\lambda^*}^-)$	15	21	31	41	46	$E(S_{\lambda^*}^-)$	33	66	97	125	153
	3.72%	5.21%	7.69%	10.17%	11.41%		1.47%	2.95%	4.33%	5.58%	6.83%
$F(S_{\lambda^*}^+)$	2.17%	4.34%	6.51%	8.68%	10.86%	$F(S_{\lambda^*}^+)$	2.15%	4.31%	6.46%	8.61%	10.77%
$F(S_{\lambda^*}^-)$	2.17%	4.34%	6.51%	8.68%	10.86%	$F(S_{\lambda^*}^-)$	2.15%	4.31%	6.46%	8.61%	10.77%
$F(T_{\lambda^*}^+)$	6.02%	6.82%	9.37%	9.68%	11.48%	$F(T_{\lambda^*}^+)$	8.29%	11.13%	13.89%	15.04%	16.77%
$F(T_{\lambda^*}^-)$	1.55%	0.87%	1.18%	1.49%	0.56%	$F(T_{\lambda^*}^-)$	-0.68%	-1.36%	-2.13%	-3.03%	-3.94%
$\widehat{FDR}^+$	26.52%	38.89%	41.02%	47.30%	48.61%	$\widehat{FDR}^+$	20.62%	27.89%	31.74%	36.42%	39.10%
$\widehat{FDR}^-$	58.33%	83.33%	84.68%	85.37%	95.11%	$\widehat{FDR}^-$	146.21%	146.21%	149.23%	154.40%	157.68%

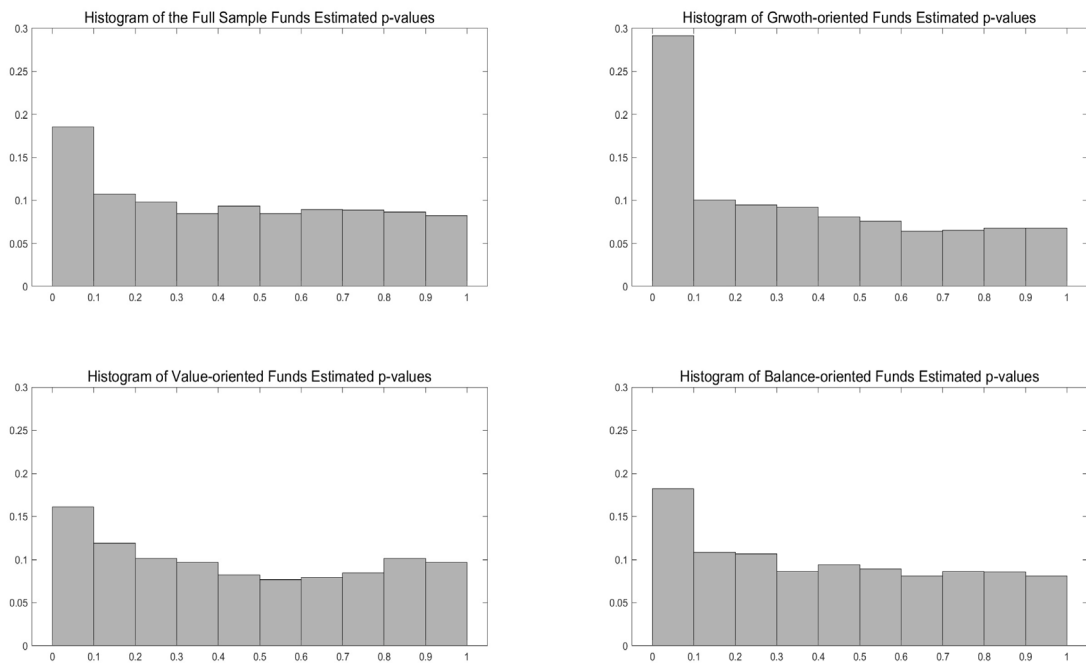
The over-modification manifests that the zero alpha funds steal some portion of the bad-skilled funds, the proportion of which is supposed to be 0%. It means that all the significantly negative funds are due to bad luck caused by false discoveries. However, the over-modification is quite slight and will not change the result, which is also illustrated by the simulation process as shown in Section 4. Compared with the result obtained from the conventional single hypothesis t-test method shown in Panel B of Table 4, we have identified more skilled funds (19.25% vs.11.88%) and fewer bad-skilled (-0.39% vs.1.66%) funds by using the bootstrap-based FDR method. Our result confirms that the bias of the conventional single hypothesis t-test method hides some truly skilled funds and falsely discovers some truly bad-skilled funds, which leads to a much more pessimistic conclusion. The bootstrap-based FDR method can remedy the bias and reveal a more realistic picture of fund performance.

Panels B–D of Table 6 show the results of growth-oriented, value-oriented, and balance-oriented funds, respectively. Growth-oriented mutual funds exhibit the highest proportion of skilled managers (27.00%). By contrast, the proportion of skilled managers for balance-oriented (14.56%) and value-oriented mutual funds (11.60%) are much lower. The same message is also conveyed in Fig. 3 since the proportion of p-values clustered around zero is the largest for growth-oriented funds.<sup>13</sup>

### 3.4. Discussion

Our empirical results are more optimistic than previous studies on some developed financial markets such as the U.S. and Germany. For example, Cuthbertson and Nitzsche (2013) find that in Germany the percentages of true zero alpha, skilled, and

<sup>13</sup> As referred to in Section 2.2.2, truly alternative features have p-values clustered around zero whereas truly null features have p-values uniformly distributed in [0, 1].



**Fig. 3.** The histogram of p-values. This figure shows the distribution histogram of estimated p-values for the full sample, growth-oriented, balance-oriented, and value-oriented funds.

bad-skilled funds are equal to 83%,  $-0.2\%$ , and  $17.1\%$ , respectively. Similarly, studies on the U.S. mutual fund market show that most mutual funds cannot beat the market for investors (e.g., Fama and French, 2010; Barras et al., 2010). In this study, we observe a relatively higher proportion of skilled funds than those reported in developed countries with  $19.25\%$  of skilled mutual funds from 2002 to 2020 across our entire sample period. This finding is consistent with Chi et al. (2022), who find that actively managed stock mutual funds in the Chinese A-share market offer good investment opportunities for investors.

We aim to explain the higher proportion of skilled mutual funds in China by examining the equilibrium accounting framework proposed by Fama and French (2010). Equilibrium accounting means that the passive investors have zero  $\alpha$  relative to passive benchmarks in terms of gross returns,<sup>14</sup> thus active investment in aggregate must have zero  $\alpha$  (Fama and French, 2010). The regression results for weighted portfolios of the U.S. active funds in Fama and French (2010) are consistent with this equilibrium accounting. Fama and French (2010) find that the  $\alpha$  estimates for gross fund returns are close to zero and the  $\alpha$  estimates for net fund returns are negative ( $-1.00\%$  for the Fama–French–Carhart four-factor model), which indicates that the total wealth invested in active funds from 1984 to 2006 in the U.S. mimics the market portfolio. By contrast, the  $\alpha$  estimates for net fund returns of the equally weighted portfolio in Panel A of Table 3 approach  $0.68\%$ , and thus the  $\alpha$  estimates for the gross returns would be even larger. In China, international investors investing on mutual funds can only do so through a specific channel which are labeled as “Qualified Foreign Institutional Investors” also known as QFII fund. International Asset management corporations (AMC, Fund family) needs to apply for a license to operate under such model. This model effectively restricts international investors to directly investing on China’s domestic stocks, or on funds through the Fund of Funds. Consequently, the result of the Chinese fund market does not echo equilibrium accounting as Fama and French (2010) does, which is in line with the extant literature (e.g., Tang et al., 2012; Chi, 2013).

We further compare the percentiles of  $t(\alpha)$  estimates for actual and simulated fund returns in which true alpha is zero. Table B.1 in Appendix B shows the  $t(\alpha)$  estimates for actual returns (in the third column) and the average of  $t(\alpha)$  estimates from the simulation of 1000 times (in the second column) at each selected percentile. The last column shows the percentage that actual return can beat the simulated counterparts among the 1000 simulations.

Regressing net fund returns by the three-factor model, Fama and French (2010) found that the value of  $t(\alpha)$  estimates for actual returns are below the corresponding average of simulations at most of the percentiles (from 1% to 97%). When they use a four-factor model, the  $t(\alpha)$  estimates for actual returns are always below the average of simulations (from 1% to 100%). Conversely, in our paper, Table B.1 shows that most of the values of  $t(\alpha)$  for actual returns are above the corresponding average of simulations at most percentiles (from 6% to 100%). In addition, the percentage of the simulation runs whose actual return can beat the simulated counterpart reaches 50% between the percentile of 6% to 7%. Our results show that some mutual funds in China have sufficient ability to generate excess returns.

<sup>14</sup> Gross fund returns refer to the returns containing the costs (fees and other expenses) while net fund returns refer to the returns deducting the costs.

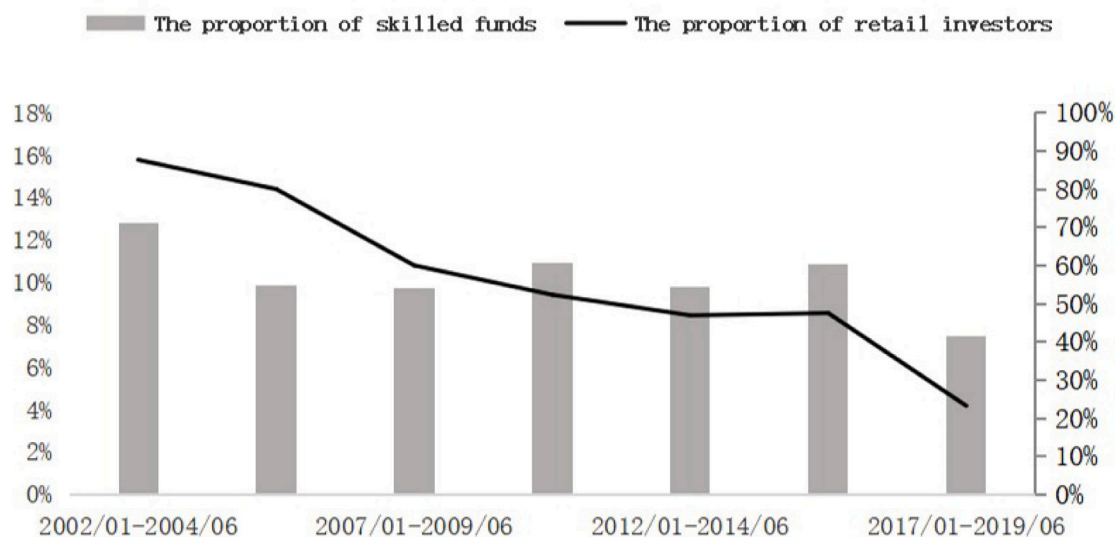


Fig. 4. Time-series variation in the proportion of skilled funds and retail investors. This figure plots the proportion of skilled mutual funds (left axis) alongside the proportion of retail investors (right axis) across four distinct periods: 2002/01–2004/06, 2007/01–2009/06, 2012/01–2014/06, and 2017/01–2019/06. The data illustrate a strong co-movement between the two series.

To summarize, the aggregate of 844 selected Chinese active equity mutual funds demonstrates a positive-sum game for net returns, in contrast to the zero-sum game for gross returns or negative-sum game for net returns observed in the U.S. mutual fund market.<sup>15</sup> Why do the selected mutual funds in China in aggregate show a positive sum result? This can be explained by the unique character of the Chinese financial market. Notably, there is a significant proportion (approximately 72% in 2014 and 52% in 2020) of free-floating shares held by retail investors, which contrasts with the dominance of institutional investors in the U.S. financial market.<sup>16</sup> This much greater proportion of retail investors in China may work to the advantage of skilled fund managers in China who better understand the trading rules, market movements and underlying fundamentals behind company valuations.

This positive aggregate alpha is a direct violation of the zero-sum condition implied by the equilibrium accounting framework. This violation serves as critical, structural evidence for our proposed channel. The accounting identity necessitates that the gains of one group must be the losses of another. In the context of China's market structure, where retail investors form the vast majority of active participants by turnover and holding, the observed positive alpha for the professional fund sector can only be logically sustained if the retail investor sector, in aggregate, is providing a corresponding subsidy through negative alpha. Therefore, this result does not merely correlate with retail investor dominance; it is a direct mathematical consequence of it, offering foundational support for our central argument. To substantiate the argument, we plot the time-series variation in the proportion of skilled funds against the proportion of retail investors across four multi-year periods. As shown in Fig. 4, the two series exhibit a strong comovement, with a correlation coefficient of 0.74. This indicates that periods with lower retail investor participation are associated with a significantly lower proportion of skilled funds. This pattern supports our argument that retail investors provide an environment in which skill is more likely to be rewarded and detected, and their withdrawal from the market diminishes both the opportunity and the incentive for fund managers to outperform.

While the dominance of retail investors provides a compelling equilibrium-based narrative for our findings, it is important to consider that this may not be the sole contributing factor. Several other characteristics of the Chinese market could concurrently shape the observed skill premium. First, industry competition dynamics appear relevant. We demonstrate that the prevalence of skilled funds is negatively related to competition within the mutual fund industry. As shown in Table 7, the proportion of skilled funds is negatively associated with the total number of funds (correlation =  $-0.57$ ). This suggests that as the market matures and competition intensifies, generating persistent excess returns becomes more challenging. Second, the attributes of fund managers themselves may play a role. The highly competitive educational landscape and the intense selection process for roles in the rapidly growing asset management industry might cultivate a cohort of managers with strong analytical skills and accelerate their accumulation of market-specific experience. However, direct testing of this channel is limited by the availability of comprehensive

<sup>15</sup> A zero-sum game means that the aggregate  $\alpha$  of the equally-weighted portfolio is equal to zero; a negative-sum game means that the aggregate  $\alpha$  of the equally-weighted portfolio is negative.

<sup>16</sup> China's A-share market has a notably high proportion of retail investors and a strong speculative atmosphere. In 2014, individual investors held 72% of the free-float market capitalization, significantly higher than in developed stock markets. However, in recent years, as the stock market matures, institutional investors in A-shares have outperformed the broader market, and their share of holdings has risen steadily. The market has shown a clear trend towards institutionalization, with the proportion held by individual investors declining to 52% by 2020 (data obtained from CICC).

**Table 7**

The number of skilled funds in different time periods. This table shows the number of skilled mutual funds, the number of mutual funds, and the proportion of skilled funds in different time periods.

Time period	Number of skilled funds	Number of funds	The proportion of skilled funds
2002/01–2004/06	17	39	43.59%
2004/07–2006/12	60	213	28.17%
2007/01–2009/06	50	494	10.12%
2009/07–2011/12	96	795	12.08%
2012/01–2014/06	95	1091	8.71%
2014/07–2016/12	291	1573	18.50%
2017/01–2019/06	69	507	13.61%

micro-data on individual manager backgrounds and capabilities. Third, unique regulatory and market micro-structure features could preserve alpha opportunities. Constraints on short-selling and the limited availability of sophisticated derivative instruments may slow down arbitrage activities that would otherwise swiftly correct mispricings in more developed markets. This, combined with the high volatility and sentiment-driven flows prevalent in a retail-dominant market, may create an environment where skilled, disciplined managers can identify and exploit anomalies over a longer horizon. Our primary analysis focuses on the investor structure channel due to its direct testability and fundamental role in cross-country comparisons. The channels outlined above are not mutually exclusive and likely interact. Future research with more granular data could productively disentangle their relative importance in shaping managerial performance in China and similar emerging markets.

#### 4. Simulation

Andrikogiannopoulou and Papakonstantinou (2019) use simulations with parameters informed by the data of BSW to examine the accuracy of FDR in fund performance evaluation, and find that the accuracy of FDR hinges on the distribution and locations of the fund alphas. In this study, we conduct simulations where we can explicitly control the specific characteristics based on the parameters informed by real data of the Chinese mutual fund market. By comparison, we also conduct a simulation study where we can explicitly control the specific characteristics based on the parameters informed by real data of the U.S. mutual fund market. We find support for the applicability of the FDR method for mutual funds in China.

##### 4.1. Simulation results for the Chinese mutual fund market

Following Andrikogiannopoulou and Papakonstantinou (2019), we generate fund returns  $r_{it}$  ( $r_{it} = R_{it} - R_{ft}$ ) using the following Data Generating Process (DGP):

$$r_{it} = \alpha_i + X_t' \beta_i + \epsilon_{it}, \text{ for } i = 1, \dots, N, t = 1, \dots, T, \quad (13)$$

where  $\beta_i$ ,  $X_t$ , and  $\epsilon_{it}$  are generated from normal distributions with parameters equal to their sample counterparts in the original data. The details of DGP are as below.

For the simulation based on the data of the Chinese mutual fund market,  $\beta_1$  to  $\beta_4$  are generated from normal distributions with the means of 0.8129,  $-0.1579$ , 0.1854, and  $-0.0037$  and the standard deviations of 0.1611, 0.4023, 0.3422 and 0.2400, respectively. Factor returns of MKT, SMB, VMG, and PMO are generated from normal distributions with the means of 0.0063, 0.0046, 0.0117, and 0.0085, and the standard deviations of 0.0778, 0.0466, 0.0382, and 0.0362, respectively. The error term  $\epsilon_{it}$  is generated from the normal distribution with mean 0 and standard deviation 0.0322. We create a balanced panel with the cross-sectional dimension  $N$  to be 4712 and the time-series dimension  $T$  to be 36. We set the proportions of funds with positive, negative, and zero alpha to be 20.00%, 0.00%, and 80.00%, respectively. Following Andrikogiannopoulou and Papakonstantinou (2019), we consider the following three scenarios for the magnitude of  $\alpha_i$ <sup>17</sup>:

A. Positive (3.00%); B. Positive (1.00%); C. Positive (0.50%).

We report the simulation results for the Chinese mutual fund market in Appendix B. Table B.2 reports the simulation results of the Chinese mutual fund market under Scenarios A, B, and C in Panels A, B, and C, respectively.

In Panel A, the estimated proportion of positive, negative, and zero alpha funds are respectively 22.03%, 1.80%, and 76.17% by the conventional single hypothesis t-test method while with the bootstrap-based FDR method, we obtain 19.97%,  $-0.07\%$  and 80.10%. Recall that in the true DGP, the proportions for those groups are 20.00%, 0.00%, and 80.00%. We can see that there exists a false discoveries phenomenon overall since that 3.83% ( $80\% - 76.17\% = 3.83\%$ ) of zero alpha funds are classified into skilled or bad-skilled funds by the conventional single hypothesis t-test method. The result of the bootstrap-based FDR method is far more accurate than the conventional single hypothesis t-test method: the estimated proportion of skilled funds is modified from 22.03% to 19.97%; the estimated proportion of bad-skilled funds is modified from 1.80% to  $-0.07\%$ ; the estimated proportion of zero alpha funds is modified from 76.17% to 80.10%. In addition, notice that the proportion of bad-skilled funds is estimated to be negative. This implies that the over-estimated zero alpha funds (80.10% vs. 80.00%) grab some portion of the bad-skilled funds ( $-0.07\%$  vs.

<sup>17</sup> Since the proportion of funds with negative alpha is 0.00%, we only set the magnitude of positive alphas.

0.00%). We can also add the over-modification bias back to obtain a result much closer to the truth: the positive, negative, and zero alpha funds are respectively 19.97%, 0.00%, and 80.03% (80.10%–0.07%). The over-modification bias is too minor to be noticed, hence we show the original result of bad-skilled funds.

Some studies on the U.S. mutual fund market argue that the FDR methodology is overly conservative and underestimates the proportion of nonzero alpha funds (Andrikogiannopoulou and Papakonstantinou, 2019). However, in the scenario when the absolute value of alpha is large enough, there exists a sufficient power of the bootstrap-based FDR method because most nonzero alpha funds are accurately accounted for and the number of undiscovered nonzero alpha funds (80.10%–80.00% = 0.10%) is negligible. Therefore, when the absolute value of alpha is large, the bootstrap-based FDR method accurately remedies the false discoveries for both sides although the estimated proportion of nonzero alpha funds is slightly overestimated.

The results of Panel B gives us some valuable information that the power dramatically declines with the reduction of the value of nonzero alpha. Low power leads to an overly conservative estimated proportion of zero alpha funds because not all nonzero alpha funds are recognized, especially for positive alpha. The missed discovery proportion for the conventional single hypothesis t-test is about 8.56%.<sup>18</sup> Only 57.2% (11.44%/20.00%) of truly nonzero alpha funds are recognized correctly by the conventional single hypothesis t-test method. Specifically, the proportion of skilled funds (9.47%) is substantially underestimated while the proportion of bad-skilled funds (1.97%) is overestimated. In this scenario when the absolute value of alpha is small, the critical problem is false nondiscovery caused by low power rather than the false discovery caused by the multiple hypothesis testing problem. However, the bootstrap-based FDR method is still more accurate than the conventional single hypothesis t-test method: the estimated proportion of skilled funds is modified from 9.47% to 12.29%; the estimated proportion of bad-skilled funds is modified from 1.97% to –0.27%; the estimated proportion of zero alpha funds is modified from 88.56% to 87.98%. Consequently, the bootstrap-based FDR method can still slightly enhance the accuracy of performance evaluation by decreasing the bias from 8.56% to 7.98%.<sup>19</sup>

Panel C reports the result of Scenario C when we set the absolute value of alpha to an even smaller value. Unlike the results in Panel A and Panel B, when the absolute number of alpha drops to a certain level, the false nondiscovery problem is quite severe and the bootstrap-based FDR method fails to correct the bias. Further, by altering the absolute value of alpha, we find that the approximate threshold for the failure of the bootstrap-based FDR method is about 0.60%. Fortunately, the absolute values of estimated alphas for all the Chinese funds in our sample are larger than 0.60%, which means that the bootstrap-based FDR method is applicable to the Chinese mutual fund market. More detailed results can be referred to Tables B.3 to B.5 in Appendix B.

In conclusion, the bootstrap-based FDR method can remedy the bias caused by multiple hypothesis testing problem for both sides in a simulation environment when the absolute value of alpha is relatively large based on the Chinese mutual fund characteristics. In addition, although the nondiscovery caused by low power is the main problem instead of the multiple hypothesis testing problem for funds with the smaller absolute value of alpha such as 1.0% and –1.0%, the bootstrap-based FDR method can balance the bias towards the left-hand (bad luck bias) and right-hand side (good luck bias) to achieve a more accurate result on the whole. The bootstrap-based FDR method still dominates the conventional single hypothesis t-test method, especially when the number of skilled alpha funds is what we focus on the most. Moreover, the location of nonzero alpha for most of the mutual funds in China is large enough to ensure the effectiveness of the bootstrap-based FDR method. Consequently, the bootstrap-based FDR method is quite feasible and applicable to the Chinese mutual fund market.

Given that the magnitude and distribution of fund alpha reported by Barras et al. (2010) are different from the Chinese empirical results, it is natural to ponder whether the bootstrap-based FDR method performs well in a totally different market. As such, we also conduct the simulation experiment using the original data of the U.S. mutual funds, and the results are presented in Appendix C. The findings are similar to the Chinese market: (i) The bias is mainly caused by the multiple hypothesis testing problem for larger absolute alpha and low power for smaller absolute alpha. (ii) The bootstrap-based FDR method can remedy luck bias for large absolute alpha. By contrast, when the absolute value of alpha is smaller, the bootstrap-based FDR method can recognize more significant alpha funds, but still underestimates the number of skilled U.S. mutual funds. Given that in fund performance evaluation, we aim to find how many funds can beat the market, the bootstrap-based FDR method is less effective for U.S. mutual funds — refer to Appendix C for more details.

#### 4.2. Simulation results for the U.S. mutual fund market

We conduct a simulation study where we can explicitly control the specific characteristics based on the parameters informed by real data of the U.S. mutual fund market. We report the simulation results for the Chinese mutual fund market in Appendix C.

Following Andrikogiannopoulou and Papakonstantinou (2019), we generate fund returns  $r_{it}$  ( $r_{it} = R_{it} - R_{ft}$ ) using the following Data Generating Processes (DGP):

$$r_{it} = \alpha_i + X_t' \beta_i + \epsilon_{it}, \text{ for } i = 1, \dots, N, t = 1, \dots, T, \quad (14)$$

<sup>18</sup> 8.56% is the gap between the true proportion of nonzero alpha funds (20.00%) and the estimated proportion of nonzero alpha funds based on conventional single hypothesis t-test method (9.47% + 1.97% = 11.44%).

<sup>19</sup> 7.98% is the gap between the true proportion of nonzero alpha funds (20.00%) and the estimated proportion of nonzero alpha funds based on the bootstrap-based FDR method (12.29% + (–0.27%) = 12.02%).

where  $\beta_i$ ,  $X_t$ , and  $\varepsilon_{it}$  are generated from normal distributions with parameters equal to their sample counterparts in the original data. The details of DGP are as below.

For the simulation based on the data of the U.S. mutual fund market,  $\beta_1$  to  $\beta_4$  are generated from normal distributions with the mean of 0.710, 0.0686,  $-0.0782$ , and  $-0.0337$ , and the standard deviation of 0.448, 0.285, 0.264, 0.130 respectively. Factor returns of MKT, SMB, HML, and MOM are generated from normal distributions with the mean of 0.00417, 0.00197, 0.00475, and 0.00921 respectively and the standard deviation of 0.0457, 0.0328, 0.0289, and 0.0404, respectively.<sup>20</sup> The error term  $\varepsilon_{it}$  is generated from the normal distribution with a mean of 0 and a standard deviation of 0.00748. We create a balanced panel with the cross-sectional dimension  $N$  to be 3313 and the time-series dimension  $T$  to be 36. Following BSW, we set the proportions of funds with positive, negative, and zero alpha to be 2%, 23%, and 75%, respectively. For the magnitude of  $\alpha_i$ , we consider the following three scenarios.

- A\* positive (3.80%), negative ( $-3.20\%$ )
- B\* positive (0.50%), negative ( $-0.50\%$ )
- C\* positive (0.10%), negative ( $-0.10\%$ )

Table C.1 reports the simulation results of the U.S. mutual fund market under Scenarios A\*, B\*, and C\* in Panels A, B, and C, respectively.

In Panel A, the estimated proportion of positive, negative, and zero alpha funds are respectively 3.59%, 24.89%, and 71.52% by the conventional single hypothesis t-test method while with the bootstrap-based FDR method, we obtain 1.87%, 23.23%, and 74.90%. Recall that in the true DGP, the proportions for those groups are 2.00%, 23.00%, and 75.00%. We can see that there exists false discoveries phenomenon overall since that 3.48% ( $75.00\% - 71.52\% = 3.48\%$ ) of zero alpha funds are classified into skilled or bad-skilled funds by the conventional single hypothesis t-test method. The result of the bootstrap-based FDR method is far more accurate than the conventional single hypothesis t-test method: the estimated proportion of skilled funds is modified from 3.59% to 1.87%; the estimated proportion of bad-skilled funds is modified from 24.89% to 23.00%; the estimated proportion of zero alpha funds is modified from 71.52% to 74.90%. Besides, there exists no overestimation for the zero alpha funds. Therefore, when the absolute value of alpha is large, the bootstrap-based FDR method can remedy the false discoveries for both sides.

Similar to the simulation results of the Chinese market, the result for Scenario B\* in Panel B shows a significant decline in the power with the reduction of the value of nonzero alpha. Low power leads to an overestimated proportion of zero alpha funds, especially for the negative alpha. The missed discovery proportion for the conventional single hypothesis t-test is about 11.5%.<sup>21</sup> Only 54% (13.5%/25.00%) of truly nonzero funds are recognized correctly by the conventional single hypothesis t-test method. Specifically, the proportion of bad-skilled funds (11.14%) is substantially underestimated while the other side (2.36%) is overestimated. In this scenario, the critical problem is also false non discovery caused by low power instead of the false discovery caused by the multiple hypothesis testing problem (MHT problem). The bootstrap-based FDR method can beat the conventional single hypothesis t-test method by decreasing the missed discovery proportion from 11.5% to 7.66%.<sup>22</sup> Therefore, the result of the bootstrap-based FDR method is more reliable than the conventional single hypothesis t-test method overall. However, it is worth noticing that the accuracy of the estimation for the zero alpha funds and the bad-skilled funds by the bootstrap-based FDR method is enhanced at the expense of the underestimation of the skilled funds (from 2.36% to 0.46%). This conclusion is different from that for the Chinese mutual fund market in Section Section 4.1 .

Panel C reports the result of Scenario C\* when we set the absolute value of alpha to an even smaller value. Similarly, when the absolute number of alpha drops to a certain level, the false non discovery problem is quite severe that the bootstrap-based FDR method fails to correct the bias and the estimated threshold for the failure of the bootstrap-based FDR method is at about 0.40%. The details can be referred to Tables C.2 and C.3.

## 5. Robustness checks

In Section 3, we employ the LSY-CH-4 factor model to conduct an evaluation framework of the bootstrap-based FDR method to distinguish genuine managerial skill from luck and find some optimistic results for Chinese mutual fund investors. In this section, we report the robustness checks to assess whether our result is sensitive to altering the number of factors in the benchmark models, alternative benchmark models and different rolling windows. The detailed results are reported in Appendix D. In general, we demonstrate that our main findings for mutual funds are robust to these variations.

<sup>20</sup> Our original data of the active U.S. equity mutual funds is obtained from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free Mutual Fund database from 02/20/1962 to 09/20/2017.

<sup>21</sup> 11.5% is the gap between the proportion of true nonzero alpha funds (25%) and the estimated proportion of nonzero alpha funds based on the bootstrap-based FDR method ( $2.36\% + 11.14\% = 13.5\%$ ).

<sup>22</sup> 7.66% is the gap between the proportion of true nonzero alpha funds (25%) and the estimated proportion of nonzero alpha funds based on the bootstrap-based FDR method ( $0.46\% + 16.88\% = 17.34\%$ ).

### 5.1. Alternative number of factors in the benchmark model

In this subsection, we use the LSY-CH-3 factor model constructed by Liu et al. (2019) to test the robustness of our results as follows:

$$r_{it} = \alpha_i + \beta_{1i} \cdot \text{MKT}_t + \beta_{2i} \cdot \text{SMB}_t + \beta_{3i} \cdot \text{VMG}_t + \epsilon_{it}, \quad (15)$$

where  $r_{it}$  denotes the excess return of fund  $i$  over the risk-free rate in month  $t$ .  $\text{MKT}_t$ ,  $\text{SMB}_t$ , and  $\text{VMG}_t$  denote the LSY-CH-3 factors, which are the Market excess return (MKT) factor, the Small-Minus-Big (SMB) size factor, and the Value-Minus-Growth (VMG) value factor. As shown in Table D.4 in Appendix D, the result of the LSY-CH-3 factor model are almost identical to that of the LSY-CH-4 factor model. The estimated proportion of zero alpha funds, skilled funds, and bad-skilled funds are 80.83%, 19.52%, and  $-0.34\%$ , respectively

### 5.2. Using the Fama–French–Carhart four-factor model as the benchmark model

Given the popularity of the Fama–French–Carhart four-factor model in mutual fund performance evaluation (e.g., Barras et al., 2010; Kim and In, 2012; Cai et al., 2018), we substitute the LSY-CH-4 factors with the Fama–French–Carhart four-factor model to perform the mutual fund performance evaluation. This model is reproduced below:

$$r_{it} = \alpha_i + \beta_{1i} \cdot \text{MKT}_t + \beta_{2i} \cdot \text{SMB}_t + \beta_{3i} \cdot \text{HML}_t + \beta_{4i} \cdot \text{MOM}_t + \epsilon_{it}, \quad (16)$$

where  $r_{it}$  denotes the excess return of fund  $i$  over the risk-free rate in month  $t$ .  $\text{MKT}_t$ ,  $\text{SMB}_t$ ,  $\text{HML}_t$  and  $\text{MOM}_t$  denote the market, size, value and momentum factors, which are constructed based on data from the Chinese stock market and are obtained from Wind Database. As shown in Table D.5 in Appendix D, the results for the full sample, growth-oriented, value-oriented, and balance-oriented funds are qualitatively similar to the results based on the LSY-CH-4 factor model. For example, by using the Fama–French–Carhart four-factor model, we find that there are 19.44% skilled funds and 2.55% bad-skilled funds for the full sample

### 5.3. Alternative benchmark models

In this subsection, we further check the robustness to our results by using alternative benchmark models such as the Fama–French three-factor model, the Fama–French five-factor model, and the Capital Asset Pricing Model (CAPM). In other words, we replace our LSY-CH-4 factor model with each of these three models. The comparable factors are obtained from WIND database and are constructed based on data from the Chinese stock market. The Fama–French three-factor model is given by:

$$r_{it} = \alpha_i + \beta_{1i} \cdot \text{MKT}_t + \beta_{2i} \cdot \text{SMB}_t + \beta_{3i} \cdot \text{HML}_t + \epsilon_{it}, \quad (17)$$

where  $r_{it}$  denotes the excess return of fund  $i$  over the risk-free rate in month  $t$ .  $\text{MKT}_t$ ,  $\text{SMB}_t$  and  $\text{HML}_t$  denote the market, size and value factors. The Fama–French five-factor model is given by:

$$r_{it} = \alpha_i + \beta_{1i} \text{MKT}_t + \beta_{2i} \text{HML}_t + \beta_{3i} \text{SMB}_t + \beta_{4i} \text{RMW}_t + \beta_{5i} \text{CMA}_t + \epsilon_{it}, \quad (18)$$

where  $\text{MKT}$  denotes the market excess return factor,  $\text{SMB}$  denotes the small-minus-big size factor,  $\text{HML}$  denotes the high-minus-low value factor,  $\text{RMW}$  denotes the robust-minus-weak probability factor, and  $\text{CMA}$  denotes the conservative-minus-aggressive investment factor.

The results of the Fama–French three-factor, Fama–French five-factor and CAPM models are reported in Tables D.6, D.7, and D.8 in Appendix D, respectively. qualitatively unchanged from the results of the LSY-CH-4 factor model. There are 20.10% (19.47%) and 1.41% (1.37%) of funds estimated to be skilled and bad-skilled funds for the full sample. The results are consistent for the three subgroups. When we use the CAPM model as the benchmark model, there is a slightly smaller number of skilled mutual funds (16.70%).

### 5.4. Using the 5-year overlapping rolling window

Another concern is whether the evaluation of mutual fund performance varies over different rolling window periods. To investigate this concern, we change the periods of overlapping rolling window to five years which is also used by Christiansen et al. (2018). The results for the full sample which are reported in Panels A–D of Table D.9 in Appendix D. Growth-oriented, value-oriented, and balance-oriented funds are that 19.47%, 29.67%, 23.23%, 15.70% skilled funds for the full, growth-oriented, value-oriented, and balance-oriented samples, respectively. The results are qualitatively similar to the results using the 3-year overlapping rolling window.

### 5.5. Using the 5-year non-overlapping rolling window

Following BSW, we also use the non-overlapping rolling window to partition our data into several subperiods of 5 years. The result of doing this, as shown in Panel A of Table D.10 in Appendix D, is that the estimated proportion of zero alpha funds, skilled funds, and bad-skilled funds are 84.04%, 16.70%, and  $-0.74\%$ , respectively. The result using the 5-year non-overlapping rolling

window lead to a slightly smaller proportion of positive alpha funds to reported result. in Section 3.3.2 but stay qualitatively unchanged.

## 6. Conclusion

Unlike the majority of the fund performance evaluation literature, which focuses on domestic funds from developed economies (mostly the U.S.), this paper evaluates the performance of 844 open-ended, domestic equity mutual funds in the case of China which is the largest emerging market in the world. The analysis spans a relatively extensive period from January 2002 to June 2020. Using the bootstrap-based FDR method based on the LSY-CH-4 factor model to distinguish skill from luck, we find a much larger portion of funds with positive alphas than is the case for developed. Specifically, using the LSY-CH-4 factor model, we find that most funds cannot beat the market but there exists a significant 19.25% of truly skilled funds with the real ability to generate positive excess risk-adjusted returns over the passive benchmarks. The proportion of truly skilled managers is the highest for growth-oriented mutual funds (27.00%) and the lowest for value-oriented mutual funds (11.60%). Our findings survive a battery of robustness tests.

We present new evidence that supports the hypothesis that a significant amount of Chinese mutual funds are able to outperform the market on a risk-adjusted basis. From the perspective of Fama and French (2010), the aggregate of the sampled Chinese funds is neither a zero-sum game for gross returns nor a negative-sum game for net returns as in the U.S. mutual fund market. In the case of Chinese mutual funds, it is possible that the positive alpha may be at the expense of less sophisticated retail traders rather than the result of the negative alpha of other mutual funds. The Chinese market has a generally higher participation rate of retail than the U.S. market with retail investors accounting for over 80% of turnover in China in 2020, well above the approximately 25% of turnover rate in the U.S. stock market. The study of Wen et al. (2021) suggests that there is evidence that institutional investors in China tend to adopt contrarian investment strategies and are able to benefit from the strong mean reversion exhibited in the Chinese stock market which is evidenced in China by Wu (2011). An interesting topic for future research would be to look at the interaction of mutual fund and retail investors investing decisions and how they influence subsequent market returns.

A final contribution of this study is that we add to the existing literature by validating the applicability of the FDR method through simulations. Our findings confirm that the FDR method significantly outperforms conventional single hypothesis t-test methods, particularly within China's mutual fund market context. Overall, our research suggests that mutual funds in China seem to have a significantly greater chance of generating a positive alpha than their developed countries counterparts. Whether this will continue in the future as the regulatory environment develops and the sophistication of Chinese retail investors improves remains to be seen.

## CRedit authorship contribution statement

**Tingting Cheng:** Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Shuo Xing:** Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Cheng Yan:** Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Keith Pilbeam:** Methodology, Formal analysis.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We are heavily indebted to three referees and the editor for providing very helpful comments and suggestions that have greatly improved the reported research.

## Appendix A

In this appendix, we present the selection procedure of optimal values of  $\lambda$  and  $\gamma$  in FDR. First, we compute  $MSE(\lambda)$  at different levels,  $\lambda = 0.3, 0.4, \dots, 0.7$ , by sampling p-values with replacement for M times and recomputing  $\hat{\pi}_0(\lambda)$ . Let  $\hat{\pi}_0^b$  denote  $\hat{\pi}_0$  for bth bootstrap, B denote the number of bootstrap steps, and  $B = 1000$ , we can get MSE for each possible value of  $\lambda$ . We choose  $\lambda^*$ , for which MSE is the minimum. The MSE for each possible  $\lambda$  is:

$$\widehat{MSE}(\lambda) = \frac{1}{B} \sum_{b=1}^B [\hat{\pi}_0^b(\lambda) - \min_{\lambda} \hat{\pi}_0(\lambda)]^2. \quad (19)$$

For  $b$ th bootstrap, let  $\hat{\pi}_A^{b+}$  denote  $\hat{\pi}_A^+$  and  $\hat{\pi}_A^{b-}$  denote  $\hat{\pi}_A^-$ . Using a similar approach, we can determine  $\gamma^*$ . Following BSW, we set  $\gamma = 0.05, 0.1, 0.15, 0.2, 0.25$  and do the bootstrap for 1000 times. The mean square error of significance level  $\gamma$  is:

$$\widehat{MSE}^+(\gamma) = \frac{1}{B} \sum_{b=1}^B [\hat{\pi}_A^{b+}(\gamma) - \max_{\gamma} \hat{\pi}_A^+(\gamma)]^2. \tag{20}$$

$$\widehat{MSE}^-(\gamma) = \frac{1}{B} \sum_{b=1}^B [\hat{\pi}_A^{b-}(\gamma) - \max_{\gamma} \hat{\pi}_A^-(\gamma)]^2. \tag{21}$$

According to the principle of least mean square error, we choose  $\gamma^+ = \arg \min_{\gamma} MSE^+(\gamma), \gamma^- = \arg \min_{\gamma} MSE^-(\gamma)$ . If  $\min_{\gamma} MSE^+(\gamma) < \min_{\gamma} MSE^-(\gamma)$ , then  $\hat{\pi}_A^+(\gamma^*) = \hat{\pi}_A^+(\gamma^+)$ , meanwhile  $\hat{\pi}_A^-(\gamma^*) = 1 - \pi_0 - \hat{\pi}_A^+(\gamma^*)$ ; Rather, if  $\min_{\gamma} MSE^+(\gamma) > \min_{\gamma} MSE^-(\gamma)$ , then  $\hat{\pi}_A^-(\gamma^*) = \hat{\pi}_A^-(\gamma^-)$ , mean while  $\hat{\pi}_A^+(\gamma^*) = 1 - \hat{\pi}_0 - \hat{\pi}_A^-(\gamma^*)$ .

**Appendix B**

This Appendix provides details of the simulation results for the mutual funds in China including the values of  $t(\alpha)$ /simulated  $t(\alpha)$  at different percentiles (Table B.1), simulation results (Table B.3) and details (Table B.4) of different alpha locations, and the simulation results for the Chinese mutual fund market (Table B.2). In addition, we present the distribution of significant alpha in Table B.5. We recognize the approximate threshold for the failure of the bootstrap-based FDR method as 0.60% around.

**Table B.1**

The values of  $t(\alpha)$  for actual returns and average values of  $t(\alpha)$  from simulation at different percentiles. This table reports the  $t(\alpha)$  estimates for actual returns (denoted by Act) and the mean value of  $t(\alpha)$  from 1000 times simulations (denoted by Sim) at each selected percentile. The percentage that actual return can beat the simulated counterparts among the 1000 times simulation is shown in the last column (denoted by % < Act). The sample is the selected 844 mutual funds from January 2002 to June 2020 and the result is shown for the LSY-CH-4 factor model.

Pct	Act	Sim	% < Act
1%	-1.1945	-0.1281	11.43%
2%	-0.9459	-0.1018	17.47%
3%	-0.6158	-0.0927	26.42%
4%	-0.3032	-0.0859	39.28%
5%	-0.1800	-0.0815	43.01%
6%	-0.0583	-0.0776	47.71%
7%	0.0051	-0.0731	51.40%
8%	0.1141	-0.0697	53.90%
9%	0.2432	-0.0675	60.42%
10%	0.3142	-0.0640	63.09%
20%	0.9164	-0.0434	82.10%
30%	1.2467	-0.0297	89.50%
40%	1.5859	-0.0181	93.82%
50%	1.9644	-0.0090	97.60%
60%	2.2779	-0.0005	98.80%
70%	2.6133	0.0092	99.60%
80%	3.0421	0.0219	99.90%
90%	3.5401	0.0384	100.00%
100%	6.5543	0.1725	100.00%

**Table B.2**

Simulation analysis for the Chinese mutual fund market. This table reports the results of the simulation process based on the data from the Chinese mutual funds. Panel A, B, and C report the simulation for Scenarios A, B, and C, respectively. DGP represents the true proportion of zero, positive and negative alpha funds obtained from the Data Generating Process. T-test represents the estimated proportion of zero, positive and negative alpha funds using the conventional single hypothesis t-test approach. FDR denotes the estimated proportion of zero, positive and negative alpha funds using the bootstrap-based FDR method. For example, in Panel A, we set the alpha of 0.03 for 20% and the alpha of 0 for 80%. By the conventional single hypothesis t-test method, we obtain 76.17% of zero alpha funds, 22.03% of positive alpha funds, and 1.80% of negative alpha funds. By the FDR method, we get 80.10% of zero alpha funds, 19.97% of positive alpha funds, and -0.07% of negative alpha funds. All the variables are defined in Table 6.

Panel A: Simulation 1					Panel B: Simulation 2					Panel C: Simulation 3							
Alpha	3.00%	-3.00%	0.00%		Alpha	1.00%	-1.00%	0.00%		Alpha	0.50%	-0.50%	0.00%				
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$			$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$			$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				
DGP	80%	20%	0%		DGP	80%	20%	0%		DGP	80%	20%	0%				
t-test	76.17%	22.03%	1.80%		t-test	88.56%	9.47%	1.97%		t-test	93.40%	4.52%	2.08%				
FDR	80.10%	19.97%	-0.07%		FDR	87.98%	12.29%	-0.27%		FDR	95.50%	5.06%	-0.56%				
At different significant level					At different significant level					At different significant level							
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	1033	1117	1213	1308	1413	$E(S_{\lambda}^+)$	454	670	848	994	1147	$E(S_{\lambda}^+)$	219	372	544	685	822
	21.92%	23.71%	25.74%	27.76%	29.99%		9.63%	14.22%	18.00%	21.10%	24.34%		4.65%	7.89%	11.54%	14.54%	17.44%
$E(S_{\lambda}^-)$	91	191	288	395	494	$E(S_{\lambda}^-)$	91	188	269	360	441	$E(S_{\lambda}^-)$	86	200	301	403	495
	1.93%	4.05%	6.11%	8.38%	10.48%		1.93%	3.99%	5.71%	7.64%	9.36%		1.83%	4.24%	6.39%	8.55%	10.51%
$E(F_{\lambda}^+)$	2.00%	4.00%	6.01%	8.01%	10.01%	$E(F_{\lambda}^+)$	2.20%	4.40%	6.60%	8.80%	11.00%	$E(F_{\lambda}^+)$	2.39%	4.78%	7.16%	9.55%	11.94%
$E(F_{\lambda}^-)$	2.00%	4.00%	6.01%	8.01%	10.01%	$E(F_{\lambda}^-)$	2.20%	4.40%	6.60%	8.80%	11.00%	$E(F_{\lambda}^-)$	2.39%	4.78%	7.16%	9.55%	11.94%
$E(T_{\lambda}^+)$	19.92%	19.70%	19.74%	19.75%	19.97%	$E(T_{\lambda}^+)$	7.44%	9.82%	11.40%	12.30%	13.34%	$E(T_{\lambda}^+)$	2.26%	3.12%	4.38%	4.99%	5.51%
$E(T_{\lambda}^-)$	-0.07%	0.05%	0.10%	0.37%	0.47%	$E(T_{\lambda}^-)$	-0.27%	-0.41%	-0.89%	-1.16%	-1.64%	$E(T_{\lambda}^-)$	-0.56%	-0.53%	-0.77%	-1.00%	-1.43%
$\widehat{FDR}^+$	9.13%	16.89%	23.34%	28.86%	33.39%	$\widehat{FDR}^+$	22.83%	30.94%	36.67%	41.71%	45.18%	$\widehat{FDR}^+$	51.37%	60.48%	62.04%	65.69%	68.43%
$\widehat{FDR}^-$	103.69%	98.80%	98.29%	95.55%	95.50%	$\widehat{FDR}^-$	113.89%	110.26%	115.59%	115.16%	117.51%	$\widehat{FDR}^-$	130.81%	112.50%	112.13%	111.66%	113.64%

**Table B.3**  
The simulation results of different alpha locations for the mutual funds in China.

		$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$
0.5 0%	DGP	80.00%	20.00%	0.00%
	t-test	93.40%	4.52%	2.08%
	FDR	95.50%	5.06%	-0.56%
0.60%	t-test	92.59%	5.41%	1.99%
	FDR	94.56%	5.83%	-0.39%
0.70%	t-test	91.00%	7.05%	1.95%
	FDR	90.90%	9.16%	-0.07%
0.80%	t-test	91.00%	7.05%	1.95%
	FDR	89.13%	11.04%	-0.17%
0.90%	t-test	89.86%	8.17%	1.97%
	FDR	89.32%	10.99%	-0.30%
1.00%	t-test	88.56%	9.47%	1.97%
	FDR	87.98%	12.29%	-0.27%

This table reports the simulation results of different alpha locations for the Chinese mutual fund market. The table shows the results obtained by conventional single hypothesis t-test and the FDR method for different locations of alpha. All the variables are defined in Table B.2.

**Table B.4**  
Additional simulation results for the Chinese mutual fund market. This table reports the results of the additional simulation process based on the data from the Chinese mutual funds. Panels A to D report the simulation results for four different values of alpha. All the variables are defined in Table B.2.

Panel A: Simulation 1						Panel B: Simulation 2					
Alpha	0.60%	-0.60%	0.00%			Alpha	0.70%	-0.70%	0.00%		
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
DGP	80.00%	20.00%	0.00%			DGP	80.00%	20.00%	0.00%		
t-test	92.59%	5.41%	1.99%			t-test	91.00%	7.05%	1.95%		
FDR	94.56%	5.83%	-0.39%			FDR	90.90%	9.16%	-0.07%		
At different significant level						At different significant level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\alpha}^+)$	248	419	566	711	867	$E(S_{\alpha}^+)$	313	506	684	840	1000
	5.26%	8.89%	12.01%	15.09%	18.40%		6.64%	10.74%	14.52%	17.83%	21.22%
$E(S_{\alpha}^-)$	93	195	294	408	512	$E(S_{\alpha}^-)$	104	203	290	381	462
	1.97%	4.14%	6.24%	8.66%	10.87%		2.21%	4.31%	6.15%	8.09%	9.80%
$E(F_{\alpha}^+)$	2.36%	4.73%	7.09%	9.46%	11.82%	$E(F_{\alpha}^+)$	2.27%	4.55%	6.82%	9.09%	11.36%
FF	2.36%	4.73%	7.09%	9.46%	11.82%	FF	2.27%	4.55%	6.82%	9.09%	11.36%
$E(T_{\alpha}^+)$	2.90%	4.16%	4.92%	5.63%	6.58%	$E(T_{\alpha}^+)$	4.37%	6.19%	7.70%	8.74%	9.86%
$E(T_{\alpha}^-)$	-0.39%	-0.59%	-0.85%	-0.80%	-0.95%	$E(T_{\alpha}^-)$	-0.07%	-0.24%	-0.66%	-1.00%	-1.56%
$\widehat{FDR}^+$	44.92%	53.17%	59.04%	62.67%	64.24%	$\widehat{FDR}^+$	34.21%	42.33%	46.97%	50.99%	53.54%
$\widehat{FDR}^-$	119.78%	114.25%	113.67%	109.21%	108.78%	$\widehat{FDR}^-$	102.96%	105.50%	110.78%	112.42%	115.89%
Panel C: Simulation 3						Panel D: Simulation 4					
Alpha	0.80%	-0.80%	0.00%			Alpha	0.90%	-0.90%	0.00%		
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
DGP	80.00%	20.00%	0.00%			DGP	80.00%	20.00%	0.00%		
t-test	91.00%	7.05%	1.95%			t-test	89.86%	8.17%	1.97%		
FDR	89.13%	11.04%	-0.17%			FDR	89.32%	10.99%	-0.30%		
At different significant level						At different significant level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\alpha}^+)$	324	535	721	895	1046	$E(S_{\alpha}^+)$	374	576	753	909	1063
	6.88%	11.35%	15.30%	18.99%	22.20%		7.94%	12.22%	15.98%	19.29%	22.56%
$E(S_{\alpha}^-)$	97	187	283	396	495	$E(S_{\alpha}^-)$	91	191	291	398	498
	2.06%	3.97%	6.01%	8.40%	10.51%		1.93%	4.05%	6.18%	8.45%	10.57%
$E(F_{\alpha}^+)$	2.23%	4.46%	6.69%	8.91%	11.14%	$E(F_{\alpha}^+)$	2.23%	4.47%	6.70%	8.93%	11.16%
$E(F_{\alpha}^-)$	2.23%	4.46%	6.69%	8.91%	11.14%	$E(F_{\alpha}^-)$	2.23%	4.47%	6.70%	8.93%	11.16%
$E(T_{\alpha}^+)$	4.65%	6.90%	8.62%	10.08%	11.06%	$E(T_{\alpha}^+)$	5.70%	7.76%	9.28%	10.36%	11.39%
$E(T_{\alpha}^-)$	-0.17%	-0.49%	-0.68%	-0.51%	-0.64%	$E(T_{\alpha}^-)$	-0.30%	-0.41%	-0.52%	-0.49%	-0.60%
$\widehat{FDR}^+$	32.41%	39.25%	43.69%	46.93%	50.19%	$\widehat{FDR}^+$	28.13%	36.53%	41.92%	46.30%	49.49%
$\widehat{FDR}^-$	108.25%	112.30%	111.31%	106.06%	106.06%	$\widehat{FDR}^-$	115.62%	110.17%	108.47%	105.74%	105.64%

**Table B.5**

The distribution of significant alpha of the mutual funds in China. This table reports the number and the proportion of funds with absolute alpha larger than 0.50%–1.00%. No. and Perc. denote the number and the percentage of funds with each level of absolute alpha. Pos. and Neg. denotes the significantly positive and negative alpha funds.

>	Pos.		Neg.	
	No.	Perc.	No.	Perc.
0.50%	472	100.00%	92	100.00%
0.60%	472	100.00%	92	100.00%
0.70%	472	100.00%	92	100.00%
0.80%	472	100.00%	92	100.00%
0.90%	468	99.15%	92	100.00%
1.00%	455	96.40%	87	94.57%

## Appendix C

This Appendix provides the results of the simulation results for the U.S. mutual funds including the simulation results (Tables C.1 and C.3) and details (Table C.2) of different alpha locations.

**Table C.1**

Simulation analysis for the U.S. mutual fund market. This table reports the results of the simulation process based on the data from the U.S. mutual fund market. Panel A, B, and C report the simulation for Scenario A\*, B\*, and C\*, respectively. DGP represents the true proportion of zero, positive and negative alpha funds obtained from the Data Generating Process. All the variables are defined in Table B.2.

Panel A: Simulation 1					Panel B: Simulation 2					Panel C: Simulation 3							
Alpha	3.80%	-3.20%	0.00%		Alpha	0.50%	-0.50%	0.00%		Alpha	0.10%	-0.10%	0.00%				
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$			$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$			$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				
DGP	75.00%	2.00%	23.00%		DGP	75.00%	2.00%	23.00%		DGP	75.00%	2.00%	23.00%				
t-test	71.52%	3.59%	24.89%		t-test	86.50%	2.36%	11.14%		t-test	95.20%	1.87%	2.93%				
FDR	74.90%	1.87%	23.23%		FDR	82.67%	0.46%	16.88%		FDR	99.71%	-0.47%	0.76%				
At different significance level					At different significance level					At different significance level							
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\hat{\alpha}}^+)$	124	173	246	301	367	$E(S_{\hat{\alpha}}^+)$	82	143	224	281	350	$E(S_{\hat{\alpha}}^+)$	67	125	216	273	352
	3.75%	5.23%	7.43%	9.09%	11.08%		2.48%	4.32%	6.77%	8.49%	10.57%		2.02%	3.78%	6.52%	8.25%	10.63%
$E(S_{\hat{\alpha}}^-)$	823	889	957	1026	1090	$E(S_{\hat{\alpha}}^-)$	374	548	679	795	895	$E(S_{\hat{\alpha}}^-)$	97	203	302	395	483
	24.86%	26.85%	28.90%	30.99%	32.92%		11.30%	16.55%	20.51%	24.01%	27.03%		2.93%	6.13%	9.12%	11.93%	14.59%
$E(F_{\hat{\alpha}}^+)$	1.87%	3.75%	5.62%	7.49%	9.36%	$E(F_{\hat{\alpha}}^+)$	2.02%	4.04%	6.06%	8.08%	10.11%	$E(F_{\hat{\alpha}}^+)$	2.49%	4.99%	7.48%	9.97%	12.46%
$E(F_{\hat{\alpha}}^-)$	1.87%	3.75%	5.62%	7.49%	9.36%	$E(F_{\hat{\alpha}}^-)$	2.02%	4.04%	6.06%	8.08%	10.11%	$E(F_{\hat{\alpha}}^-)$	2.49%	4.99%	7.48%	9.97%	12.46%
$E(T_{\hat{\alpha}}^+)$	1.87%	1.48%	1.81%	1.60%	1.72%	$E(T_{\hat{\alpha}}^+)$	0.46%	0.28%	0.70%	0.40%	0.47%	$E(T_{\hat{\alpha}}^+)$	-0.47%	-1.21%	-0.95%	-1.73%	-1.83%
$E(T_{\hat{\alpha}}^-)$	22.98%	23.10%	23.29%	23.50%	23.56%	$E(T_{\hat{\alpha}}^-)$	9.27%	12.51%	14.44%	15.93%	16.93%	$E(T_{\hat{\alpha}}^-)$	0.44%	1.15%	1.64%	1.96%	2.12%
$\widehat{FDR}^+$	50.00%	71.68%	75.61%	82.39%	84.47%	$\widehat{FDR}^+$	81.61%	93.59%	89.62%	95.26%	95.60%	$\widehat{FDR}^+$	123.19%	132.06%	114.63%	120.93%	117.24%
$\widehat{FDR}^-$	7.53%	13.95%	19.44%	24.17%	28.44%	$\widehat{FDR}^-$	17.89%	24.42%	29.57%	33.67%	37.38%	$\widehat{FDR}^-$	85.09%	81.32%	81.99%	83.58%	85.44%

**Table C.2**

The simulation results of different alpha locations for the mutual funds in US.

		$\hat{\alpha}_0$	$\hat{\alpha}^+$	$\hat{\alpha}^-$
	DGP	75.00%	2.00%	23.00%
0.10%	t-test	95.20%	1.87%	2.93%
	FDR	99.71%	-0.47%	0.76%
0.20%	t-test	93.93%	1.81%	4.26%
	FDR	96.60%	-0.48%	3.88%
0.30%	t-test	86.50%	2.36%	11.14%
	FDR	92.16%	-0.31%	8.15%
0.40%	t-test	89.73%	2.23%	8.03%
	FDR	87.03%	0.30%	12.67%
0.50%	t-test	86.50%	2.36w%	11.14%
	FDR	82.67%	0.46%	16.88%

This table reports the simulation results of different alpha locations for the US mutual fund market. The table shows the results obtained by conventional single hypothesis t-test and the FDR method for different locations of alpha. All the variables are defined in [Table B.2](#).

**Table C.3**

Additional simulation results for the U.S. mutual fund market. This table reports the results of the additional simulation process based on the data from the U.S. mutual fund. Panels A to C report the simulation results for three different values of alpha. All the variables are defined in Table B.2.

Panel A: Simulation 1					Panel B: Simulation 2					Panel C: Simulation 3							
Alpha	0.20%	-0.20%	0.00%		Alpha	0.30%	-0.30%	0.00%		Alpha	0.40%	-0.40%	0.00%				
	$\hat{\alpha}_0$	$\hat{\alpha}^+$	$\hat{\alpha}^-$			$\hat{\alpha}_0$	$\hat{\alpha}^+$	$\hat{\alpha}^-$			$\hat{\alpha}_0$	$\hat{\alpha}^+$	$\hat{\alpha}^-$				
DGP	75.00%	2.00%	23.00%		DGP	75.00%	2.00%	23.00%		DGP	75.00%	2.00%	23.00%				
t-test	93.93%	1.81%	4.26%		t-test	86.50%	2.36%	11.14%		t-test	89.73%	2.23%	8.03%				
FDR	96.60%	-0.48%	3.88%		FDR	92.16%	-0.31%	8.15%		FDR	87.03%	0.30%	12.67%				
At different significance level					At different significance level					At different significance level							
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda^+}^+)$	64	120	199	263	338	$E(S_{\lambda^+}^+)$	66	121	199	262	331	$E(S_{\lambda^+}^+)$	82	137	217	276	346
	1.93%	3.62%	6.01%	7.94%	10.21%		1.99%	3.65%	6.01%	7.91%	0.09997		2.48%	4.14%	6.55%	8.34%	10.45%
$E(S_{\lambda^-}^-)$	140	251	363	479	576	$E(S_{\lambda^-}^-)$	196	336	456	577	682	$E(S_{\lambda^-}^-)$	278	431	571	696	793
	4.23%	7.58%	10.96%	14.47%	17.40%		5.92%	10.15%	13.77%	17.43%	20.60%		8.40%	13.02%	17.25%	21.02%	23.95%
$E(F_{\lambda^+}^+)$	2.42%	4.83%	7.25%	9.66%	12.08%	$E(F_{\lambda^+}^+)$	2.30%	4.61%	6.91%	9.22%	11.52%	$E(F_{\lambda^+}^+)$	2.18%	4.35%	6.53%	8.70%	10.88%
$E(F_{\lambda^-}^-)$	2.42%	4.83%	7.25%	9.66%	12.08%	$E(F_{\lambda^-}^-)$	2.30%	4.61%	6.91%	9.22%	11.52%	$E(F_{\lambda^-}^-)$	2.18%	4.35%	6.53%	8.70%	10.88%
$E(T_{\lambda^+}^+)$	-0.48%	-1.21%	-1.24%	-1.72%	-1.87%	$E(T_{\lambda^+}^+)$	-0.31%	-0.95%	-0.90%	-1.30%	-1.52%	$E(T_{\lambda^+}^+)$	0.30%	-0.21%	0.03%	-0.37%	-0.43%
$E(T_{\lambda^-}^-)$	1.81%	2.75%	3.72%	4.81%	5.32%	$E(T_{\lambda^-}^-)$	3.62%	5.54%	6.86%	8.21%	9.08%	$E(T_{\lambda^-}^-)$	6.22%	8.67%	10.72%	12.32%	13.07%
$\widehat{FDR}^+$	124.94%	133.27%	120.55%	121.62%	118.29%	$\widehat{FDR}^+$	115.58%	126.09%	115.00%	116.47%	115.24%	$\widehat{FDR}^+$	87.85%	105.16%	99.59%	104.40%	104.10%
$\widehat{FDR}^-$	57.12%	63.72%	66.09%	66.78%	69.41%	$\widehat{FDR}^-$	38.92%	45.41%	50.19%	52.88%	55.93%	$\widehat{FDR}^-$	25.91%	33.43%	37.85%	41.40%	45.42%

Appendix D

This Appendix provides the results robustness checks in Section 5. we perform to assess whether our result is sensitive to altering the number of factors in the benchmark models, alternative benchmark models and different rolling windows.

Table D.4

The robustness check of using the LSY-CH-3 factor model. This table reports the robustness check of using the LSY-CH-3 factor model for the enlarged sample. We exhibit the results of the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panel A–D, respectively. All the variables are defined in Table 6.

Panel A: Full sample funds						Panel B: Growth-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
%	80.83%	19.52%	-0.34%			%	70.75%	29.12%	0.13%		
At different significance level						At different significance level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
SZ	612 12.99%	899 19.08%	1119 23.75%	1295 27.48%	1485 31.52%	SZ	167 19.79%	233 27.61%	272 32.23%	293 34.72%	327 38.74%
SF	79 1.68%	135 2.87%	192 4.07%	257 5.45%	307 6.52%	SF	16 1.90%	26 3.08%	32 3.79%	39 4.62%	47 5.57%
$E(F_{\lambda}^+)$	2.02%	4.04%	6.06%	8.08%	10.10%	$E(F_{\lambda}^+)$	1.77%	3.54%	5.31%	7.08%	8.84%
$E(F_{\lambda}^-)$	2.02%	4.04%	6.06%	8.08%	10.10%	$E(F_{\lambda}^-)$	1.77%	3.54%	5.31%	7.08%	8.84%
$E(T_{\lambda}^+)$	10.97%	15.04%	17.69%	19.40%	21.41%	$E(T_{\lambda}^+)$	18.02%	24.07%	26.92%	27.64%	29.90%
$E(T_{\lambda}^-)$	-0.34%	-1.18%	-1.99%	-2.63%	-3.59%	$E(T_{\lambda}^-)$	0.13%	-0.46%	-1.51%	-2.45%	-3.28%
FDR+	15.56%	21.18%	25.53%	29.41%	32.06%	FDR+	8.94%	12.81%	16.47%	20.38%	22.83%
FDR-	120.52%	141.06%	148.77%	148.19%	155.07%	FDR-	93.30%	114.84%	139.96%	153.11%	158.81%
Panel C: Value-oriented funds						Panel D: Balance-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
%	87.91%	11.31%	0.78%			%	84.72%	15.66%	-0.38%		
At different significance level						At different significance level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
SZ	34 8.44%	47 11.66%	65 16.13%	78 19.35%	92 22.83%	SZ	234 10.44%	355 15.84%	466 20.79%	552 24.63%	644 28.74%
SF	12 2.98%	17 4.22%	29 7.20%	41 10.17%	45 11.17%	SF	39 1.74%	65 2.90%	85 3.79%	110 4.91%	141 6.29%
$E(F_{\lambda}^+)$	2.20%	4.40%	6.59%	8.79%	10.99%	$E(F_{\lambda}^+)$	2.12%	4.24%	6.35%	8.47%	10.59%
$E(F_{\lambda}^-)$	2.20%	4.40%	6.59%	8.79%	10.99%	$E(F_{\lambda}^-)$	2.12%	4.24%	6.35%	8.47%	10.59%
$E(T_{\lambda}^+)$	6.24%	7.27%	9.54%	10.56%	11.84%	$E(T_{\lambda}^+)$	8.32%	11.61%	14.44%	16.16%	18.15%
$E(T_{\lambda}^-)$	0.78%	-0.18%	0.60%	1.38%	0.18%	$E(T_{\lambda}^-)$	-0.38%	-1.34%	-2.56%	-3.56%	-4.30%
FDR+	26.05%	37.69%	40.88%	45.42%	48.14%	FDR+	20.28%	26.74%	30.56%	34.39%	36.85%
FDR-	73.81%	104.20%	91.63%	86.41%	98.41%	FDR-	121.70%	146.04%	167.52%	172.60%	168.31%

Table D.5

The robustness check of using the Fama–French–Carhart four-factor model. This table reports the robustness check of using the Fama–French–Carhart four-factor model for the enlarged sample. We exhibit the results of the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panel A–D, respectively. All the variables are defined in Table 6.

Panel A: Full sample funds						Panel B: Growth-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
%	78.01%	19.44%	2.55%			%	70.07%	29.32%	0.61%		
At different significance level						At different significance level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	517 10.97%	742 15.75%	947 20.10%	1,111 23.58%	1252 26.57%	$E(S_{\lambda}^+)$	142 16.82%	193 22.87%	244 28.91%	280 33.18%	305 36.14%
$E(S_{\lambda}^-)$	212 4.50%	340 7.22%	438 9.30%	557 11.82%	645 13.69%	$E(S_{\lambda}^-)$	20 2.37%	39 4.62%	47 5.57%	62 7.35%	77 9.12%
$E(F_{\lambda}^+)$	1.95%	3.90%	5.85%	7.80%	9.75%	$E(F_{\lambda}^+)$	1.75%	3.50%	5.26%	7.01%	8.76%
$E(F_{\lambda}^-)$	1.95%	3.90%	5.85%	7.80%	9.75%	$E(F_{\lambda}^-)$	1.75%	3.50%	5.26%	7.01%	8.76%
$E(T_{\lambda}^+)$	9.02%	11.85%	14.25%	15.78%	16.82%	$E(T_{\lambda}^+)$	15.07%	19.36%	23.65%	26.17%	27.38%
$E(T_{\lambda}^-)$	2.55%	3.32%	3.44%	4.02%	3.94%	$E(T_{\lambda}^-)$	0.62%	1.12%	0.31%	0.34%	0.36%
$\widehat{FDR}^+$	17.77%	24.77%	29.11%	33.08%	36.70%	$\widehat{FDR}^+$	10.41%	15.32%	18.18%	21.12%	24.24%
$\widehat{FDR}^-$	43.35%	54.05%	62.94%	65.99%	71.23%	$\widehat{FDR}^-$	73.93%	75.82%	94.38%	95.39%	96.01%

(continued on next page)

Table D.5 (continued).

Panel C: Value-oriented funds						Panel D: Balance-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
%	72.31%	25.27%	2.41%			%	80.07%	17.03%	2.91%		
At different significance level						At different significance level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	49	64	82	98	117	$E(S_{\lambda}^+)$	258	354	459	528	575
	12.16%	15.88%	20.35%	24.32%	29.03%		11.51%	15.80%	20.48%	23.56%	25.66%
$E(S_{\lambda}^-)$	17	34	42	585	62	$E(S_{\lambda}^-)$	110	165	214	256	297
	4.22%	8.44%	10.42%	14.39%	15.38%		4.91%	7.36%	9.55%	11.42%	13.25%
$E(F_{\lambda}^+)$	1.81%	3.62%	5.42%	7.23%	9.04%	$E(F_{\lambda}^+)$	2.00%	4.00%	6.00%	8.01%	10.01%
$E(F_{\lambda}^-)$	1.81%	3.62%	5.42%	7.23%	9.04%	$E(F_{\lambda}^-)$	2.00%	4.00%	6.00%	8.01%	10.01%
$E(T_{\lambda}^+)$	10.35%	12.27%	14.92%	17.09%	19.99%	$E(T_{\lambda}^+)$	9.51%	11.79%	14.48%	15.55%	15.65%
$E(T_{\lambda}^-)$	2.41%	4.82%	5.00%	7.16%	6.35%	$E(T_{\lambda}^-)$	2.91%	3.36%	3.54%	3.42%	3.24%
$\widehat{FDR}_{\lambda}^+$	14.87%	22.77%	26.66%	29.74%	31.14%	$\widehat{FDR}_{\lambda}^+$	17.39%	25.34%	29.32%	33.98%	39.01%
$\widehat{FDR}_{\lambda}^-$	42.86%	42.86%	52.04%	50.25%	58.76%	$\widehat{FDR}_{\lambda}^-$	40.78%	54.37%	62.88%	70.09%	75.52%

Table D.6

The robustness check of using the Fama–French three-factor model. This table reports the robustness check of using the Fama–French three-factor model for the enlarged sample. We exhibit the results of the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panel A–D, respectively. All the variables are defined in Table 6.

Panel A: Full sample funds						Panel B: Growth-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
%	78.49%	20.10%	1.41%			%	67.20%	32.71%	0.10%		
At different significance level						At different significance level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	572	806	1007	1169	1316	$E(S_{\lambda}^+)$	148	215	265	310	343
	12.14%	17.11%	21.37%	24.81%	27.93%		17.54%	25.47%	31.40%	36.73%	40.64%
$E(S_{\lambda}^-)$	159	274	395	486	579	$E(S_{\lambda}^-)$	15	25	34	45	56
	3.37%	5.81%	8.38%	10.31%	12.29%		1.78%	2.96%	4.03%	5.33%	6.64%
$E(F_{\lambda}^+)$	1.96%	3.92%	5.89%	7.85%	9.81%	$E(F_{\lambda}^+)$	1.68%	3.36%	5.04%	6.72%	8.40%
$E(F_{\lambda}^-)$	1.96%	3.92%	5.89%	7.85%	9.81%	$E(F_{\lambda}^-)$	1.68%	3.36%	5.04%	6.72%	8.40%
$E(T_{\lambda}^+)$	10.18%	13.18%	15.48%	16.96%	18.12%	$E(T_{\lambda}^+)$	15.86%	22.11%	26.36%	30.01%	32.24%
$E(T_{\lambda}^-)$	1.41%	1.89%	2.50%	2.46%	2.48%	$E(T_{\lambda}^-)$	0.10%	0.00%	0.00%	0.00%	0.00%
$\widehat{FDR}_{\lambda}^+$	16.17%	22.94%	27.55%	31.64%	35.13%	$\widehat{FDR}_{\lambda}^+$	9.58%	13.19%	16.05%	18.29%	20.67%
$\widehat{FDR}_{\lambda}^-$	58.15%	67.49%	70.23%	76.10%	79.85%	$\widehat{FDR}_{\lambda}^-$	94.52%	100.00%	100.00%	100.00%	100.00%
Panel C: Value-oriented funds						Panel D: Balance-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
%	74.80%	23.33%	1.87%			%	81.91%	16.04%	2.05%		
At different significance level						At different significance level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	52	69	84	105	118	$E(S_{\lambda}^+)$	269	372	464	552	620
	12.90%	17.12%	20.84%	26.05%	29.28%		12.00%	16.60%	20.71%	24.63%	27.67%
$E(S_{\lambda}^-)$	11	22	31	40	53	$E(S_{\lambda}^-)$	71	117	161	206	248
	2.73%	5.46%	7.69%	9.93%	13.15%		3.17%	5.22%	7.18%	9.19%	11.07%
$E(F_{\lambda}^+)$	1.87%	3.74%	5.61%	7.48%	9.35%	$E(F_{\lambda}^+)$	2.05%	4.10%	6.14%	8.19%	10.24%
$E(F_{\lambda}^-)$	1.87%	3.74%	5.61%	7.48%	9.35%	$E(F_{\lambda}^-)$	2.05%	4.10%	6.14%	8.19%	10.24%
$E(T_{\lambda}^+)$	11.03%	13.38%	15.23%	18.57%	19.93%	$E(T_{\lambda}^+)$	9.96%	12.50%	14.56%	16.44%	17.43%
$E(T_{\lambda}^-)$	0.86%	1.72%	2.08%	2.45%	3.80%	$E(T_{\lambda}^-)$	1.12%	1.13%	1.04%	1.00%	0.83%
$\widehat{FDR}_{\lambda}^+$	14.49%	21.84%	26.91%	28.71%	31.93%	$\widehat{FDR}_{\lambda}^+$	17.06%	24.67%	29.67%	33.26%	37.01%
$\widehat{FDR}_{\lambda}^-$	68.51%	68.51%	72.93%	75.36%	71.09%	$\widehat{FDR}_{\lambda}^-$	64.64%	78.45%	85.51%	89.11%	92.53%

**Table D.7**

The robustness check of using the Fama–French five-factor model. This table reports the robustness check of using the Fama–French five-factor model for the enlarged sample. We exhibit the results of the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panel A–D, respectively. All the variables are defined in Table 6.

Panel A: Full sample funds						Panel B: Growth-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$			
%	79.16%	19.47%	1.37%			%	69.71%	29.67%	0.63%		
At different significance level						At different significance level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	466	694	910	1075	1234	$E(S_{\lambda}^+)$	137	196	234	270	302
	9.89%	14.73%	19.31%	22.81%	26.19%		16.23%	23.22%	27.73%	31.99%	35.78%
$E(S_{\lambda}^-)$	158	263	362	444	535	$E(S_{\lambda}^-)$	20	34	44	52	63
	3.35%	5.58%	7.68%	9.42%	11.35%		2.37%	4.03%	5.21%	6.16%	7.46%
$E(F_{\lambda}^+)$	1.98%	3.96%	5.94%	7.92%	9.89%	$E(F_{\lambda}^+)$	1.74%	3.49%	5.23%	6.97%	8.71%
$E(F_{\lambda}^-)$	1.98%	3.96%	5.94%	7.92%	9.89%	$E(F_{\lambda}^-)$	1.74%	3.49%	5.23%	6.97%	8.71%
$E(T_{\lambda}^+)$	7.91%	10.77%	13.38%	14.90%	16.29%	$E(T_{\lambda}^+)$	14.49%	19.74%	22.50%	25.02%	27.07%
$E(T_{\lambda}^-)$	1.37%	1.62%	1.75%	1.51%	1.46%	$E(T_{\lambda}^-)$	0.63%	0.54%	-0.01%	-0.81%	-1.25%
$\widehat{FDR}^+$	20.01%	26.87%	30.74%	34.70%	37.78%	$\widehat{FDR}^+$	10.74%	15.01%	18.86%	21.79%	24.35%
$\widehat{FDR}^-$	59.02%	70.91%	77.28%	84.01%	87.15%	$\widehat{FDR}^-$	73.54%	86.52%	100.28%	113.14%	116.73%
Panel C: Value-oriented funds						Panel D: Balance-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
%	75.68%	23.23%	1.09%			%	82.93%	15.70%	1.36%		
At different significance level						At different significance level					
$\gamma$	0.05	0.1	0.15	0.2	0.25	$\gamma$	0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	46	62	78	88	109	$E(S_{\lambda}^+)$	229	330	422	497	569
	11.41%	15.38%	19.35%	21.84%	27.05%		10.22%	14.73%	18.83%	22.18%	25.39%
$E(S_{\lambda}^-)$	12	26	36	45	56	$E(S_{\lambda}^-)$	77	127	176	215	258
	2.98%	6.45%	8.93%	11.17%	13.90%		3.44%	5.67%	7.85%	9.59%	11.51%
$E(F_{\lambda}^+)$	1.89%	3.78%	5.68%	7.57%	9.46%	$E(F_{\lambda}^+)$	2.07%	4.15%	6.22%	8.29%	10.37%
$E(F_{\lambda}^-)$	1.89%	3.78%	5.68%	7.57%	9.46%	$E(F_{\lambda}^-)$	2.07%	4.15%	6.22%	8.29%	10.37%
$E(T_{\lambda}^+)$	9.52%	11.60%	13.68%	14.27%	17.59%	$E(T_{\lambda}^+)$	8.15%	10.58%	12.61%	13.88%	15.02%
$E(T_{\lambda}^-)$	1.09%	2.67%	3.26%	3.60%	4.44%	$E(T_{\lambda}^-)$	1.36%	1.52%	1.63%	1.30%	1.15%
$\widehat{FDR}^+$	16.58%	24.60%	29.33%	34.66%	34.98%	$\widehat{FDR}^+$	20.29%	28.16%	33.03%	37.40%	40.83%
$\widehat{FDR}^-$	63.54%	58.65%	63.54%	67.78%	68.08%	$\widehat{FDR}^-$	60.34%	73.17%	79.20%	86.45%	90.05%

**Table D.8**

The robustness check of using the CAPM model. This table reports the robustness check of using the CAPM model for the enlarged sample. We exhibit the results of the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panel A–D, respectively. All the variables are defined in Table 6.

Panel A: Full sample funds						Panel B: Growth-oriented funds					
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		
%	84.04%	16.70%	-0.74%			%	68.13%	31.92%	-0.04%		
At different significance level						At different significance level					
	0.05	0.1	0.15	0.2	0.25		0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	622	888	1114	1292	1473	$E(S_{\lambda}^+)$	184	246	293	324	356
	13.20%	18.85%	23.64%	27.42%	31.26%		21.80%	29.15%	34.72%	38.39%	42.18%
$E(S_{\lambda}^-)$	64	117	157	204	251	$E(S_{\lambda}^-)$	14	20	28	33	40
	1.36%	2.48%	3.33%	4.33%	5.33%		1.66%	2.37%	3.32%	3.91%	4.74%
$E(F_{\lambda}^+)$	2.10%	4.20%	6.30%	8.40%	10.51%	$E(F_{\lambda}^+)$	1.70%	3.41%	5.11%	6.81%	8.52%
$E(F_{\lambda}^-)$	2.10%	4.20%	6.30%	8.40%	10.51%	$E(F_{\lambda}^-)$	1.70%	3.41%	5.11%	6.81%	8.52%
$E(T_{\lambda}^+)$	11.10%	14.64%	17.34%	19.02%	20.76%	$E(T_{\lambda}^+)$	20.10%	25.74%	29.61%	31.58%	33.66%
$E(T_{\lambda}^-)$	-0.74%	-1.72%	-2.97%	-4.07%	-5.18%	$E(T_{\lambda}^-)$	-0.04%	-1.04%	-1.79%	-2.90%	-3.78%
$\widehat{FDR}^+$	15.92%	22.30%	26.66%	30.65%	33.60%	$\widehat{FDR}^+$	7.81%	11.69%	14.72%	17.75%	20.19%
$\widehat{FDR}^-$	154.69%	169.23%	189.17%	194.12%	197.21%	$\widehat{FDR}^-$	102.68%	143.75%	154.02%	174.24%	179.69%

(continued on next page)

**Table D.8 (continued).**

Panel C: Value-oriented funds					Panel D: Balance-oriented funds						
	$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$		$\hat{\pi}_0$	$\hat{\pi}^+$	$\hat{\pi}^-$				
%	81.89%	17.93%	0.19%		%	83.19%	17.46%	-0.65%			
At different significance level					At different significance level						
	0.05	0.1	0.15	0.2	0.25	0.05	0.1	0.15	0.2	0.25	
$E(S_{\lambda}^+)$	60	74	101	112	126	$E(S_{\lambda}^+)$	299	430	535	623	712
	14.89%	18.36%	25.06%	27.79%	31.27%		13.34%	19.19%	23.87%	27.80%	31.77%
$E(S_{\lambda}^-)$	9	11	13	20	27	$E(S_{\lambda}^-)$	32	61	83	101	123
	2.23%	2.73%	3.23%	4.96%	6.70%		1.43%	2.72%	3.70%	4.51%	5.49%
$E(F_{\lambda}^+)$	2.05%	4.09%	6.14%	8.19%	10.24%	$E(F_{\lambda}^+)$	2.08%	4.16%	6.24%	8.32%	10.40%
$E(F_{\lambda}^-)$	2.05%	4.09%	6.14%	8.19%	10.24%	$E(F_{\lambda}^-)$	2.08%	4.16%	6.24%	8.32%	10.40%
$E(T_{\lambda}^+)$	12.84%	14.27%	18.92%	19.60%	21.03%	$E(T_{\lambda}^+)$	11.26%	15.03%	17.63%	19.48%	21.37%
$E(T_{\lambda}^-)$	0.19%	-1.36%	-2.92%	-3.23%	-3.54%	$E(T_{\lambda}^-)$	-0.65%	-1.44%	-2.54%	-3.81%	-4.91%
$\widehat{FDR}_{\lambda}^+$	13.75%	22.30%	24.50%	29.46%	32.74%	$\widehat{FDR}_{\lambda}^+$	15.59%	21.68%	26.13%	29.92%	32.73%
$\widehat{FDR}_{\lambda}^-$	91.67%	150.00%	190.38%	165.00%	152.78%	$\widehat{FDR}_{\lambda}^-$	145.65%	152.81%	168.46%	184.58%	189.46%

**Table D.9**

The robustness check of using the 5-year overlapping rolling window. This table reports the robustness check of using the 5-year overlapping rolling window for the enlarged sample. We exhibit the results of the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panel A–D, respectively. All the variables are defined in Table 6.

Panel A: Full sample funds					Panel B: Growth-oriented funds						
	$\pi_0$	$\pi_+$	$\pi_-$		$\pi_0$	$\pi_+$	$\pi_-$				
%	81.57%	19.55%	-1.12%		%	72.94%	28.61%	-1.56%			
$\gamma$	At different significant level					$\gamma$	At different significant level				
	0.05	0.1	0.15	0.2	0.25	0.05	0.1	0.15	0.2	0.25	
$E(S_{\lambda}^+)$	434	638	807	957	1066	$E(S_{\lambda}^+)$	85	104	125	138	153
	13.30%	19.55%	24.73%	29.33%	32.67%		22.55%	27.59%	33.16%	36.60%	40.58%
$E(S_{\lambda}^-)$	30	58	93	123	169	$E(S_{\lambda}^-)$	1	7	11	16	22
	0.92%	1.78%	2.85%	3.77%	5.18%		0.27%	1.86%	2.92%	4.24%	5.84%
$F(S_{\lambda}^+)$	2.04%	4.08%	6.12%	8.16%	10.20%	$F(S_{\lambda}^+)$	1.82%	3.65%	5.47%	7.29%	9.12%
$F(S_{\lambda}^-)$	2.04%	4.08%	6.12%	8.16%	10.20%	$F(S_{\lambda}^-)$	1.82%	3.65%	5.47%	7.29%	9.12%
$F(T_{\lambda}^+)$	11.26%	15.47%	18.61%	21.17%	22.47%	$F(T_{\lambda}^+)$	20.72%	23.94%	27.69%	29.31%	31.47%
$F(T_{\lambda}^-)$	-1.12%	-2.30%	-3.27%	-4.39%	-5.02%	$F(T_{\lambda}^-)$	-1.56%	-1.79%	-2.55%	-3.05%	-3.28%
$\widehat{FDR}_{\lambda}^+$	15.33%	20.86%	24.74%	27.81%	31.21%	$\widehat{FDR}_{\lambda}^+$	8.09%	13.22%	16.50%	19.93%	22.47%
$\widehat{FDR}_{\lambda}^-$	221.81%	229.45%	214.65%	216.40%	196.87%	$\widehat{FDR}_{\lambda}^-$	687.50%	196.43%	187.50%	171.88%	156.25%
Panel C: Value-oriented funds					Panel D: Balance-oriented funds						
	$\pi_0$	$\pi_+$	$\pi_-$		$\pi_0$	$\pi_+$	$\pi_-$				
%	94.87%	5.61%	-0.48%		%	86.22%	12.13%	1.65%			
$\gamma$	At different significant level					$\gamma$	At different significant level				
	0.05	0.1	0.15	0.2	0.25	0.05	0.1	0.15	0.2	0.25	
$E(S_{\lambda}^+)$	27	37	45	55	62	$E(S_{\lambda}^+)$	181	273	345	417	469
	10.42%	14.29%	17.37%	21.24%	23.94%		50.56%	76.26%	96.37%	116.48%	131.01%
$E(S_{\lambda}^-)$	5	9	10	13	16	$E(S_{\lambda}^-)$	14	27	49	64	88
	1.93%	3.47%	3.86%	5.02%	6.18%		3.80%	7.34%	13.32%	17.39%	23.91%
$F(S_{\lambda}^+)$	2.37%	4.74%	7.12%	9.49%	11.86%	$F(S_{\lambda}^+)$	2.16%	4.31%	6.47%	8.62%	10.78%
$F(S_{\lambda}^-)$	2.37%	4.74%	7.12%	9.49%	11.86%	$F(S_{\lambda}^-)$	2.16%	4.31%	6.47%	8.62%	10.78%
$F(T_{\lambda}^+)$	8.05%	9.54%	10.26%	11.75%	12.08%	$F(T_{\lambda}^+)$	48.40%	71.95%	89.90%	107.86%	120.23%
$F(T_{\lambda}^-)$	-0.44%	-1.27%	-3.25%	-4.47%	-5.68%	$F(T_{\lambda}^-)$	1.65%	3.03%	6.85%	8.77%	13.14%
$\widehat{FDR}_{\lambda}^+$	22.75%	33.20%	40.95%	44.68%	49.54%	$\widehat{FDR}_{\lambda}^+$	4.26%	5.65%	6.71%	7.40%	8.23%
$\widehat{FDR}_{\lambda}^-$	122.86%	136.51%	184.29%	189.01%	191.96%	$\widehat{FDR}_{\lambda}^-$	56.66%	58.75%	48.56%	49.57%	45.07%

**Table D.10**

The robustness check of using the 5-year non-overlapping rolling window. This table reports the robustness check of using the 5-year non-overlapping rolling window for the enlarged sample by following BSW. We exhibit the results of the full sample, growth-oriented, value-oriented, and balance-oriented funds in Panel A–D, respectively. All the variables are defined in Table 6.

Panel A: Full sample funds						Panel B: Growth-oriented funds					
	$\pi_0$	$\pi_+$	$\pi_-$			$\pi_0$	$\pi_+$	$\pi_-$			
%	78.32%	23.21%	−1.52%			78.45%	23.51%	−1.96%			
$\gamma$	At different significant level					$\gamma$	At different significant level				
	0.05	0.1	0.15	0.2	0.25		0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	141	211	264	296	328	$E(S_{\lambda}^+)$	39	48	59	63	67
	15.28%	22.86%	28.60%	32.07%	35.54%		22.54%	27.75%	34.10%	36.42%	38.73%
$E(S_{\lambda}^-)$	4	14	24	33	47	$E(S_{\lambda}^-)$	0	1	4	6	7
	0.43%	1.52%	2.60%	3.58%	5.09%		0.00%	0.58%	2.31%	3.47%	4.05%
$F(S_{\lambda}^+)$	1.96%	3.92%	5.87%	7.83%	9.79%	$F(S_{\lambda}^+)$	1.96%	3.92%	5.88%	7.84%	9.81%
	1.96%	3.92%	5.87%	7.83%	9.79%		1.96%	3.92%	5.88%	7.84%	9.81%
$F(S_{\lambda}^-)$	13.32%	18.94%	22.73%	24.24%	25.75%	$F(S_{\lambda}^-)$	20.58%	23.82%	28.22%	28.57%	28.92%
	13.32%	18.94%	22.73%	24.24%	25.75%		20.58%	23.82%	28.22%	28.57%	28.92%
$F(T_{\lambda}^+)$	−1.52%	−2.40%	−3.27%	−4.26%	−4.70%	$F(T_{\lambda}^+)$	−1.96%	−3.34%	−3.57%	−4.38%	−5.76%
	−1.52%	−2.40%	−3.27%	−4.26%	−4.70%		−1.96%	−3.34%	−3.57%	−4.38%	−5.76%
$\widehat{FDR}^+$	12.82%	17.13%	20.54%	24.42%	27.55%	$\widehat{FDR}^+$	8.70%	14.14%	17.25%	21.54%	25.32%
	12.82%	17.13%	20.54%	24.42%	27.55%		8.70%	14.14%	17.25%	21.54%	25.32%
$\widehat{FDR}^-$	451.79%	258.16%	225.89%	219.05%	192.25%	$\widehat{FDR}^-$	–	678.57%	254.46%	226.19%	242.35%
	451.79%	258.16%	225.89%	219.05%	192.25%		–	678.57%	254.46%	226.19%	242.35%
Panel C: Value-oriented funds						Panel D: Balance-oriented funds					
	$\pi_0$	$\pi_+$	$\pi_-$			$\pi_0$	$\pi_+$	$\pi_-$			
%	93.47%	2.69%	3.84%			85.58%	13.33%	1.09%			
$\gamma$	At different significant level					$\gamma$	At different significant level				
	0.05	0.1	0.15	0.2	0.25		0.05	0.1	0.15	0.2	0.25
$E(S_{\lambda}^+)$	27	37	45	55	62	$E(S_{\lambda}^+)$	181	273	345	417	469
	33.33%	45.68%	55.56%	67.90%	76.54%		41.71%	62.90%	79.49%	96.08%	108.06%
$E(S_{\lambda}^-)$	5	9	10	13	16	$E(S_{\lambda}^-)$	14	27	49	64	88
	6.17%	11.11%	12.35%	16.05%	19.75%		3.23%	6.22%	11.29%	14.75%	20.28%
$F(S_{\lambda}^+)$	2.34%	4.67%	7.01%	9.35%	11.68%	$F(S_{\lambda}^+)$	2.14%	4.28%	6.42%	8.56%	10.70%
	2.34%	4.67%	7.01%	9.35%	11.68%		2.14%	4.28%	6.42%	8.56%	10.70%
$F(S_{\lambda}^-)$	31.00%	41.01%	48.54%	58.55%	64.86%	$F(S_{\lambda}^-)$	39.57%	58.62%	73.07%	87.52%	97.37%
	31.00%	41.01%	48.54%	58.55%	64.86%		39.57%	58.62%	73.07%	87.52%	97.37%
$F(T_{\lambda}^+)$	3.84%	6.44%	5.34%	6.70%	8.07%	$F(T_{\lambda}^+)$	1.09%	1.94%	4.87%	6.19%	9.58%
	3.84%	6.44%	5.34%	6.70%	8.07%		1.09%	1.94%	4.87%	6.19%	9.58%
$\widehat{FDR}^+$	7.01%	10.23%	12.62%	13.77%	15.26%	$\widehat{FDR}^+$	5.13%	6.80%	8.07%	8.91%	9.90%
	7.01%	10.23%	12.62%	13.77%	15.26%		5.13%	6.80%	8.07%	8.91%	9.90%
$\widehat{FDR}^-$	37.86%	42.06%	56.79%	58.24%	59.15%	$\widehat{FDR}^-$	66.33%	68.78%	56.85%	58.04%	52.76%
	37.86%	42.06%	56.79%	58.24%	59.15%		66.33%	68.78%	56.85%	58.04%	52.76%

## Data availability

The authors do not have permission to share data.

## References

- Allen, F., Gu, X., Li, C.W., Qian, Y., et al., 2023. Implicit guarantees and the rise of shadow banking: The case of trust products. *J. Financial Econ.* 149 (2), 115–141.
- Allen, F., Qian, J., Shan, C., Zhu, J., et al., 2022. Dissecting the long-term performance of the Chinese stock market. Available At SSRN 2880021.
- Allen, F., Qian, Y., Tu, G., Yu, F., 2019. Entrusted loans: A close look at China's shadow banking system. *J. Financial Econ.* 133 (1), 18–41.
- Ammer, J., Rogers, J., Wang, G., Yu, Y., 2022. Visible Hands: Professional Asset Managers' Expectations and the Stock Market in China. *International Finance Discussion Paper*, pp. 1–52.
- Andrikiannopoulou, A., Papakonstantinou, F., 2019. Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power? *J. Finance* 74 (5), 2667–2688.
- Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. *J. Financial Econ.* 106 (3), 473–491.
- Baks, K.P., Metrick, A., Wachter, J., 2001. Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation. *J. Finance* 56 (1), 45–85.
- Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *J. Finance* 65 (1), 179–216.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57 (1), 289–300.
- Blake, D., Caulfield, T., Ioannidis, C., Tonks, I., 2017. New evidence on mutual fund performance: A comparison of alternative bootstrap methods. *J. Financial Quant. Anal.* 52 (3), 1279–1299.
- Cai, B., Cheng, T., Yan, C., 2018. Time-varying skills (versus luck) in US active mutual funds and hedge funds. *J. Empir. Finance* 49, 81–106.
- Cheng, T., Yan, C., 2017. Evaluating the size of the bootstrap method for fund performance evaluation. *Econom. Lett.* 156, 36–41.
- Cheng, T., Yan, C., Yan, Y., 2021. Improved inference for fund alphas using high-dimensional cross-sectional tests. *J. Empir. Finance* 61 (2).
- Chi, Y., 2013. Performance Evaluation of Chinese Actively Managed Stock Mutual Funds. Chicago Booth Research Paper No. 13-55, Fama-Miller Working Paper, Vol. 20, (13-55).

- Chi, Y., Liu, Y., Qiao, X., 2022. Performance evaluation, factor models, and portfolio strategies: Evidence from Chinese mutual funds. *J. Portf. Manag.* 48 (8), 159–176.
- Christiansen, C., Grnborg, N.S., Nielsen, O.L., 2018. Mutual Fund Selection for Realistically Short Samples. CREATES Research Papers.
- Cuthbertson, K., Nitzsche, D., 2013. Performance, stock selection and market timing of the German equity mutual fund industry. *J. Empir. Finance* 21, 86–101.
- Cuthbertson, K., Nitzsche, D., O'Sullivan, N., 2008. UK mutual fund performance: Skill or luck? *J. Empir. Finance* 15 (4), 613–634.
- Cuthbertson, K., Nitzsche, D., O'Sullivan, N., 2012. False discoveries in UK mutual fund performance. *Eur. Financial Manag.* 18 (3), 444–463.
- Dyck, A., Lins, K.V., Pomorski, L., 2013. Does active management pay? New international evidence. *Rev. Asset Pricing Stud.* 3, 200–228.
- Fama, E.F., French, K.R., 2010. Luck versus skill in the cross-section of mutual fund returns. *J. Finance* 65 (5), 1915–1947.
- Glode, V., Green, R.C., 2011. Information spillovers and performance persistence for hedge funds. *J. Financial Econ.* 101 (1), 1–17.
- Harvey, C.R., Liu, Y., 2022. Luck versus skill in the cross section of mutual fund returns: Reexamining the evidence. *J. Finance* 77 (3), 1921–1966.
- Harvey, C.R., Liu, Y., Saretto, A., 2020. An evaluation of alternative multiple testing methods for finance applications. *Rev. Asset Pricing Stud.* 10 (2), 199–248.
- Hsu, P.-H., Kyriakou, I., Ma, T., Sermpinis, G., 2025. Mutual funds' conditional performance free of data snooping bias. *J. Financial Quant. Anal.* 60 (3), 1373–1400.
- Huang, R., Asteriou, D., Pouliot, W., 2020. A reappraisal of luck versus skill in the cross-section of mutual fund returns. *J. Econ. Behav. Organ.* 176, 166–187.
- Huang, R., Pilbeam, K., Pouliot, W., 2021. Do actively managed US mutual funds produce positive alpha? *J. Econ. Behav. Organ.* 182, 472–492.
- Hunter, D., Kandel, E., Kandel, S., Wermers, R., 2014. Mutual fund performance evaluation with active peer benchmarks. *J. Financial Econ.* 112 (1), 1–29.
- Jacobs, H., 2016. Market maturity and mispricing. *J. Financial Econ.* 122 (2), 270–287.
- Kim, S., In, F., 2012. False discoveries in volatility timing of mutual funds. *J. Bank. Finance* 36 (7), 2083–2094.
- Kim, S., In, F., Ji, P.I., Park, R.J., 2014. False discoveries in the performance of Australian managed funds. *Pacific-Basin Finance J.* 26, 244–256.
- Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund “stars” really pick stocks? New evidence from a bootstrap analysis. *J. Finance* 61 (6), 2551–2595.
- Koutmos, D., Wu, B., Zhang, Q., 2020. In search of winning mutual funds in the Chinese stock market. 54, 589–616.
- Li, Z., Liu, L.X., Liu, X., Wei, K., 2024. Replicating and digesting anomalies in the Chinese A-share market. *Manage. Sci.* 70 (8), 5066–5090.
- Liu, J., Stambaugh, R.F., Yuan, Y., 2019. Size and value in China. *J. Financial Econ.* 134 (1), 48–69.
- Pástor, L., Stambaugh, R.F., 2002. Investing in equity mutual funds. *J. Finance Econ.* 63 (3), 351–380.
- Storey, J.D., 2002. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64 (3), 479–498.
- Tang, K., Wang, W., Xu, R., 2012. Size and performance of Chinese mutual funds: The role of economy of scale and liquidity. *Pacific-Basin Finance J.* 20 (2), 228–246.
- Wen, F., Zou, Q., Xiong, W., 2021. The contrarian strategy of institutional investors in Chinese stock market. *Finance Res. Lett.* 41, 101845.
- Wu, Y., 2011. Momentum trading, mean reversal and overreaction in Chinese stock market. *Rev. Quant. Finance Account.* 37 (3), 301–323.
- Yan, C., Cheng, T., 2019. In search of the optimal number of fund subgroups. *J. Empir. Finance* 50, 78–92.
- Yan, C., Zhang, H., 2017. Mean-variance versus Naïve diversification: The role of mispricing. *J. Int. Financial Mark. Institutions Money* 48, 61–81.
- Yang, B., Xue, F., Su, Y., Yan, C., 2019. Is informational inefficiency priced in stock markets? A comparison between the U.S. and Chinese cases. *Pacific-Basin Finance J.* 55, 222–238.
- Zhang, H., Yan, C., 2018. Modelling fundamental analysis in portfolio selection. *Quant. Finance* 18 (8), 1315–1326.