Accepted Manuscript

Time-varying skills (versus luck) in U.S. active mutual funds and hedge funds

Tingting Cheng, Biqing Cai, Cheng Yan

PII: S0927-5398(18)30068-9
DOI: https://doi.org/10.1016/j.jempfin.2018.09.001
Reference: EMPFIN 1072

To appear in: *Journal of Empirical Finance*

Received date : 31 January 2018
Revised date : 22 August 2018
Accepted date : 5 September 2018

Please cite this article as: Cheng T., et al., Time-varying skills (versus luck) in U.S. active mutual funds and hedge funds. *Journal of Empirical Finance* (2018), https://doi.org/10.1016/j.jempfin.2018.09.001

Time-varying skills (versus luck) in U.S. active mutual funds and hedge funds

Tingting Cheng[a], Biqing Cai[b], Cheng Yan[c]

[a] *School of Finance, Nankai University, Tianjin City, 300350, PRC.*

[b] *Huazhong University of Science and Technology, Wuhan City 430074, PRC.*

[c] *Essex Business School, Colchester, CO4 3SQ, UK.*

Abstract

In this paper, we develop a nonparametric methodology for estimating and testing time-varying fund alphas and betas as well as their long-run counterparts (i.e., their time-series averages). Traditional linear factor model arises as a special case without time variation in coefficients. Monte Carlo simulation evidence suggests that our methodology performs well in finite samples. Applying our methodology to U.S. mutual funds and hedge funds, we find most fund alphas decrease with time. Combining our methodology with the bootstrap method which controls for 'luck', positive long-run alphas of mutual funds but hedge funds disappear, while negative long-run alphas of both mutual and hedge funds remain. We further check the robustness of our results by altering benchmarks, fund skill indicators and samples.

Keywords: Fund performance evaluation; Mutual fund and hedge fund; Skill vs. luck;

☆Corresponding author: Cheng Yan, Tel: +44 (0) 191 3345 197, Email: Cheng.Yan@durham.ac.uk. Tingting Cheng, Tel: +86 156 0216 6061, Email: Tingting.Cheng@nankai.edu.cn. We are grateful to Jushan Bai, David Blake, Yong Chen, Ken French, Jiti Gao, Robert Kosowski, Yan Liu, Lubos Pastor and participants in the workshops in ShanghaiTech University, University of Durham, University of Glasgow, University of Leicester, University of Kent, and University of Sydney for helpful comments. We are grateful to Xiangdong Wang for his excellent research assistance, and grants from the National Natural Science Foundation of China (NNSFC Grant No.71703045, 71803091, 71801117, 71803196).

Time-varying coefficient model

JEL Classification: C1; G1; G2

HIGHLIGHTS

- Challenge the view a time-invariant scalar (e.g., alpha) can capture fund skill
- A nonparametric method to estimate and testing time-varying fund alphas and betas.
- Our methodology performs well in various cases of Monte Carlo Simulations
- Combine our method with bootstrap to control for `luck' in time dimension.

Time-varying skills (versus luck) in U.S. active mutual funds and hedge funds

Abstract

In this paper, we develop a nonparametric methodology for estimating and testing time-varying fund alphas and betas as well as their long-run counterparts (i.e., their time-series averages). Traditional linear factor model arises as a special case without time variation in coefficients. Monte Carlo simulation evidence suggests that our methodology performs well in finite samples. Applying our methodology to U.S. mutual funds and hedge funds, we find most fund alphas decrease with time. Combining our methodology with the bootstrap method which controls for 'luck', positive long-run alphas of mutual funds but hedge funds disappear, while negative long-run alphas of both mutual and hedge funds remain. We further check the robustness of our results by altering benchmarks, fund skill indicators and samples.

Keywords: Fund performance evaluation; Mutual fund and hedge fund; Skill vs. luck; Time-varying coefficient model

JEL Classification: C1; G1; G2

1. Introduction

The Efficient Market Hypothesis (EMH) implies that funds should not have skills to persistently beat the market, which raises two classical questions regarding fund performance evaluation: i) Do (a group of, or on average) funds have skills or not? ii) If they have skills, are these skills persistent? Taking fund alphas as the fund skill indicator, the traditional linear factor models have been widely employed in the literature[1]. Unfortunately, the inherent assumption of constant alphas and factor loadings (betas), does not empirically holds at either the asset or portfolio level (see Ang and Kristensen (2012) and the references therein), and hence may distort the validity of the standard factor models with misleading inferences.

We suggest these two questions can, however, be answered simultaneously if the alphas (and betas) are viewed as time-varying coefficients at every time point, instead of period-specific constants. To this end, we propose a nonparametric methodology to estimate and test time-varying fund alphas and betas, which imposes no parametric assumptions on the time variation of alphas and betas[2]. Specifically, we first present an estimator for the time-varying fund alphas and betas, and then construct a Generalized Likelihood Ratio (GLR) statistic to test whether the estimated alphas are indeed time-varying or not. To evaluate the overall performance of funds, we also construct their long-run counterparts: the time-series averages of time-varying fund alphas and betas. To illustrate the flexibility of our methodology, we combine

[1]Following the main literature, we use the net alphas (i.e., fund alpha net of all management expenses and 12b-fees) as fund skill indicator. Our main perspective is therefore in line with many other studies that are primarily concerned with the abnormal return that fund investors can earn by investing in mutual funds, see e.g., Fama and French (2010).

[2]The nonparametric method has been previously used in this area either via bootstrap methods (e.g., Kosowski et al. (2007); Blake et al. (2014)), or to help construct approaches to detect false discoveries (e.g., Barras et al. (2010); Bajgrowicz and Scaillet (2012); Bajgrowicz et al. (2015)).

our methodology with the bootstrap approach to design three new fund bootstrap schemes to control for 'luck'. Thus, we are able to distinguish whether the superior fund performance is due to luck or to skill.

Our approach has several advantages. First, we have proposed a formal statistical approach to identify time-varying alphas and betas for an individual fund from a general perspective, which constitutes a methodological novelty and a unique contribution to fund performance evaluation. The time-series averages of the obtained time-varying alphas are a more accurate indicator of fund skills than OLS alphas, which can be further cross-sectionally refined using any approach applying to the OLS alphas.

Second, due to the nonparametric nature of our estimates, we do not need to assume any *ex ante* linear or nonlinear relationship between the fund returns and the factors. Instead, our methodology adopts a 'let-data-speak' approach to reveal the relationship between the fund returns and the factors, which means that our methodology does not only gauge the magnitude of alpha for an individual fund at each time point, but also uncover the plausible source of the alpha: stock-picking or market-timing (see, e.g., Kacperczyk et al. (2014)). The former is often represented by the unexplained alpha in a linear model of market-related factors, while the latter is captured by including an additional squared market returns (Treynor and Mazuy (1966); Jiang et al. (2007); Chen et al. (2010); Blake et al. (2013)).

Third, our estimates only require a series of kernel-weighted least squares regressions. If the genuine alphas and betas are indeed time-invariant rather than time-varying, our methodology degenerates to its special case: the standard linear factor models.

Fourth, we extend the traditional bootstrap methods for fund performance evaluation (e.g., Kosowski et al. (2006); Fama and French (2010)) by capturing the time variations. Kosowski et al. (2006) and Fama and French (2010) distinguish alphas from luck and skills by comparing the estimates from bootstrap simulations of the cross-section of funds with zero alphas to the actual cross-section of fund alphas. The idea is that the returns of the funds in a simulation run have the properties of actual fund returns, except they set true alpha to zero in the return population from which simulation samples are drawn. The simulations thus describe the distribution of alpha estimates when there is no abnormal performance in fund returns. Comparing the distribution of alpha estimates from the simulations to the cross-section of alpha estimates for actual fund returns allows them to draw inferences about the existence of skilled managers. However, as noted in the literature (Fama and French, 2010, page 1925), one major caveat of this method is that *'Because we randomly sample months, we also lose any effects of variation through time in the regression slopes in (1). (The issues posed by time-varying slopes are discussed by Ferson and Schadt (1996).) Capturing time variation in the regression slopes poses thorny problems, and we leave this potentially important issue for future research'*. Adapted from Kosowski et al. (2006) and Fama and French (2010), we compare estimates of long-run alphas from bootstrap simulations of the cross-section of funds with zero alphas across time periods to the actual cross-section of fund long-run alphas. Our method thus provides a remedy for the drawback of the existing literature.

Even in simulated funds with only 200 time-series observations, our methodology performs well in different cases, including i) constant α and constant β; ii) constant α but time-varying β, iii) time-varying α but constant β, and iv) time-varying α and time-varying β.

Applying our methodology to the legendary Fidelity Magellan fund, we find a positive and significant long-run alpha using the time-varying Fama-French-Carhart 4-factor model. Furthermore, the alpha is time-varying and decreases from positive before the 1980s to insignificant and further to negative in the 2000s. We conclude that though it was once a star fund, it definitely is not any longer.

Applying our methodology to the whole mutual and hedge funds industry, we find that most net fund alphas are time-varying and in general, with a decreasing trend. Only 1% (19%)

of mutual (hedge) funds have positive and significant long-run net alphas, while 9% (9%) have negative and significant long-run net alphas. Combining our methodology with the bootstrap approach to control for 'luck', positive long-run alphas of mutual funds but hedge funds disappear, while negative long-run alphas of both funds remain[3].

Our results are robust in altering the number of factors in our benchmark fund performance evaluation model, adding back fees and expenses to fund returns, as well as sub-sample analysis. They are not robust when using the value-added measure from Berk and Van Binsbergen (2015) instead of fund alphas, or using the Vanguard index fund as the passive benchmark portfolio alternative to the traditional Fama-French factors[4]. This is not surprising as this is also the case in existing performance evaluation studies (e.g., Kosowski et al. (2006); Barras et al. (2010); Fama and French (2010); Berk and Van Binsbergen (2015)).

Our idea of time-varying alphas can be traced back to the earlier conditional beta model (e.g., Ferson and Schadt (1996)) and the conditional alpha and beta model (e.g., Christopherson et al. (1998)), which add a factor conditional on the state of the economy to the original unconditional model of Jensen (1968)[5]. Perhaps due to its simplicity, this setup has been used with few doubts (for recent examples, see, Fung et al. (2008); Ferson and Lin (2014); Kacperczyk et al. (2014)). Using kernel-based method, our methodology uses all the data in an efficient way to estimate time-varying alphas and betas, and hence nests almost all extant

---

[3]We use "*fund alpha*" and "*alpha*" interchangeably in this paper. We focus on the fund skills in this paper, but our results hold when we limit our sample to the periods after the current portfolio manager taking control. Put differently, we focus on fund skills in general instead of specific manager skills.

[4]Benchmark means the next best investment opportunity available to investors rather than the fund at the same time. For fund performance evaluation, we must compare the fund performance with the performance of its benchmark (Berk and Van Binsbergen (2015)).

[5]Typically, this strand of literature assumes a different but constant beta (and/or an alpha) according to whether the returns of the market factor is below or above the risk-free rate.

conditional beta models and conditional alpha and beta models as special cases.

Our work complements Mamaysky et al. (2008), which justifies the existence of time-varying alphas and betas for mutual funds. Unlike their Kalman filter approach, which is specifically developed for identifying the market-timing abilities, our methodology aims to identify the overall time-varying skills of funds. While their model hinges on the assumption that assets under management within a fund are reallocated on the basis of some unobserved factor (market-timing), we follow the mainstream literature and build up our model on the observable factors.

Compared with the model in Ang and Kristensen (2012), our approach is specifically designed for fund performance evaluation. We argue that this kind of nonparametric methodology is more suitable for measuring fund alphas than assets, given the wide variety of dynamic complex strategies used in funds (see, e.g., Brown and Goetzmann (1997) for discussion). Unlike their study, we have i) used the GLR statistic to test whether the fund alphas are indeed time-varying or not; ii) quantified the finite sample properties of our time-varying alpha and beta estimates relative to their OLS and/or rolling OLS counterparts via Monte Carlo simulation; iii) combined our time-varying model with the bootstrap approach to control for 'luck'.

The remainder of the paper proceeds as follows. In Section 2, we introduce our nonparametric time-varying methodology, the associated estimators and various model specifications. Section 3 investigates the finite sample properties of our methodology using Monte Carlo simulations. Section 4 describes our data and applies our methodology to a representative 'star' fund, and to the whole mutual fund and hedge fund industry, respectively. Section 5 combines our time-varying estimators with the extant refinements (using the popular bootstrap approach as an example) to further distinguish skill from luck for fund performance evaluation. Section 6 explores further robustness of the mutual fund results in Section 5 by using a variety of alternative fund skill indicators as well as benchmark portfolios. Section 7 concludes.

2. Methodology

As mentioned in the Introduction, the traditional fund performance evaluation methodology assumes constant alpha and betas which is rejected by real data. In this section, we propose a time-varying coefficient model which is able to catch the time variation of alpha and betas. In addition, we use the local linear method to estimate the model and construct specification tests for the model.

2.1. Nonparametric time-varying coefficient model

For a balanced or unbalanced panel of N funds, at any time point t we use r_{it} to denote the excess return for the ith fund, for $i = 1, \cdots, N$ and $t = 1, \cdots, T_i$, where T_i denotes the time length for the i-th fund. For the simplicity of notations, here we simply use T to denote the length of time. Let $x_t = (x_{1t}, \cdots, x_{Kt})'$ to denote the K observable and common tradable factors (e.g., Fama-French type of factors). To capture the potential time variations of the fund alphas and betas, we propose the following time-varying coefficient factor model to evaluate fund performance:

$$r_{it} = \alpha_i(t) + \sum_{j=1}^{K} \beta_{ij}(t) x_{jt} + e_{it}, \text{ for } i = 1, \cdots, N; t = 1, \cdots, T_i; j = 1, \cdots, K. \tag{1}$$

where $\alpha_i(t)$ is the vector of time-varying alphas across fund i and $\beta_{ij}(t)$, $j = 1, \cdots, K$ are the corresponding time-varying factor loadings. We do not need to impose any *ex ante* parametric constraints on the dynamics of alphas and betas, which allows for a richer set of time paths for alphas and betas. Importantly, although we maintain the canonical assumption that errors and factor returns are orthogonal for comparison reasons, we do not rule out any potential relationship between alphas/betas and the factor returns.

If for a specific fund i, its alpha and betas are indeed time-invariant rather than time-varying, model (1) degenerates to its special case which has been widely used in the existing literature since Jensen (1968): $r_{it} = \alpha_i + \sum_{j=1}^{K} \beta_{ij} x_{jt} + e_{it}$. Our method therefore nests the traditional one as a special case.

2.2. *Estimating the time-varying alphas and betas*

Following the econometrics literature on time-varying coefficient models (e.g., Cai (2007), Cheng et al. (2017)), we employ the local linear estimation method to estimate the unknown coefficient functions in model (1) .

For the sake of notational simplicity, we denote $\gamma_i(\tau_t) = (\alpha_i(t), \beta_{i1}(t), \cdots, \beta_{iK}(t))'$ where $\tau_t = t/T$. This scaled time is required for the justification of asymptotic properties of the local linear estimator. Please refer to Robinson (1989) for more details. Under the assumption that $\gamma_i(\cdot)$ has a continuous second order derivative, we approximate $\gamma_i(\tau_t)$ by a linear function of $\tau \in [0,1]$ given as

$$\gamma_i(\tau_t) \approx \gamma_i(\tau) + \gamma_i^{(1)}(\tau)(\tau_t - \tau),$$

where $\gamma_i^{(1)}(\tau)$ is the first order derivative of $\gamma_i(\tau)$.

Let z_t be a column vector whose elements are respectively x_t and $x_t(\tau_t - \tau)$. Let $\theta_i(\tau) = \left(\gamma_i(\tau)', \gamma_i^{(1)}(\tau)'\right)'$, model (1) can be written as

$$r_{it} \approx z_t' \theta_i(\tau) + e_{it}. \tag{2}$$

The parameter vector $\theta_i(\tau)$ can be estimated by minimizing the locally weighted sum of squares:

$$\sum_{t=1}^{T} (r_{it} - z_t' \theta_i(\tau))^2 K_h(\tau_t - \tau), \tag{3}$$

where $K_h(u) = K(u/h)/h$, $K(\cdot)$ is a kernel function, and $h > 0$ is the bandwidth satisfying that $h \to 0$ and $Th \to \infty$ as $T \to \infty$.

The local linear estimator of $\theta_i(\tau)$ is the minimizer of (3) and is given by

$$\widehat{\theta}_i(\tau;h) = \begin{pmatrix} S_{T,0}(\tau) & S'_{T,1}(\tau) \\ S_{T,1}(\tau) & S_{T,2}(\tau) \end{pmatrix}^{-1} \begin{pmatrix} R_{T,0}(\tau) \\ R_{T,1}(\tau) \end{pmatrix},$$

where

$$S_{T,j}(\tau) = \frac{1}{T}\sum_{t=1}^{T} x_t x_t'(\tau_t - \tau)^j K_h(\tau_t - \tau), \;\; R_{T,j}(\tau) = \frac{1}{T}\sum_{t=1}^{T} x_t(\tau_t - \tau)^j K_h(\tau_t - \tau) r_{it}.$$

The first $K+1$ components of $\widehat{\theta}_i(\tau;h)$ are the estimators of fund alpha and betas at τ, i.e., the estimators of $(\alpha_i(\tau), \beta_{i1}(\tau), \cdots, \beta_{iK}(\tau))'$. We present the detailed steps to construct confidence intervals for the time-varying alphas and betas in Appendix A.

For practical use of the proposed estimator, we need to choose a kernel function and a bandwidth. In the nonparametric literature, there is almost a consensus that the choice of kernel function is trivial but the bandwidth is critical. In this study, we use the Epanechnikov kernel function which is widely used in empirical applications. For the bandwidth selection, throughout the paper we use the normal reference rule that $h = 1.06\omega T^{-1/5}$, where ω denotes the standard deviation of the smoothing variable[6].

2.3. *Testing time-varying alphas*

After we obtain the estimates of time-varying fund alphas and betas, financial economists may further ask whether the estimated fund alphas are indeed time-varying or not. To answer this question, we construct a GLR test statistic following Fan et al. (2001). The GLR method has

[6]Our results are robust to some other kernel functions including the Gaussian function, and several mainstream bandwidth selection methods (for example, cross-validation or plug-in method). The results are available upon request.

been widely used in nonparametric hypothesis testing, see for example Fan and Jiang (2005), Cai (2007) and Cheng (2017).

Let us consider the hypothesis testing problem

$$H_0 : \alpha_i(t) = \alpha_i \quad \text{versus} \quad H_1 : \alpha_i(t) \neq \alpha_i, \quad t = 1, 2, \cdots, T,$$

where α_i is a constant (e.g., $\alpha_i = 0$ if we want to test whether $\alpha_i(t) = 0$).

Assuming the error distribution is normal $N(0, \sigma^2)$, the log-likelihood function for model (1) is $-\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(r_{it} - \alpha_i(t) - x_t'\beta_i(t)\right)^2$. Replace the unknown function $\alpha_i(t)$ and $\beta_i(t)$ by the local linear estimators $\widehat{\alpha}_i(t)$ and $\widehat{\beta}_i(t)$, respectively and then define $RSS_1 = \sum_{t=1}^{T}\left(r_{it} - \widehat{\alpha}_i(t) - x_t'\widehat{\beta}_i(t)\right)^2$. We can obtain the log-likelihood under H_1 by maximizing over the parameter σ^2,

$$\ell(H_1) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(RSS_1) - \frac{T}{2}.$$

Similarly, the log-likelihood under H_0 is given by

$$\ell(H_0) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(RSS_0) - \frac{T}{2},$$

where $RSS_0 = \sum_{t=1}^{T}\left(r_{it} - \widetilde{\alpha}_i - x_t'\widetilde{\beta}_i(t)\right)^2$ and $\widetilde{\alpha}_i$, $\widetilde{\beta}_i(t)$ are the estimators of α_i and $\beta_i(t)$ under H_0, respectively.

Following Fan et al. (2001), we define the following GLR statistic:

$$TS = \ell(H_1) - \ell(H_0) = \frac{T}{2}\log\frac{RSS_0}{RSS_1} \approx \frac{RSS_0 - RSS_1}{RSS_1}. \tag{4}$$

It can be shown that TS asymptotically follows standard normal distribution (e.g., Fan et al. (2001)) under H_0, and thus critical values can be obtained accordingly. The normal approximation usually does not perform well in finite samples. Thus we adopt a bootstrap refinement, which has been outlined in Appendix B.

2.4. Long-run alpha analysis

The question then arises of how to evaluate a fund's overall performance during a specific time period, especially when its performance spans across both positive and negative values, which is a typical case in our fund sample. As a logical choice, we develop a long-run alpha estimator which is the average of the time-varying alphas across time. This is more accurate than the traditional OLS alpha estimator because our long-run alpha estimator is directly derived from the time-varying alphas, and is more robust due to the nonparametric nature of the methodology.

To measure the long-run performance for fund managers, we first estimate model (1) to obtain the the estimate for unknown function $\alpha_i(t)$ for the i-th fund, $i = 1, 2, \cdots, N$, with the proposed local linear estimation method. Let $\widehat{\alpha}_i(t)$ denote the estimate of $\alpha_i(t)$ and then following Ang and Kristensen (2012), we obtain the long-run alpha for the i-th fund by computing the average $\frac{1}{T_i}\sum_{t=1}^{T_i}\widehat{\alpha}_i(t)$, in which T_i denotes the length of time series for the i-th fund. Note that in Section 2.1, we simply assume each fund has the same time length T, which is a special case here. We allow that each fund may have different durations. Following Theorem 2 in Ang and Kristensen (2012), we can compute the asymptotic variance for the estimate of long-run alpha, based on which we can construct the confidence interval of long-run alpha.

3. Simulation Study

In this section, we examine the finite sample performance of our time-varying coefficient time series model using fund returns simulated from the one-factor CAPM model below.

$$r_{it} = \alpha_i(t) + \beta_i(t)x_t + e_{it}, \tag{5}$$

where x_t is generated from a Gaussian distribution with mean 0.08/12 and standard deviation $\sqrt{0.15^2/12}$ and e_{it} is generated from a Gaussian distribution with mean 0 and standard deviation 0.02, $\alpha_i(t)$ and $\beta_i(t)$ are generated from the following four Data Generating Processes (DGPs).

DGP1: Constant α and constant β. α_i and β_i are generated from $U[-0.1/12, 0.1/12]$ and $U[0.5, 1.5]$ respectively, where $U[a, b]$ denotes uniform distribution over the support of $[a, b]$.

DGP2: Constant α but time-varying β. $\alpha_i(t) = \alpha_i$, $\beta_i(t) = 1 + 100\sin(2\pi t/T)/T$, where α_i is generated from $U[-0.1/12, 0.1/12]$.

DGP3: Time-varying α but constant β. $\alpha_i(t) = a_i + 0.1\sin(\pi t/T)$, where a_i is generated from $U[-0.1/12, 0.1/12]$. β_i is generated from $U[0.5, 1.5]$.

DGP4: Time-varying α and time-varying β. $\alpha_i(t) = a_i + 0.1\sin(\pi t/T)$, $\beta_i(t) = 1 + 100\sin(2\pi t/T)/T$, where a_i is generated from $U[-0.1/12, 0.1/12]$.

We can see that DGP1 and DGP2 focus on the cases where the fund alphas are constants and DGP3 and DGP4 consider the cases where the fund alphas are time-varying. In particular, since $t/T \in [0, 1]$, the functional form of fund alpha in DGP 3 and DGP4 will be increasing and then decreasing which is consistent with the prediction of Berk and Green (2004).

For each DGP, we generated $N = 200$ funds and we set the time length T to be 200, 400 and 800, respectively. We measure the accuracy of local linear estimation method by computing Mean Squared Error (MSE) for $\widehat{\alpha}_i(t)$ and $\widehat{\beta}_i(t)$ separately as follows:

$$\mathrm{MSE}(\widehat{\alpha}_i(t)) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{\alpha}_i(t) - \alpha_i(t))^2, \ \mathrm{MSE}(\widehat{\beta}_i(t)) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{\beta}_i(t) - \beta_i(t)\right)^2.$$

The results of mean squared errors for coefficient functions in DGPs 1-4 are presented in Table 1, from which we find that the values of MSE are quite small. For instance, the MSE of $\widehat{\alpha}_i(t)$ is less than one basis point in all cases. When the sample size increases, the MSE decreases, indicating that the local linear estimation method has very good finite sample performance. We plot the median of estimated coefficient functions $\widehat{\alpha}_i(t)$ ($\widehat{\beta}_i(t)$) under DGPs 1-4 with $T = 200, 400$ and 800 in Figure 1 (Figure 2). Under each DGP, the estimated curve comes closer to the true curve with the increase of sample size.

4. Time-varying Fund Performance Evaluation

We now apply the time-varying fund performance evaluation methodology to real data of fund returns net of all management expenses and 12b-fees in two scenarios: for a single representative mutual fund, all U.S. mutual funds and all U.S. hedge funds, respectively. We first introduce our fund data sets as follows.

4.1. Data and descriptive statistics

We first apply our methodology to a representative mutual fund: Fidelity Magellan fund (NASDAQ Ticker Symbol: FMAGX; CRSP Fund Identifier: 11943). Of course, there are many other candidates as the single representative mutual fund, but we choose Fidelity Magellan fund for several reasons. First of all, due to its prolonged superior performance, the Fidelity Magellan fund has been deemed as a skilled 'star' fund extensively in academic studies (e.g., Wermers (2000); Kosowski et al. (2006); Huang et al. (2007); Berk and Van Binsbergen (2015)). Moreover, the former manager of the Fidelity Magellan fund, Peter Lynch, has been considered a legendary fund manager for a long time by practitioners and the press (e.g., Marcus (1990)) and this fund is possibly the world's best known actively managed mutual funds. In addition, although its Asset Under Management (AUM) historically varies substantially, the

Fidelity Magellan fund had been the single largest mutual fund in the world for a long period with the AUM over $100 billion before it was overtaken by Vanguard's passive S&P 500 index fund in April 2000. Finally, the Fidelity fund family is arguably one of the longest-lived active mutual fund families, with an inception date as early as in 1930. Accordingly, the fund offers us a long sample from June 1963 to March 2017 in monthly frequency[7].

We then apply our methodology to all the active U.S. mutual funds similar to Harvey and Liu (2018); Ferson and Chen (2017). We obtain active U.S. equity mutual funds data from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free Mutual Fund database for the 1984-2011 period. The sample period is exactly the same as that of Harvey and Liu (2018) and Ferson and Chen (2017) for comparison reasons. We exclude the index funds. To mitigate omission bias (Elton et al. (2001)) and incubation and back-fill bias (Evans (2010)), we exclude observations prior to the reported year when the mutual funds were entered into the database, and the funds which do not report a year of organization. We only include the funds which have initial Total Net Assets (TNA) above $10 million and more than 80% of their holdings in equity markets. To avoid the look-ahead bias, we do not exclude funds whose TNA subsequently fall below $10 million. These screens leave us with a sample of 2557 mutual funds with at least 30 months of fund returns[8].

After that, we apply our methodology to U.S. equity-oriented hedge funds similar to Ferson and Chen (2017). To be specific, we obtain U.S. equity-oriented hedge funds data from Lipper TASS for the 1994-2011 period. The sample period is identical to that of Ferson and Chen

[7]We have used the Fidelity Magellan fund in daily frequency from September 1st, 1998 to December 30th, 2016, and obtained qualitatively similar results.

[8]Similarly, Harvey and Liu (2018) and Ferson and Chen (2017) have obtained a sample of 3619 and 3716 mutual funds with at least 8 months of returns over the same period, respectively. We follow Hunter et al. (2014) by using 30 months as our threshold as it adds robustness to our results.

(2017) for comparison reasons. To mitigate back-fill bias, we remove the first 24 months of returns and returns before the dates when funds first entered into the database, and funds with missing values in the field for the add date (Ferson and Chen (2017)). We only include those categorized for a given month as either dedicated short bias, event-driven, equity market neutral, fund-of-funds or long/short equity hedge. Similar to the mutual fund sample, we require that a fund to have an initial TNA above \$10 million as of the first date of entry. These screens leave us with a sample of 2072 mutual funds with at least 30 months of fund returns.

Table 2 presents summary statistics of the mutual fund and hedge fund data in our paper. We find that they share similar characteristics with the data sample used in Ferson and Chen (2017). The main characteristics are summarized as follows.

- The range of average returns across funds is much greater in the hedge fund sample ($-0.114 \sim 0.173$) than that in the mutual fund sample ($-0.09 \sim 0.06$).
- The median of estimated alpha from the Fung-Hsieh seven-factor (Fung and Hsieh (1997, 2001) model[9] for the hedge funds is positive, while the one from the Fama-French-Carhart four-factor (Carhart (1997)) for the mutual funds it is slightly negative. The tails of the cross-sectional alpha distributions extend to larger values for the hedge funds. For example, the upper 5% tail value for the alphas in the hedge fund sample is 1.2% per month, while for the mutual funds it is only 0.4%. In the left tails the two types of funds also present different alpha distributions, with a thicker lower tail for the alphas in the hedge fund sample.

[9]These seven factors (i.e., Bond Trend-Following Factor, Currency Trend-Following Factor, Commodity Trend-Following Factor, Equity Market Factor, Size Spread Factor constructed from Russell 2000 index and S&P500, Bond Market Factor and Credit Spread Factor) proposed by Fung and Hsieh (1997, 2001) are arguably more suitable for the hedge funds than the Fama-French-Carhart four factors.

- The sample volatility of the median hedge fund return (2.8% per month) is smaller than for the median mutual fund (5.3%). The range of volatilities across the hedge funds is greater, with more mass in the lower tail. Between the 10% and 90% quantiles, the volatility range is 1.2% - 7.5% (1.2% - 6.7% in Ferson and Chen (2017)) for hedge funds, and 3.6% - 7.8% (4.2% - 7.0% in Ferson and Chen (2017)) for mutual funds.

- The return autocorrelation is slightly higher for the hedge funds than mutual funds. The median autocorrelation for the hedge (mutual) funds is 0.127 (0.121), and some of the hedge funds have substantially higher autocorrelations.

4.2. Performance of individual funds

In this subsection, we examine the performance of the Fidelity Magellan fund based on the monthly data from June 1963 to March 2017. We allow the coefficients in the Fama-French-Carhart four-factor model to vary with time, that is, we consider the following time-varying coefficient four factor model as our benchmark:

$$r_{it} = \alpha_i(t) + \beta_{i1}(t)MKT_t + \beta_{i2}(t)SMB_t + \beta_{i3}(t)HML_t + \beta_{i4}(t)MOM_t + e_{it}, \ t = 1, \cdots, T \quad (6)$$

where r_{it} denotes the excess return of fund i at time t. MKT_t, SMB_t, HML_t and MOM_t denote the Fama-French-Carhart four factors, which are the Market excess return (MKT) factor, the Small-Minus-Big (SMB) size factor, the High-Minus-Low (HML) value factor and the Momentum (MOM) factor at time t, respectively. $\alpha_i(t)$ and $\beta_{i1}(t), \beta_{i2}(t), \beta_{i3}(t), \beta_{i4}(t)$ are unknown time-varying functions.

We first obtain $\widehat{\alpha}_i(t)$ and $\widehat{\beta}_{ij}(t)$ by local linear estimation, for $j = 1, 2, 3, 4$. We also compute the 95% confidence interval of estimated coefficients using the bootstrap procedure described in Section 2.2. We compare the nonparametric estimation results with the traditional OLS estimation and overlapping rolling estimations with a rolling window of 24 months (Figure 3), 36 months (Figure 4) and 60 months (Figure 5), respectively. Several interesting observations can be made.

First of all, we find a time-varying alpha which decreases from positive before the 1980s to insignificant and then to negative in the 2000s. The declining trend of alpha mirrors the fact that Peter Lynch generated a 2% monthly alpha on average AUM of roughly $40 million over his first five-year managing period from 1977 to 1981, but only 20 basis points(bp) when his AUM exceeded $40 billion over his last five-year managing period from 1986 to 1990. We conclude that it was a star fund then, but definitely is not one now. Instead of asking the question whether the fund has skills or not, we suggest that it might be a more proper question to ask when the fund did have skills. A typical fund marketing strategy for funds is to advertise the long history and positive historical OLS alpha. This may be misleading according to our results, as neither the long history nor the positive historical OLS alpha necessarily signals good contemporaneous fund performance. Our time-varying approach offers a remedy.

The magnitude of the four factor loadings make sense, as the beta for the MKT (HML) factor surrounds one (zero), while the betas for the SMB size factor and the Momentum factor decrease from one to zero, which indicates a decreasing explanatory power of these Fama-French-Carhart factors on the alpha for the Fidelity Magellan fund. Overall, we conclude that the Fidelity Magellan fund indeed had superior skills during its early period, but these skills gradually vanished over time and have become inferior in recent years. This is consistent with the stylized fact that the skilled funds have been decreasing since the 20th century (Kosowski et al. (2006); Barras et al. (2010); Fama and French (2010); Pástor and Stambaugh (2012); Pástor et al. (2015); Jones and Mo (2016); Ferson and Chen (2017)).

Compared with the rolling window estimation which is widely used in the finance literature, the nonparametric method has several advantages. First of all, our nonparametric estimation method uses the observations in an efficient way as the kernel function automatically assigns larger weights to the observations closer to the estimated time point, and smaller weights to the observations further away. Moreover, the choice of window length in traditional rolling window estimation is mostly arbitrary, while the bandwidth in our method hinges on the data. Finally, due to the flexible nature of nonparametric methodology, our model is extremely useful when the true relationship between fund returns and factor returns is unknown.

In order to answer the two classical questions in the literature, namely - do funds have skills? and if so, are these skills persistent?- we further conduct two formal statistical tests for the potential constancy of the time-varying fund alphas. For the first question, we employ the GLR test statistic in Section 2.3 to test the null hypothesis that the fund alphas are (constantly) zero. We obtain a p value of 0, which strongly rejects the null hypothesis that the fund alpha for the Fidelity Magellan fund is zero over our sample period.

To investigate the persistence of fund alphas and to test the null hypothesis that the fund alpha is a constant, we use the GLR test by comparing the nonparametric model with a semi-parametric model where the alpha is a constant while the betas are time-varying. We obtain a p value of 0, which strongly rejects the null hypothesis that the fund alpha for the Fidelity Magellan fund is a constant over the sample and justifies the use of our time-varying fund performance evaluation methodology.

4.3. Long-run performance of mutual funds

In Section 4.2, we have justified the use of our methodology by investigating whether the Fidelity Magellan fund outperforms the market and if so, whether the performance is persistent or time-varying. Another question arising now concerns how many mutual funds with significant alphas exist in the mutual fund industry? To answer this question, in this subsection we examine the long-run performance of the cross-section of 2557 U.S. domestic equity mutual funds, instead of considering only the time-varying performance of one particular fund, to answer this question.

We measure the long-run performance for all the 2557 U.S. mutual funds following the procedure outlined in Section 2.4 and plot the long-run alpha for all the 2557 mutual funds in Figure 6. The range of long-run alpha is approximately between -0.047 and 0.064; and as expected, most mutual funds have long-run alpha around zero. In contrast, the range of OLS alpha is much wider, which is between -0.14 and 0.03. There is a much smaller amount of mutual funds with significant alphas after accounting for time-variations, which means that our nonparametric estimator has reduced the range of mutual fund alphas towards zero.

We plot the confidence intervals of the long-run alpha estimates for all the 2557 mutual funds in the left top panel of Figure 6 based on the theoretical results (Theorem 2) in Ang and Kristensen (2012). To see the long-run alpha estimates and the corresponding confidence intervals more clearly, we present the results of the bottom 50 mutual funds with the smallest long-run alphas and the top mutual 50 funds in the top panel of Figure 6. By checking whether the confidence interval contains zero or not, we find that only 32 (about 1% of 2557) mutual funds significantly outperform the market in the long-run, and 229 (about 9% of 2557) mutual funds significantly underperform the market. For most mutual funds in our sample, the long-run alphas are not significantly different from zero.

By checking whether the zero line is constantly inside the estimated confidence interval or not, we find that among the top 5 and bottom 5 mutual funds, we reject the null hypothesis $H_0: \ \alpha_i(t) = 0$ for most of the mutual funds at the conventional 5% level. In addition, similar to the Fidelity Magellan fund, the alphas of most mutual funds have a decreasing trend.

4.4. Long-run performance of hedge funds

Similar to Section 4.3, we now examine the long-run performance of a cross-section of 2072 U.S. equity-oriented hedge funds in this subsection. Our question is analogical to one of the research questions in Ferson and Chen (2017), i.e.: how many hedge funds are there in the hedge fund industry with significant alphas? Instead of the Fama-French-Carhart four factors, we present the results from the time-varying Fung-Hsieh seven-factor model. Our results remain qualitatively the same when we replace the Fung-Hsieh seven-factor model with the Fama-French-Carhart four-factor model.

Compared with the results for the mutual fund industry in Section 4.3, we find a much smaller portion of hedge funds have long-run alpha around zero, which is also as expected (see, e.g., Ferson and Chen (2017)). By checking whether the confidence interval contains zero or not, we find that only 389 (about 19% of 2072) hedge funds significantly outperform the market in the long-run and 193 (about 9% of 2072) mutual funds significantly underperform the market. For the rest of hedge funds in our sample, the long-run alphas are not significantly different from zero.

For the hedge funds, the range of long-run alphas is approximately between -0.051 and 0.047, wider than the one of OLS alphas, which is between -0.031 and 0.045. There is a larger amount of hedge funds with significant alphas after accounting for time-variations, which means that our nonparametric estimator has expanded the range of mutual fund alphas away from zero.

Similar to our mutual fund results, we reject the null hypothesis of a constant alpha for most of the hedge funds at the 5% level of significance using the GLR test and find most hedge fund alphas are decreasing with time as well. For brevity, we do not plot the long-run alphas for the hedge fund industry and the time-varying alphas for the top 5 and the bottom 5 hedge funds in the market, as they are qualitatively similar to Figure 7 and Figure 8, respectively.

It is beyond the aim and scope of this paper to investigate the reasons behind the decreasing alphas. Based on the empirical evidence we have obtained so far, we speculate that fund age is likely to be one reason (Pástor et al. (2015)), although more detailed examination with all other possible related variables is required. Fund scale/size (e.g., Berk and Green (2004), Chen et al. (2004)) and flows (Lou (2012)) may help explain the decreasing alpha for the Fidelity Magellan fund during its early period when its AUM dramatically increases from about \$20 million in 1963 to \$100 billion in 2000, but perhaps not the continued deterioration of alpha in the 21st century when its AUM drops back to about \$15 billion until 2011. Other potential explanations include manager changes (Khorana (1996), Dangl et al. (2008)), industry size and competition (Pástor and Stambaugh (2012))), macroeconomic factors (Ferson and Schadt (1996); Avramov et al. (2011); Glode (2011)), etc.

5. Skills versus 'Luck'

As we have shown in Section 4, the fund alphas are time-varying and the long-run alphas are significantly different from zero. The next question is: are these significant (and positive) alphas due to genuine managerial skills or pure sampling variability, that is, to 'luck'? To answer this question, we combine our nonparametric model with the cross-sectional bootstrap approach in earlier studies (e.g., Kosowski et al. (2006); Fama and French (2010)), and develop three new bootstrap schemes. After that, we apply the proposed bootstrap schemes to the cross-section of U.S. mutual funds and hedge funds, respectively. Put differently, in this section we tackle the challenge of distinguishing skill from 'luck' in the framework of time-varying coefficient models. We focus on the long-run performance of these funds, with special attention paid to the funds with superior or inferior skills.

5.1. Three new bootstrap schemes with time-varying alphas

To distinguish the true skills from 'luck', we first estimate model (6) by local linear estimation method and obtain $\widehat{\alpha}_i(t)$ and $\widehat{\beta}_{ji}(t)$, for $j = 1, 2, 3, 4$. We also compute the bootstrapped alphas and betas using the following bootstrap scheme. As we only use the intra-fund information when bootstrapping, we call it the intra-fund bootstrap. Different from the aforementioned literature, we allow both alphas and betas to be time-varying. Our intra-fund bootstrap scheme is as follows:

(1) Estimate the time-varying coefficient models for the N funds with estimators $\{\widehat{\alpha}_i(t), \widehat{\beta}_{ij}(t)\}$ and residuals $\{\widehat{e}_{it}\}$. For each fund i, we obtain the long-run alpha as $\frac{1}{T_i}\sum_{t=1}^{T_i}\widehat{\alpha}_i(t)$. Sorting the funds based on their long-run alphas, we can obtain the 1% to 99% quantiles accordingly.

(2) For each fund i, generate the bootstrap residuals $\{e_{it}^*\}_{t=1}^{T}$ from the empirical distribution of the residuals $\{\widehat{e}_{it}\}_{t=1}^{T}$, and then generate $r_{it}^* = \sum_{j=1}^{K} x_{jt}\widehat{\beta}_{ij}(t) + e_{it}^*$. In other words, the alphas are imposed to be 0. We then re-estimate the model based on r_{it}^* and obtain the N long-run simulated alphas.

(3) Replicate step 2 for $B(=200)$ times.

(4) Obtain the quantiles of the long-run alphas based on the simulated samples.

Thus steps (2) and (3) generate an artificial world where the alphas are zero across funds and time periods. Funds with positive and negative long-run alphas exist in the bootstrap samples but are due to sample variability (pure luck). Step (4) compares the distribution of true long-run alphas with their bootstrap counterparts which allows us to make inference of true skills.

We further propose an inter-fund bootstrap, which only differs from the intra-fund bootstrap in the residual generation step. In the second step of the inter-fund bootstrap, we generate the bootstrap residuals $\{e_{it}^*\}$ from the empirical distribution of the residuals $\{\widehat{e}_{it}\}_{i=1}^{N}$, that is, when we generate the residual at time t for the i-th fund, we actually use the cross-sectional information from the residual series of all the N fund at time t. Then we generate bootstrap sample by $r_{it}^* = x_t'\widehat{\beta}_i(t) + e_{it}^*$. The other steps are the same as that in the intra-fund bootstrap.

Moreover, we consider a pooled bootstrap scheme, which means that we re-sample each residual from the pool of all the residuals $\{\widehat{e}_{it}\}_{i=1,t=1}^{N,T}$. The rest is as the same as the intra-fund bootstrap and the inter-fund bootstrap approach.

5.2. Results for mutual funds

In this subsection, we apply the proposed intra-fund bootstrap, pooled bootstrap and inter-fund bootstrap schemes to the cross-section of 2557 U.S. mutual funds to distinguish skill from 'luck'. The results are presented in Table 3. The first two columns report the selected quantiles and the Cumulative Distribution Function of the actual (Act) long-run alphas at selected quantiles when they are ranked from the highest to the lowest. The next two columns report the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution as well as the p values that correspond to the selected quantiles of the distribution of the simulated long-run alphas generated by the intra-fund bootstrap scheme. Analogically, the remaining four columns report the results generated by the pooled bootstrap and inter-fund bootstrap scheme, respectively.

Our results are both similar to, and different from, the existing literature in a number of ways. For instance, consistent with Carhart (1997) and Kosowski et al. (2006), we find that the median mutual fund in our sample generates a Fama-French-Carhart risk-adjusted monthly long-run alpha of -0.1% (annualized alpha of -1.2%), while the top and bottom 1% mutual funds generate a Fama-French-Carhart risk-adjusted monthly long-run alpha of 0.7% and -1.2%, respectively. Also, as further examples, we find the net-of-costs negative alphas of all mutual funds below the median cannot be simply attributed to sampling variability (i.e., 'luck'), as all bootstrapped p values strongly reject this null hypothesis across the intra-fund bootstrap, pooled bootstrap, and inter-fund bootstrap schemes. This finding indicates that the investors, who have invested in the below-median mutual funds, would be much better off if they put their investments in the low-cost index funds.

On the other hand, except for the top two mutual funds, we cannot reject the null hypothesis that the performance of the above-median mutual funds is an artifact of sampling variability (i.e., 'luck'), either at 5% or 10% significance level. However, in our sample the top two mutual funds only have a short time-series span of 39 and 41 months, respectively. In untabulated results which are available from the authors upon request, we arrived at the same conclusion for the top two mutual funds as other above-median funds when we extend their sample period from December 2011 to March 2017. That's to say, even the performance of the top two mutual funds are subject to the critique of sampling variability (i.e., 'luck') and our results are disheartening for mutual fund investors.

Overall, we find little evidence that the significant and positive alphas are due to genuine superior managerial skills, but strong evidence suggesting that the significant and negative alphas are due to inferior managerial skills, at the conventional level of significance for most of the mutual funds by our bootstrap schemes.

5.3. Results for hedge funds

This subsection applies the proposed intra-fund bootstrap, pooled bootstrap and inter-fund bootstrap procedures to 2072 U.S. hedge funds to distinguish skill from 'luck'. The results are presented in Table 4, which is analogical to Table 3.

Consistent with page 12 in Ferson and Chen (2017), we find that the median hedge fund in our sample generates a slightly positive Fama-French-Carhart risk-adjusted long-run alpha which is negative for the median mutual fund as we have already shown in Section 5.3. Relative to our mutual fund sample, the tails of the hedge fund cross-sectional alpha distribution include much larger values, as the top and bottom 1% hedge funds generate a Fung-Hsieh seven-factor risk-adjusted monthly long-run alpha of 2.1% and -1.8%, respectively.

Like mutual funds, the net-of-costs negative alphas of all hedge funds below the median cannot be simply attributed to sampling variability (i.e., 'luck'), as all bootstrapped p-values strongly reject this null across the three bootstrap schemes. This finding indicates that the investors, who have invested in the below-median hedge funds, would be much better off if they put their investments in the low-cost index funds instead. Unlike mutual funds, we find that the performance of the median hedge funds is not subject to the critique of sampling variability (i.e., 'luck'), as we can reject the null hypothesis that the performance of the majority of the top 20% hedge funds is an artifact of sampling variability (i.e., 'luck'), albeit with a few exceptions.

Overall, at the conventional level of significance, we find strong evidence that the significant and positive long-run alphas are due to genuine superior managerial skills, and the significant and negative long-run alphas are due to inferior managerial skills, especially using our intra-fund bootstrap and inter-fund bootstrap procedures. Our results are roughly consistent with those of Ferson and Chen (2017) but from a distinct perspective.

6. Robustness Checks

In Section 5, we use the net fund alpha and the Fama-French-Carhart 4-factor model to conduct a large set of bootstrap tests to distinguish genuine managerial skills from luck and find disheartening results for mutual funds investors. While net fund alpha measures the abnormal return earned by fund investors, gross fund alphas measures the return the fund earns, and the value-added measure from Berk and Van Binsbergen (2015) evaluates the money/value that the fund extracts from capital markets[10]. In this section, we perform robustness checks to test whether our results are sensitive to altering the number of factors in our benchmark model,

[10]By no means we have the intention to be involved in the re-heated debate on which measure is the right/better measure of fund skill. They have all been used in the literature as indicators of fund skill, depending on whether the researchers take the perspective of the fund investors, fund managers, etc. Nevertheless, almost all existing fund indicators (including the ones we list here) are time-invariant constants, which are subject to our key critique in this paper. Regarding this, our nonparametric technique offers more flexibility and has better potential to capture the time variation in risk-taking of the funds.

adding back fees and expenses to fund returns, using the value-added measure from Berk and Van Binsbergen (2015) instead of fund alphas, using the Vanguard index fund as the passive benchmark portfolio alternative to the traditional Fama-French factors, as well as sub-sample analysis. For brevity, we report the results selectively to be reader-friendly. In general, we demonstrate that our main findings for mutual funds are robust to these changes except for using the value-added measure from Berk and Van Binsbergen (2015) instead of fund alphas, and using the Vanguard index fund from Berk and Van Binsbergen (2015) as the benchmark portfolio alternative to the traditional Fama-French-Carhart factors. This is unsurprising as it is also the case in the literature relating to fund performance evaluation (e.g., Kosowski et al. (2006); Barras et al. (2010); Fama and French (2010); Berk and Van Binsbergen (2015)). We are aware of other benchmarks, fund skill indicators, and data sources, but since the main contribution of this paper is methodological, we leave them for future research.

6.1. Altering the number of factors in the benchmark model

We use the Fama-French-Carhart 4-factor model as our benchmark simply due to the fact that it is the most popular benchmark in the literature on mutual fund performance evaluation. We do, however, fully acknowledge that there may be a certain degree of arbitrariness in choosing such a benchmark. Other benchmark models such as CAPM and Fama-French 3-factor model could also have been selected. To allay the concerns that our conclusion may hinge on the choice of Fama-French-Carhart 4-factor model as our benchmark, we also substitute our benchmark with CAPM and Fama-French 3-factor model, and selectively report the results from the pooled bootstrap scheme in Panel A and B of Table 5, respectively. The results are almost identical when we replace our benchmark of Fama-French-Carhart 4-factor model with Fama-French 3-factor model. However, there are a greater number of under-performing mutual funds surviving in our luck tests when we replace our benchmark of Fama-French-Carhart 4-factor model with CAPM. In general, our main results stay qualitatively unchanged.

6.2. Adding back fees and expenses to fund returns

Another concern arises from a minor strand of literature (e.g., Fama and French (2010)), which have proposed the gross alphas in place of net alphas as the indicator of fund skills. To alleviate this concern, we follow the procedures of Fama and French (2010) and add back fees and expenses to net fund returns and re-run our bootstrapping tests. Specifically, we calculate gross returns as the net returns plus 1/12th of the fund's annual expense ratio. When a fund's annual expense ratio is missing, we follow Fama and French (2010) and assume it is the same as the average of other active mutual funds with similar Assets Under Management (AUM)[11]. That's to say, our gross fund returns include the costs in expense ratios but exclude other costs such as trading costs, in light of the highlighted measurement issues in trading costs in the appendix A of Fama and French (2010). Panel A, B, and C of Table 6 report the pooled bootstrap results generated by using the long-run alphas estimated from the Fama-French-Carhart 4-factor model, Fama and French 3-factor model and CAPM, respectively. Interestingly, we find almost identical results from our bootstrap tests (for both top quantiles and bottom quantiles) to those of net fund returns. If anything, the median actual long-run alphas (i.e., -0.000, -0.000, 0.000) estimated from the gross fund returns via the Fama-French-Carhart 4-factor model, Fama and French 3-factor model and CAPM are generally higher than the ones (-0.001, -0.001, -0.000) from net fund returns, which is not surprising due to the fact that we have added back fees and expenses.

[11]Taking a tack different from Fama and French (2010), Berk and Van Binsbergen (2015) deal with the missing expense ratios in a much more complex way using additional information from Morningstar, the Securities and Exchange Commission (SEC) website, and the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system.

6.3. Using the value-added measure instead of fund alphas

The indicators of fund skills also include non-alpha indicators that we cannot entirely ignore for the sake of robustness reasons. Perhaps the strongest competitor so far to net alphas and gross alphas is the value-added measure proposed by Berk and Van Binsbergen (2015), although on page 4 immediately after their equation 1, they admit "*The most commonly used measure of skill in the literature is the unconditional mean of ε_{it}, or the net alpha, ...*". Specifically, Berk and Van Binsbergen (2015) have proposed using the value-added measure, which is the product of AUM and gross alphas, to measure fund performance, and argued that it is a better measure than both net alphas and gross alphas[12]. Based on their suggestion, we construct our value-added measure as the product of the gross alphas obtained in Section 6.2 (estimated from the Fama-French-Carhart 4-factor model, Fama and French 3-factor model and CAPM, respectively) and the natural logarithm of Total Net Asset Value (i.e., the variable "*Mtna*" in CRSP database) for each mutual fund in our sample[13]. After that, we re-run our bootstrap tests with our new value-added measure and report our results in Panel A, B, and C of Table 7. The results for mutual funds look better, as our value-added measure suggests that both the outperforming and underperforming mutual funds are due to luck and all the median values of the actual value-added measure turn positive (0.001, 0.001, 0.003) no matter we use the Fama-French-Carhart 4-factor model, Fama-French 3-factor model or the CAPM to estimate

[12]A possible extension of the value-added measure of Berk and Van Binsbergen (2015) is to use time-varying gross alphas instead of their time-invariant gross alphas, which theoretically has better potential to capture the time variation of the value-added.

[13]Strictly speaking, we should have used lagged AUM to make our value-added measure identical to Berk and Van Binsbergen (2015). However, due to the persistence in AUM, the sample correlation coefficient between our value-added measure and the measure constructed from lagged AUM is above 0.98 (no matter whether we lag one month or one year) and hence we ignore this very subtle difference in this subsection as well as the next subsection. Alternative results are available upon request.

the gross alphas before constructing the value-added measure. This is expected, as Berk and Van Binsbergen (2015) have demonstrated that their value-added measure is fundamentally different from the fund alpha indicators (including both net fund alphas and gross fund alphas).

6.4. Using Vanguard index fund as the benchmark

Although the standard practice in fund performance evaluation is to simply adjust for risk using the traditional factor models (e.g., the Fama-French-Carhart 4-factor model, Fama and French 3-factor model or the CAPM), Berk and Van Binsbergen (2015) argue that it is worth constructing an alternative passive investment opportunity itself. One reason for this is that there has recently been an extensive debate on the extent to which the traditional factor models accurately adjust for risk and which factor model is the best to achieve this goal (e.g., Hou et al. (2015) and others). Another reason is that in some cases these traditional (e.g., Fama-French) factors may either be unknown/unavailable to fund investors (in the early periods), or involves intensive transaction costs (e.g., the momentum factor).

To explore this concern, we use the net return of the Vanguard S&P 500 index fund as the alternative benchmark to the traditional factor models. We only use the Vanguard S&P 500 index fund instead of all 11 Vanguard index funds in Berk and Van Binsbergen (2015) for two reasons: i) The number of mutual funds (i.e., 2557) in our sample is much smaller than the one (i.e., 6054) in Berk and Van Binsbergen (2015), as we only take data from CRSP and focus on funds that hold mainly U.S. equity (like Harvey and Liu (2018); Ferson and Chen (2017) as well as many other studies). ii) Our sample starts from 1984 and the Vanguard S&P 500 index fund is the only one available at that time. We re-construct the net alpha (alternative net alpha), gross alpha (alternative gross alpha) and value-added measure (alternative value-added measure) via CAPM (using the net return of Vanguard S&P 500 index fund as the new MKT factor) and report our pooled bootstrap test results in Panel A, B and C of Table 8, respectively. Consistent with Section 9 of Berk and Van Binsbergen (2015), a strikingly different picture emerges[14]. The median value for net alphas, gross alphas, value-added measure are 0.003, 0.004 and 0.013, respectively, which are much larger than all the previous ones using the traditional factor models as the benchmark. Moreover, the under-performing mutual funds are no longer due to bad skills, and the outperformed mutual funds are no longer due to luck.

6.5. Sub-sample analysis

To further examine whether the cross-sectional distribution of the mutual fund industry varies over our sample period, we implement a subsample analysis for the 1984–2001 period since Harvey and Liu (2018) find "*noticeable differences between the parameter estimates for the 1984–2001 period and for full sample period*". Although we share the sample periods with Harvey and Liu (2018), we find few differences in our results from the pooled bootstrap tests for the 1984–2001 period (see Tables 9 to 12) and for full sample period (1984–2011), thereby justifying the robustness of our methodology. Arguably, the mutual funds look slightly better for the 1984–2001 period than for our full sample period, as our pooled bootstrap test suggests that in many cases the top 1% of them are due to genuine managerial skills rather than luck which does not show up for the full sample period[15]. To a lesser extent, this tendency is consistent with the existing literature (e.g., Kosowski et al. (2006); Barras et al. (2010); Fama and French (2010); Ferson and Chen (2017)), and may be due to the fact that in the new century the capital markets have become more efficient (cf., Bai et al. (2016)), or the increasing size of the fund industry has made competition more intense and hence trading profits scarcer (Pástor and Stambaugh (2012); Pástor et al. (2015)), or both.

[14]Interestingly, Fama and French (2010) have also considered this alternative but conclude differently on page 1922 that "*The bottom line is that for efficiently managed passive funds, the costs missed in expense ratios are close to zero. Thus, adjusting the benchmarks produced by (1) for estimates of these costs is unnecessary*".

[15]Bear in mind that in the subsample period the number of mutual funds is smaller than 2557, so here the top 1% includes a smaller number of funds compared to our previous full sample analysis.

7. Conclusion

In this paper, we introduce a time-varying coefficient model to evaluate fund performance. The model allows us to capture the time variation in alphas and betas and thus allows us to answer the two important questions: whether the funds have skills and if so, whether these skills are persistent.

Applying our methodology to the legendary Fidelity Magellan fund, we find a time-varying alpha which goes from being positive before the 1980s to being insignificant and then to being negative in the 2000s. We conclude that although it was once a star fund, it no longer is.

Applying our methodology to the mutual and hedge funds industry, we find that most net fund alphas are time-varying, with a general decreasing trend. We find that only 1% (19%) of mutual (hedge) funds have positive and significant long-run net alphas, while 9% (9%) have negative and significant long-run net alphas. Combining our methodology with the bootstrap method which controls for 'luck', positive long-run alphas of mutual funds but hedge funds disappear, while negative long-run alphas of both mutual and hedge funds remain.

We find benchmarks, fund skill indicators, and even data matter, which is not surprising as this is also the case in the literature evaluating fund performance (e.g., Kosowski et al. (2006); Barras et al. (2010); Fama and French (2010); Berk and Van Binsbergen (2015)).

Our work can be further extended in the following dimensions. First of all, although we have shown that the fund performance generally decreases with time, we are yet to figure out the mechanism for this phenomenon. Secondly, since benchmarks and fund skill indicators are important to our results, it would be interesting to provide a more detailed comparison and discussion. Last but not the least, we can further extend the time-varying fund evaluation method to panel data setup which may help to catch cross-sectional dependence. We leave this work for future research.

Reference

Ang, A. and Kristensen, D. (2012), 'Testing conditional factor models', *Journal of Financial Economics* **106**(1), 132–156.

Avramov, D., Kosowski, R., Naik, N. Y. and Teo, M. (2011), 'Hedge funds, managerial skill, and macroeconomic variables', *Journal of Financial Economics* **99**(3), 672–692.

Bai, J., Philippon, T. and Savov, A. (2016), 'Have financial markets become more informative?', *Journal of Financial Economics* **122**(3), 625–654.

Bajgrowicz, P. and Scaillet, O. (2012), 'Technical trading revisited: False discoveries, persistence tests, and transaction costs', *Journal of Financial Economics* **106**(3), 473–491.

Bajgrowicz, P., Scaillet, O. and Treccani, A. (2015), 'Jumps in high-frequency data: Spurious detections, dynamics, and news', *Management Science* **62**(8), 2198–2217.

Barras, L., Scaillet, O. and Wermers, R. (2010), 'False discoveries in mutual fund performance: Measuring luck in estimated alphas', *Journal of Finance* **65**(1), 179–216.

Berk, J. B. and Green, R. C. (2004), 'Mutual fund flows and performance in rational markets', *Journal of Political Economy* **112**(6), 1269–1295.

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

Blake, D., Caulfield, T., Ioannidis, C. and Tonks, I. (2014), 'Improved inference in the evaluation of mutual fund performance using panel bootstrap methods', *Journal of Econometrics* **183**(2), 202–210.

Blake, D., Rossi, A. G., Timmermann, A., Tonks, I. and Wermers, R. (2013), 'Decentralized investment management: Evidence from the pension fund industry', *Journal of Finance* **68**(3), 1133–1178.

Brown, S. J. and Goetzmann, W. N. (1997), 'Mutual fund styles', *Journal of Financial Economics* **43**(3), 373–399.

Cai, Z. (2007), 'Trending time–varying coefficient time series models with serially correlated errors', *Journal of Econometrics* **136**(1), 163–188.

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *Journal of Finance* **52**(1), 57–82.

Chen, J., Hong, H., Huang, M. and Kubik, J. D. (2004), 'Does fund size erode mutual fund performance? the role of liquidity and organization', *American Economic Review* **94**(5), 1276–1302.

Chen, Y., Ferson, W. and Peters, H. (2010), 'Measuring the timing ability and performance of bond mutual funds', *Journal of Financial Economics* **98**(1), 72–89.

Cheng, T. (2017), 'Functional coefficient time series models with trending regressors', *Econometric Reviews* **forthcoming**.

Cheng, T., Gao, J. and Zhang, X. (2017), 'Bayesian bandwidth estimation in nonparametric time–varying coefficient models', *Journal of Business and Economic Statistics* **forthcoming**.

Christopherson, J. A., Ferson, W. E. and Glassman, D. A. (1998), 'Conditioning manager alphas on economic information: Another look at the persistence of performance', *Review of Financial Studies* **11**(1), 111–142.

Dangl, T., Wu, Y. and Zechner, J. (2008), 'Market discipline and internal governance in the mutual fund industry', *Review of Financial Studies* **21**(5), 2307–2343.

Elton, E. J., Gruber, M. J. and Blake, C. R. (2001), 'A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases', *Journal of Finance* **56**(6), 2415–2430.

Evans, R. B. (2010), 'Mutual fund incubation', *Journal of Finance* **65**(4), 1581–1611.

Fama, E. F. and French, K. R. (2010), 'Luck versus skill in the cross-section of mutual fund returns', *Journal of Finance* **65**(5), 1915–1947.

Fan, J. and Jiang, J. (2005), 'Nonparametric inferences for additive models', *Journal of the American Statistical Association* **100**(471), 890–907.

Fan, J., Zhang, C. and Zhang, J. (2001), 'Generalized likelihood ratio statistics and wilks phenomenon', *Annals of Statistics* **29**(1), 153–193.

Ferson, W. and Chen, Y. (2017), 'How many good and bad fund managers are there, really?', *University of Southern California working paper* .

Ferson, W. E. and Schadt, R. W. (1996), 'Measuring fund strategy and performance in changing economic conditions', *Journal of Finance* **51**(2), 425–461.

Ferson, W. and Lin, J. (2014), 'Alpha and performance measurement: The effects of investor disagreement and heterogeneity', *Journal of Finance* **69**(4), 1565–1596.

Fung, W. and Hsieh, D. A. (1997), 'Empirical characteristics of dynamic trading strategies: The case of hedge funds', *Review of Financial Studies* **10**(2), 275–302.

Fung, W. and Hsieh, D. A. (2001), 'The risk in hedge fund strategies: Theory and evidence from trend followers', *Review of Financial Studies* **14**(2), 313–341.

Fung, W., Hsieh, D. A., Naik, N. Y. and Ramadorai, T. (2008), 'Hedge funds: Performance, risk, and capital formation', *Journal of Finance* **63**(4), 1777–1803.

Glode, V. (2011), 'Why mutual funds "underperform"', *Journal of Financial Economics* **99**(3), 546–559.

Harvey, C. R. and Liu, Y. (2018), 'Detecting repeatable performance', *Review of Financial Studies* **31**(7), 2499–2552.

Hou, K., Xue, C. and Zhang, L. (2015), 'Digesting anomalies: An investment approach', *Review of Financial Studies* **28**(3), 650–705.

Huang, J., Wei, K. D. and Yan, H. (2007), 'Participation costs and the sensitivity of fund flows to past performance', *Journal of Finance* **62**(3), 1273–1311.

Hunter, D., Kandel, E., Kandel, S. and Wermers, R. (2014), 'Mutual fund performance evaluation with active peer benchmarks', *Journal of Financial economics* **112**(1), 1–29.

Jensen, M. C. (1968), 'The performance of mutual funds in the period 1945–1964', *Journal of Finance* **23**(2), 389–416.

Jiang, G. J., Yao, T. and Yu, T. (2007), 'Do mutual funds time the market? evidence from portfolio holdings', *Journal of Financial Economics* **86**(3), 724–758.

Jones, C. S. and Mo, H. (2016), 'Out-of-sample performance of mutual fund predictors', *University of Southern California working paper* .

Kacperczyk, M., Van Nieuwerburgh, S., Veldkamp, L. et al. (2014), 'Time-varying fund manager skill', *Journal of Finance* **69**(4), 1455–1484.

Khorana, A. (1996), 'Top management turnover an empirical investigation of mutual fund managers', *Journal of Financial Economics* **40**(3), 403–427.

Kosowski, R., Naik, N. Y. and Teo, M. (2007), 'Do hedge funds deliver alpha? a bayesian and bootstrap analysis', *Journal of Financial Economics* **84**(1), 229–264.

Kosowski, R., Timmermann, A., Wermers, R. and White, H. (2006), 'Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis', *Journal of Finance* **61**(6), 2551–2595.

Lou, D. (2012), 'A flow-based explanation for return predictability', *Review of Financial Studies* **25**(12), 3457–3489.

Mamaysky, H., Spiegel, M. and Zhang, H. (2008), 'Estimating the dynamics of mutual fund alphas and betas', *Review of Financial Studies* **21**(1), 233–264.

Marcus, A. J. (1990), 'The magellan fund and market efficiency', *Journal of Portfolio Management* **17**(1), 85–88.

Pástor, L. and Stambaugh, R. F. (2012), 'On the size of the active management industry', *Journal of Political Economy* **120**(4), 740–781.

Pástor, L., Stambaugh, R. F. and Taylor, L. A. (2015), 'Scale and skill in active management', *Journal of Financial Economics* **116**(1), 23–45.

Robinson, P. M. (1989), Nonparametric estimation of time–varying parameters, *in* P. Hackl, ed., 'Statistical Analysis and Forecasting of Economis Structural Change', Springer, Berlin, pp. 253–264.

Treynor, J. and Mazuy, K. (1966), 'Can mutual funds outguess the market', *Harvard Business Review* **44**(4), 131–136.

Wermers, R. (2000), 'Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses', *Journal of Finance* **55**(4), 1655–1695.

Table 1: **Mean Squared Errors (MSE) of the nonparametric estimates of the $\alpha_i(t)$ and $\beta_{ji}(t)$.** This table reports the Mean Squared Errors (MSE) of nonparametric estimates of the $\alpha_i(t)$ and $\beta_{ji}(t)$ from DGP1, DGP2, DGP3 and DGP4, respectively. T is the time length.

DGP	T	MSE($\widehat{\alpha}_i(t)$)	MSE($\widehat{\beta}_i(t)$)
DGP1	200	0.000015	0.008736
	400	0.000008	0.005192
	800	0.000005	0.002522
DGP2	200	0.000016	0.009680
	400	0.000009	0.004768
	800	0.000004	0.002423
DGP3	200	0.000016	0.008743
	400	0.000009	0.005194
	800	0.000005	0.002522
DGP4	200	0.000016	0.009708
	400	0.000009	0.004774
	800	0.000005	0.002425

Table 2: **Summary statistics.** Monthly returns are summarized for mutual funds (top panel) and hedge funds (bottom panel), measured in excess of the one-month return of a three-month Treasury bill. The values at the cutoff points for various quantiles of the cross-sectional distributions of the sample of funds are reported. Each column is sorted on the statistic shown. Nobs is the number of available monthly returns, where for the top left (and bottom left) panel, there is no restriction while a minimum of 30 is required for the top right (and bottom right) panel. Mean is the sample mean return, Std is the sample standard deviation of return, and Rho1 is the first order sample autocorrelation. The alpha estimates are based on OLS regressions using the Fama-French-Carhart four factors (Carhart (1997)) for mutual funds, while the Fung-Hsieh seven factors (Fung and Hsieh (1997, 2001) are used for the hedge funds.

Quantiles	Mutual funds (full sample)					Mutual funds (minimum 30 obs)				
	Nobs	Mean	Std	Rho1	$\widehat{\alpha}_{ols}$	Nobs	Mean	Std	Rho1	$\widehat{\alpha}_{ols}$
Top	335	0.060	0.512	0.688	0.032	335	0.060	0.512	0.688	0.024
1%	333	0.021	0.117	0.406	0.008	335	0.018	0.114	0.361	0.008
5%	263	0.013	0.088	0.303	0.004	277	0.012	0.087	0.284	0.004
10%	223	0.010	0.078	0.254	0.003	232	0.010	0.077	0.243	0.003
20%	178	0.008	0.069	0.207	0.001	190	0.007	0.068	0.205	0.001
30%	149	0.006	0.062	0.172	0.001	163	0.006	0.062	0.173	0.001
Median	97	0.004	0.053	0.121	-0.000	118	0.004	0.054	0.127	-0.000
30%	53	0.002	0.046	0.062	-0.001	76	0.002	0.047	0.079	-0.001
20%	38	-0.000	0.042	0.020	-0.002	58	0.001	0.043	0.049	-0.002
10%	22	-0.003	0.036	-0.057	-0.003	44	-0.002	0.038	0.000	-0.003
5%	13	-0.008	0.030	-0.121	-0.005	38	-0.004	0.034	-0.052	-0.005
1%	9	-0.023	0.018	-0.287	-0.010	32	-0.010	0.022	-0.149	-0.009
Bottom	8	-0.090	0.002	-0.627	-0.141	31	-0.035	0.004	-0.551	-0.049
Quantiles	Hedge funds (full sample)					Hedge funds (minimum 30 obs)				
	Nobs	Mean	Std	Rho1	$\widehat{\alpha}_{ols}$	Nobs	Mean	Std	Rho1	$\widehat{\alpha}_{ols}$
Top	192	0.173	0.695	0.814	0.868	192	0.051	0.324	0.814	0.045
1%	172	0.026	0.173	0.579	0.024	182	0.021	0.156	0.584	0.020
5%	126	0.014	0.098	0.457	0.012	147	0.012	0.090	0.479	0.011
10%	102	0.009	0.075	0.390	0.008	124	0.009	0.071	0.409	0.008
20%	73	0.006	0.053	0.296	0.005	96	0.006	0.052	0.323	0.005
30%	56	0.004	0.042	0.234	0.004	78	0.004	0.042	0.265	0.004
Median	38	0.001	0.028	0.127	0.002	57	0.002	0.029	0.170	0.002
30%	22	-0.002	0.020	0.009	-0.000	46	-0.000	0.022	0.078	0.000
20%	16	-0.005	0.016	-0.072	-0.001	40	-0.002	0.018	0.021	-0.001
10%	11	-0.011	0.012	-0.188	-0.004	36	-0.005	0.014	-0.071	-0.003
5%	8	-0.018	0.009	-0.304	-0.009	33	-0.008	0.010	-0.133	-0.006
1%	3	-0.043	0.005	-0.518	-0.024	31	-0.018	0.007	-0.284	-0.014
Bottom	1	-0.114	0.000	-0.794	-1.513	31	-0.038	0.001	-0.492	-0.031

Table 3: **Net long-run alpha estimates from Fama-French-Carhart 4-factor model for mutual funds.** This table reports the Fama-French-Carhart risk-adjusted monthly alphas for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest long-run alphas on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest long-run alphas at the bottom. We report the results generated by the intra-fund bootstrap, pooled bootstrap and inter-fund bootstrap schemes, in Panel A, B and C, respectively. The first two columns prior to the panels report the selected quantiles and the Cumulative Distribution Function of the actual (Act) long-run alphas at selected quantiles when they are ranked from the highest to the lowest, while the two columns in each panel report the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution as well as the p values that correspond to the selected quantiles of the distribution of the simulated long-run alphas generated by each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

		Panel A: Intra-fund		Panel B: Pooled		Panel C: Inter-fund	
Quantiles	Act	Sim	p value	Sim	p value	Sim	p value
Top	0.064	0.028	0.000	0.016	0.000	0.028	0.095
2.	0.038	0.022	0.000	0.014	0.000	0.017	0.050
3.	0.013	0.020	0.990	0.014	0.485	0.013	0.295
4.	0.013	0.018	0.980	0.013	0.425	0.012	0.205
5.	0.012	0.017	0.980	0.012	0.405	0.011	0.195
1%	0.007	0.010	1.000	0.009	0.840	0.006	0.110
3%	0.005	0.006	1.000	0.006	0.965	0.004	0.150
5%	0.004	0.005	1.000	0.005	0.995	0.004	0.330
10%	0.002	0.003	1.000	0.004	0.995	0.002	0.750
20%	0.001	0.002	1.000	0.002	0.995	0.002	1.000
30%	0.000	0.001	1.000	0.002	0.995	0.001	1.000
40%	-0.000	0.000	1.000	0.001	0.995	0.000	1.000
Median	-0.001						
40%	-0.002	-0.000	0.000	-0.000	0.000	-0.000	0.000
30%	-0.002	-0.001	0.000	-0.001	0.000	-0.001	0.000
20%	-0.003	-0.001	0.000	-0.002	0.000	-0.001	0.000
10%	-0.005	-0.003	0.000	-0.003	0.000	-0.002	0.000
5%	-0.007	-0.004	0.000	-0.005	0.000	-0.003	0.000
3%	-0.009	-0.005	0.000	-0.006	0.000	-0.004	0.000
1%	-0.012	-0.008	0.000	-0.008	0.015	-0.006	0.000
5.	-0.031	-0.015	0.000	-0.012	0.000	-0.009	0.000
4.	-0.035	-0.016	0.000	-0.012	0.000	-0.010	0.000
3.	-0.043	-0.018	0.000	-0.012	0.000	-0.010	0.000
2.	-0.043	-0.020	0.000	-0.013	0.000	-0.012	0.000
Bottom	-0.047	-0.027	0.030	-0.015	0.000	-0.014	0.000

Table 4: **Net long-run alpha estimates from the Fung-Hsieh 7-factor model for hedge funds.** This table reports the Fung-Hsieh seven-factor risk-adjusted monthly alphas for both actual and simulated hedge funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for hedge funds with the five highest long-run alphas on the top, followed by results for marginal hedge funds at different percentiles in the right and left tail of the distribution respectively, as well as results for hedge funds with the five lowest long-run alphas at the bottom. We report the results generated by the intra-fund bootstrap, pooled bootstrap and inter-fund bootstrap schemes, in Panel A, B and C, respectively. The first two columns prior to the panels report the selected quantiles and the Cumulative Distribution Function of the actual (Act) long-run alphas at selected quantiles when they are ranked from the highest to the lowest, while the two columns in each panel report the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution as well as the p values that correspond to the selected quantiles of the distribution of the simulated long-run alphas generated by the each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

		Panel A: Intra-fund		Panel B: Pooled		Panel C: Inter-fund	
Quantiles	Act	Sim	p value	Sim	p value	Sim	p value
Top	0.047	0.067	0.865	0.031	0.045	0.031	0.130
2.	0.042	0.051	0.800	0.028	0.040	0.024	0.080
3.	0.040	0.045	0.710	0.026	0.025	0.021	0.060
4.	0.038	0.041	0.630	0.025	0.025	0.019	0.045
5.	0.038	0.038	0.465	0.024	0.020	0.018	0.025
1%	0.021	0.019	0.050	0.016	0.025	0.009	0.000
3%	0.012	0.012	0.185	0.011	0.085	0.006	0.000
5%	0.010	0.009	0.010	0.009	0.145	0.005	0.000
10%	0.007	0.006	0.000	0.007	0.380	0.003	0.000
20%	0.004	0.003	0.000	0.004	0.815	0.002	0.000
30%	0.002	0.001	0.015	0.002	0.855	0.001	0.000
40%	0.000	0.000	1.000	0.000	1.000	0.000	1.000
Median	0.000						
40%	0.000	0.000	0.000	0.000	0.000	0.000	0.000
30%	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20%	0.000	-0.000	0.715	0.000	0.130	-0.000	0.980
10%	-0.003	-0.003	0.040	-0.002	0.045	-0.002	0.000
5%	-0.007	-0.005	0.000	-0.004	0.005	-0.003	0.000
3%	-0.009	-0.008	0.005	-0.006	0.010	-0.004	0.000
1%	-0.018	-0.014	0.000	-0.010	0.005	-0.007	0.000
5.	-0.036	-0.030	0.060	-0.018	0.030	-0.013	0.000
4.	-0.039	-0.032	0.085	-0.020	0.035	-0.014	0.000
3.	-0.044	-0.036	0.090	-0.021	0.020	-0.015	0.000
2.	-0.047	-0.042	0.240	-0.023	0.020	-0.016	0.005
Bottom	-0.051	-0.059	0.575	-0.027	0.050	-0.020	0.005

Table 5: **Net long-run alpha estimates from 3-factor model and CAPM for mutual funds via pooled bootstrap.** This table reports the risk-adjusted monthly alphas for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest long-run alphas on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest long-run alphas at the bottom. Panel A and B report the results generated by using the alphas estimated from the Fama and French 3-factor model and CAPM, respectively. The first column prior to the panels reports the selected quantiles. The three columns in each panel report the Cumulative Distribution Function of the actual (Act) long-run alphas at selected quantiles when they are ranked from the highest to the lowest, the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution, as well as the p values that correspond to the selected quantiles of the distribution of the simulated long-run alphas generated by the each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

	Panel A: 3-factor net α			Panel B: CAPM net α		
Quantiles	Act	Sim	p value	Act	Sim	p value
Top	0.049	0.015	0.000	0.019	0.015	0.160
2	0.031	0.014	0.000	0.018	0.014	0.090
3	0.013	0.013	0.420	0.018	0.013	0.070
4	0.013	0.012	0.375	0.016	0.013	0.150
5	0.013	0.012	0.345	0.015	0.013	0.145
1%	0.008	0.008	0.570	0.009	0.009	0.480
3%	0.005	0.006	0.825	0.007	0.007	0.440
5%	0.004	0.005	0.920	0.005	0.006	0.695
10%	0.003	0.004	0.975	0.004	0.004	0.830
20%	0.001	0.002	0.995	0.002	0.003	0.960
30%	0.000	0.002	0.995	0.001	0.002	0.955
40%	-0.000	0.001	1.000	0.000	0.001	0.955
Median	-0.001			-0.000		
40%	-0.002	-0.000	0.000	-0.001	-0.000	0.070
30%	-0.002	-0.001	0.015	-0.002	-0.001	0.205
20%	-0.003	-0.002	0.020	-0.002	-0.002	0.415
10%	-0.005	-0.003	0.010	-0.004	-0.004	0.245
5%	-0.007	-0.005	0.005	-0.006	-0.005	0.100
3%	-0.009	-0.006	0.000	-0.007	-0.006	0.135
1%	-0.012	-0.008	0.025	-0.012	-0.008	0.015
5	-0.029	-0.011	0.000	-0.029	-0.011	0.000
4	-0.032	-0.012	0.000	-0.030	-0.011	0.000
3	-0.035	-0.012	0.000	-0.035	-0.011	0.000
2	-0.046	-0.013	0.000	-0.036	-0.012	0.000
Bottom	-0.049	-0.014	0.000	-0.038	-0.013	0.000

Table 6: **Gross long-run alpha estimates for mutual funds via pooled bootstrap.** This table reports the risk-adjusted monthly gross alphas for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest long-run gross alphas on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest gross long-run alphas at the bottom. Panel A, B, and C report the results generated by using the alphas estimated from the Fama-French-Carhart 4-factor model, Fama and French 3-factor model and CAPM, respectively. The first column prior to the panels reports the selected quantiles. The three columns in each panel report the Cumulative Distribution Function of the actual (Act) long-run alphas at selected quantiles when they are ranked from the highest to the lowest, the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution, as well as the p values that correspond to the selected quantiles of the distribution of the simulated gross long-run alphas generated by the each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

	Panel A: 4-factor gross α			Panel B: 3-factor gross α			Panel C: CAPM gross α		
Quantiles	Act	Sim	p value	Act	Sim	p value	Act	Sim	p value
Top	0.064	0.017	0.000	0.049	0.016	0.000	0.020	0.015	0.075
2	0.039	0.015	0.000	0.031	0.014	0.000	0.019	0.013	0.050
3	0.014	0.014	0.405	0.014	0.013	0.340	0.019	0.013	0.015
4	0.014	0.013	0.335	0.013	0.013	0.370	0.016	0.012	0.045
5	0.013	0.013	0.335	0.013	0.012	0.315	0.016	0.012	0.035
1%	0.009	0.009	0.645	0.009	0.009	0.350	0.010	0.010	0.365
3%	0.006	0.007	0.800	0.006	0.007	0.630	0.008	0.007	0.255
5%	0.005	0.006	0.935	0.005	0.006	0.755	0.007	0.006	0.235
10%	0.003	0.004	0.975	0.004	0.004	0.880	0.005	0.005	0.250
20%	0.002	0.003	0.985	0.002	0.003	0.990	0.003	0.003	0.410
30%	0.001	0.002	0.970	0.001	0.002	0.970	0.002	0.002	0.545
40%	0.000	0.001	0.990	0.001	0.001	0.990	0.001	0.001	0.555
Median	-0.000			-0.000			0.000		
40%	-0.001	-0.000	0.000	-0.001	-0.000	0.000	-0.000	-0.000	0.665
30%	-0.002	-0.001	0.000	-0.002	-0.001	0.000	-0.001	-0.001	0.730
20%	-0.003	-0.002	0.000	-0.003	-0.002	0.005	-0.002	-0.002	0.745
10%	-0.005	-0.003	0.000	-0.004	-0.003	0.015	-0.004	-0.004	0.555
5%	-0.007	-0.005	0.000	-0.006	-0.005	0.010	-0.005	-0.005	0.270
3%	-0.009	-0.006	0.000	-0.009	-0.006	0.005	-0.007	-0.006	0.175
1%	-0.012	-0.009	0.050	-0.012	-0.008	0.040	-0.012	-0.009	0.050
5	-0.025	-0.012	0.005	-0.021	-0.011	0.015	-0.021	-0.011	0.010
4	-0.025	-0.012	0.010	-0.021	-0.012	0.025	-0.021	-0.011	0.010
3	-0.030	-0.013	0.005	-0.029	-0.013	0.005	-0.028	-0.012	0.000
2	-0.043	-0.014	0.000	-0.032	-0.014	0.010	-0.030	-0.012	0.005
Bottom	-0.046	-0.017	0.005	-0.048	-0.016	0.000	-0.036	-0.014	0.005

Table 7: **Value-added (VA) measure for mutual funds via pooled bootstrap.** This table reports the value-added (VA) measures for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest value-added (VA) measures on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest value-added (VA) measures at the bottom. Panel A, B, and C report the results generated by using the alphas estimated from the Fama-French-Carhart 4-factor model, Fama and French 3-factor model and CAPM, respectively. The first column prior to the panels reports the selected quantiles. The three columns in each panel report the Cumulative Distribution Function of the actual (Act) value-added (VA) measures at selected quantiles when they are ranked from the highest to the lowest, the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution, as well as the p values that correspond to the selected quantiles of the distribution of the simulated value-added (VA) measures generated by each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

	Panel A: 4-factor VA			Panel B: 3-factor VA			Panel C: CAPM VA		
Quantiles	Act	Sim	p value	Act	Sim	p value	Act	Sim	p value
Top	0.201	0.082	0.000	0.156	0.079	0.005	0.097	0.074	0.120
2	0.083	0.071	0.175	0.086	0.069	0.155	0.094	0.067	0.070
3	0.073	0.066	0.210	0.084	0.063	0.120	0.074	0.063	0.195
4	0.066	0.062	0.260	0.081	0.060	0.095	0.072	0.061	0.185
5	0.063	0.059	0.265	0.081	0.057	0.060	0.070	0.059	0.160
1%	0.046	0.043	0.305	0.049	0.043	0.175	0.051	0.047	0.260
3%	0.032	0.032	0.495	0.034	0.032	0.315	0.037	0.035	0.315
5%	0.024	0.027	0.805	0.027	0.027	0.500	0.032	0.029	0.220
10%	0.017	0.020	0.875	0.018	0.020	0.735	0.023	0.022	0.285
20%	0.010	0.012	0.910	0.011	0.013	0.910	0.014	0.014	0.510
30%	0.006	0.008	0.860	0.006	0.008	0.945	0.009	0.009	0.520
40%	0.003	0.004	0.815	0.004	0.005	0.870	0.005	0.006	0.505
Median	0.001			0.001			0.003		
40%	-0.001	-0.001	0.510	-0.002	-0.001	0.285	0.000	-0.001	0.855
30%	-0.004	-0.005	0.730	-0.004	-0.004	0.555	-0.002	-0.005	0.990
20%	-0.007	-0.009	0.825	-0.007	-0.008	0.695	-0.005	-0.009	0.990
10%	-0.013	-0.015	0.910	-0.013	-0.015	0.870	-0.010	-0.016	0.990
5%	-0.020	-0.022	0.710	-0.018	-0.021	0.890	-0.015	-0.023	0.990
3%	-0.025	-0.027	0.700	-0.025	-0.026	0.640	-0.021	-0.028	0.970
1%	-0.043	-0.040	0.275	-0.037	-0.038	0.435	-0.036	-0.039	0.510
5	-0.057	-0.054	0.310	-0.052	-0.053	0.415	-0.057	-0.051	0.220
4	-0.068	-0.057	0.180	-0.053	-0.056	0.440	-0.079	-0.053	0.070
3	-0.085	-0.061	0.095	-0.060	-0.060	0.395	-0.082	-0.055	0.085
2	-0.100	-0.066	0.080	-0.084	-0.067	0.180	-0.088	-0.059	0.095
Bottom	-0.143	-0.079	0.035	-0.141	-0.077	0.060	-0.093	-0.065	0.100

Table 8: **Alternative benchmark for mutual funds via pooled bootstrap and CAPM.** This table reports the alternative net alphas, alternative gross alphas, alternative value-added measure for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest alpha/VA measures on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest alpha/VA measures at the bottom. Panel A, B, and C report the results generated via CAPM by using the alternative net alphas, alternative gross alphas, alternative value-added measure, respectively. The first column prior to the panels reports the selected quantiles. The three columns in each panel report the Cumulative Distribution Function of the actual (Act) long-run alpha/VA measures at selected quantiles when they are ranked from the highest to the lowest, the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution, as well as the p values that correspond to the selected quantiles of the distribution of the simulated long-run alpha/VA measures generated by each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

	Panel A: alternative net α			Panel B: alternative gross α			Panel C: alternative VA		
Quantiles	Act	Sim	p value	Act	Sim	p value	Act	Sim	p value
Top	0.021	0.018	0.190	0.022	0.016	0.075	0.123	0.084	0.075
2	0.021	0.016	0.130	0.021	0.015	0.035	0.104	0.075	0.050
3	0.021	0.015	0.110	0.021	0.014	0.015	0.086	0.071	0.155
4	0.019	0.015	0.110	0.020	0.014	0.025	0.086	0.069	0.120
5	0.019	0.014	0.100	0.020	0.013	0.020	0.085	0.066	0.095
1%	0.013	0.010	0.145	0.014	0.011	0.065	0.070	0.053	0.080
3%	0.010	0.007	0.010	0.011	0.008	0.015	0.057	0.039	0.010
5%	0.009	0.006	0.000	0.010	0.007	0.000	0.050	0.032	0.000
10%	0.007	0.004	0.000	0.008	0.005	0.000	0.039	0.024	0.000
20%	0.006	0.003	0.000	0.006	0.003	0.000	0.029	0.015	0.000
30%	0.005	0.002	0.000	0.005	0.002	0.000	0.023	0.010	0.000
40%	0.004	0.001	0.000	0.004	0.001	0.000	0.017	0.006	0.000
Median	0.003			0.004			0.013		
40%	0.002	-0.001	0.995	0.003	-0.000	1.000	0.010	-0.002	1.000
30%	0.002	-0.001	0.995	0.002	-0.001	1.000	0.006	-0.005	1.000
20%	0.001	-0.002	0.995	0.001	-0.002	1.000	0.002	-0.010	1.000
10%	-0.001	-0.004	0.995	-0.000	-0.004	1.000	-0.002	-0.018	1.000
5%	-0.003	-0.006	0.990	-0.003	-0.006	0.995	-0.008	-0.025	1.000
3%	-0.004	-0.007	0.975	-0.004	-0.007	0.985	-0.011	-0.031	0.995
1%	-0.009	-0.009	0.530	-0.009	-0.010	0.640	-0.024	-0.043	1.000
5	-0.024	-0.012	0.005	-0.017	-0.012	0.060	-0.044	-0.056	0.830
4	-0.025	-0.012	0.005	-0.018	-0.013	0.075	-0.065	-0.058	0.260
3	-0.032	-0.013	0.005	-0.023	-0.013	0.015	-0.069	-0.061	0.205
2	-0.035	-0.013	0.000	-0.025	-0.014	0.010	-0.072	-0.064	0.240
Bottom	-0.037	-0.014	0.000	-0.035	-0.015	0.005	-0.078	-0.072	0.295

Table 9: **Net long-run alpha estimates for mutual funds via pooled bootstrap (1984-2001).** This table reports the risk-adjusted monthly alphas for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest long-run alphas on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest long-run alphas at the bottom. Panel A, B and C report the results generated by using the alphas estimated from the Fama-French-Carhart 4-factor model, Fama and French 3-factor model and CAPM, respectively. The first column prior to the panels reports the selected quantiles. The three columns in each panel report the Cumulative Distribution Function of the actual (Act) long-run alphas at selected quantiles when they are ranked from the highest to the lowest, the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution, as well as the p values that correspond to the selected quantiles of the distribution of the simulated long-run alphas generated by the each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

	Panel A: 4-factor net α			Panel B: 3-factor net α			Panel C: CAPM net α		
Quantiles	Act	Sim	p value	Act	Sim	p value	Act	Sim	p value
Top	0.034	0.021	0.045	0.029	0.021	0.080	0.030	0.018	0.020
2	0.033	0.019	0.045	0.029	0.018	0.050	0.027	0.016	0.035
3	0.030	0.017	0.025	0.028	0.017	0.035	0.027	0.015	0.025
4	0.028	0.017	0.035	0.023	0.016	0.100	0.025	0.015	0.025
5	0.023	0.016	0.090	0.021	0.016	0.125	0.024	0.015	0.035
1%	0.016	0.012	0.160	0.015	0.012	0.165	0.014	0.012	0.190
3%	0.009	0.009	0.445	0.010	0.009	0.245	0.010	0.009	0.170
5%	0.006	0.007	0.625	0.007	0.007	0.525	0.008	0.007	0.370
10%	0.004	0.005	0.840	0.004	0.005	0.740	0.005	0.005	0.545
20%	0.002	0.003	0.905	0.002	0.003	0.910	0.003	0.003	0.585
30%	0.001	0.002	0.955	0.001	0.002	0.970	0.002	0.002	0.710
40%	0.000	0.001	0.980	0.000	0.001	0.985	0.001	0.001	0.610
Median	-0.001			-0.001			-0.000		
40%	-0.002	-0.001	0.040	-0.002	-0.001	0.025	-0.001	-0.001	0.435
30%	-0.003	-0.002	0.025	-0.003	-0.002	0.070	-0.002	-0.002	0.490
20%	-0.004	-0.003	0.030	-0.004	-0.003	0.010	-0.004	-0.004	0.470
10%	-0.007	-0.005	0.000	-0.007	-0.005	0.005	-0.006	-0.005	0.195
5%	-0.010	-0.006	0.010	-0.010	-0.007	0.000	-0.009	-0.007	0.110
3%	-0.012	-0.008	0.005	-0.013	-0.008	0.000	-0.011	-0.008	0.055
1%	-0.016	-0.011	0.015	-0.016	-0.010	0.015	-0.015	-0.010	0.045
5	-0.030	-0.013	0.000	-0.032	-0.013	0.000	-0.021	-0.012	0.010
4	-0.031	-0.014	0.000	-0.033	-0.013	0.000	-0.029	-0.013	0.000
3	-0.043	-0.015	0.000	-0.033	-0.014	0.000	-0.030	-0.013	0.000
2	-0.043	-0.016	0.000	-0.046	-0.015	0.000	-0.036	-0.013	0.000
Bottom	-0.047	-0.018	0.000	-0.049	-0.016	0.000	-0.038	-0.014	0.000

Table 10: **Gross long-run alpha estimates for mutual funds via pooled bootstrap (1984-2001).** This table reports the risk-adjusted monthly gross alphas for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest gross long-run alphas on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest gross long-run alphas at the bottom. Panel A, B and C report the results generated by using the alphas estimated from the Fama-French-Carhart 4-factor model, Fama and French 3-factor model and CAPM, respectively. The first column prior to the panels reports the selected quantiles. The three columns in each panel report the Cumulative Distribution Function of the actual (Act) gross long-run alphas at selected quantiles when they are ranked from the highest to the lowest, the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution, as well as the p values that correspond to the selected quantiles of the distribution of the simulated gross long-run alphas generated by the each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

	Panel A: 4-factor net α			Panel B: 3-factor net α			Panel C: CAPM net α		
Quantiles	Act	Sim	p value	Act	Sim	p value	Act	Sim	p value
Top	0.034	0.016	0.000	0.030	0.016	0.005	0.039	0.022	0.005
2	0.034	0.015	0.000	0.029	0.015	0.005	0.039	0.019	0.000
3	0.030	0.014	0.000	0.029	0.014	0.000	0.032	0.018	0.005
4	0.029	0.013	0.000	0.024	0.013	0.000	0.029	0.017	0.000
5	0.024	0.013	0.005	0.022	0.013	0.010	0.027	0.015	0.000
1%	0.020	0.011	0.005	0.017	0.011	0.030	0.022	0.013	0.000
3%	0.010	0.009	0.205	0.011	0.009	0.090	0.015	0.010	0.010
5%	0.008	0.007	0.260	0.009	0.007	0.165	0.012	0.008	0.015
10%	0.005	0.006	0.500	0.006	0.006	0.440	0.007	0.006	0.225
20%	0.003	0.004	0.670	0.003	0.004	0.650	0.004	0.004	0.480
30%	0.002	0.002	0.745	0.002	0.002	0.675	0.002	0.003	0.570
40%	0.001	0.001	0.635	0.001	0.001	0.755	0.001	0.001	0.510
Median	0.000			0.000			0.000		
40%	-0.001	-0.001	0.425	-0.001	-0.001	0.315	-0.001	-0.001	0.565
30%	-0.002	-0.002	0.355	-0.002	-0.002	0.295	-0.002	-0.002	0.480
20%	-0.004	-0.003	0.250	-0.004	-0.003	0.120	-0.004	-0.003	0.360
10%	-0.007	-0.005	0.030	-0.007	-0.005	0.035	-0.006	-0.005	0.155
5%	-0.010	-0.007	0.005	-0.010	-0.007	0.030	-0.010	-0.007	0.020
3%	-0.011	-0.008	0.015	-0.012	-0.008	0.035	-0.012	-0.009	0.015
1%	-0.015	-0.010	0.035	-0.016	-0.010	0.015	-0.018	-0.011	0.000
5	-0.029	-0.013	0.000	-0.029	-0.012	0.000	-0.030	-0.013	0.000
4	-0.030	-0.013	0.000	-0.032	-0.013	0.000	-0.031	-0.014	0.000
3	-0.030	-0.014	0.000	-0.033	-0.013	0.000	-0.031	-0.015	0.000
2	-0.043	-0.015	0.000	-0.033	-0.014	0.000	-0.032	-0.016	0.000
Bottom	-0.046	-0.016	0.000	-0.048	-0.015	0.000	-0.033	-0.018	0.000

Table 11: **value-added (VA) measure for mutual funds via pooled bootstrap (1984-2001).** This table reports the value-added (VA) measures for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest value-added (VA) measures on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest value-added (VA) measures at the bottom. Panel A, B and C report the results generated by using the alphas estimated from the Fama-French-Carhart 4-factor model, Fama and French 3-factor model and CAPM, respectively. The first column prior to the panels reports the selected quantiles. The three columns in each panel report the Cumulative Distribution Function of the actual (Act) value-added (VA) measures at selected quantiles when they are ranked from the highest to the lowest, the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution, as well as the p values that correspond to the selected quantiles of the distribution of the simulated value-added (VA) measures generated by the each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

	Panel A: 4-factor VA			Panel B: 3-factor VA			Panel C: CAPM VA		
Quantiles	Act	Sim	p value	Act	Sim	p value	Act	Sim	p value
Top	0.150	0.073	0.005	0.161	0.075	0.000	0.147	0.083	0.030
2	0.147	0.064	0.000	0.150	0.067	0.000	0.119	0.075	0.075
3	0.133	0.061	0.000	0.135	0.063	0.000	0.118	0.072	0.055
4	0.133	0.058	0.000	0.122	0.060	0.000	0.114	0.070	0.050
5	0.129	0.055	0.000	0.118	0.058	0.000	0.105	0.069	0.070
1%	0.093	0.047	0.010	0.088	0.050	0.015	0.087	0.059	0.070
3%	0.051	0.037	0.035	0.051	0.038	0.040	0.057	0.046	0.125
5%	0.038	0.031	0.130	0.041	0.033	0.085	0.048	0.038	0.160
10%	0.025	0.023	0.340	0.027	0.025	0.300	0.034	0.028	0.190
20%	0.015	0.015	0.475	0.015	0.016	0.605	0.021	0.017	0.225
30%	0.009	0.009	0.500	0.009	0.010	0.730	0.013	0.010	0.235
40%	0.005	0.005	0.480	0.004	0.005	0.695	0.008	0.004	0.155
Median	0.001			0.001			0.003		
40%	-0.002	-0.003	0.785	-0.002	-0.003	0.695	0.000	-0.005	0.950
30%	-0.005	-0.008	0.860	-0.005	-0.008	0.835	-0.003	-0.011	0.990
20%	-0.011	-0.013	0.830	-0.011	-0.014	0.820	-0.009	-0.017	0.980
10%	-0.020	-0.022	0.725	-0.021	-0.022	0.600	-0.017	-0.026	0.995
5%	-0.029	-0.029	0.450	-0.030	-0.030	0.440	-0.026	-0.035	0.960
3%	-0.037	-0.035	0.305	-0.040	-0.036	0.220	-0.034	-0.040	0.830
1%	-0.056	-0.046	0.135	-0.054	-0.046	0.180	-0.052	-0.049	0.305
5	-0.062	-0.054	0.205	-0.071	-0.054	0.095	-0.066	-0.054	0.130
4	-0.062	-0.056	0.290	-0.072	-0.056	0.105	-0.077	-0.056	0.045
3	-0.072	-0.059	0.130	-0.076	-0.059	0.100	-0.082	-0.058	0.045
2	-0.085	-0.062	0.080	-0.084	-0.062	0.095	-0.088	-0.060	0.045
Bottom	-0.143	-0.070	0.000	-0.141	-0.069	0.005	-0.093	-0.065	0.045

Table 12: **Alternative benchmark for mutual funds via pooled bootstrap and CAPM (1984-2001).** This table reports the alternative net alphas, alternative gross alphas, alternative value-added (VA) measures for both actual and simulated mutual funds, ranked from highest (Top) to lowest (Bottom). In each column, we firstly report results for mutual funds with the five highest long-run alphas/VAs on the top, followed by results for marginal mutual funds at different percentiles in the right and left tail of the distribution respectively, as well as results for mutual funds with the five lowest long-run alphas/VAs at the bottom. Panel A, B and C report the results generated via CAPM by using the alternative net alphas, alternative gross alphas, alternative value-added measure, respectively. The first column prior to the panels reports the selected quantiles. The three columns in each panel report the Cumulative Distribution Function of the actual (Act) long-run alphas/VAs at selected quantiles when they are ranked from the highest to the lowest, the Cumulative Distribution Function of the simulated (Sim) 'luck' distribution, as well as the p values that correspond to the selected quantiles of the distribution of the simulated long-run alphas/VAs generated by the each bootstrap scheme, respectively. The p values for the three bootstrap schemes are based on the distribution of the best (worst) funds in 200 bootstrap resamples.

	Panel A: alternative net α			Panel B: alternative gross α			Panel C: alternative VA		
Quantiles	Act	Sim	p value	Act	Sim	p value	Act	Sim	p value
Top	0.032	0.021	0.045	0.032	0.019	0.015	0.160	0.100	0.045
2	0.031	0.019	0.040	0.031	0.017	0.005	0.136	0.089	0.080
3	0.030	0.018	0.025	0.031	0.017	0.005	0.126	0.084	0.085
4	0.029	0.017	0.015	0.029	0.016	0.005	0.120	0.081	0.095
5	0.027	0.017	0.020	0.028	0.016	0.010	0.117	0.080	0.085
1%	0.018	0.014	0.130	0.020	0.014	0.045	0.102	0.069	0.060
3%	0.013	0.011	0.125	0.016	0.011	0.055	0.073	0.053	0.080
5%	0.011	0.009	0.080	0.013	0.010	0.030	0.062	0.045	0.065
10%	0.009	0.006	0.045	0.010	0.007	0.040	0.051	0.032	0.010
20%	0.007	0.004	0.005	0.008	0.004	0.000	0.034	0.019	0.000
30%	0.005	0.002	0.000	0.006	0.003	0.000	0.024	0.010	0.000
40%	0.004	0.001	0.000	0.005	0.001	0.000	0.019	0.004	0.000
Median	0.003			0.003			0.014		
40%	0.002	-0.001	1.000	0.002	-0.001	1.000	0.008	-0.008	1.000
30%	0.001	-0.003	0.995	0.001	-0.003	1.000	0.003	-0.014	1.000
20%	-0.001	-0.004	0.995	-0.000	-0.004	1.000	-0.001	-0.021	1.000
10%	-0.003	-0.006	0.995	-0.003	-0.007	1.000	-0.009	-0.032	1.000
5%	-0.006	-0.008	0.975	-0.005	-0.009	0.990	-0.018	-0.041	1.000
3%	-0.008	-0.009	0.830	-0.008	-0.010	0.945	-0.024	-0.047	1.000
1%	-0.012	-0.012	0.390	-0.012	-0.012	0.565	-0.040	-0.057	0.975
5	-0.019	-0.014	0.035	-0.017	-0.014	0.110	-0.056	-0.065	0.770
4	-0.024	-0.014	0.000	-0.018	-0.014	0.100	-0.067	-0.066	0.425
3	-0.025	-0.014	0.000	-0.023	-0.015	0.015	-0.069	-0.068	0.400
2	-0.035	-0.015	0.000	-0.025	-0.015	0.015	-0.072	-0.071	0.345
Bottom	-0.037	-0.016	0.000	-0.035	-0.017	0.000	-0.078	-0.078	0.380

Figure 1: **The nonparametric estimates of alphas from simulated data.** This figure plots the estimated $\widehat{\alpha}_i(t)$ from simulated data in each row (column) with T=200, 400 and 800 (from DGP1, DGP2, DGP3 and DGP4), respectively.

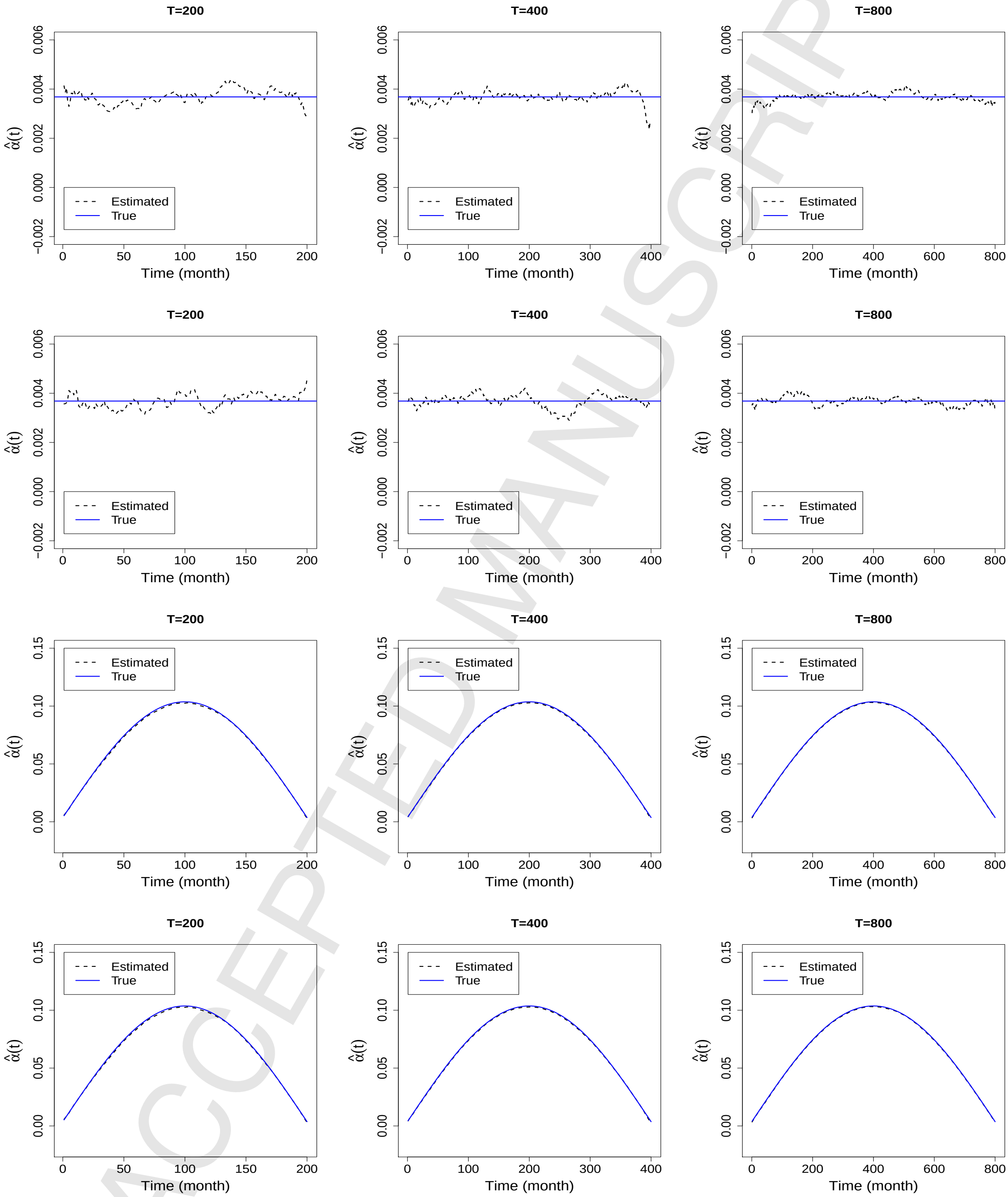


Figure 2: **The nonparametric estimates of betas from simulated data.** This figure plots the estimated $\widehat{\beta}_i(t)$ from simulated data in each row (column) with T=200, 400 and 800 (from DGP1, DGP2, DGP3 and DGP4), respectively.

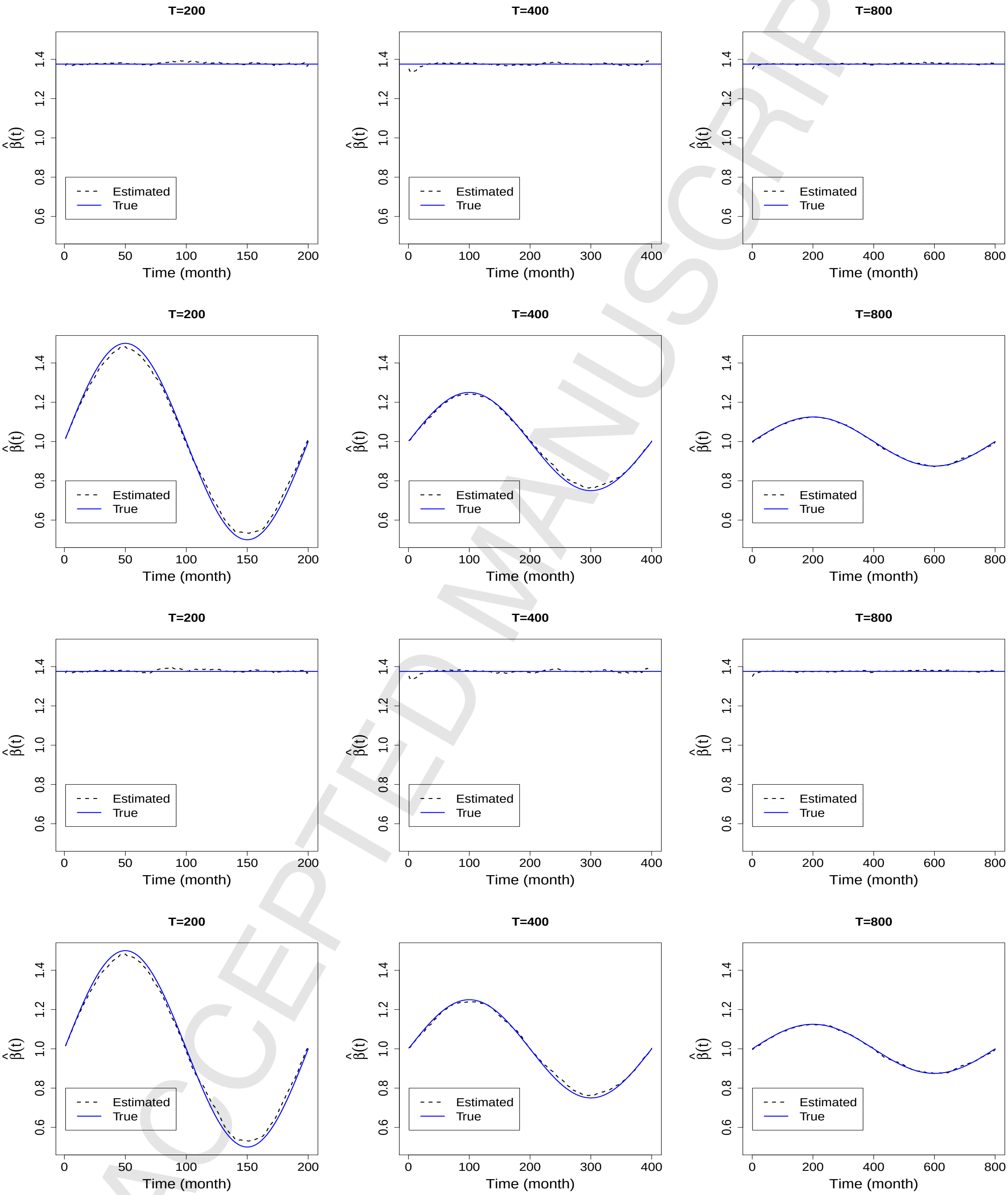


Figure 3: **The estimated net alphas and betas of the Fidelity Magellan fund.** This figure plots the Fidelity Magellan fund performance evaluation results using monthly returns. In each graph, the black solid line denotes the corresponding nonparametric estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$ and the black dotted line denotes their corresponding 95% confidence interval. The red dashed line denotes the corresponding time-invariant OLS estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$, while the blue dash-doted line denotes the corresponding rolling OLS estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$ with a rolling window of 24 months, respectively.

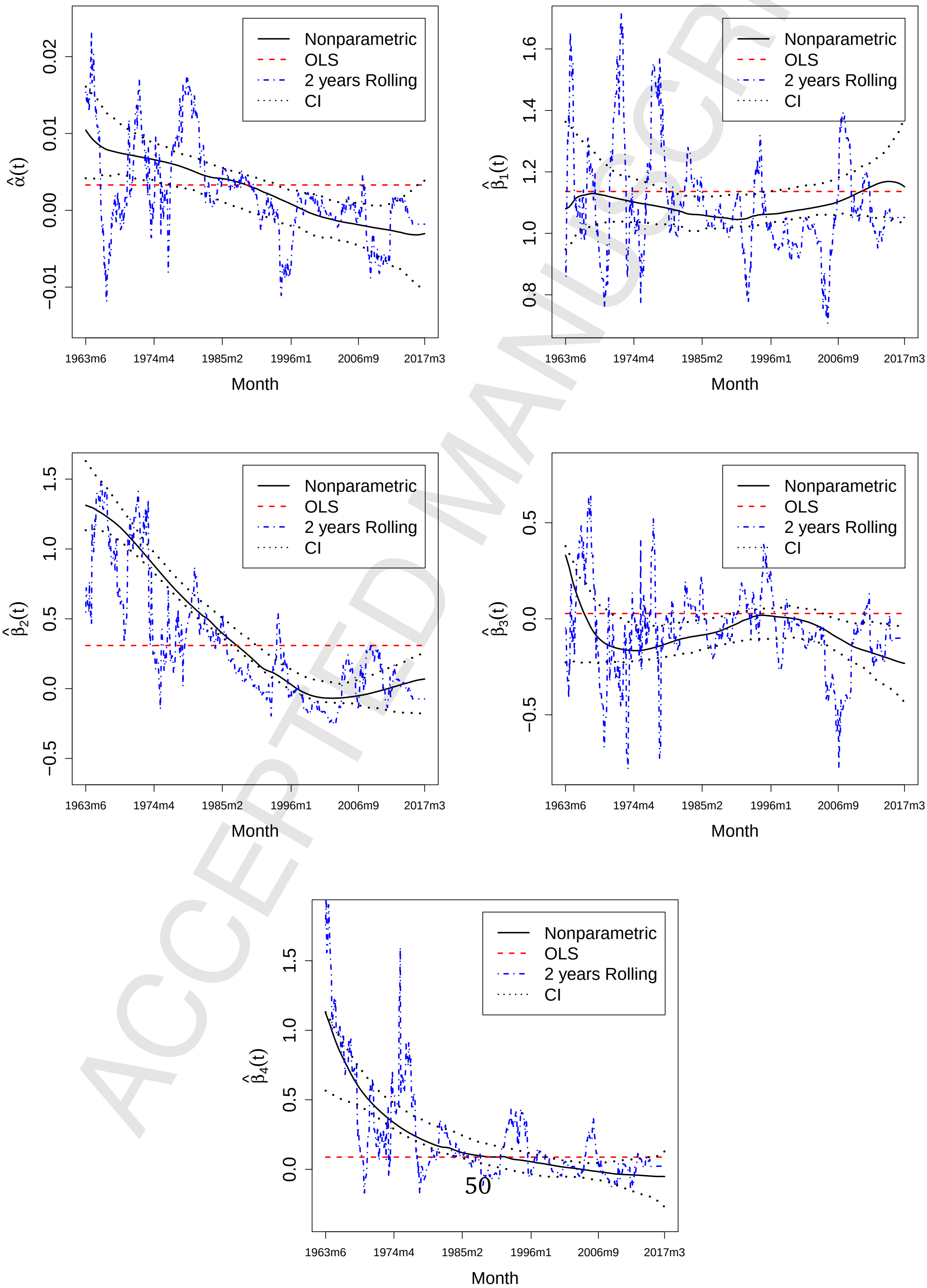


Figure 4: **The estimated net alphas and betas of the Fidelity Magellan fund.** This figure plots the Fidelity Magellan fund performance evaluation results using monthly returns. In each graph, the black solid line denotes the corresponding nonparametric estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$ and the black dotted line denotes their corresponding 95% confidence interval. The red dashed line denotes the corresponding time-invariant OLS estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$, while the blue dash-doted line denotes the corresponding rolling OLS estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$ with a rolling window of 36 months, respectively.

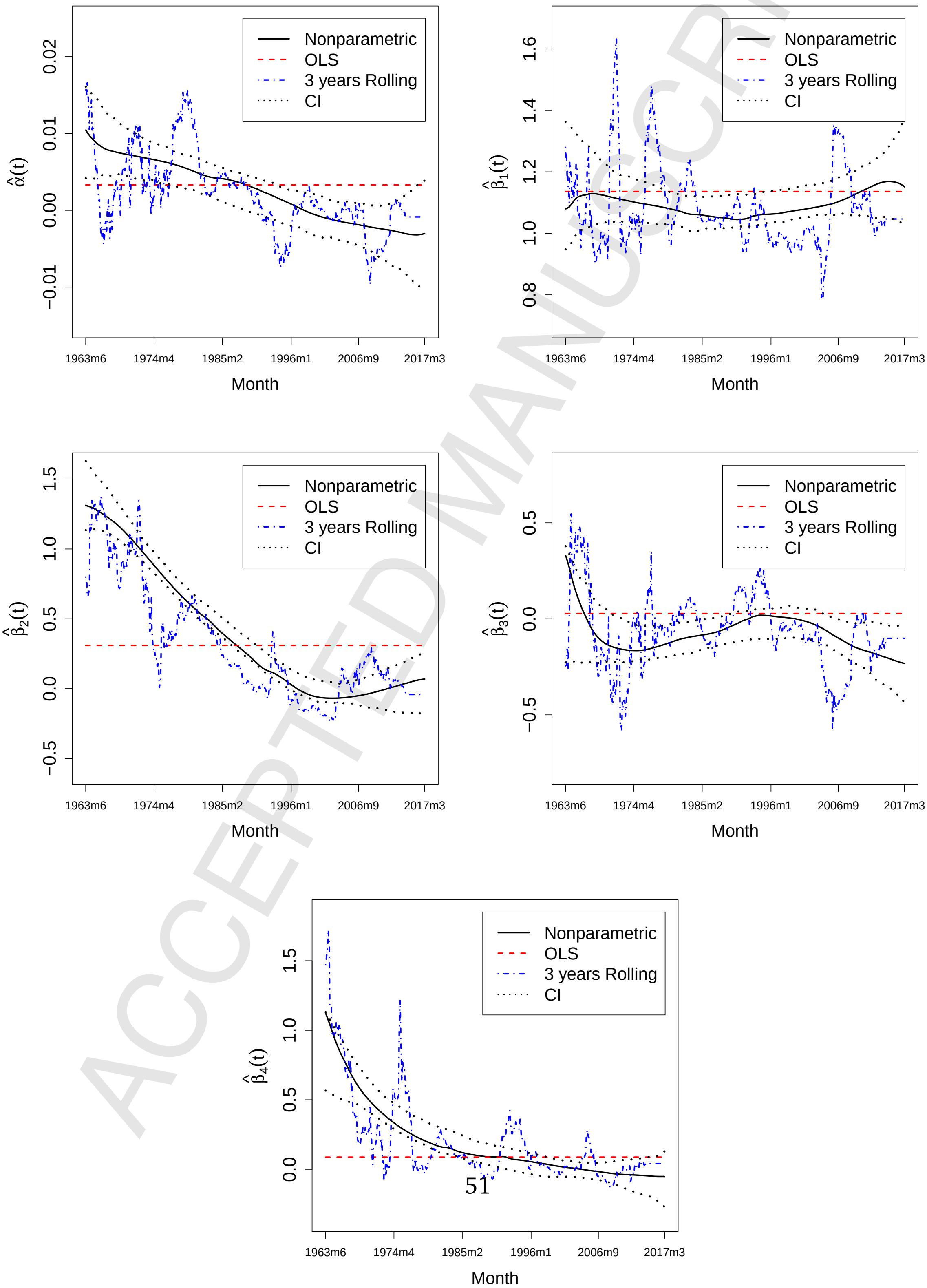


Figure 5: **The estimated net alphas and betas of the Fidelity Magellan fund.** This figure plots the Fidelity Magellan fund performance evaluation results using monthly returns. In each graph, the black solid line denotes the corresponding nonparametric estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$ and the black dotted line denotes their corresponding 95% confidence interval. The red dashed line denotes the corresponding time-invariant OLS estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$, while the blue dash-doted line denotes the corresponding rolling OLS estimates of $\alpha_i(t)$ or $\beta_{ji}(t)$ with a rolling window of 60 months, respectively.

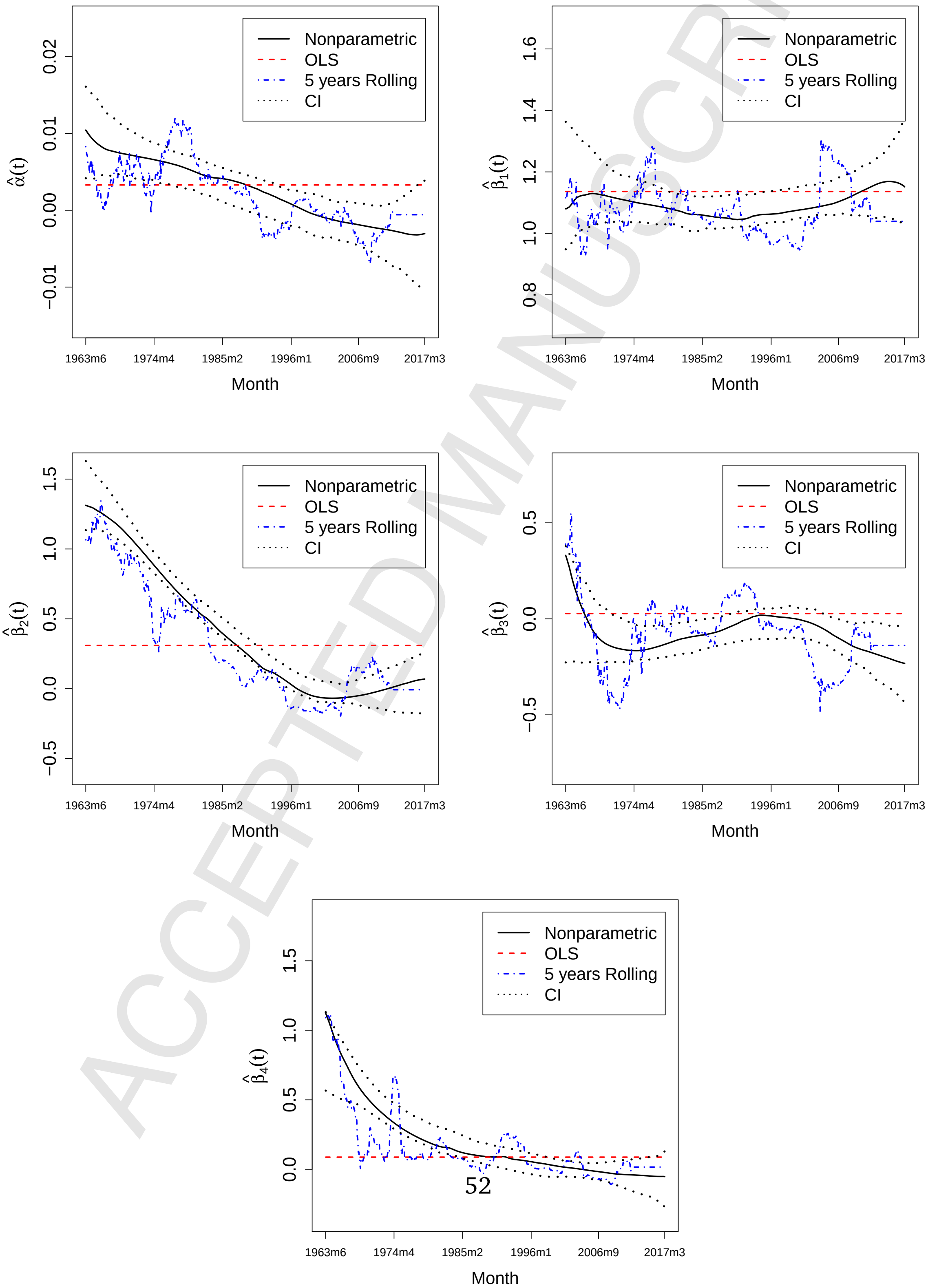


Figure 6: **The estimated net alphas of the whole mutual fund industry.** This figure plots the long-run alpha estimates and 95% confidence intervals by both nonparametric local linear estimation method and traditional OLS estimation method for the whole cross-section of mutual funds. The top panel shows the nonparametric estimates of long-run alpha for all the 2557 funds, the top 50 funds and the bottom 50 funds, respectively, while the bottom panel shows long-run alpha estimates and 95% confidence intervals by both nonparametric local linear estimation method and traditional OLS estimation method for all the 2557 funds, the top 50 funds (with the largest long-run alphas) and the bottom 50 funds (with the smallest long-run alphas), respectively.

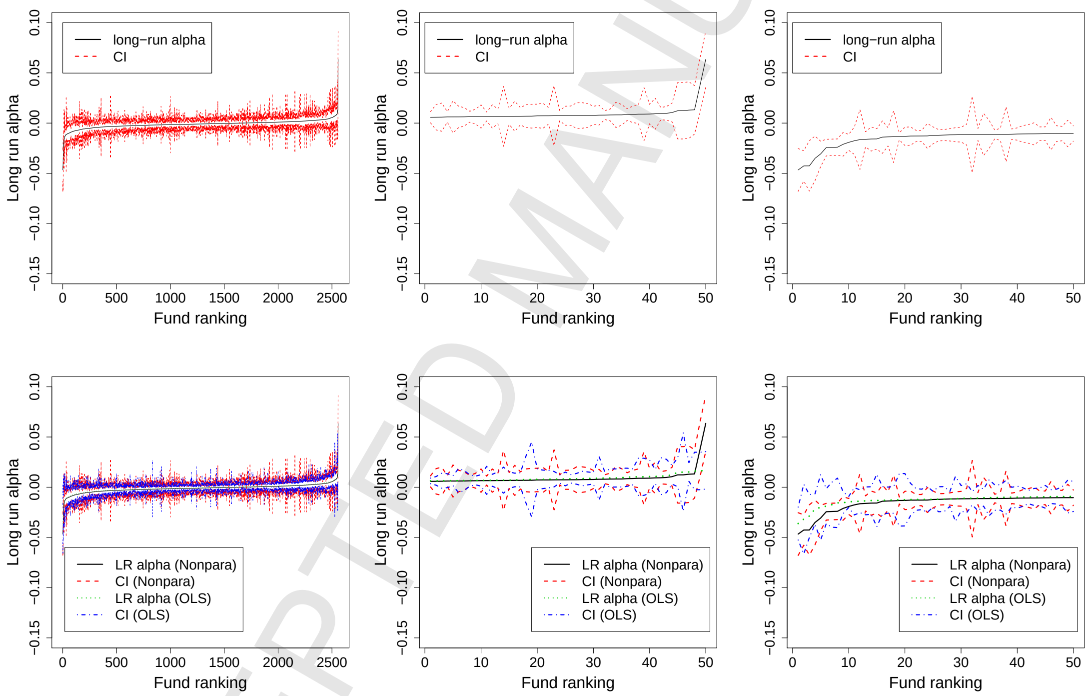


Figure 7: **The estimated time-varying net alphas of the top 5 mutual funds.** This figure plots the confidence interval of the estimated alpha curve for the top 5 funds with CRSP fund number =37271, 37640, 31182, 8432, 8434 respectively. In each graph, the black solid line denotes the corresponding nonparametric estimates of $\alpha_i(t)$; the blue dashed line denotes the 95% confidence interval and the green dotted line denotes the zeros.

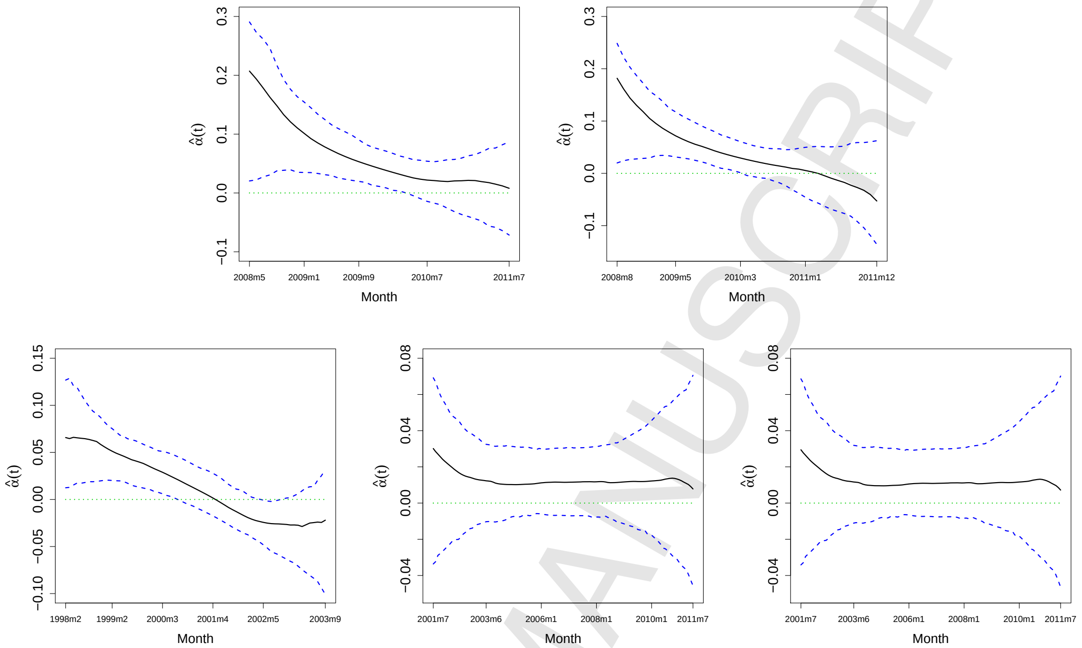
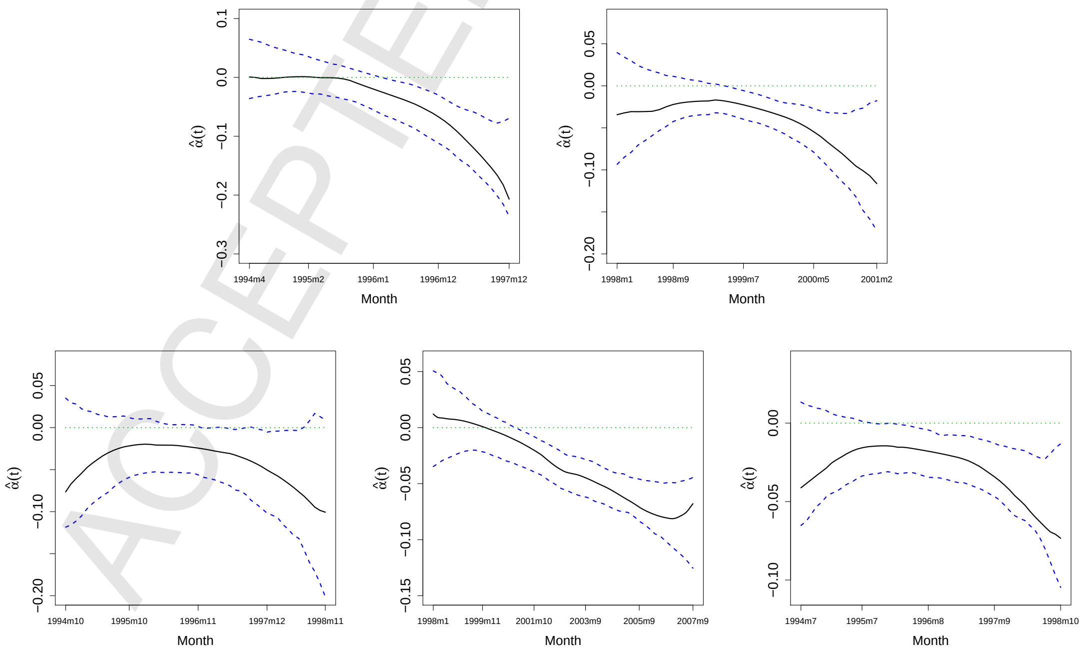


Figure 8: **The estimated time-varying net alphas of the bottom 5 mutual funds.** This figure plots the confidence interval of the estimated alpha curve for the bottom 5 funds with CRSP fund number = 34474, 28964, 2620, 5369, 2117 respectively. In each graph, the black solid line denotes the corresponding nonparametric estimates of $\alpha_i(t)$; the blue dashed line denotes the 95% confidence interval and the green dotted line denotes the zeros.



Appendix

The content of this Appendix section is as follows. We present the detailed steps to construct confidence intervals for the time-varying alphas and betas defined in subsection 2.2 in Appendix A. We then present in Appendix B the procedure we adapt to obtain the p value of our proposed GLR statistic defined in subsection 2.3.

Appendix A. Constructing confidence intervals for the time-varying alphas and betas To further conduct more statistical inference, we construct the 95% confidence interval for the local linear estimator using the following bootstrap procedure. Although we could alternatively use the asymptotic results in Cai (2007) to construct the confidence interval, in finite samples it has been shown that bootstrap procedure can result in more satisfactory results.

(1) Based on the local linear estimator, we can first compute the residuals by

$$\widehat{e}_{it} = r_{it} - \widehat{\alpha}_i(t) - \sum_{j=1}^{K} \widehat{\beta}_{ij}(t) x_{jt}.$$

(2) We sample e_{it}^* from $\{\widehat{e}_{it}\}_{t=1}^T$ and then generate a bootstrap sample by

$$r_{it}^* = \widehat{\alpha}_i(t) + \sum_{j=1}^{K} \widehat{\beta}_{ij}(t) x_{jt} + e_{it}^*.$$

(3) Based on the bootstrap sample $\{r_{it}^*, x_{1t}, \cdots, x_{Kt}\}$, we estimate the unknown coefficient functions in model (1) and save the estimated coefficient function as $\widehat{\alpha}_i^{(1)*}(t)$ and $\widehat{\beta}_{ji}^{(1)*}(t)$, for $j = 1, \cdots, K$.

(4) Repeat steps (2) and (3) for B times and obtain $\widehat{\alpha}_i^{(b)*}(t)$ and $\widehat{\beta}_{ji}^{(b)*}(t)$, for $b = 1, 2, \cdots, B$.

(5) Calculate the 2.5% and 97.5% quantiles of the series $\{\widehat{\alpha}_i^{(b)*}(t)\}_{b=1}^B$ to obtain the 95% confidence interval of $\widehat{\alpha}_i(t)$. Similarly we can obtain the confidence interval for $\widehat{\beta}_{ji}(t)$.

Appendix B. Obtaining the p value of the GLR statistic

Here, we briefly explain the nonparametric bootstrap approach to obtain the p value of our proposed GLR statistic defined in Section 2.3.

(1) Generate the bootstrap residuals $\{e^*_{it}\}_{t=1}^T$ from the empirical distribution of the residuals $\{\widetilde{e}_{it}\}_{t=1}^T$, where $\widetilde{e}_{it} = r_{it} - \widetilde{\alpha}_i - x'_t\widetilde{\beta}_i(t)$. Define $r^*_{it} = \widetilde{\alpha}_i + x'_t\widetilde{\beta}_i(t) + e^*_{it}$.

(2) Calculate the bootstrap test statistic TS^* based on the sample $\{x_t, r^*_{it}\}_{t=1}^T$.

(3) Repeat the above steps (1) and (2) for B times.

(4) Compute TS based on the original sample. Reject the null hypothesis H_0 when TS is greater than the upper-α point of the conditional distribution of TS^* given $\{x_t, r^*_{it}\}_{t=1}^T$.

The p value of the test is simply the relative frequency of the event $\{TS^* \geq TS\}$ in the replications of the bootstrap sampling.