Statistica Sinica Preprint No: SS-2022-0416	
Title	Testing for Zero Skill In Stock Picking or Market Timing
Manuscript ID	SS-2022-0416
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0416
Complete List of Authors	Qingsong Shan, Lei Jiang, Liang Peng and Zhongling Qin
Corresponding Authors	Zhongling Qin
E-mails	zzq0018@auburn.edu
Notice: Accepted version subject to English editing.	

Testing for Zero Skill in Stock Picking or Market Timing

Qingsong Shan[a] Lei Jiang[b] Liang Peng[c] Zhongling Qin[d]

[a]*School of Statistics, Jiangxi University of Economics and Finance*

[b]*Department of Finance, School of Economics and Management, Tsinghua University*

[c]*Department of Risk Management and Insurance, Georgia State University*

[d]*Department of Finance, Auburn University*

Abstract: To identify funds skilled in both stock picking and market timing, we develop a test for the zero product of these two skills to first single out funds with at least one zero skill. Our simulations confirm the test's accurate size and good power. We apply our test to active U.S. equity mutual funds to exclude zero skill funds and classify remaining funds based on stock picking and market timing. We find that the 1% of funds with both skills are the only group with significant risk-adjusted performance. We also provide evidence for stock picking and market timing trade-offs along multiple dimensions.

Keywords and phrases: ARMA-GARCH model, bootstrap, heteroscedasticity, hypothesis testing, weighted inference

1. Introduction

Following seminal work by Jensen (1968), Treynor and Mazuy (1966), and Henriksson and Merton (1981), researchers and practitioners often evaluate mutual fund performance by measures of stock picking and market timing skills inferred from fund returns and common risk factors. On one hand, if such measures are rooted in funds' superior human capital,

then top funds should exhibit skills in both stock picking and market timing.[1] On the other hand, top funds may face trade-offs when applying the two types of skills as Kon (1983), Henriksson (1984), Jagannathan and Korajczyk (1986), and Goetzmann et al. (2000) empirically find a negative association between market timing and stock picking skill.[2] Thus, the negative empirical association between market timing and stock picking poses the question of identifying the mutual funds with both skills, if they exist at all.

Our approach is motivated by the well-known underperformance of the majority of active management industry relative to passive index benchmarks. Thus, most mutual funds should have zero skill in stock picking or market timing, and as a result, studies of average investment performance can be misleading when including zero skill funds. To address this issue, we propose a new approach that tests whether a fund has zero skill in either stock picking or market timing.

This hypothesis test is equivalent to a composite test for a zero product between the two skill parameters. We show such a test is non-trivial to implement since a naive test based on independently estimating either stock picking or market timing skill may theoretically fail. Instead, our proposed direct inference for the product of these two skills leads to a unified test regardless of whether both skills are zero or only one skill is zero. Our framework starts with a general factor model for both skills based on observed fund returns. To incorporate the econometric features of daily data, we model errors

[1] Back et al. (2018) also focus on the trade-offs faced by mutual funds.

[2] One economic explanation for this negative association is proposed by Kacperczyk et al. (2014, 2016), who argue that stock picking and market timing are not talents but tasks that trade off each other. They present evidence consistent with mutual fund managers allocating their time to focusing on either stock picking or market timing depending on economic conditions.

by a GARCH sequence to account for heteroscedasticity or an ARMA-GARCH sequence for serial correlation and heteroscedasticity.[3] We further develop a weighted inference to reduce the heavy tail effect of daily returns. Because the proposed inference avoids estimating the GARCH model, it is robust against heteroscedasticity and applicable to monthly returns. We quantify the inference uncertainty using a random weighted bootstrap method. Our simulation studies confirm our test's accurate size and good power across various settings.

Empirically, using our test, we quantify the prevalence of stock picking and market timing skills among all actively managed mutual funds in the United States in a formal econometric way. While Kacperczyk et al. (2014) find that the top 25% of managers exhibit stock picking and market timing skills at different times, our novel statistical test finds that the co-existence of both skills is far less prevalent at about 1% of the time. Overall, our proposed test and findings may prove to be a valuable aid for mutual fund investment allocation decisions.

The rest of the paper is organized as follows. Section 2 introduces the proposed methodologies. Section 3 reports on the simulation study results. Section 4 describes our data analysis and main findings with the other supporting results and an extension to correlated and heteroscedastic errors provided in the online supplementary file. Section 5 concludes. We put theoretical proofs in the appendix.

[3] ARMA-GARCH models have become standard in modeling heteroscedasticity since Engle (1982) and Bollerslev (1986).

2. Models, Tests, and Theoretical Results

Suppose Y_t is a fund's excess return[4] at time t, $\boldsymbol{X}_t = (X_{t,1}, \cdots, X_{t,d})^\tau$ represents common factors with $X_{t,1}$ being the market excess return, and A^τ denotes the transpose of the matrix or vector A. For evaluating fund performance, the literature employs the following model:

$$Y_t = \alpha + \boldsymbol{\beta}^\tau \boldsymbol{X}_t + \gamma H(X_{t,1}) + \varepsilon_t, \ t = 1, \cdots, n, \tag{2.1}$$

where α and γ measure a fund's stock picking and market timing skills, respectively, and H is a known function related to the market volatility. For example, Treynor and Mazuy (1966) use $H(X_{t,1}) = X_{t,1}^2$, Henriksson and Merton (1981) use $H(X_{t,1}) = \max(0, X_{t,1})$, Busse (1999) uses the conditional standard deviation of $X_{t,1}$ as $H(X_{t,1})$, and Goetzmann et al. (2000) use $H(X_{t,1})$ as a monthly quantity computed from daily returns when the above model is applied to monthly data. We refer readers to Bollen and Busse (2001) for a comparison study of these measures.

Previous literature, such as Carhart (1997), reports that most funds have zero skill in either stock picking or market timing. The inclusion of these zero skill funds in a study would then introduce noise or even bias into the process of identifying funds with stock picking and market timing skills and any analysis of fund skill trade-offs. The effects of estimation uncertainty suggest that excluding funds with at least one zero skill is important for more meaningfully evaluating mutual fund performance. To do so, we note that identifying and then excluding funds with at least one zero skill is equivalent to testing the composite null hypothesis,

$$H_0 : \alpha = 0 \text{ or } \gamma = 0. \tag{2.2}$$

[4]Excess returns are fund net returns minus the risk-free rate.

Put $\theta = \alpha\gamma$, then H_0 is equivalent to $H_0 : \theta = 0$. Therefore, one may use the naive estimator $\widehat{\theta}_{LSE} = \widehat{\alpha}_{LSE}\widehat{\gamma}_{LSE}$, where $\widehat{\alpha}_{LSE}$ and $\widehat{\gamma}_{LSE}$ are the least-squares estimators for model (2.1), i.e.,

$$(\widehat{\alpha}_{LSE}, \widehat{\boldsymbol{\beta}}_{LSE}^{\tau}, \widehat{\gamma}_{LSE})^{\tau} = \arg\min_{\alpha,\boldsymbol{\beta},\gamma} \sum_{t=1}^{n} \{Y_t - \alpha - \boldsymbol{\beta}^{\tau}\boldsymbol{X}_t - \gamma H(X_{t,1})\}^2.$$

However, this estimator's asymptotic limit depends on whether one skill or both are zero. When $\alpha = 0$ and $\gamma = 0$, $\widehat{\theta}_{LSE}$ not only has a convergence rate of n^{-1} rather than the standard $n^{-1/2}$ rate but also has a limiting distribution that is non-normal. Conversely, when only one of α or γ is zero, $\widehat{\theta}_{LSE}$ has the standard convergence rate with a normal limit. Thus, it is challenging to test H_0 based on $\widehat{\theta}_{LSE}$ without distinguishing between these two cases. This difficulty has been noticed by Nguyen and Jiang (2020) in a different context. To develop a test for H_0 with the asymptotically correct size, we propose to estimate θ directly by constructing a model with the parameter θ.

Put $Z_t = Y_t - \boldsymbol{\beta}^{\tau}\boldsymbol{X}_t$ for $t = 1, \cdots, n$. Then, model (2.1) implies that

$$Z_t^2 = \alpha^2 + E(\varepsilon_t^2) + 2\alpha\gamma H(X_{t,1}) + \gamma^2 H^2(X_{t,1}) + 2\varepsilon_t(\alpha + \gamma H(X_{t,1})) + \{\varepsilon_t^2 - E(\varepsilon_t^2)\},$$

which motivates directly estimating $\theta = \alpha\gamma$ by minimizing

$$\sum_{t=1}^{n} \{\widehat{Z}_{t,LSE}^2 - \alpha^* - \theta 2H(X_{t,1}) - \gamma^* H^2(X_{t,1})\}^2, \tag{2.3}$$

where $\alpha^* = \alpha^2 + E(\varepsilon_t^2)$, $\gamma^* = \gamma^2$, and

$$\widehat{Z}_{t,LSE} = Y_t - \widehat{\boldsymbol{\beta}}_{LSE}^{\tau}\boldsymbol{X}_t \text{ for } t = 1, \cdots, n.$$

We note that $\varepsilon_t^2 - E(\varepsilon_t^2) = \sigma_t^2(\eta_t^2 - 1) + \sigma_t^2 - E(\sigma_t^2)$, which means minimizing (2.3) can lead to an inconsistent inference if $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\sigma_t^2 - E(\sigma_t^2)\}$ does not converge in distribution due to a lack of finite moments.

To avoid the higher moments of Z_t^2, we propose to split the data into two parts and use a product to directly estimate θ by noting that

$$\begin{aligned}\varepsilon_t\varepsilon_{t+m} &= \{Z_t - \alpha - \gamma H(X_{t,1})\}\{Z_{t+m} - \alpha - \gamma H(X_{t+m,1})\} \\ &= Z_tZ_{t+m} - (\alpha + \gamma H(X_{t,1}))Z_{t+m} - (\alpha + \gamma H(X_{t+m,1}))Z_t \\ &\quad +\alpha_1 + \theta(H(X_{t,1}) + H(X_{t+m,1})) + \gamma_1 H(X_{t,1})H(X_{t+m,1})\end{aligned}$$

for $t = 1, \cdots, m$, where $m = [n/2]$, $\alpha_1 = \alpha^2$, and $\gamma_1 = \gamma^2$. Unfortunately, when heteroscedasticity exists, the asymptotic normality of the above inference requires $E(\sigma_t^4\bar{\sigma}_{t,1}^4) < \infty$, which may need $E(\varepsilon_t^8) < \infty$ and $E(X_{t,1}^8) < \infty$. To get rid of these higher finite moment requirements caused by heteroscedasticity, we propose to employ the following weighted inference via modeling the risk factors by the following ARMA-GARCH models:

$$\begin{cases} X_{t,l} = \mu_l + \sum_{i=1}^{s_l} \phi_{i,l}X_{t-i,l} + \sum_{j=1}^{r_l} \psi_{j,l}\bar{\varepsilon}_{t-j,l} + \bar{\varepsilon}_{t,l},\ \bar{\varepsilon}_{t,l} = \bar{\eta}_{t,l}\bar{\sigma}_{t,l}, \\ \bar{\sigma}_{t,l}^2 = w_l + \sum_{i=1}^{p_l} a_{i,l}\bar{\varepsilon}_{t-i,l}^2 + \sum_{j=1}^{q_l} b_{j,l}\bar{\sigma}_{t-j,l}^2,\ l = 1, \cdots, d, \end{cases} \tag{2.4}$$

and assuming that the regression errors follow the following GARCH model:

$$\varepsilon_t = \eta_t\sigma_t,\ \sigma_t^2 = w + \sum_{i=1}^{p} a_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q} b_j\sigma_{t-j}^2, \tag{2.5}$$

where $\{(\eta_t, \bar{\eta}_{t,1}, \cdots, \bar{\eta}_{t,d})^\top\}_{t=1}^n$ is a sequence of independent and identically distributed random vectors with means zero and variances one.

First, we estimate $\alpha, \boldsymbol{\beta}$, and γ by weighted least squares:

$$(\widehat{\alpha}_{WLSE}, \widehat{\boldsymbol{\beta}}_{WLSE}, \widehat{\gamma}_{WLSE})^\tau = \arg\min_{\alpha,\boldsymbol{\beta},\gamma} \sum_{t=1}^{n} \{Y_t - \alpha - \boldsymbol{\beta}^\tau \boldsymbol{X}_t - \gamma H(X_{t,1})\}^2 w_{t,1},$$

where

$$w_{t,1}^{-1} = 1 + \max\{||\boldsymbol{X}_t||, H(X_{t,1})\} \text{ with } ||\boldsymbol{X}_t|| = \max_{1\le i\le d} |X_{t,i}|. \tag{2.6}$$

Next, we define $\widehat{Z}_{t,WLSE} = Y_t - \widehat{\boldsymbol{\beta}}^{\tau}_{WLSE}\boldsymbol{X}_t$ for $t = 1, \cdots, n$, and estimate θ by

$$\begin{aligned}
&(\widehat{\alpha}_{1,w}, \widehat{\theta}_w, \widehat{\gamma}_{1,w})^{\tau} \\
= \;& \arg\min_{\alpha_1,\theta,\gamma_1} \textstyle\sum_{t=1}^{m} \{\widehat{Z}_{t,WLSE}\widehat{Z}_{t+m,WLSE} - (\widehat{\alpha}_{WLSE} + \widehat{\gamma}_{WLSE}H(X_{t,1}))\widehat{Z}_{t+m,WLSE} \\
& -(\widehat{\alpha}_{WLSE} + \widehat{\gamma}_{WLSE}H(X_{t+m,1}))\widehat{Z}_{t,WLSE} + \alpha_1 + \theta(H(X_{t,1}) + H(X_{t+m,1})) \\
& +\gamma_1 H(X_{t,1})H(X_{t+m,1})\}^2 w_{t,2},
\end{aligned}$$

where

$$w_{t,2}^{-1} = 1 + \max\{|Y_t|, ||\boldsymbol{X}_t||, ||\boldsymbol{X}_{t+m}||, H(X_{t,1}), H(X_{t+m,1}), H(X_{t,1})H(X_{t+m,1})\}. \tag{2.7}$$

Like Ling (2007), the weight functions of (2.6) and (2.7) are used to reduce the heavy tail effect due to heteroscedasticity and bound the factors in the score equations to ensure a normal limit when $E(\eta_t^2) < \infty$. There are many different choices of weight functions, but our simulation study confirms the good finite sample performance of using (2.6) and (2.7).

To establish the asymptotic behavior of the estimator, we employ the following regularity conditions.

- C1) $\{\varepsilon_t\}$ and $\{\boldsymbol{X}_t\}$ are strictly stationary and ergodic with finite variance; See conditions in Theorem 3.1 of Basrak et al. (2002).
- C2) $\{(\eta_t, \bar{\eta}_{t,1}, \cdots, \bar{\eta}_{t,d})^{\top}\}$ is a sequence of independent and identically distributed random vectors with means zero and variances one.
- C3) Assume

$$E(\eta_t|\bar{\eta}_{t,1}, \cdots, \bar{\eta}_{t,d}) = 0 \tag{2.8}$$

  and there exists $\delta > 0$ such that $E|\eta_t|^{2+\delta} < \infty$.

- C4) Assume the covariance matrices of $\{w_{t,1}\varepsilon_t(1, \boldsymbol{X}_t^\tau, H(X_{t,1}))^\tau\}_{t=1}^n$ and

$$\{w_{t,2}\varepsilon_t\varepsilon_{t+m}(1, H(X_{t,1}) + H(X_{t+m,1}), H(X_{t,1})H(X_{t+m,1}))^\tau\}_{t=1}^m$$

  are positive definite.

Theorem 1. *Suppose models (2.1), (2.4), and (2.5) hold with conditions C1)–C4). Then, as $n \to \infty$, $\sqrt{n}(\widehat{\theta}_w - \theta) \xrightarrow{d} N(0, \sigma_0^2)$, where σ_0^2 has a complicated formula given in the proof.*

To avoid estimating the complicated σ_0^2, we adopt the random weighted bootstrap method from Jin et al. (2001) and Zhu (2016) as follows.[5]

- Step Ai) Draw a random sample of size n from the standard exponential distribution. Denote these random draws by $\xi_1^b, \cdots, \xi_n^b$.

- Step Aii) Compute

$$(\widehat{\alpha}_{WLSE}^b, \widehat{\boldsymbol{\beta}}_{WLSE}^{b\tau}, \widehat{\gamma}_{WLSE}^b)^\tau = \arg\min_{(\alpha, \boldsymbol{\beta}^\tau, \gamma)^\tau} \sum_{t=1}^n \xi_t^b \{Y_t - \alpha - \boldsymbol{\beta}^\tau \boldsymbol{X}_t - \gamma H(X_{t,1})\}^2 w_{t,1}.$$

- Step Aiii) Define

$$\widehat{Z}_{t,WLSE}^b = Y_t - \widehat{\boldsymbol{\beta}}_{WLSE}^{b\tau} \boldsymbol{X}_t, \ t = 1, \cdots, n,$$

  and calculate

$$\begin{aligned}
&(\widehat{\alpha}_{1,w}^b, \widehat{\theta}_w^b, \widehat{\gamma}_{1,w}^b)^\tau \\
= \ & \arg\min_{\alpha_1, \theta, \gamma_1} \textstyle\sum_{t=1}^m \xi_{t+m}^b \{\widehat{Z}_{t,WLSE}^b \widehat{Z}_{t+m,WLSE}^b - (\widehat{\alpha}_{WLSE}^b + \widehat{\gamma}_{WLSE}^b H(X_{t,1}))\widehat{Z}_{t+m,WLSE}^b \\
& -(\widehat{\alpha}_{WLSE}^b + \widehat{\gamma}_{WLSE}^b H(X_{t+m,1}))\widehat{Z}_{t,WLSE}^b + \alpha_1 + \theta(H(X_{t,1}) + H(X_{t+m,1})) \\
& +\gamma_1 H(X_{t,1})H(X_{t+m,1})\}^2 w_{t,2},
\end{aligned}$$

[5] We comment that the conventional residual-based bootstrap method (see Hall (1992)) is not applicable in our approach as we do not infer the GARCH model of regression errors.

- Step Aiv) Repeat the above three steps B times to get $\{\widehat{\theta}_w^b\}_{b=1}^B$ and estimate the asymptotic variance of $\widehat{\theta}_w$ by

$$\widehat{\sigma}_0^2 = \frac{n}{B}\sum_{b=1}^{B}(\widehat{\theta}_w^b - \widehat{\theta}_w)^2.$$

Theorem 2. *Under the conditions of Theorem 1, $\widehat{\sigma}_0^2/\sigma_0^2$ converges in probability to one as $B \to \infty$ and then $n \to \infty$.*

Using Theorems 1 and 2, we reject the null hypothesis of (2.2) at level a if $\widehat{\theta}_w^2/\widehat{\sigma}_0^2 > \chi^2_{1,1-a}$, where $\chi^2_{1,1-a}$ is the $(1-a)$-th quantile of a chi-squared distribution with one degree of freedom. The online supplement generalizes this method to correlated and heteroscedastic ε_t's. Alternatively, we can compute the pvalue for testing H_0 in (2.2) by $\frac{1}{B}\sum_{b=1}^{B} I(|\widehat{\theta}_w| < |\widehat{\theta}_w^b - \widehat{\theta}_w|)$, which leads to the asymptotically correct size by taking $B \to \infty$ and then $n \to \infty$.

3. Simulation Study

In this section, we investigate the finite sample performance of the proposed test in terms of size and power. To mimic the results of realistic mutual fund investing, we simulate fund returns under a factor model fitted to the empirical features of the mutual funds in the dataset that we study in Section 4. We then analyze the test's performance for different simulated settings depending on stock picking skill (α), market timing skill (γ), sample size, and data-generating process. As a benchmark, we also compare the performance of our proposed estimator with that of the naive ordinary least squares estimator $\widehat{\theta}_{LSE} = \widehat{\alpha}_{LSE}\widehat{\gamma}_{LSE}$. To be more comparable, we employ a similar random weighted bootstrap method to estimate the asymptotic variance of the naive estimator $\widehat{\theta}_{LSE}$.

We draw random samples from the following four-factor model:

$$Y_t = \alpha + \boldsymbol{\beta}^\tau \boldsymbol{X}_t + \gamma H(X_{t,1}) + \varepsilon_t, \ t = 1, \cdots, n, \tag{3.9}$$

where $\boldsymbol{X}_t = (X_{t,1}, \cdots, X_{t,4})^\tau$ represents the four factors from Carhart (1997): the market excess return (MKT), size (SMB), book-to-market (HML), and momentum (UMD) factors, respectively, and $H(X_{t,1}) = X_{t,1}^2$, as defined by Treynor and Mazuy (1966). We set $\boldsymbol{\beta} = (0.9757290, -0.1010977, 0.1064889, -0.2045018)^\tau$ based on the vector of empirical regression estimates of the four-factor model for a representative fund in our dataset. We set the true stock picking (α) and market timing (γ) parameters to be either 0, 0.01, or 0.05.

We model the factors $X_{t,1}, \cdots, X_{t,4}$ by independent AR(1)-GARCH(1,1) processes. We generate ε_t's in (3.9) independently from these four factors using three different scenarios: a sequence of independent random variables with normal distributions, a GARCH(1,1) process, and an AR(1)-GARCH(1,1) process. To make our simulation more realistic, the coefficients of these models are obtained from actual data. Specifically, we use the 'fGarch' R package to fit an ARMA(1,0)-GARCH(1,1) model to each of the four factors and the residuals from model (3.9) using a representative fund in our dataset from September 1, 1998, to December 31, 2018. Tables 1 and 2 below summarize the coefficients of the four factors $X_{t,1}, \cdots, X_{t,4}$ and errors ε_t in the three scenarios, respectively.

We conduct the hypothesis test of $H_0 : \theta = 0$ (i.e., $H_0 : \alpha\gamma = 0$) at the 10% significance level. Using 1000 repetitions and $B = 1000$ bootstrap iterations for the random weighted bootstrap method, we compute and compare the simulated size and power of the hypothesis tests using our proposed estimator $\widehat{\theta}_w$ and the naive estimator $\widehat{\theta}_{LSE}$. Tables 3 and 4 report on the size and power, respectively, and with the following

Table 1: **ARMA-GARCH Coefficients for the Four Factors**

	μ	ϕ	ω	a	b
$X_{t,1}$	0.10467535	-0.07631574	0.04075129	0.14544129	0.7657024
$X_{t,2}$	-0.00238095	0.01933178	0.01296749	0.06118342	0.8908670
$X_{t,3}$	-0.04445163	0.00072542	0.02254173	0.09856252	0.7906360
$X_{t,4}$	0.05034941	0.01640715	0.02024227	0.14370110	0.8036767

Table 2: **Coefficients for Error**

ε_t	μ	ϕ	ω	a	b
Scenario 1: i.i.d. $N(0, 0.1)$	0	0	0	0	0
Scenario 2: GARCH(1,1)	0	0	0.03155241	0.11454071	0.61573066
Scenario 3: AR(1)-GARCH(1,1)	0	-0.08733819	0.03155241	0.11454071	0.61573066

observations.

- i) The test using the naive estimator $\widehat{\theta}_{LSE}$ generally has distorted size, consistent with its asymptotic limit being a non-normal distribution when both α and γ are close to zero. The test size is generally below 0.01 across all our simulation settings, with the sole exception of when the sample size is large at 1000, and α is non-zero at 0.05.

- ii) The test using the proposed estimator $\widehat{\theta}_w$ has the accurate size for all the cases we consider. The proposed estimator is also a meaningful improvement relative to the naive estimator.

- iii) The power of the test using the proposed estimator $\widehat{\theta}_w$ increases as the sample

size becomes larger or when α and γ are bigger than zero. The test under Scenario 1 has much higher power than under the other two scenarios.

- iv) Our approach of splitting the data impacts the test's power when the sample size is small.

In summary, it is challenging to test $H_0 : \alpha = 0$ or $\gamma = 0$ as exemplified by the naive ordinary least squares estimator. The proposed technique of splitting data to test the product of skill parameters provides a test with accurate size and good power. However, it does impact test power when the sample size is small.

4. Data Analysis

This section applies our test to identifying mutual funds with stock picking and/or market timing skill. We start by describing our dataset of actively managed equity mutual funds. Then, we apply our test to exclude zero skill funds from the sample and classify the remaining funds into various skill groups. We use these classifications to examine each skill group's prevalence and returns and if there are skill trade-offs.

4.1 Data and Implementation of Test

We obtain U.S. open-end mutual fund returns and their characteristics from CRSP (the Center for Research in Security Prices) Survivor-Bias-Free US Mutual Fund Database. Funds' daily and monthly returns are value-weighted averages across all fund share classes (using total net assets of different share classes as the weight). We collect the risk-free rate and risk factor data from the Ken French data library.

The actively managed mutual funds sample is constructed following Kacperczyk et al.

Table 3: **Simulation Study for Comparing Test Size**

This table reports the results of our simulation study comparing the test sizes of the proposed estimator and the naive estimator at level 10%.

Case	α	γ	Sample Size	Method	Scenario 1	Scenario 2	Scenario 3
One	0.00	0.00	100	Naive	0.002	0.001	0.002
			100	New	0.107	0.114	0.119
			200	Naive	0.000	0.002	0.001
			200	New	0.106	0.088	0.094
			1000	Naive	0.000	0.000	0.000
			1000	New	0.107	0.101	0.116
Two	0.00	0.01	100	Naive	0.003	0.002	0.002
			100	New	0.130	0.083	0.109
			200	Naive	0.000	0.004	0.000
			200	New	0.100	0.110	0.099
			1000	Naive	0.002	0.003	0.000
			1000	New	0.094	0.108	0.118
Three	0.00	0.05	100	Naive	0.001	0.007	0.003
			100	New	0.120	0.101	0.091
			200	Naive	0.000	0.004	0.007
			200	New	0.099	0.113	0.109
			1000	Naive	0.002	0.051	0.057
			1000	New	0.099	0.086	0.104
Four	0.01	0.00	100	Naive	0.000	0.003	0.003
			100	New	0.110	0.083	0.107
			200	Naive	0.000	0.004	0.001
			200	New	0.104	0.093	0.098
			1000	Naive	0.000	0.002	0.002
			1000	New	0.097	0.111	0.120
Five	0.05	0.00	100	Naive	0.001	0.009	0.008
			100	New	0.094	0.099	0.100
			200	Naive	0.002	0.014	0.015
			200	New	0.090	0.116	0.093
			1000	Naive	0.002	0.086	0.084
			1000	New	0.093	0.101	0.110

Table 4: **Simulation Study for Comparing Test Power**

This table reports the results of our simulation study comparing the test powers of the proposed estimator at the level 10%.

Case	α	γ	Sample Size	Scenario 1	Scenario 2	Scenario 3
One	0.05	0.05	100	0.292	0.095	0.099
			200	0.479	0.110	0.113
			1000	0.962	0.125	0.127
Two	0.05	0.10	100	0.580	0.113	0.097
			200	0.829	0.103	0.118
			1000	1.000	0.165	0.201
Three	0.10	0.05	100	0.573	0.125	0.103
			200	0.791	0.130	0.109
			1000	1.000	0.171	0.210
Four	0.10	0.10	100	0.893	0.126	0.130
			200	0.987	0.143	0.156
			1000	1.000	0.366	0.394
Five	0.20	0.20	100	1.000	0.442	0.451
			200	1.000	0.631	0.696
			1000	1.000	0.998	0.993

(2008). We begin by using the investment objective codes from CRSP. We exclude ETFs, annuities, and index funds based on their indicator variables or fund names from CRSP following Busse et al. (2021). Since we focus on equity funds, we require 80% of the assets under management to be invested in common stocks. We restrict our sample to funds of at least one year old and have at least $15 million in assets under management. We address incubation bias as in Evans (2010). Our final sample includes 3,569 U.S. actively managed domestic equity funds from January 1980 to December 2018.

To test the null hypothesis of zero skill, $H_0 : \alpha = 0$ or $\gamma = 0$, we use daily data available from 1998 to 2018 to fit model (2.1) for each fund. To be consistent with our simulation study, we estimate (2.1) based on the four-factor specification from Carhart (1997) that includes the daily market excess return (MKT), size (SMB), value (HML), and momentum (UMD) factors.[6] Our tests are based on the AR-GARCH model, where we use the AIC model criterion to select the best AR model. Then, we utilize 1000 bootstrap iterations for each fund to compute p-values against the null hypothesis.

To create our mutual fund skill classifications, at the 10% level, we first sort funds with either zero stock picking or market timing skill based on a failure to reject the null into a benchmark zero skill group. Then, among funds with either non-zero picking or market timing skill, we classify funds into those that have positive stock picking ($\widehat{\alpha} > 0$) and market timing skill ($\widehat{\gamma} > 0$), the two groups of funds with only one of either positive stock picking or market timing skill, and a group that has negative stock picking and market timing skills. Our estimates of stock picking and market timing come from the weighted least squares estimation as in Theorem S1 with the weight function described

[6]We also run our tests using the CAPM one-factor model as in Jensen (1968). We find qualitatively similar results, with a slightly lower number of funds with both stock picking and market timing skill.

in (S2) in the online supplement. Lastly, we note that, although we use daily data to implement the test, the conclusions apply at the fund level, some of which have monthly returns back to 1980. So to minimize survivorship bias, we also put funds that failed before 1998 into the zero skill group.

4.2 Empirical Results

The first line of Table 5 reports the results of our classification of funds into mutually exclusive skill groups. At the 10% level, we find that 3,159 out of 3,569 funds have zero skill in either stock picking or market timing. Conversely, a very small subset of 36 funds has positive stock picking and market timing skills. We also find that a larger group of funds have one skill but not the other: 120 funds have positive (negative) stock picking (market timing) skill, and 146 funds have the positive (negative) market timing (stock picking) skill. Lastly, 108 funds have neither stock picking nor market timing skill. As a result, funds that possess both stock picking and market timing abilities are rare, occurring only 1.0% of the time. This is in contrast to the finding of Kacperczyk et al. (2014) that the top 25% of managers exhibit both abilities.

Table 5 displays the returns distribution for each skill group, computed by summarizing the equally-weighted average monthly returns of all funds in each group. We expect funds with stock picking skills to have better performance based on risk-return trade-offs. Indeed, funds with both skills have the highest Sharpe ratio at 0.69, followed by pure stock picking funds at 0.55, zero skill funds at 0.52, neither skill group at 0.51, and pure market timing funds at 0.48. Funds with market timing skills should adjust their market exposure during market expansions and downturns. We find that funds in the timing skill

group reduce their overall market beta from 1.00 to 0.96 during bear market states.[7] In contrast, the other four skill groups have higher market beta during bear market states. We further find that pure market timing funds have the lowest volatility (i.e., standard deviation) during bear market states, the least negative return skewness, and the smallest kurtosis among all the skill groups. Hence, pure market timing funds appear to manage downside risks as well.

Given that our novel test does indeed appear to classify funds well based on their skill, do there appear to be stock picking and market timing trade-offs? We approach this question from three perspectives. First, comparing pure stock picking and pure market timing funds in the third and fourth columns of Table 5, we do observe such a trade-off. Pure market timing sacrifices the higher risk-return profile of pure stock picking to have a better market and downside risk management. Furthermore, these two skill groups do appear to generate different types of value for investors as pure stock picking has higher Sharpe ratios than zero skill funds. In contrast, pure market timing does manage risk during downturns more than zero skill funds do.

Second, our evidence also suggests that funds with both skills favor utilizing stock picking over market timing. We observe that, for funds with stock picking and market timing skills, their return distributions are similar to that of the pure stock picking funds. The funds with both skills generate by far the best Sharpe ratios while failing to scale back their market exposure (i.e., market beta) during bear market states. They also incur the most negative skewness and the highest positive kurtosis, meaning they are more exposed to heavy-tailed outcomes. At the same time, relative to funds in the pure stock picking, neither skill, and zero skill groups, funds with both skills do appear to have some market

[7] We define bear market states as the 10% of months with the lowest market return.

timing skill. Funds with both skills have the overall lowest volatility. Compared to the pure market timing funds, they also have lower market exposure during bear market states than the other groups have. Finally, they manage downside risks a bit better than pure stock-pickers do, as their volatility during bear market states is lower.

Third, we do not find stock picking and market timing trade-offs for funds with negative stock picking and market timing skills. While this neither group has a risk profile in terms of volatility and market beta similar to the pure stock picking group, they do not generate similarly high returns. They also fail to decrease market exposure and overall volatility during bear states. Finally, they exhibit higher tail risks, with the most negative skewness and return during bear markets.

So, why then is stock picking so heavily weighted over market timing by funds with both skills? The answer appears to be that funds with both skills particularly excel at stock picking. Among all four groups, they generate the highest return at 9.81% per year, with the lowest volatility at 14.29% per year. To further delve into this stock picking ability, we explore whether standard factor exposures can explain this risk-return profile. In Table 6, we regress the average monthly excess return of each skill group on either 1) the market excess return (MKT) or 2) the market excess return plus the size (SMB), value (HML), and momentum (UMD) factors, following Carhart (1997). Because both the independent and dependent variables in these regressions are returns, we can interpret the constant, i.e., alpha, as the average abnormal return unexplained by factor exposures. At the 1% level, we find funds with both skills are the only group generating statistically significant and positive alphas, earning 3.0% or 2.7% per year relative to the market or Carhart factor models. In sharp contrast, the alphas of the other four groups are

insignificant.

Overall, our novel test points toward a meaningful classification of funds based on stock picking and market timing skill. We can then identify funds with either stock picking or market timing skill relative to zero skill funds. Importantly, these are the only funds with the ability to generate particularly attractive risk-adjusted returns. Finally, our online supplement reports additional analyses to validate these skill classifications based on each group's active management characteristics and stock-holding styles.

5. Conclusions

This paper introduces a novel statistical test for at least one zero skill in stock picking and market timing, because direct inference for the product of these two skills is infeasible. Applying the developed test to the universe of actively managed U.S. mutual funds, we are able to exclude zero skill funds and find among the remaining skilled funds that there are clear trade-offs between stock picking and market timing along multiple dimensions related to a fund's risk-return profile, market timing, active management, and stock holding style. Importantly, we find only 1% of funds can optimize these trade-offs and be simultaneously skilled. These funds are the only group of funds that generate abnormal risk-adjusted returns at around 3% per year while also managing their market risk exposures.

Table 5: **Risk-Return Summary Statistics**

This table presents summary statistics for the return distribution of the gross monthly excess returns over the risk-free rate for active funds in each skill group. At the 10% level among funds rejecting $H_0 : \alpha = 0$ or $\gamma = 0$, the "Both" group comprises funds with positive picking ($\alpha > 0$) and timing ($\gamma > 0$) skills, the "Picking" group comprises funds with positive picking and negative timing skills, the "Timing" group comprises funds with positive timing and negative picking skills, and the "Neither" group comprises funds with negative picking and timing skills. The "Zero Skill" group comprises all other funds that fail to reject H_0. The mean monthly return is annualized by multiplying by 12, the standard deviation is annualized by multiplying by $\sqrt{12}$, and the Sharpe ratio is annualized by multiplying by $\sqrt{12}$. Beta is the coefficient from an ordinary least squares regression of the monthly excess return on the market excess return. Bear market states are defined as months when the market return is in its lowest decile during the sample period. The sample period is from 1980 to 2018.

	Zero Skill	Both	Picking	Timing	Neither
Number of Funds	3159	36	120	146	108
Annualized Mean Return (%)	8.09	9.81	9.20	7.42	8.31
Annualized Std. Dev. (%)	15.64	14.29	16.78	15.45	16.34
Annualized Sharpe Ratio	0.52	0.69	0.55	0.48	0.51
Skewness	-0.80	-0.87	-0.75	-0.68	-0.88
Kurtosis	5.62	6.15	5.82	4.90	5.67
Beta	1.01	0.89	1.05	1.00	1.04
Worst Monthly Return (%)	-24.03	-21.55	-25.57	-21.53	-25.23
% of Months w/ Negative Return	38.68	37.18	39.53	38.89	38.89
Annualized Mean Return in Bear Markets (%)	-97.34	-84.54	-100.12	-96.59	-102.21
Annualized Std. Dev. in Bear Markets (%)	13.08	13.75	14.83	11.89	14.14
Beta in Bear Markets	1.05	1.01	1.14	0.96	1.12

Table 6: **Factor Exposures of Different Funds**

This table presents the factor exposures of the gross monthly excess returns over the risk-free rate for active funds in each skill group. At the 10% level among funds rejecting $H_0 : \alpha = 0$ or $\gamma = 0$, the "Both" group comprises funds with positive picking ($\alpha > 0$) and timing ($\gamma > 0$) skills, the "Picking" group comprises funds with positive picking and negative timing skills, the "Timing" group comprises funds with positive timing and negative picking skills, and the "Neither" group comprises funds with negative picking and timing skills. The "Zero Skill" group comprises all other funds that fail to reject H_0. For each skill group, we compute the equally-weighted monthly return (R_t) each month, which is then, in a time-series regression, regressed on the contemporaneous 1) market excess return (MKT_t) or 2) the market excess return, size (SMB_t), value (HML_t) and momentum (UMD_t) factors. All returns are annualized by multiplying by 12. In the parentheses below the coefficient estimates, we report Newey and West (1987) t-statistics with 12 lags. The sample period is from 1980 to 2018, with N representing the number of months. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Zero Skill		Both		Picking		Timing		Neither	
					Dep Var: R_t					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MKT_t	1.008***	0.980***	0.887***	0.885***	1.052***	1.000***	0.997***	0.972***	1.040***	1.007***
	(96.923)	(105.322)	(30.475)	(46.793)	(61.618)	(70.425)	(87.595)	(74.150)	(0.012)	(0.009)
SMB_t		0.194***		0.136***		0.275***		0.169***		0.241***
		(10.563)		(4.850)		(9.064)		(10.456)		(0.014)
HML_t		0.002		0.127***		−0.064		−0.004		−0.007
		(0.088)		(4.314)		(−1.339)		(−0.201)		(0.015)
UMD_t		0.008		−0.035**		0.020		0.012		0.032***
		(0.587)		(−2.448)		(0.847)		(0.989)		(0.009)
Constant	0.327	0.263	2.982***	2.701***	1.100	1.244	−0.264	−0.332	0.297	0.080
	(0.556)	(0.624)	(2.963)	(3.379)	(1.026)	(1.533)	(−0.609)	(−0.861)	(0.625)	(0.483)
N	468	468	468	468	468	468	468	468	468	468
R^2	0.967	0.983	0.897	0.913	0.915	0.948	0.970	0.983	0.944	0.969

Appendix: Theoretical Proofs

Define $\mathcal{F}_t$ as the σ-field generated by $\{\eta_u, \bar{\eta}_{v,1}, \cdots, \bar{\eta}_{v,d} : u \leq t, v \leq t+1\}$, $\mathbf{0}_d$ as the d-dimensional zero vector, and put

$$\boldsymbol{W}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} w_{t,1} \varepsilon_t (1, \boldsymbol{X}_t^\tau, H(X_{t,1}))^\tau,$$

$$\boldsymbol{\Gamma}_n = \frac{1}{n} \sum_{t=1}^{n} w_{t,1} \begin{pmatrix} 1 & \boldsymbol{X}_t^\tau & H(X_{t,1}) \\ \boldsymbol{X}_t & \boldsymbol{X}_t \boldsymbol{X}_t^\tau & \boldsymbol{X}_t H(X_{t,1}) \\ H(X_{t,1}) & \boldsymbol{X}_t^\tau H(X_{t,1}) & H^2(X_{t,1}) \end{pmatrix},$$

$$\boldsymbol{\Gamma} = \begin{pmatrix} E(w_{t,1}) & E(w_{t,1}\boldsymbol{X}_t^\tau) & E(w_{t,1}H(X_{t,1})) \\ E(w_{t,1}\boldsymbol{X}_t) & E(w_{t,1}\boldsymbol{X}_t\boldsymbol{X}_t^\tau) & E(w_{t,1}\boldsymbol{X}_t H(X_{t,1})) \\ E(w_{t,1}H(X_{t,1})) & E(w_{t,1}\boldsymbol{X}_t^\tau H(X_{t,1})) & E(w_{t,1}H^2(X_{t,1})) \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} E(w_{t,1}^2\varepsilon_t^2) & E(w_{t,1}^2\varepsilon_t^2\boldsymbol{X}_t^\tau) & E(w_{t,1}^2\varepsilon_t^2 H(X_{t,1})) \\ E(w_{t,1}^2\varepsilon_t^2\boldsymbol{X}_t) & E(w_{t,1}^2\varepsilon_t^2\boldsymbol{X}_t\boldsymbol{X}_t^\tau) & E(w_{t,1}^2\varepsilon_t^2\boldsymbol{X}_t H(X_{t,1})) \\ E(w_{t,1}^2\varepsilon_t^2 H(X_{t,1})) & E(w_{t,1}^2\varepsilon_t^2\boldsymbol{X}_t^\tau H(X_{t,1})) & E(w_{t,1}^2\varepsilon_t^2 H^2(X_{t,1})) \end{pmatrix},$$

$$H_t = H(X_{t,1}) + H(X_{t+m,1}), \ \widetilde{H}_t = H(X_{t,1})H(X_{t+m,1}),$$

$$\widetilde{\boldsymbol{W}}_m = \frac{1}{\sqrt{m}} \sum_{t=1}^{m} w_{t,2} \varepsilon_t \varepsilon_{t+m} (1, H_t, \widetilde{H}_t)^\tau,$$

$$\widetilde{\boldsymbol{\Gamma}}_m = \frac{1}{m} \sum_{t=1}^{m} w_{t,2} \begin{pmatrix} 1 & H_t & \widetilde{H}_t \\ H_t & H_t^2 & \widetilde{H}_t H_t \\ \widetilde{H}_t & H_t\widetilde{H}_t & \widetilde{H}_t^2 \end{pmatrix},$$

$$\widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2) & E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2 H_t) & E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2\widetilde{H}_t) \\ E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2 H_t) & E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2 H_t^2) & E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2 H_t\widetilde{H}_t) \\ E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2\widetilde{H}_t) & E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2\widetilde{H}_t H_t) & E(w_{t,2}^2\varepsilon_t^2\varepsilon_{t+m}^2\widetilde{H}_t^2) \end{pmatrix},$$

$$\widetilde{\boldsymbol{\Gamma}} = \begin{pmatrix} E(w_{t,2}) & E(w_{t,2}H_t) & E(w_{t,2}\widetilde{H}_t) \\ E(w_{t,2}H_t) & E(w_{t,2}H_t^2) & E(w_{t,2}\widetilde{H}_tH_t) \\ E(w_{t,2}\widetilde{H}_t) & E(w_{t,2}\widetilde{H}_tH_t) & E(w_{t,2}\widetilde{H}_t^2) \end{pmatrix},$$

$$\widetilde{\boldsymbol{S}}_1 = E\{w_{t,2}\varepsilon_t(1, \boldsymbol{X}_t^\tau, H(X_{t,1}))^\tau(1, H_t, \widetilde{H}_t)^\tau\},$$

$$\widetilde{\boldsymbol{S}}_2 = E\{w_{t,2}(2\alpha + \gamma H_t, \mathbf{0}_d^\tau, \alpha H_t + 2\gamma\widetilde{H}_t)^\tau(1, H_t, \widetilde{H}_t)^\tau\}.$$

Throughout, we use $\xrightarrow{p}$ and $\xrightarrow{d}$ to denote the convergence in probability and in distribution, respectively.

Lemma 1. *Define* $\boldsymbol{\xi} = (\alpha, \boldsymbol{\beta}^\tau, \gamma)^\tau$ *and* $\widehat{\boldsymbol{\xi}} = (\widehat{\alpha}_{WLSE}, \widehat{\boldsymbol{\beta}}_{WLSE}^\tau, \widehat{\gamma}_{WLSE})^\tau$. *Under conditions of Theorem 1, as* $n \to \infty$, *we have*

$$\boldsymbol{\Gamma}_n \xrightarrow{p} \boldsymbol{\Gamma}, \ \boldsymbol{W}_n \xrightarrow{d} N(\mathbf{0}_{d+2}, \boldsymbol{\Sigma}), \ \sqrt{n}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = -\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_n + o_p(1). \tag{5.10}$$

Proof. It follows from ergodicity of $\{\boldsymbol{X}_t\}$ that $\boldsymbol{\Gamma}_n \xrightarrow{p} \boldsymbol{\Gamma}$ as $n \to \infty$. Use (2.8) and the fact that $\{(\eta_t, \bar{\eta}_{t,1}, \cdots, \bar{\eta}_{t,d})^\tau\}$ is a sequence of independent and identically distributed random variables, we have $E(\boldsymbol{W}_t|\mathcal{F}_{t-1}) = \mathbf{0}_{d+2}$. Hence, it follows from the central limit theorem for martingale differences in Hall and Heyde (1980) that $\boldsymbol{W}_n \xrightarrow{d} N(\mathbf{0}_{d+2}, \boldsymbol{\Sigma})$. Because $\sqrt{n}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = -\boldsymbol{\Gamma}_n^{-1}\boldsymbol{W}_n$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = -\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_n + o_p(1).$$

□

Lemma 2. *Define* $\boldsymbol{\xi}_w = (\alpha_1, \theta, \gamma_1)^\tau$ *and* $\widehat{\boldsymbol{\xi}}_w = (\widehat{\alpha}_{1,w}, \widehat{\theta}_w, \widehat{\gamma}_{1,w})^\tau$. *Under conditions of Theorem 1,* $\alpha_1 = \alpha^2$, $\theta = \alpha\gamma$, *and* $\gamma_1 = \gamma^2$, *as* $n \to \infty$, *we have*

$$\widetilde{\boldsymbol{\Gamma}}_m \xrightarrow{p} \widetilde{\boldsymbol{\Gamma}}, \ \widetilde{\boldsymbol{W}}_m \xrightarrow{d} N(\mathbf{0}_3, \widetilde{\boldsymbol{\Sigma}}), \tag{5.11}$$

$$\sqrt{m}(\widehat{\boldsymbol{\xi}}_w - \boldsymbol{\xi}_w) = -\widetilde{\boldsymbol{\Gamma}}^{-1}\{\widetilde{\boldsymbol{W}}_m + \frac{1}{\sqrt{2}}(\widetilde{\boldsymbol{S}}_1 + \widetilde{\boldsymbol{S}}_2)\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_n\} + o_p(1), \tag{5.12}$$

$$\boldsymbol{W}_n\widetilde{\boldsymbol{W}}_m^{\tau} = \sqrt{2}E\{w_{t+m,1}w_{t,2}\varepsilon_t\varepsilon_{t+m}^2(1, \boldsymbol{X}_{t+m}^{\tau}, H(X_{t+m,1}))^{\tau}(1, H_t, \widetilde{H}_t)\} + o_p(1). \tag{5.13}$$

Proof. Proofs of (5.11) and (5.13) follow the same arguments in proving (5.10). For proving (5.12), write

$$\begin{aligned}
& \widehat{Z}_{t,WLSE}\widehat{Z}_{t+m,WLSE} - (\widehat{\alpha}_{WLSE} + \widehat{\gamma}H(X_{t,1}))\widehat{Z}_{t+m,WLSE} \\
& -(\widehat{\alpha} + \widehat{\gamma}H(X_{t+m,1}))\widehat{Z}_{t,WLSE} + \alpha_1 + \theta H_t + \gamma_1\widetilde{H}_t \\
= \; & (\widehat{Z}_{t,WLSE} - \widehat{\alpha}_{WLSE} - \widehat{\gamma}_{WLSE}H(X_{t,1}))(\widehat{Z}_{t+m,WLSE} - \widehat{\alpha}_{WLSE} - \widehat{\gamma}_{WLSE}H(X_{t+m,1})) \\
& -(\widehat{\alpha}^2 - \alpha^2) - (\widehat{\alpha}\widehat{\gamma} - \alpha\gamma)H_t - (\widehat{\gamma}^2 - \gamma^2)\widetilde{H}_t \\
= \; & \varepsilon_t\varepsilon_{t+m} - \{(\widehat{\alpha}_{WLSE} - \alpha) + (\widehat{\boldsymbol{\beta}}_{WLSE} - \boldsymbol{\beta})^{\tau}\boldsymbol{X}_t + (\widehat{\gamma}_{WLSE} - \gamma)H(X_{t,1})\}\varepsilon_{t+m} \\
& -\{(\widehat{\alpha}_{WLSE} - \alpha) + (\widehat{\boldsymbol{\beta}}_{WLSE} - \boldsymbol{\beta})^{\tau}\boldsymbol{X}_{t+m} + (\widehat{\gamma} - \gamma)H(X_{t,1})\}\varepsilon_t \\
& -\{(\widehat{\alpha}_{WLSE}^2 - \alpha^2) + (\widehat{\alpha}\widehat{\gamma} - \alpha\gamma)H_t + (\widehat{\gamma}^2 - \gamma^2)\widetilde{H}_t\} + o_p(1/\sqrt{n}) \\
= \; & \varepsilon_t\varepsilon_{t+m} + R_{t,1} + R_{t,2} + R_{t,3} + o_p(1/\sqrt{n}).
\end{aligned}$$

Hence,

$$\sqrt{m}(\widehat{\boldsymbol{\xi}}_w - \boldsymbol{\xi}_w) = -\widetilde{\boldsymbol{\Gamma}}^{-1}\{\widetilde{\boldsymbol{W}}_m + \sum_{j=1}^{3}\frac{1}{\sqrt{m}}\sum_{t=1}^{m}w_{t,2}R_{t,j}(1, H_t, \widetilde{H}_t)^{\tau}\} + o_p(1).$$

Using (2.8), (5.11), and ergodicity, we have

$$\begin{aligned}
& \tfrac{1}{\sqrt{m}}\textstyle\sum_{t=1}^{m}w_{t,2}R_{t,1}(1, H_t, \widetilde{H}_t)^{\tau} \\
= \; & -\tfrac{1}{m}\textstyle\sum_{t=1}^{m}w_{t,2}\varepsilon_{t+m}(1, H_t, \widetilde{H}_t)^{\tau}(1, \boldsymbol{X}_t^{\tau}, H(X_{t,1}))\sqrt{m}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\
= \; & o_p(1),
\end{aligned}$$

$$
\begin{aligned}
& \tfrac{1}{\sqrt{m}}\textstyle\sum_{t=1}^{m} w_{t,2}R_{t,2}(1,H_t,\widetilde{H}_t)^\tau \\
= & -\tfrac{1}{m}\textstyle\sum_{t=1}^{m} w_{t,2}\varepsilon_t(1,H_t,\widetilde{H}_t)^\tau(1,\boldsymbol{X}_t^\tau,H(X_{t,1}))\sqrt{m}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}) \\
= & -E\{w_{t,2}\varepsilon_t(1,H_t,\widetilde{H}_t)^\tau(1,\boldsymbol{X}_t^\tau,H(X_{t,1}))\}\sqrt{m}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})+o_p(1) \\
= & -\widetilde{\boldsymbol{S}}_1\sqrt{m}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})+o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
& \tfrac{1}{\sqrt{m}}\textstyle\sum_{t=1}^{m} w_{t,2}R_{t,3}(1,H_t,\widetilde{H}_t)^\tau \\
= & -\tfrac{1}{m}\textstyle\sum_{t=1}^{m} w_{t,2}(1,H_t,\widetilde{H}_t)^\tau(2\alpha+\gamma H_t,\mathbf{0}_d^\tau,\alpha H_t+2\gamma\widetilde{H}_t)\sqrt{m}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})+o_p(1) \\
= & -\widetilde{\boldsymbol{S}}_2\sqrt{m}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})+o_p(1).
\end{aligned}
$$

Therefore, (5.12) follows from the equations above. □

Proof of Theorem 1. The theorem follows from Lemmas 1 and 2 and the fact that $\hat{\theta}_w-\theta=(0,1,0)(\widehat{\boldsymbol{\xi}}_w-\boldsymbol{\xi}_w)$, where σ_0^2 can be calculated explicitly, which we skip deriving the formula as we will use the random weighted bootstrap method to estimate it later. □

Proof of Theorem 2. Define $\widehat{\boldsymbol{\xi}}^b=(\widehat{\alpha}^b_{WLSE},\widehat{\boldsymbol{\beta}}^{b\tau}_{WLSE},\widehat{\gamma}^b_{WLSE})^\tau$, $\widehat{\boldsymbol{\xi}}^b_w=(\widehat{\alpha}^b_{1,w},\widehat{\theta}^b_w,\widehat{\gamma}^b_{1,w})^\tau$,

$$
\boldsymbol{W}_n^b=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\xi_t^b-1)w_{t,1}\varepsilon_t(1,\boldsymbol{X}_t^\tau,H(X_{t,1}))^\tau,
$$

$$
\boldsymbol{\Gamma}_n^b=\frac{1}{n}\sum_{t=1}^{n}\xi_t^b w_{t,1}\begin{pmatrix} 1 & \boldsymbol{X}_t^\tau & H(X_{t,1}) \\ \boldsymbol{X}_t & \boldsymbol{X}_t\boldsymbol{X}_t^\tau & \boldsymbol{X}_tH(X_{t,1}) \\ H(X_{t,1}) & \boldsymbol{X}_t^\tau H(X_{t,1}) & H^2(X_{t,1})\end{pmatrix},
$$

$$
\widetilde{\boldsymbol{W}}_m^b=\frac{1}{\sqrt{m}}\sum_{t=1}^{m}(\xi_{t+m}^b-1)w_{t,2}\varepsilon_t\varepsilon_{t+m}(1,H_t,\widetilde{H}_t)^\tau.
$$

Therefore,

$$
\begin{aligned}
& \sqrt{n}(\widehat{\boldsymbol{\xi}}^b-\boldsymbol{\xi}) \\
= & -(\boldsymbol{\Gamma}_n^b)^{-1}\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\xi_t^b w_{t,1}\varepsilon_t(1,\boldsymbol{X}_t^\tau,H(X_{t,1}))^\tau \\
= & -\boldsymbol{\Gamma}^{-1}\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\xi_t^b w_{t,1}\varepsilon_t(1,\boldsymbol{X}_t^\tau,H(X_{t,1}))^\tau+o_p(1),
\end{aligned}
$$

implying that

$$\sqrt{n}(\widehat{\boldsymbol{\xi}}^b - \widehat{\boldsymbol{\xi}}) = -\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_n^b + o_p(1).$$

Similarly,

$$\sqrt{m}(\widehat{\boldsymbol{\xi}}_w^b - \widehat{\boldsymbol{\xi}}_w) = -\widetilde{\boldsymbol{\Gamma}}^{-1}\{\widetilde{\boldsymbol{W}}_m^b + \frac{1}{\sqrt{2}}(\widetilde{\boldsymbol{S}}_1 + \widetilde{\boldsymbol{S}}_2)\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_n^b\} + o_p(1). \tag{5.14}$$

We can show that

$$\boldsymbol{W}_n^b \xrightarrow{d} N(\mathbf{0}_{d+2}, \boldsymbol{\Sigma}),\ \widetilde{\boldsymbol{W}}_m^b \xrightarrow{d} N(\mathbf{0}_3, \widetilde{\boldsymbol{\Sigma}}),$$

$$\boldsymbol{W}_n^b(\widetilde{\boldsymbol{W}}_m^b)^\tau = \sqrt{2}E\{w_{t+m,1}w_{t,2}\varepsilon_t\varepsilon_{t+m}^2(1, \boldsymbol{X}_{t+m}^\tau, H(X_{t+m,1}))^\tau(1, H_t, \widetilde{H}_t)\} + o_p(1).$$

Hence, both $\sqrt{m}(\widehat{\theta}_w - \theta)$ and $\sqrt{m}(\widehat{\theta}_w^b - \widehat{\theta}_w)$ have a normal limit with the same asymptotic variance. Further, we can show that $\frac{m}{B}\sum_{b=1}^B(\widehat{\theta}_w^b - \widehat{\theta}_w)^2$ converges in probability to the asymptotic variance of $\sqrt{m}(\widehat{\theta}_w - \theta)$ as $B \to \infty$ and $n \to \infty$ by using the independence between $\{\xi_t^b\}$ and $\{\boldsymbol{X}_t, Y_t\}$. That is, the theorem follows.

□

Supplementary Materials

The online supplement generalizes the method to correlated and heteroscedastic ε_t's and reports additional data analysis validating our classification of mutual funds based on stock picking and market timing skills.

Acknowledgements

We thank an associate editor and two reviewers for their helpful comments. Jiang's research is supported by the Tsinghua University Initiative Scientific Research Program and Institute for Industrial Innovation and Finance at Tsinghua University. Peng's research is partly supported by the Simons Foundation and the NSF grant of DMS-2012448.

References

Back, K., A. D. Crane, and K. Crotty (2018, December). Skewness Consequences of Seeking Alpha. *The Review of Financial Studies 31*(12), 4720–4761.

Basrak, B., R. A. Davis, and T. Mikosch (2002, May). Regular variation of GARCH processes. *Stochastic Processes and their Applications 99*(1), 95–115.

Bollen, N. P. B. and J. A. Busse (2001). On the Timing Ability of Mutual Fund Managers. *The Journal of Finance 56*(3), 1075–1094.

Bollerslev, T. (1986, April). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Busse, J. A. (1999). Volatility timing in mutual funds: Evidence from daily returns. *The Review of Financial Studies 12*(5), 1009–1041.

Busse, J. A., L. Jiang, and Y. Tang (2021). Double-Adjusted Mutual Fund Performance. *The Review of Asset Pricing Studies 11*(1), 169–208.

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

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*(4), 987–1007.

Evans, R. B. (2010). Mutual Fund Incubation. *The Journal of Finance 65*(4), 1581–1611.

Goetzmann, W. N., J. Ingersoll, and Z. Ivković (2000). Monthly measurement of daily timers. *Journal of Financial and Quantitative Analysis 35*(3), 257–290.

Hall, P. (1992, January). *The Bootstrap and Edgeworth Expansion* (1992nd edition ed.). New York: Springer.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. Academic Press.

Henriksson, R. D. (1984). Market timing and mutual fund performance: An empirical investigation. *Journal of*

Business, 73–96.

Henriksson, R. D. and R. C. Merton (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, 513–533.

Jagannathan, R. and R. A. Korajczyk (1986). Assessing the market timing performance of managed portfolios. *Journal of Business*, 217–235.

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

Jin, Z., Z. Ying, and L. J. Wei (2001). A Simple Resampling Method by Perturbing the Minimand. *Biometrika 88*(2), 381–390.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp (2014). Time-Varying Fund Manager Skill. *The Journal of Finance 69*(4), 1455–1484.

Kacperczyk, M., C. Sialm, and L. Zheng (2008, November). Unobserved Actions of Mutual Funds. *The Review of Financial Studies 21*(6), 2379–2416.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A Rational Theory of Mutual Funds' Attention Allocation. *Econometrica 84*(2), 571–626.

Kon, S. J. (1983). The market-timing performance of mutual fund managers. *Journal of Business*, 323–347.

Ling, S. (2007, October). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics 140*(2), 849–873.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–708.

Nguyen, T. and J. Jiang (2020, January). Power analysis, sample size calculation for testing the largest binomial probability. *Statistical Theory and Related Fields 4*(1), 78–83.

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

131–136.

Zhu, K. (2016). Bootstrapping the portmanteau tests in weak auto-regressive moving average models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(2), 463–485.

Qingsong Shan, Jiangxi University of Economics and Finance

E-mail: (mathhuman@outlook.com)

Lei Jiang, Tsinghua University

E-mail: (jianglei@sem.tsinghua.edu.cn)

Liang Peng, Georgia State University

E-mail: (lpeng@gsu.edu)

Zhongling Qin, Auburn University

E-mail: (zzq0018@auburn.edu)