Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Test for Market Timing Using Daily Fund Returns

Lei Jiang, Weimin Liu & Liang Peng

To cite this article: Lei Jiang, Weimin Liu & Liang Peng (2023) Test for Market Timing Using Daily Fund Returns, Journal of Business & Economic Statistics, 41:1, 184-196, DOI: 10.1080/07350015.2021.2006670

To link to this article: https://doi.org/10.1080/07350015.2021.2006670

Published online: 21 Dec 2021.

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2023, VOL. 41, NO. 1, 184–196
https://doi.org/10.1080/07350015.2021.2006670

Test for Market Timing Using Daily Fund Returns

Lei Jiang[a], Weimin Liu[b], and Liang Peng[c]

[a]Department of Finance, Tsinghua University, Beijing, China; [b]Department of Mathematics and Statistics, Georgia State University, Atlanta, GA; [c]Department of Risk Management and Insurance, Georgia State University, Atlanta, GA

ABSTRACT
Using daily mutual fund returns to estimate market timing, some econometric issues, including heteroscedasticity, correlated errors, and heavy tails, make the traditional least-squares estimate in Treynor–Mazuy and Henriksson–Merton models biased and severely distort the *t*-test size. Using ARMA-GARCH models, weighted least-squares estimate to ensure a normal limit, and random weighted bootstrap method to quantify uncertainty, we find more funds with positive timing ability than the Newey–West *t*-test. Empirical evidence indicates that funds with perverse timing ability have high fund turnovers and funds tradeoff between timing and stock picking skills.

ARTICLE HISTORY
Received July 2021
Accepted November 2021

KEYWORDS
ARMA-GARCH model; Heavy tails; Market timing; Mutual funds

1. Introduction

After seminal works in Treynor and Mazuy (1966) and Henriksson and Merton (1981), partly because of the limited availability of comprehensive daily fund return data, researchers such as Chang and Lewellen (1984), Henriksson (1984), Grinblatt and Titman (1988), and Becker et al. (1999) use monthly returns of U.S. mutual funds to estimate market timing skill parametrically. The general conclusion is the lack of market timing skill on average in the U.S. mutual fund industry based on the timing parameter for each fund. Goetzmann, Ingersoll, and Ivkovic (2000) identify the downward estimation bias of timing skill using monthly return when the mutual funds' timing happens daily. Bollen and Busse (2001) use daily fund returns for a sample of 230 actively managed funds and find the positive timing skill on average. Simulating fake funds with no timing skill proves that using daily returns is more powerful than monthly returns to detect timing. Chance and Hemler (2001), Bollen and Busse (2005), and Mamaysky, Speigel, and Zhang (2008) all estimate fund managers' timing ability based on daily fund returns and make inferences using least-square estimates in both Treynor–Mazuy and Henriksson–Merton models.

This article first empirically shows that the model errors in Treynor and Mazuy (1966) and Henriksson and Merton (1981) are serially correlated and heteroscedastic, where Breen, Jagannathan and Ofer (1986) find correcting heteroscedasticity is important when inferring timing ability in Henriksson and Merton (1981) model. Second, the error terms empirically depend on the timing variables in Treynor and Mazuy (1966) and Henriksson and Merton (1981) for most of the U.S. mutual funds, which requires a higher finite moment for errors and market returns to ensure the normality of the least-squares estimation. Third, we apply the Hill tail index estimation procedure in Hill (1975), finding that the product of the error and the timing variables in Treynor and Mazuy (1966) and Henriksson and Merton (1981) has a heavy tail making the least-squares estimate for timing parameter have a nonnormal limit. The lack of finite moments severely distorts the size of the Newey–West *t*-test with variance correction for timing detection. Therefore, the judgment about timing for different mutual fund managers using the Newey–West *t*-test does not accurately reflect the fund managers' timing skill and thus bias the fund investors' decision.

For accounting heteroscedasticity of fund returns and the dependence between residuals in the factor model and factors, we propose to model fund daily excess returns and daily risk factors by ARMA-GARCH processes in Engle (1982) and Bollerslev (1986) and use the ARMA structure in the factor model to estimate the timing parameter effectively. Because the method does not infer the GARCH part in the factor model and the ARMA-GARCH models for risk factors, the proposed method is robust against heteroscedasticity. To ensure a normal limit for estimating the timing parameter, we use a weighted least-square (WLS) estimation in the spirit of Ling (2007) and a random weighted bootstrap method to quantify the estimation uncertainty. Hence, the proposed method is robust against heavy tails too. A simulation study confirms the proposed test's accurate size and good power for zero timing ability based on the developed weighted least-squares estimation and random weighted bootstrap method.

We apply the developed zero timing ability test to each of the U.S. actively managed domestic equity funds with daily fund returns in the Center for Research in Security Prices (CRSP) from September 1, 1998 to December 31, 2018. To achieve robust conclusions, we focus on funds with at least 1000 daily fund returns, resulting in 2610 funds. Using Treynor–Mazuy timing ability measure and the traditional Newey–West *t*-test with kernel variance correction at the level 10%, we find that 1775, 2105, and 2156 out of the total 2610 funds have no timing

CONTACT Liang Peng lpeng@gsu.edu Department of Risk Management and Insurance, Georgia State University, 35 Broad St., Atlanta, GA 30302-3965.

ability for the one-factor, three-factor, and four-factor model, respectively. After removing these identified zero timing funds, timing ability (i.e., positive least-squares estimate) exists for 165, 188, and 176 funds for the one-factor, three-factor, and four-factor models. In contrast, using AIC to choose the best AR(p) model with $p \le 15$ and the proposed test for zero timing ability, we find that 1922, 1845, and 1914 funds have no timing ability, and 195, 295, and 284 funds have timing ability (i.e., positive weighted least-squares estimate) after removing the identified zero timing funds for the one-factor, three-factor, and four-factor model, respectively. These significant differences may be because the least-squares estimate and Newey–West t-test do not have a normal limit due to the lack of finite moments.

Because the proposed method models the factor model's errors by an ARMA-GARCH process and estimates the ARMA part but does not estimate the GARCH part, it becomes vital to check if the employed ARMA model catches enough error correlations. After using AIC to pick the best AR(p) model with $p \le 15$ and the one-factor, three-factor, and four-factor model, respectively, the proposed zero correlations test finds that 411, 400, and 350 out of the total 2610 funds reject the null hypothesis for the lag 5 and level 10%. Therefore, we repeat our analysis for timing ability by removing these 411, 400, and 350 funds for the one-factor, three-factor, and four-factor models, respectively. Results remain pretty different between the proposed method and the Newey–West t-test with variance correction as before. We use the developed test to group funds with positive, zero, and negative timing skills based on the entire sample of funds or funds passing the zero correlations test. Then we compare the characteristic difference between positive and negative timing funds and find that funds with positive skills are smaller in a small fund family and charge less expense to their investors. In contrast, funds with negative timing skills have a much higher turnover than funds with positive timing skills.

Last, we further study the hypothesis of a tradeoff between the stock picking skill and the market timing skill studied in Kon (1983), Henriksson (1984), Jagannathan and Korajczyk (1986), and Goetzmann, Ingersoll, and Ivkovich (2000). Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014) find fund managers focus on market timing during a market downturn and focus on stock picking during a market upturn because of limited attention. Recently, Back, Crane, and Crotty (2018) show that funds tradeoff relatively high alpha for low coskewness, which results in an inverse empirical relation between alpha and market timing. Based on our method, we empirically estimate stock picking and market timing parameters, summarize the percentage of funds with different types of skills, and find supporting evidence on the inverse relation between alpha and market timing. The results are especially evident after excluding funds with zero timing ability.

Our article contributes to the literature of measuring the market timing ability of mutual funds using daily fund returns. The previous work by Treynor and Mazuy (1966), Henriksson and Merton (1981), and others use monthly fund returns to estimate the timing parameter. Because monthly fund returns underestimate timing ability when funds are daily timers (Goetzmann et al. 2000; Bollen and Busse 2001), it may be informative to using daily data as argued by Chance and Hemler (2001), Bollen and Busse (2005), and Mamaysky, Speigel, and Zhang (2008). Our article identifies econometric issues such as heteroscedasticity, correlated errors, and heavy tails in using daily fund returns to estimate timing ability. These issues lead to a nonnormal limit of the least-squares estimate and the Newey–West t-test with variance correction due to the lack of finite moments. We solve these problems using an ARMA-GARCH process to model the factor model's errors, a weighted least-squares estimator to ensure a normal limit, and a random weighted bootstrap method to quantify the estimation uncertainty. The proposed test for zero timing ability has an accurate size and is powerful, and thus fund investors can make more precise investment decisions based on our method. Unlike our fund-by-fund analysis, Fan, Liao, and Yao (2015) develop a simultaneous test for zero alphas using thresholding technique and sparsity structure for cross-sectional dependence, and Fan et al. (2019) propose a generic factor-adjusted robust multiple test for means. Because our article tests zero timing for each individual fund by taking correlated residuals and heavy tails into account, where the Newey–West test fails, we neither assume sparsity nor explore sparse alphas.

We organize the article as follows. Section 2 presents our model and method for testing zero market timing ability. Sections 3 and 4 are our simulation study and empirical analysis of U.S. equity funds, respectively. Section 5 concludes. Proofs are in the Appendix.

2. Models, Tests, and Theoretical Results

Let Y_t be a mutual fund's excess return at time t, and $X_t = (X_{t,1}, \ldots, X_{t,d})^\tau$ be the benchmark factors with $X_{t,1}$ being the market excess return. We use A^τ to denote the transpose of the matrix or vector A. Some popular factor models include the one-factor model (Capital Asset Pricing Model) in Jensen (1968), the three-factor model in Fama and French (1996), and the four-factor model in Carhart (1997). Researchers evaluate the fund's performance by measuring the stock picking and market timing skills. A simple way of measuring these two skills is to employ the following model:

$$Y_t = \alpha + \beta^\tau X_t + \gamma H(X_{t,1}) + \varepsilon_t, \tag{1}$$

where α and γ measure a fund manager's stock picking skill and market timing skill, respectively, and H is a known function. For example, Treynor and Mazyu (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 $H(X_{t,1})$ being the conditional standard deviation of $X_{t,1}$, and Goetzmann, Ingersoll, and Ivković (2000) use $H(X_{t,1})$ as a quantity computed from daily returns in that month. A comparison study between the power to detect timing based on monthly fund return and daily fund return is given in Bollen and Busse (2001).

After fitting model (1) by the least-squares estimation, one can test for $H_0 : \alpha = 0$ for each fund to find if a fund manager has the stock picking skill. Alternatively, one can test for $H_0 : \gamma = 0$ to see if a fund manager has the market timing skill. Also, one can examine the association between these two skills by comparing the signs of the least-squares estimators for α and γ. To allow series correlation and heteroscedasticity in $\{\varepsilon_t\}$, one often uses the Newey–West t-test with a variance correction, which requires $E(\varepsilon_t^2 H^2(X_{t,1})) < \infty$ at least to ensure a normal limit. However, when the study uses daily returns, heteroscedasticity in $\{\varepsilon_t\}$ and the dependence between $\{X_t\}$ and $\{\varepsilon_t\}$ cause $\varepsilon_t H(X_{t,1})$ to have a heavier tail with an infinite second moment, invalidating the Newey–West t-test. To confirm these econometric issues, we conduct the following brief data analysis. We refer to Section 4 for a detailed description of our data.

First, we apply the Box.test with lag=10 and fitdf=$d+2$ in the R statistical software to the residuals and their absolute values after fitting model (1) to the daily mutual fund excess returns in our data analysis below. We obtain and plot the p-values of testing uncorrelated errors and uncorrelated absolute values of errors for each fund in Figure 1 for $H(X_{t,1}) = X_{t,1}^2$. This figure shows that ε_t's in (1) are correlated and heteroscedastic for our daily mutual fund excess returns. Hence, it is better to use the Newey–West t-test to correct the asymptotic variance, which still needs $E(\varepsilon_t^2 H^2(X_{t,1})) < \infty$ at least to ensure a normal limit. To save space, we do not report the analysis for $H(X_{t,1}) = \max(0, X_{t,1})$.

Secondly, we compute the p-values for testing zero correlation between $|\hat{\varepsilon}_t|$ and $H(X_{t,1})$ in Figure 2, where $\hat{\varepsilon}_t$ is the residual obtained from the least-squares estimation. Figure 2 shows that $|\varepsilon_t|$ and $H(X_{t,1})$ are correlated for most funds, implying that ε_t and $H(X_{t,1})$ are dependent for most funds. Both the dependence between $\{\varepsilon_t\}$ and $\{X_t\}$ and the heteroscedasticity in $\{\varepsilon_t\}$ will lead to heavy tailed $\varepsilon_t H(X_{t,1})$ causing $E(\varepsilon_t^2 H^2(X_{t,1})) = \infty$ and invalidating the Newey–West test.

Thirdly, for examining the above finite moment requirement, we estimate the tail index of $|\varepsilon_t H(X_{t,1})|$ by assuming that its distribution has a heavy tail, that is,

$$\lim_{t\to\infty} \frac{P(|\varepsilon_t H(X_{t,1})| > tx)}{P(|\varepsilon_t H(X_{t,1})| > t)} = x^{-a} \text{ for } x > 0 \text{ and some } a > 0.$$

We employ the well-known Hill estimate in Hill (1975) with upper order statistics $k = 50$, 100, and 150. We plot these Hill estimates for $H(X_{t,1}) = X_{t,1}^2$ and $H(X_{t,1}) = \max(0, X_{t,1})$, which shows that many funds have a tail index less than 2, that is, $E(\varepsilon_t^2 H^2(X_{t,1})) = \infty$. So, the least-squares estimate of γ for many funds will have a nonnormal limit, which cautions the application of the Newey–West t-test. To save space, we do not report these two figures.

We remark that the above analysis of heteroscedasticity and dependence between errors and factors ignores the necessary moment conditions, which may bias the calculated p-values, but it is vital to take these issues into account for evaluating mutual funds' timing. Therefore, we propose to model ε_t's and each of $X_{t,i}$'s by ARMA-GARCH models:

$$\begin{cases} \varepsilon_t = \sum_{i=1}^{s} \phi_i \varepsilon_{t-i} + \sum_{j=1}^{r} \psi_j U_{t-j} + U_t, \ U_t = \eta_t \sigma_t, \\ \quad \sigma_t^2 = w + \sum_{i=1}^{p} a_i U_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \\ X_{t,l} = \mu_l + \sum_{i=1}^{s_l} \phi_{i,l} X_{t-i,l} + \sum_{j=1}^{r_l} \psi_{j,l} \bar{\varepsilon}_{t-j,l} + \bar{\varepsilon}_{t,l}, \\ \quad \bar{\varepsilon}_{t,l} = \bar{\eta}_{t,l} \bar{\sigma}_{t,l}, \\ \bar{\sigma}_{t,l}^2 = w_l + \sum_{i=1}^{p_l} a_{i,l} \bar{\varepsilon}_{t-i,l}^2 + \sum_{j=1}^{q_l} b_{j,l} \bar{\sigma}_{t-j,l}^2, \ l = 1, \ldots, d, \end{cases} \tag{2}$$

where $\{(\eta_t, \bar{\eta}_{t,1}, \ldots, \bar{\eta}_{t,d})^\tau\}_{t=1}^n$ is a sequence of independent and identically distributed random vectors with means zero and variances one. For estimating $\boldsymbol{\beta}$ and γ consistently, we need that ε_t is uncorrelated with X_t and $H(X_{t,1})$. When η_t and $\bar{\eta}_t = (\bar{\eta}_{t,1}, \ldots, \bar{\eta}_{t,d})^\tau$ are uncorrelated, ε_t does not correlate with X_t and $X_{t,1}^2$ but may still correlate with $H(X_{t,1}) = \max(0, X_{t,1})$. To allow ε_t and $H(X_{t,1})$ to be dependent but uncorrelated for any H, we assume that

$$E(\eta_t | \bar{\eta}_{t,1}, \ldots, \bar{\eta}_{t,d}) = 0. \tag{3}$$

An example satisfying (3) is $\eta_t = \bar{\eta}_{t,0} G(\bar{\eta}_t)$ for some unknown function G, where $\{\bar{\eta}_{t,0}\}$ is a sequence of independent and identically random variables with zero mean and finite variance and independent of $\{\bar{\eta}_t\}$.

Write $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_s)^\tau$, $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_r)^\tau$, $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\tau, \gamma, \boldsymbol{\phi}^\tau, \boldsymbol{\psi}^\tau)^\tau$, $\varepsilon_t(\alpha, \boldsymbol{\beta}, \gamma) = Y_t - \alpha - \boldsymbol{\beta}^\tau X_t - \gamma H(X_{t,1})$, and $U_t(\boldsymbol{\theta}) = \varepsilon_t(\alpha, \boldsymbol{\beta}, \gamma) - \sum_{i=1}^{s} \phi_i \varepsilon_{t-i}(\alpha, \boldsymbol{\beta}, \gamma) - \sum_{j=1}^{r} \psi_j U_{t-j}(\boldsymbol{\theta})$. To infer $\boldsymbol{\theta}$ effectively, we follow the idea in Xiao et al. (2003) and Liu, Chen, and Yao (2010) to take the ARMA structure into account and use the weighted idea in Ling (2007) to reduce the moment effect of σ_t in U_t and X_t and $H(X_{t,1})$ in the score equations. Because of (3), we use the following weight function

$$\begin{cases} w_t = w_{t,X}(w_{t,X} + w_{t-1,X,Y}), \\ w_{t,X} = \max\{1, C_X^{-1} \sum_{i=1}^{t} \exp(\log(h) \log^2(i)) \\ \qquad \max(|X_{t-i+1,1}|, \ldots, |X_{t-i+1,d}|, |H(X_{t-i+1,1})|)\}, \\ w_{t,Y} = \max\{1, C_Y^{-1} \sum_{i=1}^{t} \exp(\log(h) \log^2(i)) |Y_{t-i+1}|\}, \\ w_{t,X,Y} = \max(w_{t,X}, w_{t,Y}), \end{cases} \tag{4}$$

for $t = 1, \ldots, n$, where $h \in (0, 1)$, C_X and C_Y are the 90% quantile of the distribution function of $\max(|X_{t,1}|, \ldots, |X_{t,d}|, |H(X_{t,1})|)$ and $|Y_t|$, respectively, to avoid overweight. Define $w_{t,X} = 1$ and $w_{t,Y} = 1$ for $t \le 0$ or $h = 0$ throughout. That is, $h = 0$ means unweighted inference. We use $w_{t-1,X,Y}$ instead of $w_{t,X,Y}$ because we try to control the moment effect of σ_t, not U_t, and we add $w_{t,X}$ to effectively control X_t and $H(X_{t,1})$ because of (3). In practice, we replace C_X and C_Y by the corresponding sample quantile, and the asymptotic limit of our proposed inference remains unchanged, which can be proved like He et al. (2020). Also, we use $h = 0.2$ in our simulation study and data analysis. In conclusion, we propose to estimate $\boldsymbol{\theta}$ by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{n} U_t^2(\boldsymbol{\theta}) w_t^{-1}, \tag{5}$$

where w_t is given in (4). We remark there are other choices of weight functions. Finding an optimal weight is challenging for our proposed test for zero market timing ability as it requires developing Edgeworth expansions for evaluating the accuracy of the test size.

To derive the asymptotic limit of the proposed weighted least-squares estimation, we use the following regularity conditions.

- C1. $\{\varepsilon_t\}$ and $\{X_t\}$ are strictly stationary and ergodic; See conditions in Theorem 3.1 of Basrak, Davis, and Mikosch (2002).
- C2. $\{(\eta_t, \bar{\eta}_{t,1}, \ldots, \bar{\eta}_{t,d})^\tau\}$ is a sequence of independent and identically distributed random vectors with means zero and variances one.
- C3. Assume condition (3) holds, and there exists $\delta > 0$ such that $E|\eta_t|^{2+\delta} < \infty$.

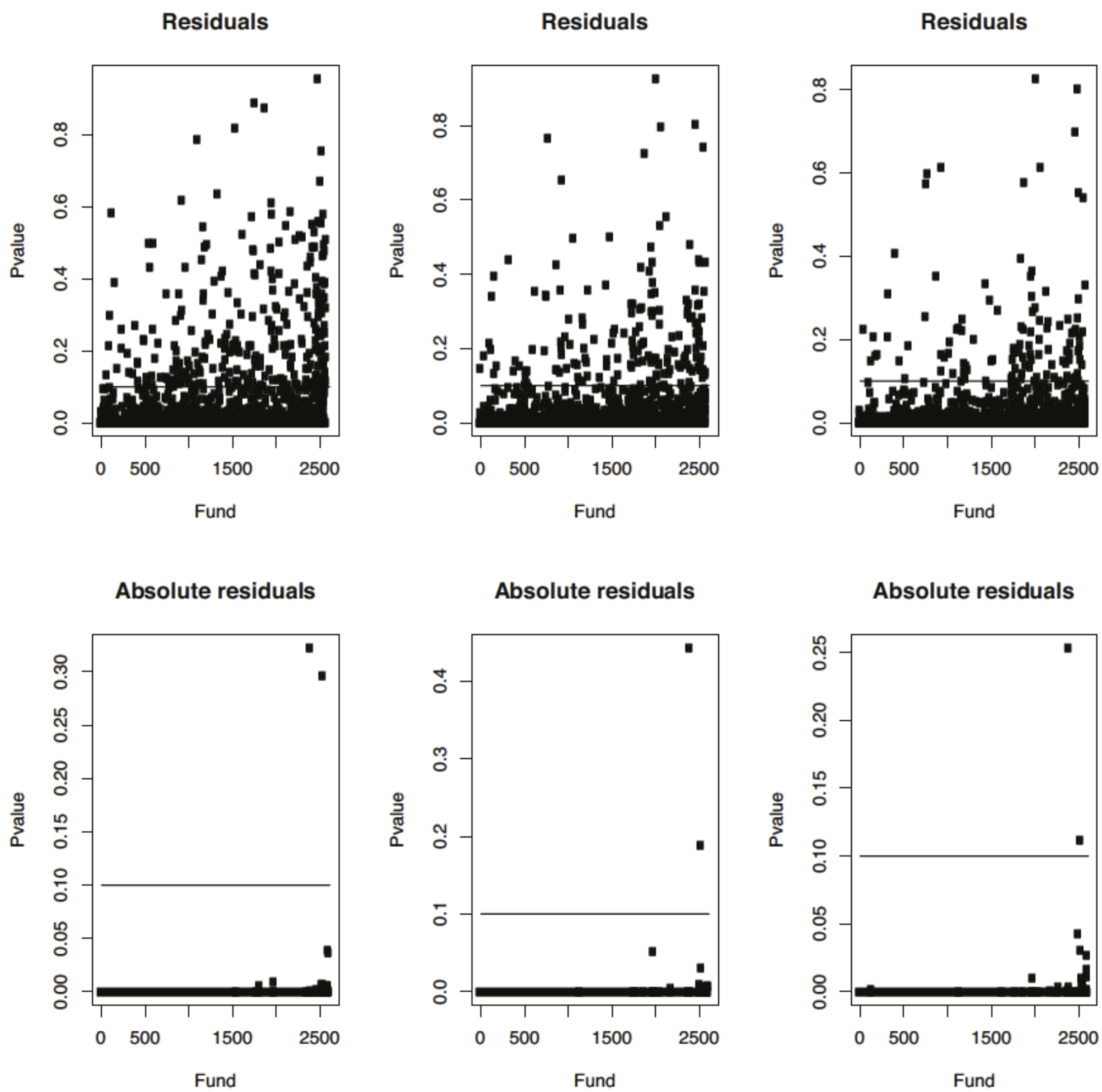

Figure 1. *P*-values for testing uncorrelated errors and uncorrelated absolute values of errors.
Applying the Box.test with lag=10 and fitdf=$d+2$ in the R statistical software to the residuals and their absolute values in (1) with $H(X_{t,1}) = X_{t,1}^2$, we plot the *p*-values of testing uncorrelated errors (top panels) and uncorrelated absolute values (bottom panels) of errors for each fund. From left to right panels, they are the one-factor, three-factor, and four-factor models.

- C4. Assume the covariance matrix of $w_t^{-1}\varepsilon_t(1, X_t^\tau, H(X_{t,1}))^\tau$ is positive definite.

Theorem 1. Suppose models (1) and (2) hold with conditions C1)–C4). Then, as $n \to \infty$, $\sqrt{n}(\widehat{\theta} - \theta_0)$ converges in distribution to a normal limit with zero means and a complicated covariance matrix given in the proof, where θ_0 is the true value of θ.

To test for the zero market timing skill without estimating the asymptotic covariance, we propose to use the following random weighted bootstrap method in Zhu (2016). Note that the residual based bootstrap method does not apply because we do not infer the GARCH models, and we allow U_t and X_t to be dependent.

- *(A1)* Draw a random sample with sample size n from a distribution with mean one and variance one and denote them by $\xi_1^b, \ldots, \xi_n^b$. Our simulation study employs the standard exponential distribution and shows good finite performance.
- *(A2)* Minimize $\sum_{t=1}^n \xi_t^b U_t^2(\theta) w_t^{-1}$ and denote the estimator by $\widehat{\theta}^b$.
- *(A3)* Repeat the above two steps B times to get $\{\widehat{\theta}^b\}_{b=1}^B$.

Hence, we estimate the asymptotic variance σ_γ^2 of $\sqrt{n}\widehat{\gamma}$ by $\widehat{\sigma}_\gamma^2 = \frac{n}{B}\sum_{b=1}^B(\widehat{\gamma}^b - \widehat{\gamma})^2$.

Theorem 2. Under conditions of Theorem 1, $\widehat{\sigma}_\gamma/\sigma_\gamma$ converges in probability to one as both $B \to \infty$ and $n \to \infty$.

Using the above theorem, we reject $H_0 : \gamma = 0$ at level a if $n\widehat{\gamma}^2/\widehat{\sigma}_\gamma^2 > \chi_{1,1-a}^2$, where $\chi_{1,1-a}^2$ is the (1-a)-th quantile of a chi-squared distribution function with one degree of freedom.

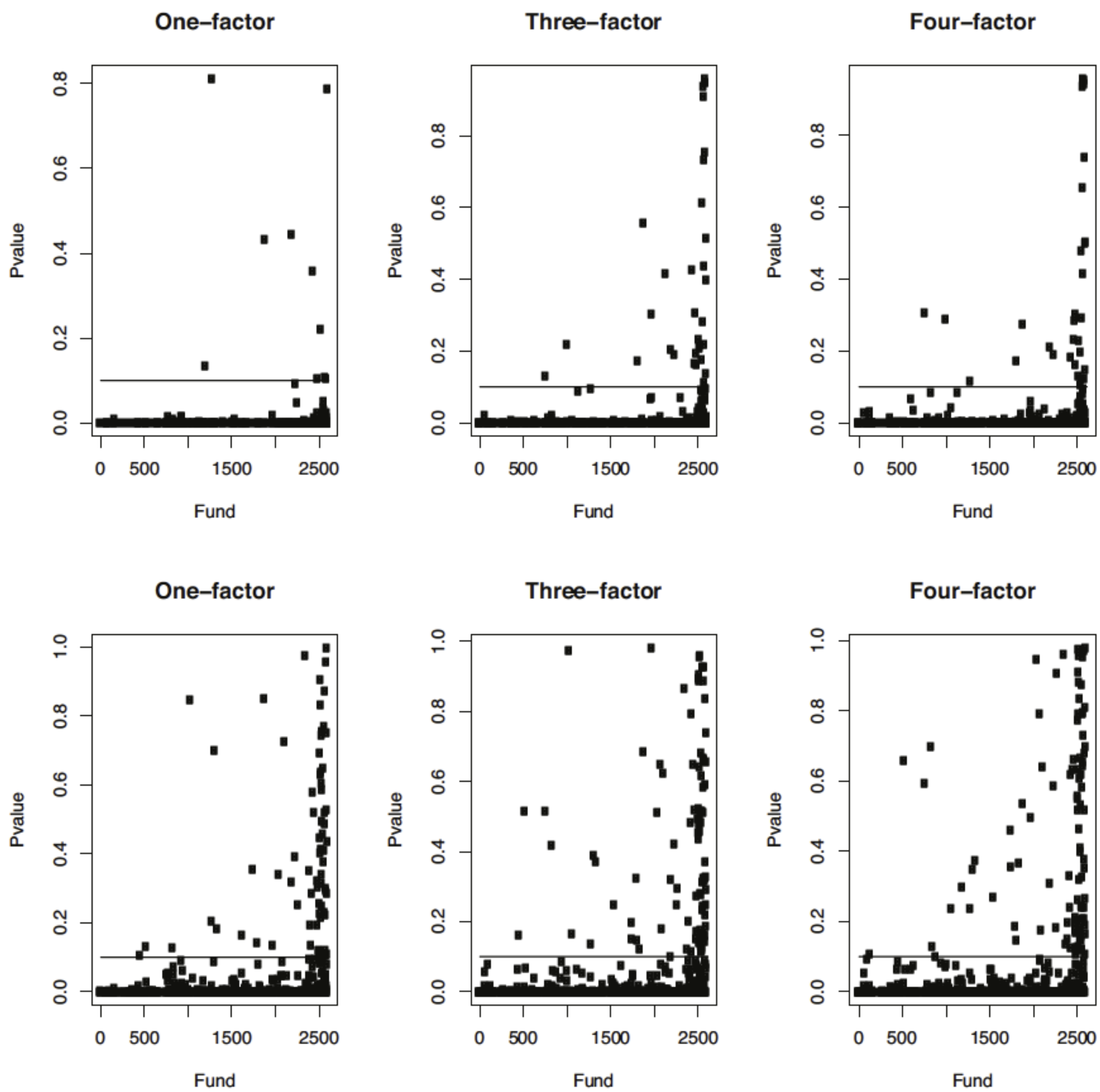

Figure 2. P-values for testing zero correlation between $|\hat{\varepsilon}_t|$ and $H(X_{t,1})$.
We plot the P-values for testing zero correlation between $|\hat{\varepsilon}_t|$ and $H(X_{t,1}) = X_{t,1}^2$ in the upper panels and between $|\hat{\varepsilon}_t|$ and $H(X_{t,1}) = \max(0, X_{t,1})$ in the bottom panels for each fund, where $\hat{\varepsilon}_t$ is the residual computed from the least-squares estimation. From left to right panels, they are the one-factor, three-factor, and four-factor models.

Because our inference uses the ARMA structure of U_t's in (2), it is crucial to test the assumption that $H_0 : E(U_t U_{t-1}) = \dots = E(U_t U_{t-l}) = 0$ for a pre-assigned lag l. To allow for fewer finite moments, we consider the weighted correlations $\rho = (\rho_1, \dots, \rho_l)$ with $\rho_i = E(\frac{U_t}{w_{t,X}+w_{t-1,X,Y}} \frac{U_{t-i}}{w_{t-i,X,Y}})$ rather than $E(U_t U_{t-i})$. Again, we use $w_{t,X}+w_{t-1,X,Y}$ and $w_{t-i,X,Y}$ to control the moment effect of σ_t in U_t and U_{t-i}, respectively. Also, we add $w_{t,X}$ because of (3). Therefore, we estimate ρ by $\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_l)^\tau$ with

$$\hat{\rho}_i = \frac{1}{n-l+1} \sum_{t=l+1}^{n} \frac{U_t(\hat{\theta})}{w_{t,X} + w_{t-1,X,Y}} \frac{U_{t-i}(\hat{\theta})}{w_{t-i,X,Y}} \text{ for } i = 1, \dots, l.$$

To test for $H_0 : \rho = \mathbf{0}$ without estimating the asymptotic covariance of $\hat{\rho}$, we employ a similar random weighted bootstrap method as before.

- ($B1$) Draw a random sample with sample size n from a distribution with mean one and variance one and denote them by $\xi_1^b, \dots, \xi_n^b$.
- ($B2$) Minimize $\sum_{t=1}^n \xi_t^b U_t^2(\theta) w_t^{-1}$ and denote the resulting estimator by $\hat{\theta}^b$. Therefore, we have $\hat{\rho}^b = (\hat{\rho}_1^b, \dots, \hat{\rho}_l^b)^\tau$, where $\hat{\rho}_i^b = \frac{\sum_{t=l+1}^n \xi_t^b U_t(\hat{\theta}^b)\{w_{t,X}+w_{t-1,X,Y}\}^{-1} U_{t-i}(\hat{\theta}^b)\{w_{t-i,X,Y}\}^{-1}}{\sum_{t=l+1}^n \xi_t^b}$.
- ($B3$) Repeat the above two steps B times to get $\{\hat{\rho}^b\}_{b=1}^B$.

Therefore, we estimate the asymptotic covariance of $\hat{\rho}$ by $\hat{\Sigma}_\rho = \frac{1}{B}\sum_{b=1}^B (\hat{\rho}^b - \hat{\rho})(\hat{\rho}^b - \hat{\rho})^\tau$ and reject $H_0 : \rho = \mathbf{0}$ at level a whenever $\hat{\rho}^\tau \hat{\Sigma}_\rho^{-1} \hat{\rho} > \chi_{l,1-a}^2$, where $\chi_{l,1-a}^2$ is the $(1-a)$-th quantile of chi-squared distribution with l degrees of freedom. In the empirical analysis, around four hundred funds reject zero

Table 1. Test sizes for $H_0: \gamma = 0$ and $H_0: \rho = 0$ at level 10%.

n	Case	ε_t	Newey–West	$\hat{\gamma}$ with $h=0$	$\hat{\gamma}$ with $h=0.2$	$\hat{\rho}$ with $h=0$	$\hat{\rho}$ with $h=0.2$
500	A	GARCH	0.1284	0.1226	0.1064	0.1126	0.1108
1000	A	GARCH	0.1192	0.1138	0.1022	0.0974	0.1000
2000	A	GARCH	0.1192	0.1146	0.1052	0.1018	0.1038
500	B	GARCH	0.1678	0.1142	0.0948	0.0858	0.0922
1000	B	GARCH	0.1558	0.1046	0.0966	0.0876	0.0906
2000	B	GARCH	0.1384	0.0930	0.1006	0.0876	0.0892
500	C	GARCH	0.2166	0.1604	0.0978	0.0930	0.0930
1000	C	GARCH	0.1962	0.1462	0.0954	0.0904	0.0898
2000	C	GARCH	0.1862	0.1392	0.0982	0.0868	0.0906
500	D	GARCH	0.2842	0.2366	0.1138	0.0884	0.1018
1000	D	GARCH	0.2660	0.2272	0.1082	0.0838	0.0964
2000	D	GARCH	0.2452	0.2054	0.1086	0.0974	0.0918
500	A	AR-GARCH	0.1270	0.1216	0.1016	0.0702	0.0986
1000	A	AR-GARCH	0.1198	0.1154	0.0986	0.0618	0.0942
2000	A	AR-GARCH	0.1186	0.1151	0.1022	0.0631	0.0997
500	B	AR-GARCH	0.1664	0.1094	0.0956	0.0558	0.0856
1000	B	AR-GARCH	0.1558	0.1034	0.0972	0.0524	0.0864
2000	B	AR-GARCH	0.1385	0.0929	0.0993	0.0478	0.0887
500	C	AR-GARCH	0.2158	0.1590	0.1036	0.0568	0.0818
1000	C	AR-GARCH	0.2002	0.1502	0.0994	0.0550	0.0874
2000	C	AR-GARCH	0.1882	0.1442	0.0974	0.0522	0.0932
500	D	AR-GARCH	0.2840	0.2410	0.1058	0.0550	0.0988
1000	D	AR-GARCH	0.2656	0.2274	0.1138	0.0494	0.0906
2000	D	AR-GARCH	0.2460	0.2064	0.1088	0.0544	0.0912

NOTE: The table presents the empirical size of testing zero timing and zero correlation at the 10% significance level. The simulated mutual fund data are with the number of time series observations of 500, 1000, and 2000. Cases A, B, C, and D are defined in Section 3.

Table 2. Test powers for $H_0: \gamma = 0$ at level 10%.

n	Case	ε_t	Newey–West	$\hat{\gamma}$ with $h=0$	$\hat{\gamma}$ with $h=0.2$
500	A	GARCH	0.7694	0.7708	0.5980
1000	A	GARCH	0.9018	0.9018	0.7366
2000	A	GARCH	0.9890	0.9682	0.8628
500	B	GARCH	0.8580	0.8397	0.3923
1000	B	GARCH	0.9509	0.9464	0.4817
2000	B	GARCH	0.9876	0.9866	0.5930
500	C	GARCH	0.7320	0.6994	0.2972
1000	C	GARCH	0.8644	0.8578	0.3582
2000	C	GARCH	0.9514	0.9488	0.4428
500	D	GARCH	0.5954	0.5744	0.2868
1000	D	GARCH	0.7230	0.7150	0.3410
2000	D	GARCH	0.8536	0.8524	0.4432

NOTE: The table presents the empirical power of zero timing at the 10% significance level. The simulated mutual fund data are with the number of time series observations of 500, 1000, and 2000. Cases A, B, C, and D are defined in Section 3.

correlation even adjusting Auto-regression up to 15 periods. We separate the funds with nonzero correlation from others when empirically estimating the timing skills.

3. Simulation Study

We draw 5000 random samples from models (1) and (2) with $d = 1, \alpha = -0.008743, \gamma = 0$ for the study of test size, $\gamma = 0.01$ for test power, $\beta = 0.969277$, $s_1 = p_1 = q_1 = 1$, $r_1 = 0$, $\mu_1 = 0.07726$, $\psi_{1,1} = -003865$, $w_1 = 0.01028$, $a_{1,1} = 0.09749$, $b_{1,1} = 0.90001$, $s = 0$ or 1, $\phi_1 = 0.01167$ for $s = 1$, $r = 0$, $p = q = 1$, $w = 0.00016$, $a = 0.06851$, and $b = 0.93084$. That is, we model $X_{t,1}$ by an AR(1)-GARCH(1,1) process and model ε_t by either a GARCH(1,1) or an AR(1)-GARCH(1,1) process. Note that the settings in the above AR-GARCH models are calibrated from the dataset analyzed in Section 4. We take

Table 3. Number of funds in different portfolios using all funds.

		Weighted least-square estimate			
		Positive	Zero	Negative	Total
Panel A: One-factor model					
Least-square estimate	Positive	91	74	0	165
	Zero	104	1482	189	1775
	Negative	0	366	304	670
	Total	195	1922	493	2610
Panel B: Three-factor model					
Least-square estimate	Positive	137	51	0	188
	Zero	158	1676	271	2105
	Negative	0	118	199	317
	Total	295	1845	470	2610
Panel C: Four-factor model					
Least-square estimate	Positive	132	44	0	176
	Zero	152	1763	241	2156
	Negative	0	107	171	278
	Total	284	1914	412	2610

NOTE: We compare our proposed estimation with the traditional least-square estimation of the timing coefficients in the Treynor–Mazuy approach based on the one-factor model (Panel A), three-factor model (Panel B), and four-factor model (Panel C), using daily data for 2610 funds from September 1, 1998 to December 31, 2018. We report the number of funds in different portfolios.

Table 4. Number of funds in different portfolios using funds passing zero correlations test.

		Weighted least-square estimate			
		Positive	Zero	Negative	Total
Panel A: One-factor model					
Least-square estimate	Positive	80	60	0	140
	Zero	86	1265	146	1497
	Negative	0	320	242	562
	Total	166	1645	388	2199
Panel B: Three-factor model					
Least-square estimate	Positive	110	44	0	154
	Zero	135	1416	232	1783
	Negative	0	103	170	273
	Total	245	1563	402	2210
Panel C: Four-factor model					
Least-square estimate	Positive	111	36	0	147
	Zero	129	1529	209	1867
	Negative	0	97	149	246
	Total	240	1662	358	2260

NOTE: Using funds passing zero correlation test, we compare our proposed estimation with the traditional least-square estimation of the timing coefficients in the Treynor–Mazuy approach based on the one-factor (Panel A), three-factor (Panel B), and four-factor models (Panel C), using daily data for funds passing the zero correlations test from September 1, 1998 to December 31, 2018. We report the number of funds in different portfolios.

$n = 500$, or 1000, or 2000, and consider the following four scenarios for $(\eta_t, \bar{\eta}_{t,1})^\tau$:

- Case A. η_t and $\bar{\eta}_{t,1}$ are independent with standard normal distribution.

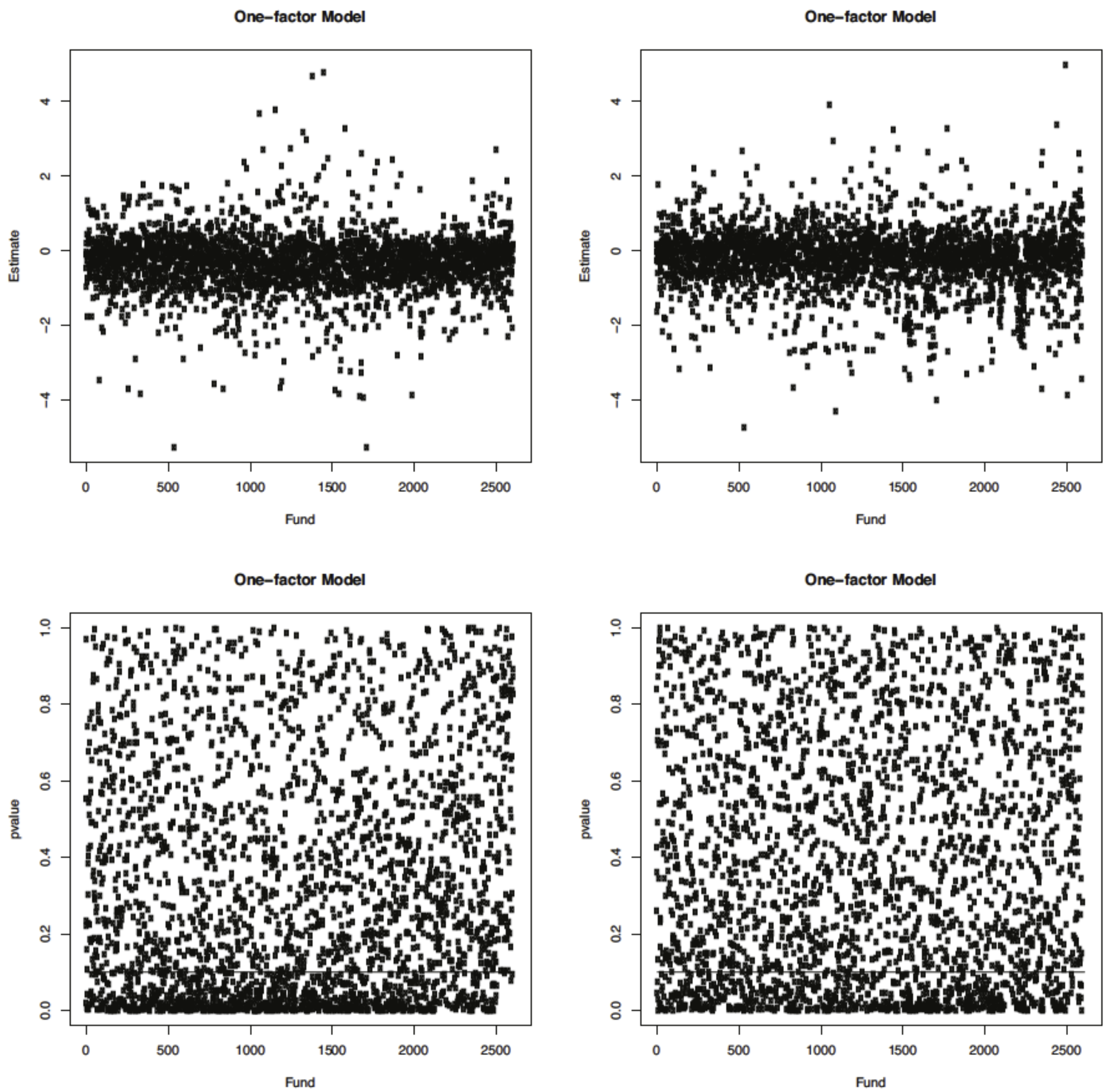

Figure 3. Estimates and p-values for testing zero Treynor–Mazuy market timing skill.
Using the one-factor model, we plot estimates for γ and p-values for testing $H_0 : \gamma = 0$ for each fund. The left panels are the least-squares estimate and the Newey–West t-test using heteroscedasticity and autocorrelation consistent variance estimation with Parzen kernel. The right panels are the proposed weighted estimation and test.

- Case B. $\eta_t = V_t(Z_t - 1)$ and $\bar{\eta}_{t,1} = \bar{V}_t(\bar{Z}_t - 1)$, where $V_t, Z_t, \bar{V}_t, \bar{Z}_t$ are independent with the standard normal distribution, standard exponential distribution, standard normal distribution, and standard exponential distribution, respectively. Hence, $E\eta_t^2 = 1$ and $E\bar{\eta}_{t,1}^2 = 1$.
- Case C. $\eta_t = V_t(Z_t - 1)$ and $\bar{\eta}_{t,1} = V_t(\bar{Z}_t - 1)$, where $V_t, Z_t, \bar{Z}_t$ are independent with the standard normal distribution, standard exponential distribution, and standard exponential distribution, respectively. Hence, $E\eta_t^2 = 1$ and $E\bar{\eta}_{t,1}^2 = 1$.
- Case D. $\eta_t = V_t(Z_t - 1)$ and $\bar{\eta}_{t,1} = \bar{V}_t(Z_t - 1)$, where $V_t, Z_t, \bar{V}_t$ are independent with the standard normal distribution, standard exponential distribution, and standard normal distribution, respectively. Hence, $E\eta_t^2 = 1$ and $E\bar{\eta}_{t,1}^2 = 1$.

Clearly, $X_{t,1}$ and U_t are independent in Cases A and B but dependent in Cases C and D.

We employ the weight function in (4) with $h = 0$ (unweighted estimation), $h = 0.2$ (weighted estimation), $B = 5000$ and the standard exponential distribution for the bootstrap method, and C_X and C_Y replaced by the corresponding 90% sample quantile. For comparison with the conventional method, we also implement the Newey–West t-test using heteroscedasticity and autocorrelation consistent variance estimation with Parzen kernel in the R package "sandwich," which is unable to reduce the heavy tail effect of $H(X_t)\varepsilon_t$ caused by the dependence between η_t and $\bar{\eta}_t$ and the heteroscedasticity in ε_t.

Table 1 reports the test size for zero market timing ability. It shows that (i) the Newey–West t-test is oversized for Cases B, C, and D, (ii) the unweighted test is oversized when η_t and $\bar{\eta}_{t,1}$ are dependent (i.e., Cases C and D), and (iii) the weighted test provides an accurate size for all cases. In general, when $H(X_{t,1})$ has a heavier tail, the Newey–West t-test and unweighted test

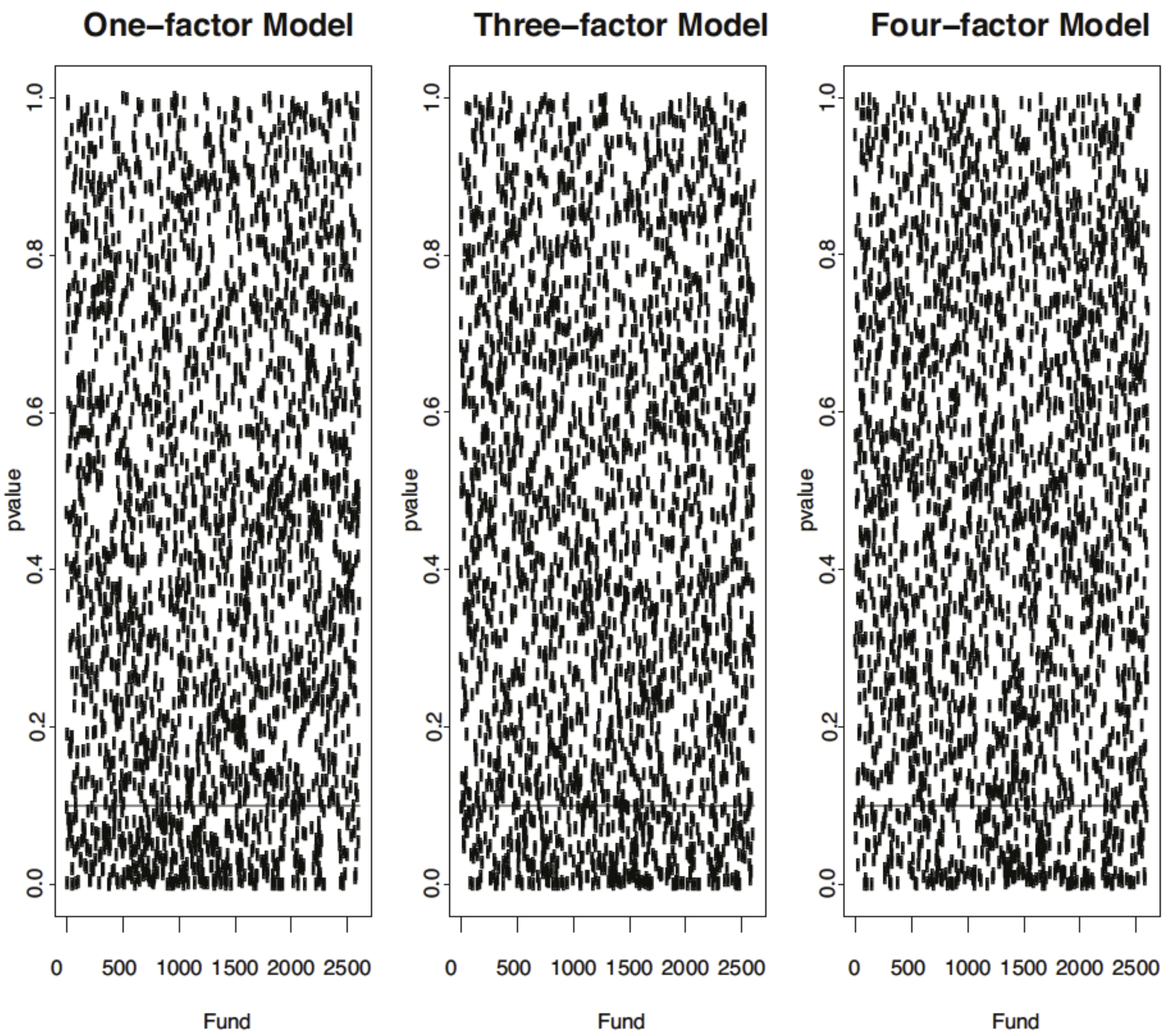

Figure 4. *P*-values for testing zero correlations.
Using the one-factor model (left panel), the three-factor model (middle panel), and the four-factor model (right panel), we plot *p*-values for testing $H_0 : \rho = \mathbf{0}$ with lag $l = 5$ for each fund based on using auto.arima to choose the AR order. The number of funds below the straight line is 411, 400, and 350 for the one-factor, three-factor, and four-factor models, respectively.

have a distorted size due to underestimating uncertainty. Hence, it is vital to take the heavy tails into account in testing zero market timing ability. Table 1 also reports the size for testing zero correlations of U_t's with pre-assigned lag $l = 5$, which shows that the weighted inference is slightly undersized, and the unweighted inference has a severely distorted size for most cases. We report the test power for GARCH errors in Table 2, which shows the weighted test has nontrivial power. Note that it does not make sense to compare the weighted test with the Newey–West t-test and unweighted test in terms of power as the last two tests have a severely distorted test size.

In summary, our simulation study indicates that the proposed test for zero market timing skill has an accurate size. In contrast, the commonly employed Newey–West t-test may have a severely distorted size due to the lack of finite moments.

4. Empirical Analysis

We consider U.S. actively managed mutual funds from September 1998 to December 2018. At the end of each month, we obtain total net assets (TNA), turnover ratios, expense ratios, and other fund characteristics for each share class and aggregate multiple share classes using MFLINK1 provided by Wharton Research Data Services. Fund TNA is the sum of TNA across all share classes of the fund. The fund age is the years of the oldest share class in the fund. Family size is the sum of total net assets under the management of all other funds in the same fund family. The turnover ratio and expense ratio are the averages of them in different share classes of the funds weighted by TNA. Monthly fund flows are the net growth of TNA beyond capital gains and dividends. We focus on equity funds by excluding funds with an average percentage of common stocks lower than 80% of the total net asset. We also exclude index funds, Exchange Traded Funds, and other non-actively-managed funds based on their names and index fund identifiers in the Center for Research in Security Prices (CRSP) Survivorship Bias Free Mutual Fund Database following Ferson and Lin (2014). From the Thomson Reuters Mutual Fund Holdings database, we can exclude nondiversified funds (holding less than 10 stocks) and filter out funds with the Investment Objective Codes, including International, Municipal Bonds, Bond and Preferred, Balanced, and Metals. We further select actively-managed funds following Kacperczyk

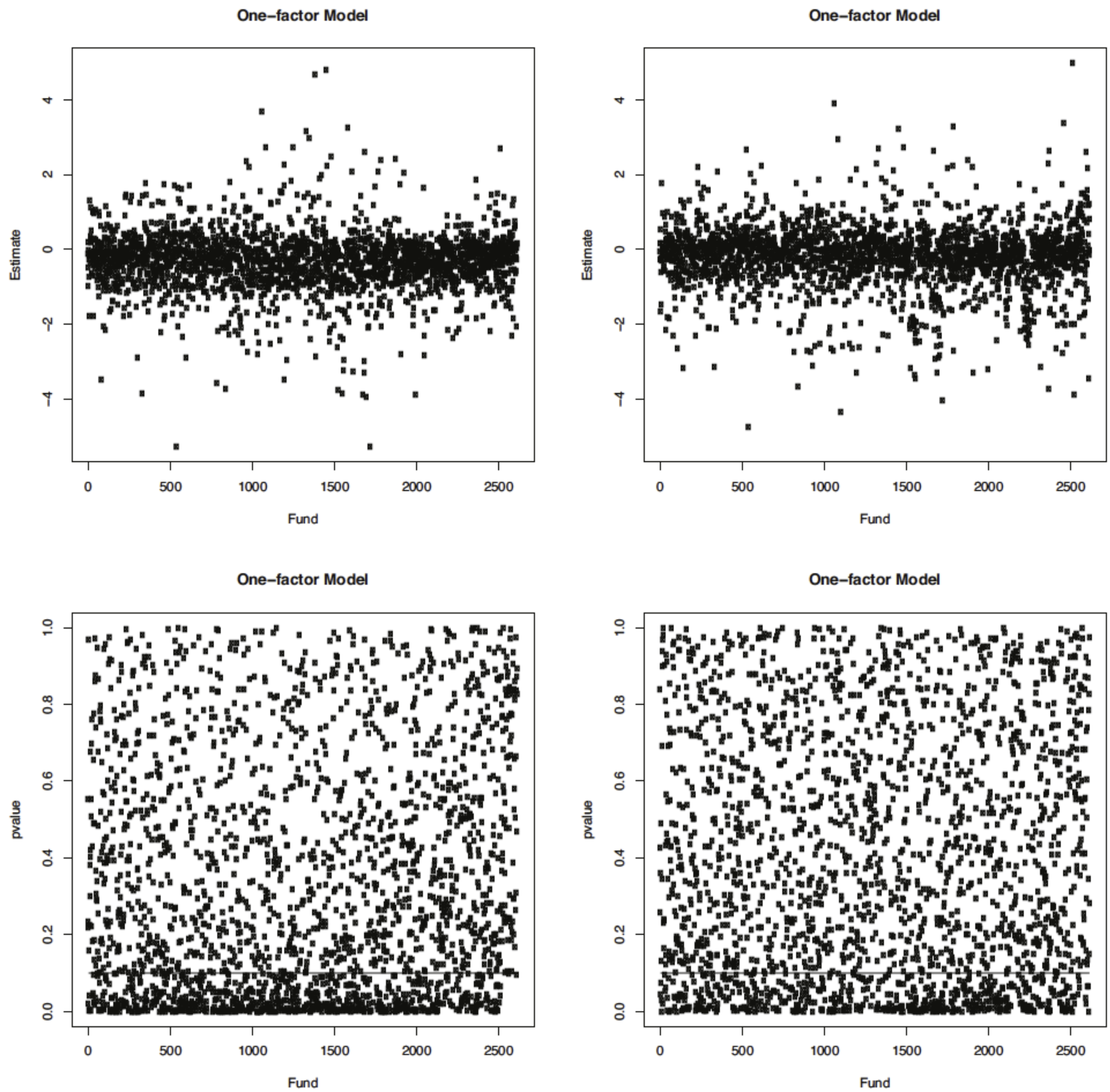

Figure 5. Estimates and p-values for testing zero Treynor–Mazuy market timing skill for 2199 funds.
Using the one-factor model, we plot estimates for γ and p-values for testing $H_0 : \gamma = 0$ for the total 2199 funds, which donot reject the zero correlations test at level 10% by using auto.arima to choose the AR order. The left panels are the least-squares estimate and the Newey–West t-test using heteroscedasticity and autocorrelation consistent variance estimation with Parzen kernel. The right panels are the proposed weighted estimation and test.

et al. (2008), exclude funds with less than $15 million in total net asset, and address the incubation bias issue following Evans (2010). We restrict funds with at least 1000 daily fund returns to ensure robust estimates for timing, which results in 2610 funds in total. For each of these funds, we fit model (1) with $d = 1$ for the one-factor model in Jensen (1968), $d = 3$ for the three-factor model in Fama and French (1996), and $d = 4$ for the four-factor model in Carhart (1997). To save space, we report our empirical analysis for the one-factor model and the Treynor–Mazuy market timing ability, that is, $H(X_{t,1}) = X_{t,1}^2$ in (1). We summarize findings for the three-factor and four-factor models.

To implement the proposed test for zero market timing skill, we need to decide the ARMA model's orders in (2), although we do not need to estimate the GARCH model. We use an AR-GARCH model and select the order in the AR model by using the auto.arima with maximum order 15 in the R statistical software. After selecting the order in the AR model, we compute the pvalue of the proposed test by employing the weight function in (4) with $h = 0.2$, $B = 1000$ for the random weighted bootstrap method, and C_X and C_Y in the weight function replaced by the corresponding 90% sample quantile. Compared with the conventional method, we also implement the Newey–West t-test using heteroscedasticity and autocorrelation consistent variance estimation with Parzen kernel in the R package "sandwich."

4.1. Analysis of Market Timing Ability

We plot the estimates for γ and p-values for testing $H_0 : \gamma = 0$ for each fund by using Newey–West t-test with variance

Table 5. Mutual fund characteristics using all funds.

Variable	Pos. γ	Zero γ	Neg. γ	Pos.-Neg.	t-stat
Panel A: One-factor model					
Variable	Pos. γ	Zero γ	Neg. γ	Pos.-Neg.	t-stat
(Number of funds)	195	1922	493		
Fund characteristic					
logtna	5.6311	5.8505	5.8671	−0.236***	(−4.48)
logage	2.4540	2.4115	2.4757	−0.022	(−1.21)
logtna_family	8.5099	9.5775	9.9334	−1.423***	(−28.07)
turn_ratio	62.9813	76.0953	94.4506	−31.469***	(−15.61)
flow_pct	0.3344	0.3119	0.2020	0.132	(0.85)
exp_ratio	1.2224	1.1736	1.3083	−0.086***	(−10.10)
Panel B: Three-factor model					
Variable	Pos. γ	Zero γ	Neg. γ	Pos.-Neg.	t-stat
(Number of funds)	295	1845	470		
Fund characteristic					
logtna	5.6816	5.8684	5.8116	−0.130***	(−4.16)
logage	2.4070	2.4373	2.4051	0.002	(0.16)
logtna_family	8.9230	9.6195	9.7860	−0.863***	(−33.28)
turn_ratio	68.3027	75.2283	98.7238	−30.421***	(−19.52)
flow_pct	0.2259	0.2748	0.3884	−0.162	(−1.03)
exp_ratio	1.2283	1.1716	1.3143	−0.086***	(−31.98)
Panel C: Four-factor model					
Variable	Pos. γ	Zero γ	Neg. γ	Pos.-Neg.	t-stat
(Number of funds)	284	1914	412		
Fund characteristic					
logtna	5.6734	5.8697	5.7979	−0.125***	(−3.95)
logage	2.4473	2.4373	2.3805	0.06***	(4.87)
logtna_family	9.0647	9.6217	9.6935	−0.629***	(−19.85)
turn_ratio	73.0280	76.6933	93.2254	−20.197***	(−17.00)
flow_pct	0.1740	0.2716	0.4575	−0.284	(−1.64)
exp_ratio	1.2295	1.1808	1.2998	−0.070***	(−24.25)

NOTE: After grouping these 2610 funds into positive, zero, and negative timing ability by using the proposed test for zero timing ability and the sign of weighted least-squares estimate, we report the time-series averages of the monthly cross-sectional means in each portfolio and the differences in means between the two extreme portfolios at the 10% significance level. We compute t-statistics of the differences between the positive and negative timing funds with Newey–West (1987) correction for time-series correlation with six lags. We take the log of the total net assets ($ million), for the age of the fund's oldest share class (in years), and for the fund family's total net assets ($ million). Annual turnover and expense ratio (both in percentage points) are the value-weighted averages across all fund share classes. Fund flow (%) is the TNA-weighted average of flow across all fund share classes. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

Table 6. Mutual fund characteristics using funds passing zero correlations test.

Variable	Pos. γ	Zero γ	Neg. γ	Pos.-Neg.	t-stat
Panel A: One-factor model					
Variable	Pos. γ	Zero γ	Neg. γ	Pos.-Neg.	t-stat
(Number of funds)	166	1645	388		
Fund characteristic					
logtna	5.5788	5.8003	5.7974	−0.219***	(−3.77)
logage	2.3977	2.3828	2.4483	−0.051***	(−2.71)
logtna_family	8.4947	9.4802	9.8885	−1.394***	(−22.47)
turn_ratio	61.7136	76.3961	91.8813	−30.168***	(−12.53)
flow_pct	0.3727	0.3399	0.2160	0.157	(0.91)
exp_ratio	1.2209	1.1789	1.3016	−0.081***	(−9.16)
Panel B: Three-factor model					
Variable	Pos. γ	Zero γ	Neg. γ	Pos.-Neg.	t-stat
(Number of funds)	245	1563	402		
Fund characteristic					
logtna	5.6450	5.8464	5.8027	−0.158***	(−4.40)
logage	2.3670	2.4306	2.3993	−0.032**	(−2.50)
logtna_family	8.9586	9.5292	9.8160	−0.857***	(−22.06)
turn_ratio	66.8987	75.7926	99.4454	−32.547***	(−22.41)
flow_pct	0.1832	0.2826	0.3942	−0.211	(−1.32)
exp_ratio	1.2325	1.1769	1.3066	−0.074***	(−27.78)
Panel C: Four-factor model					
Variable	Pos. γ	Zero γ	Neg. γ	Pos.-Neg.	t-stat
(Number of funds)	240	1662	358		
Fund characteristic					
logtna	5.6264	5.8785	5.7639	−0.137***	(−4.02)
logage	2.4335	2.4402	2.3636	0.070***	(−5.13)
logtna_family	8.9776	9.6062	9.6596	−0.682***	(−18.31)
turn_ratio	74.3336	76.6095	95.3396	−21.006***	(−16.91)
flow_pct	0.1441	0.2750	0.4613	−0.317*	(−1.69)
exp_ratio	1.2236	1.1812	1.2962	−0.073***	(−22.05)

NOTE: After grouping those funds, passing the zero correlations test into positive, zero, and negative timing ability by using the proposed test for zero timing ability and the sign of weighted least-squares estimate, we report the time-series averages of the monthly cross- sectional means in each portfolio and the differences in means between the two extreme portfolios at the 10% significance level. We compute t-statistics of the differences between the positive and negative timing funds with Newey–West (1987) correction for time-series correlation with 6 lags. We take the log of the total net assets ($ million), for the age of the fund's oldest share class (in years), and for the fund family's total net assets ($ million). Annual turnover and expense ratio (both in percentage points) are the value-weighted averages across all fund share classes. Fund flow (%) is the TNA-weighted average of flow across all fund share classes. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

correction (see left panels) and the proposed test (see right panels) and employing the one-factor, three-factor, and four-factor model, respectively. To save space, we only show Figure 3 for the one-factor model. Based on the Newey–West t-test with Parzen kernel variance correction, we find that (i) 1775, 2105, and 2156 out of 2610 funds have zero market timing skill for the one-factor, three-factor, and four-factor models, respectively at 10% significant level, (ii) and 165, 188, and 176 funds for the one-factor, three-factor, and four-factor models, respectively, have positive market timing skill after excluding zero timing funds. In contrast, based on the proposed test, we find that (i) 1922, 1845, and 1914 out of 2610 funds have zero market timing skill for the one-factor, three-factor, and four-factor models, respectively at 10% significant level, (ii) and 195, 295, and 284 funds for the one-factor, three-factor, and four-factor models, respectively have a positive market timing skill after excluding zero timing funds. Hence, the proposed new test finds more funds with a significantly positive market timing skill than the Newey–West t-test, suggesting the importance of taking moment conditions into account.

4.2. Zero Correlations Test

Because the proposed method estimates the ARMA model, and the above analysis fits an AR model using AIC to choose the order, it is vital to check if the employed AR model catches enough correlations in the errors of the factor model. Using the developed test for zero correlations, Figure 4 plots p-values for testing $H_0 : \boldsymbol{\rho} = \mathbf{0}$ with lag $l = 5$ for each fund, which shows 411, 400, and 350 out of 2610 funds reject the zero correlations assumption for the one-factor, three-factor, and four-factor models, respectively at 10% level. This indicates that some funds have very strong series dependence, and the four-factor model is preferred to model the error correlations of a factor model.

4.3. Market Timing Analysis After Zero Correlations Test

We remove these 411, 400, and 350 funds from the total 2610 funds for the one-factor, three-factor, and four-factor models, respectively, which reject the null hypothesis of zero correlations. We redo the market timing analysis above by using the rest of 2199, 2210, and 2260 funds and the one-factor, three-factor,

Table 7. Trade-offs between timing and stock picking skills based on Treynor–Mazuy market timing measure.

Panel A: Including zero timing funds						
One-factor		$\gamma > 0$	$\gamma < 0$		$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	4.37%	53.10%	$\alpha > 0$	8.43%	44.44%
	$\alpha < 0$	25.44%	17.09%	$\alpha < 0$	31.30%	15.82%
Three-factor		$\gamma > 0$	$\gamma < 0$		$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	8.05%	33.14%	$\alpha > 0$	6.48%	35.63%
	$\alpha < 0$	34.25%	24.56%	$\alpha < 0$	37.13%	20.77%
Four-factor		$\gamma > 0$	$\gamma < 0$		$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	8.97%	30.73%	$\alpha > 0$	7.55%	32.64%
	$\alpha < 0$	33.91%	26.40%	$\alpha < 0$	36.63%	23.18%
Panel B: Excluding zero timing funds						
One-factor	835	$\gamma > 0$	$\gamma < 0$	688	$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	0.48%	74.25%	$\alpha > 0$	1.74%	66.28%
	$\alpha < 0$	19.28%	5.99%	$\alpha < 0$	26.60%	5.38%
Three-factor	505	$\gamma > 0$	$\gamma < 0$	765	$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	1.98%	47.92%	$\alpha > 0$	2.09%	53.46%
	$\alpha < 0$	35.25%	14.85%	$\alpha < 0$	36.47%	7.97%
Four-factor	454	$\gamma > 0$	$\gamma < 0$	696	$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	3.08%	43.61%	$\alpha > 0$	3.59%	47.99%
	$\alpha < 0$	35.68%	17.62%	$\alpha < 0$	37.21%	11.21%

NOTE: In this table, we compute the percentage of 2610 funds from September 1, 1998 to December 31, 2018 with each possible combination of signs on the estimates for α and γ using daily fund returns. The left and right parts are the least-squares estimation and the proposed weighted estimation, respectively. In panel B, we use only funds with nonzero market timing measures. The integers denote the total number of funds with nonzero nonparametric Treynor–Mazuy market timing measures based on the Newey–West t-test using heteroscedasticity and autocorrelation consistent variance estimation with Parzen kernel for the left part and proposed weighted test for the right part. We use 10% significance level for testing zero market timing skill.

Table 8. Trade-offs between timing and stock picking skills using funds passing the zero correlation test.

Panel A: Including zero timing funds						
One-factor		$\gamma > 0$	$\gamma < 0$		$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	4.23%	52.34%	$\alpha > 0$	8.50%	42.56%
	$\alpha < 0$	25.92%	17.51%	$\alpha < 0$	32.20%	16.73%
Three-factor		$\gamma > 0$	$\gamma < 0$		$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	8.05%	33.44%	$\alpha > 0$	6.65%	35.75%
	$\alpha < 0$	33.76%	24.75%	$\alpha < 0$	36.83%	20.77%
Four-factor		$\gamma > 0$	$\gamma < 0$		$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	8.85%	31.02%	$\alpha > 0$	7.35%	32.88%
	$\alpha < 0$	33.67%	26.46%	$\alpha < 0$	36.77%	23.01%
Panel B: Excluding zero timing funds						
One-factor	702	$\gamma > 0$	$\gamma < 0$	554	$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	0.28%	73.93%	$\alpha > 0$	1.99%	64.08%
	$\alpha < 0$	19.66%	6.13%	$\alpha < 0$	27.98%	5.96%
Three-factor	427	$\gamma > 0$	$\gamma < 0$	647	$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	1.64%	48.24%	$\alpha > 0$	2.01%	54.25%
	$\alpha < 0$	34.43%	15.69%	$\alpha < 0$	35.86%	7.88%
Four-factor	393	$\gamma > 0$	$\gamma < 0$	598	$\gamma > 0$	$\gamma < 0$
	$\alpha > 0$	2.54%	44.78%	$\alpha > 0$	3.01%	48.16%
	$\alpha < 0$	34.86%	17.81%	$\alpha < 0$	37.12%	11.71%

NOTE: In this table, we compute the percentage of 2199 (2210, 2260) funds passing the zero correlation test for the one-factor, three-factor, and four-factor model from September 1, 1998 to December 31, 2018 with each possible combination of signs on the estimates for α and γ. The left and right parts are the least-squares estimation and the proposed weighted estimation, respectively. In panel B, we use only funds with nonzero market timing measures. The integers denote the total number of funds with nonzero nonparametric Treynor–Mazuy market timing measures based on the Newey–West t-test using heteroscedasticity and autocorrelation consistent variance estimation with Parzen kernel for the left part and proposed weighted test for the right part. We use 10% significance level for testing zero market timing skill.

and four-factor models, respectively. We plot the estimates for γ and p-values for testing $H_0 : \gamma = 0$. To save space, we only show Figure 5 for the one-factor model. Using Newey–West t-test with variance correction, we find that (i) 1497, 1783, and 1867 funds have zero market timing skill for the one-factor, three-factor, and four-factor models, respectively at 10% significant level, (ii) and 140, 154, and 147 funds for the one-factor, three-factor, and four-factor models, respectively have market timing skill after excluding funds with zero market timing skill. On the other hand, using the proposed test, we find that (i) 1645, 1563, and 1662 funds have zero market timing skill for the one-factor, three-factor, and four-factor models, respectively at 10% significant level, (ii) and 166, 245, and 240 funds for the one-factor, three-factor, and four-factor models, respectively have market timing skill after excluding funds with zero market timing skill. As before, results confirm that our new method identifies more funds with a significantly positive market timing skill than the traditional method.

4.4. Fund Characteristics Analysis

Using the proposed test for zero timing ability, we first identify funds with different timing abilities among the total 2610 funds and group them into three portfolios: zero timing, negative timing, and positive timing. Next, using the sign of the proposed weight least-squares estimate, we classify funds with nonzero timing into positive and negative timing groups. For the portfolios with positive, zero, and negative timing ability, we calculate the cross-sectional mean in each month for commonly used fund characteristics, including total net asset, the age of the fund, the size of fund families the funds are in, turnover ratio, fund flow, and expense ratio. Table 5 reports the time series average of the means using the one-factor (panel A), three-factor (panel B), and four-factor (panel C) model. Based on the popular Carhart 4 factor model, we find that compared with funds with negative timing skills, funds with timing ability are smaller, older, and in a small fund family. Interestingly, those funds do not trade as much as funds with perverse timing ability, and they charge less fee to fund investors. The results are robust to different factor models.

Next, we repeat the above analysis only for funds passing the zero correlations test with lag 5 at level 10%. Results are reported in Table 6, similar to the above analysis using the entire sample of 2610 funds.

4.5. Association Study

We apply our method to test market timing and revisit the association between stock picking skill and market timing skill previously investigated by Kon (1983) and others. In Table 7, we compute the percentage of 2610 funds with each possible combination of signs on the estimates for stock picking skill and timing skill. The left and right parts are the least-squares estimation and the proposed weighted estimation, respectively. We report the results separately for the one-factor, three-factor, and four-factor models. In panel B, we use only funds with nonzero market timing measures. We find strong evidence of a tradeoff between timing and stock picking. If a fund has a positive timing skill, it is more likely to present a negative alpha. The tradeoff is especially strong when we remove funds with zero timing ability, as in Panel B in Table 7. We also repeat this analysis using funds, which do not reject the null hypothesis of zero correlations with lag 5 at 10% level. Results are reported in Table 8, reaching the same conclusions as above.

5. Conclusions

It is argued to be decisive in using daily fund returns to estimate market timing ability for actively managed U.S. equity funds. This article identifies some econometric issues, including heteroscedasticity, correlated errors, and heavy tails. These issues challenge using the Newey–West t-test with variance correction due to the lack of finite moments. We propose to model the errors in a factor model by an ARMA-GARCH process, use a weighted least-squares estimation to ensure a normal limit, and employ a random weighted bootstrap method to quantify the estimation uncertainty. We apply the proposed test for zero market timing to the sample of actively managed U.S. equity funds and find more funds with a positive timing skill than the traditional method. Funds with positive timing ability are smaller, older, and in a small fund family, they do not trade as frequently as funds with perverse timing ability, and they charge fewer fees to fund investors. We also revisit the problem of a tradeoff between timing and stock picking skills empirically. After removing funds with zero timing ability, we find robust results to support that when focusing on timing, fund managers may sacrifice stock picking and vice versa. Our future research will develop a simultaneous test to group funds into three categories with zero, negative, and positive market timing skill by allowing heavy tails, serial correlation, heteroscedasticity, and cross-sectional dependence.

Appendix: Theoretical Derivations

Let $\mathcal{F}_t$ denote the σ-field generated by $\{\eta_{s_1}, \bar{\eta}_{s_2}^\tau : s_1 \le t, s_2 \le t+1\}$. Note that we include $\bar{\eta}_{t+1}$ in $\mathcal{F}_t$ because of condition (3). We use $\xrightarrow{d}$ and $\xrightarrow{p}$ to denote the convergence in distribution and in probability, respectively.

Proof of Theorem 1. For $t = 1, \ldots, n$, define

$$Z_t(\boldsymbol{\theta}) = w_t^{-1} U_t(\boldsymbol{\theta}) \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \text{ and}$$

$$\boldsymbol{\Gamma}_t(\boldsymbol{\theta}) = w_t^{-1} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\tau} + w_t^{-1} U_t(\boldsymbol{\theta}) \frac{\partial^2 U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau}.$$

Because $\frac{1}{n}\sum_{t=1}^n Z_t(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, it follows from Taylor expansion that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \{\frac{1}{n}\sum_{t=1}^n \boldsymbol{\Gamma}_t(\boldsymbol{\theta}_0)\}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t(\boldsymbol{\theta}_0) + o_p(1).$$

By (3), $\{Z_t, \mathcal{F}_t\}_{t=1}^n$ is a martingale difference sequence. Like Ling (2007) and He et al. (2020), we can show that $E||Z_t(\boldsymbol{\theta}_0)||^{2+\delta} < \infty$. Hence, by the central limit theorem for a martingale difference sequence in Hall and Heyde (1980), we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{6}$$

where $\boldsymbol{\Sigma} = \lim_{t\to\infty} E\{w_t^{-2} U_t^2(\boldsymbol{\theta}_0) \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\}$. Using the weak law of large numbers for a martingale difference sequence in Hall and Heyde (1980) and the ergodicity for $\{\varepsilon_t\}$ and $\{X_t\}$, we can show that

$$\frac{1}{n}\sum_{t=1}^n w_t^{-1} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\tau}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{p} \lim_{t\to\infty} E\{w_t^{-1} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\tau}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\} := \boldsymbol{\Gamma}$$

and $\frac{1}{n}\sum_{t=1}^n w_t^{-1} U_t(\boldsymbol{\theta}_0) \frac{\partial^2 U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{p} \mathbf{0}$, that is,

$$\frac{1}{n}\sum_{t=1}^n \boldsymbol{\Gamma}_t \xrightarrow{p} \boldsymbol{\Gamma}. \tag{7}$$

By (6) and (7), we have $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1})$, that is, Theorem 1 holds. □

Proof of Theorem 2. For $t = 1, \ldots, n$, and $b = 1, \ldots, B$, define

$$Z_t^b(\boldsymbol{\theta}) = \xi_t^b w_t^{-1} U_t(\boldsymbol{\theta}) \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

$$\boldsymbol{\Gamma}_t^b(\boldsymbol{\theta}) = \xi_t^b w_t^{-1} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\tau} + \xi_t^b w_t^{-1} U_t(\boldsymbol{\theta}) \frac{\partial^2 U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau},$$

$$W_t^b(\boldsymbol{\theta}) = (\xi_t^b - 1) w_t^{-1} U_t(\boldsymbol{\theta}) \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad W_t(\boldsymbol{\theta}) = w_t^{-1} U_t(\boldsymbol{\theta}) \frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Like the proof of Theorem 1, we have

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}}^b - \boldsymbol{\theta}_0) &= \{\tfrac{1}{n}\textstyle\sum_{t=1}^n \boldsymbol{\Gamma}_t^b(\boldsymbol{\theta}_0)\}^{-1} \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n Z_t^b(\boldsymbol{\theta}_0) + o_p(1) \\ &= \{\tfrac{1}{n}\textstyle\sum_{t=1}^n \boldsymbol{\Gamma}_t(\boldsymbol{\theta}_0)\}^{-1} \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n Z_t^b(\boldsymbol{\theta}_0) + o_p(1) \\ &= \boldsymbol{\Gamma}^{-1} \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n Z_t^b(\boldsymbol{\theta}_0) + o_p(1). \end{aligned}$$

Hence, as $n \to \infty$,

$$\begin{cases} \sqrt{n}(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}}) = \boldsymbol{\Gamma}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n W_t^b(\boldsymbol{\theta}_0) + o_p(1), \\ n(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}})^\tau = \boldsymbol{\Gamma}^{-1} \frac{1}{n} \sum_{t=1}^n W_t^b(\boldsymbol{\theta}_0)(W_t^b(\boldsymbol{\theta}_0))^\tau \boldsymbol{\Gamma}^{-1} + o_p(1), \end{cases}$$

implying that

$$\begin{aligned} &\tfrac{1}{B}\textstyle\sum_{b=1}^B n(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}})^\tau \\ &\quad= \boldsymbol{\Gamma}^{-1} \tfrac{1}{n} \textstyle\sum_{t=1}^n \tfrac{1}{B} \sum_{b=1}^B W_t^b(\boldsymbol{\theta}_0)(W_t^b(\boldsymbol{\theta}_0))^\tau \boldsymbol{\Gamma}^{-1} + o_p(1) \\ &\quad= \boldsymbol{\Gamma}^{-1} \tfrac{1}{n} \textstyle\sum_{t=1}^n W_t(\boldsymbol{\theta}_0) W_t^\tau(\boldsymbol{\theta}_0) \boldsymbol{\Gamma}^{-1} + o_p(1) \\ &\quad= \boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1} + o_p(1), \end{aligned}$$

that is, the theorem holds.

□

Acknowledgments

The authors appreciate the editor, Professor Jianqing Fan, an associate editor, and two reviewers for helpful comments.

Funding

Peng's research was partly supported by the Simons Foundation and the NSF grant of DMS-2012448. Jiang's research was partly supported by Tsinghua University Initiative Scientific Research Program and Institute for Industrial Innovation and Finance at Tsinghua University and by Tsinghua National Laboratory for Information Science and Technology.

References

Back, K., Crane, A. D., and Crotty, K. (2018), "Skewness Consequences of Seeking Alpha," *Review of Financial Studies*, 31, 4720–4761. [185]

Basrak, B., Davis, R. A., and Mikosch, T. (2002), "Regular Variation of GARCH Processes," *Stochastic Processes and their Applications*, 99, 95–115. [186]

Becker, C., Ferson, W., Myers, D. H., and Schill, M. J. (1999), "Conditional Market Timing with Benchmark Investors," *Journal of Financial Economics*, 52, 119–148. [184]

Bollen, N. P., and Busse, J. A. (2001), "On the Timing Ability of Mutual Fund Managers," *Journal of Finance*, 56, 1075–1094. [184,185]

Bollen, N. P., and Busse, J. A. (2005), "Short-Term Persistence in Mutual Fund Performance," *Review of Financial Studies*, 18, 569–597. [184,185]

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. [184]

Breen, W., Jagannathan, R., and Ofer, A. R. (1986), "Correcting for Heteroscedasticity in Tests for Market Timing Ability," *Journal of Business*, 59, 585–598. [184]

Busse, J. A. (1999), "Volatility Timing in Mutual Funds: Evidence from Daily Returns," *Review of Financial Studies*, 12, 1009–1041. [185]

Carhart, M. M. (1997), "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82. [185,192]

Chance, D. M., and Hemler, M. L. (2001), "The Performance of Professional Market Timers: Daily Evidence from Executed Strategies," *Journal of Financial Economics*, 62, 377–411. [184,185]

Chang, E. C., and Lewellen, W. G. (1984), "Market Timing and Mutual Fund Investment Performance," *Journal of Business*, 57, 57–72. [184]

Crane, A. D., and Crotty, K. (2018), "Passive Versus Active Fund Performance: Do Index Funds Have Skill?," *Journal of Financial and Quantitative Analysis*, 53, 33–64.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007. [184]

Evans, R. (2010), "Mutual Fund Incubation," *Journal of Finance*, 65, 1581–1611. [192]

Fama, E. F., and French, K. R. (1996), "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 55–84. [185,192]

Fan, J., Ke, Y., Sun, Q., and Zhou, W. X. (2019), "FarmTest: Factor-Adjusted Robust Multiple Tests with False Discovery Control," *Journal of the American Statistical Association*, 114, 1880–1893. [185]

Fan, J., Liao, Y., and Yao, J. (2015), "Power Enhancement in High Dimensional Cross Sectional Tests," *Econometrica*, 83, 1497-1541. [185]

Ferson, W., and Lin, J. (2014), "Alpha and Performance Measurement: The Effects of Investor Disagreement and Heterogeneity," *The Journal of Finance*, 69, 1565–1596. [191]

Goetzmann, W. N., Ingersoll, J., and Ivković, Z. (2000), "Monthly Measurement of Daily Timers," *Journal of Financial and Quantitative Analysis*, 35, 257–290. [184,185]

Grinblatt, M., and Titman, S. (1988), *The Evaluation of Mutual Fund Performance: An Analysis of Monthly Returns*. John E. Anderson Graduate School of Management, University of California, California. [184]

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Applications*, New York: Academic Press. [195]

He, Y., Hou, Y., Peng, L., and Shen, H. (2020), "Inference for Conditional Value-at-Risk of a Predictive Regression," *Annals of Statistics*, 48, 3442–3464. [186,195]

Henriksson, R. D. (1984), "Market Timing and Mutual fund Performance: An Empirical Investigation," *Journal of Business*, 57, 73–96. [184,185]

Henriksson, R. D., and Merton, R. C. (1981), "On Market Timing and Investment Performance II. Statistical Procedures for Evaluating Skills," *Journal of Business*, 54, 513–533. [184,185]

Hill, B. M. (1975), "A Simple General Approach to Inference About the Tail of a Distribution," *Annals of Statistics*, 3, 1163–1174. [184]

Jensen, M. C. (1968), "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Finance*, 23, 389–416. [185,192]

Jagannathan, R., and Korajczyk, R. A. (1986), "Assessing the Market Timing Performance of Managed Portfolios," *Journal of Business*, 59, 217–235. [185]

Kacperczyk, M., Sialm, C., and Zheng, L. (2008), "Unobserved Actions of Mutual Funds," *The Review of Financial Studies*, 21, 2379–2416. [192]

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2014), "Time-Varying Fund Manager Skill," *Journal of Finance*, 69, 1455–1484. [185]

Kon, S. J. (1983), "The Market-Timing Performance of Mutual Fund Managers," *Journal of Business*, 56, 323–347. [185,194]

Ling, S. (2007), "Self-Weighted and Local Quasi-Maximun Likelihood Estimators for ARMA-GARCH/IGARCH Models," *Journal of Econometrics*, 140, 849–873. [184,186,195]

Liu, J. M., Chen, R., and Yao, Q. (2010), "Nonparametric Transfer Function Models," *Journal of Econometrics*, 157, 151–164. [186]

Mamaysky, H., Spiegel, M., and Zhang, H. (2008), "Estimating the Dynamics of Mutual Fund Alphas and Betas," *The Review of Financial Studies*, 21, 233–264. [184,185]

Treynor, J., and Mazuy, K. (1966), "Can Mutual Funds Outguess the Market?," *Harvard Business Review*, 44, 131–146. [184,185]

Xiao, Z., Linton, O. B., Carroll, R. J., and Mammen, E. (2003), "More Efficient Local Polynomial Estimation in Nonparametric Regression with Auto Correlated Errors," *Journal of American Statistical Association*, 98, 980–992. [186]

Zhu, K. (2016), "Bootstrapping the Portmanteau Tests in Weak Auto-Regressive Moving Average Models," *Journal of the Royal Statistical Society*, 78, 463–485. [187]