Journal of Economic Dynamics & Control 150 (2023) 104635

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Nonparametric tests for market timing ability using daily mutual fund returns☆

Jing Ding[a], Lei Jiang[b,*], Xiaohui Liu[c], Liang Peng[d]

[a] School of Economics and Management, Tongji University, Shanghai, 200092, China
[b] School of Economics and Management, Tsinghua University, Beijing, 100084, China; Ambassador Crawford College of Business and Entrepreneurship, Kent State University, Kent, OH, 44240, USA.
[c] School of Statistics, and Key Laboratory of Data Science in Finance and Economics, Jiangxi University of Finance and Economics, Nanchang, 330013, China
[d] Department of Risk Management and Insurance, Georgia State University, Atlanta, GA, 30303, USA

ARTICLE INFO

Article history:
Received 19 January 2022
Revised 22 January 2023
Accepted 6 March 2023
Available online 9 March 2023

JEL classification:
G11
G23
C58

Keywords:
Market timing
Mutual fund
Weighted nonparametric measure

ABSTRACT

When using daily mutual fund returns to study market timing ability, heavy tails and heteroscedasticity significantly challenge the existing methods. We propose a weighted nonparametric measure and test for market timing. The test finds different results from the traditional parametric inference concerning timing. By examining the holding characteristics of the funds with different levels of timing ability, we find that funds with positive timing ability hold stocks with lower trading frictions. We find evidence of a tradeoff between market timing ability and stock picking skill after excluding funds with zero timing ability, which is robust to different benchmark models.

1. Introduction

Whether mutual funds possess timing ability has been debated since Treynor and Mazuy (1966), Henriksson and Merton (1981) published seminal papers. Applying parametric measures (i.e., a coefficient related to market volatility in a factor model) to monthly fund returns, researchers find that, on average, U.S. equity funds have negative market timing ability; see, for example, Chang and Lewellen (1984), Henriksson (1984), Grinblatt and Titman (1988), Becker et al. (1999). Jiang (2003) finds similar results using ratios of change in fund return to that in market return or change in residual of modeling fund returns to that of modeling market returns. The test in Jiang (2003) assumes independent ratios, which requires modeling and estimating both fund returns and market returns.

☆ We thank two reviewers for their helpful comments. Ding's work was supported by the National Natural Science Foundation of China (NSFC) Grant (72121001). Liu's work was supported by the National Natural Science Foundation of China (NSFC) Grant (11971208) and the NSF of Jiangxi Province (No. 2018ACB21002, 20224ACB211003). Peng's research was partly supported by the Simons Foundation and the NSF grant of DMS-2012448. The research was also supported by Tsinghua University Initiative Scientific Research Program, Institute for Industrial Innovation and Finance at Tsinghua University and Tsinghua National Laboratory for information Science and Technology.

* Corresponding author.
E-mail address: jianglei@sem.tsinghua.edu.cn (L. Jiang).

https://doi.org/10.1016/j.jedc.2023.104635
0165-1889/

Since most well-diversified equity funds hold and trade hundreds of stocks, we expect that mutual funds will time the market and trade very frequently in response to continuous information, time-varying stock market liquidity, and daily capital flow.[1] Thus, compared with monthly data, daily fund returns should better capture the timing ability of mutual funds. Goetzmann et al. (2000) find that estimating market timing ability based on monthly returns results in a downward bias when funds time the market daily. Bollen and Busse (2001) indicate that using daily fund returns provides better statistical power when testing for the absence of timing ability than using monthly fund returns, even if mutual funds time market at a monthly frequency. Over time, more and more researchers have reached the consensus that, because of the dynamic trading effect, using daily fund returns to evaluate fund timing ability should be as accurate as, if not more accurate than, using monthly returns. For example, Chance and Hemler (2001), Bollen and Busse (2005), Mamaysky et al. (2008) all estimate fund managers' timing ability based on daily fund returns and a factor model with a coefficient for market timing measure.[2] Recently, Back et al. (2018) use daily fund returns and find a negative association between nonparametrically measured market timing ability[3] and stock picking skill measured by alpha (see Fama and French 2010). However, these papers do not explicitly consider the stylized facts of mutual fund daily returns, such as heavy tails and heteroscedasticity, making the conventional test fail as the limit is not normal. Breen et al. (1986) show the importance of heteroscedasticity in parametrically inferring and evaluating the timing ability of mutual fund managers and find that heteroscedasticity affects the power of the timing test.

By estimating the tail indexes of daily returns of U.S. equity funds, we find that empirically many funds do not have enough finite moments, which confirms heavy tails in those funds. Hence, the coskewness (a nonparametric market timing measure) in Back et al. (2018) may not be well-defined (i.e., finite), or its estimator has a nonnormal limit. This paper first refines their measure and proposes nonparametric market timing measures for individual funds requiring fewer finite moments. We also develop robust and efficient methods to test whether a fund has zero timing ability based on our measure. To precisely capture heteroscedasticity and volatility persistence of fund daily returns and to consider autocorrelation of daily risk factors, we assume that the factor model for fund daily excess return has GARCH errors (see Bollerslev 1986; Engle 1982), and each risk factor follows an ARMA-GARCH process.[4] This assumption is more realistic and different from most previous empirical studies of mutual funds, such as Carhart (1997), in which researchers use the pure white noise assumption.[5]

We define a weighted nonparametric market timing measure to reduce the heavy tail effect. Since the weighted and unweighted measures have the same sign, using either one leads to the same negative, zero, and positive timing classification. Therefore, the critical issue is to quantify the estimation uncertainty. Uncertainty quantification becomes challenging when using the unweighted measure without enough finite moments as the limit is nonnormal and standard techniques provide an incorrect estimation. We derive the asymptotic normality of the proposed weighted measure estimator. We use a random weighted bootstrap method to avoid estimating the complicated asymptotic variance due to GARCH errors for daily returns, the ARMA-GARCH models for factors, and the weighted inference. We also provide an extension to generalized GARCH errors in a remark.

Applying the proposed test for the absence of market timing ability to all actively managed U.S. equity funds using daily fund returns from September 1, 1998 to December 31, 2018, we find that out of 534 funds with positive Treynor-Mazuy timing ability based on the traditional parametric method, 344 of them have zero or perverse timing ability using our weighted nonparametric method. The traditional parametric method identifies 1837 funds with zero Treynor-Mazuy timing ability, while the weighted nonparametric method identifies 2426 funds with zero timing, and 1522 of them overlap. These significant differences may come from the fact that the traditional least squares estimate and t-test do not take the heavy tail and heteroscedasticity into account, or the parametric model is misspecified.

What kind of funds is likely to show timing ability? After grouping funds into three categories of positive, zero, and negative nonparametric market timing ability, we find that different groups hold stocks with different characteristics. For example, funds with significantly positive ability hold big stocks and stocks with lower values of Amihud ratio and volatility of liquidity based on dollar trading volume. They also hold stocks with fewer zero trading days than funds with negative timing ability. Therefore, funds with timing ability tend to have stocks with lower trading frictions (according to the classification in Hou et al. 2015). The results hold even after controlling for spurious timing bias. One potential explanation for this phenomenon is that, when facing the dynamics of market returns, only funds holding liquid stocks time the market and show significantly positive timing ability. In contrast, other funds are constrained by the high trading costs and give up their opportunity to time the market. Another reason could be that fund managers with timing ability intentionally hold stocks with low trading frictions because they expect to time the market and generate high turnover in the future.

[1] Busse et al. (2019) find that 20% of funds with the highest trade regularity execute 1.66 daily trades for each stock in their portfolios, on average.

[2] Some researchers propose a significantly different approach based on stock holding by mutual funds to study the market timing and use daily stock returns and stock holdings by mutual funds to estimate betas (see Elton et al. 2012; Jiang et al. 2007).

[3] As proved in their paper, the coskewness is proportional to the Treynor-Mazuy market timing measure. Furthermore, Back et al. (2018) also indicate that part of the coskewness may be attributed to "passive timing", i.e., investing stocks with option-like returns (Jagannathan and Korajczyk, 1986).

[4] Because we test for zero market timing individually rather than simultaneously, cross-sectional dependence plays no role in our study. We refer to Jiang et al. (2021) for an excellent review of statistical issues in mutual fund market timing tests.

[5] A notable exception is Busse (1999), who models daily stock market returns by an EGARCH process for studying the volatility timing of mutual funds.

Given our statistically identified distinction between funds with zero and nonzero timing ability, we reexamine the association between fund alpha and market timing for each fund. Back et al. (2018) find a tradeoff between alpha and timing, which may be driven by the negative effect of seeking alpha on coskewness. Kacperczyk et al. (2014) provide empirical evidence that market timing and stock picking may appear separately during different market conditions. Both papers include funds with zero market timing ability in their sample, which may blur the association. Motivated by Hansen (2005), Hansen et al. (2011), we reexamine the association by excluding funds with zero timing ability and find that the negative association between market timing ability and stock picking skill is more substantial. The tradeoff is robust to the selection of the factor models. Finally, different from stock pricking skill, alpha, we find that mutual funds with significantly positive timing ability based on our weighted nonparametric measure provide a higher net return to the investors in the long run than poor timers, which is consistent with the expectation that market timing ability of fund managers is valuable to investors. However, we can not see such a pattern based on the traditional parametric approach.

Our paper is related to but different from Back et al. (2018). First, our unweighted nonparametric Treynor-Mazuy market timing measure is the same as the coskewness measure in Back et al. (2018). Back et al. (2018) find a tradeoff between stock picking skill and market timing ability for the one-factor model using daily data. However, they do not derive a test for zero coskewness and can not exclude funds with insignificant market timing ability from their samples when analyzing the tradeoff. Our paper develops such a test and finds evidence of the tradeoff after excluding funds with zero market timing ability, regardless of the factor model selected. Second, this paper finds that the coskewness in Back et al. (2018) is ill-defined for funds without enough finite moments and that its estimator has a nonnormal limit for those funds. Therefore, this paper further defines a weighted nonparametric market timing measure and develops a test for zero timing ability for daily returns with fewer finite moments.[6]

Our paper is also related to the literature measuring mutual fund market timing ability: the standard Treynor-Mazuy model and Henriksson-Merton model test the nonlinear relation between fund returns and contemporaneous market returns. However, the nonlinear relation can also be induced by reasons other than active market timing ability. Possible reasons are the spurious timing from holding stocks with option-like features (Jagannathan and Korajczyk, 1986) and the artificial timing caused by funds changing their beta in response to the previous market return. Our paper uses daily returns to overcome the dynamic trading effect and solves underestimating timing ability using monthly fund returns when funds are daily timers (Bollen and Busse, 2001; Goetzmann et al., 2000). Our developed method is robust against heavy tails and heteroscedasticity of fund daily returns.

We organize the paper as follows. Section 2 presents the definitions of the nonparametric market timing measure and its weighted version, the model, and methods for testing the absence of market timing ability. Sections 3 and 4 are our simulation study and empirical analysis of U.S. equity funds. Section 5 concludes. The choice of weights, regularity conditions, and theoretical proofs are in Appendices A, B, and C, respectively.

2. Measures, models, and methods

Suppose Y_t is the daily fund excess return at time t and $\boldsymbol{X}_t = (X_{t,1}, \dots, X_{t,d})^\top$ represents d factors with $X_{t,1}$ being the daily market excess return, where $A^\top$ denotes the transpose of the matrix or vector A. As we use daily returns, according to Breen et al. (1986), it is vital to consider the heavy tail, heteroscedasticity, and volatility persistence features when estimating market timing measures. Since the seminar papers of Engle (1982), Bollerslev (1986), ARMA-GARCH models have been popular in modeling both correlation and heteroscedasticity, which motivates the use of the following model for daily fund returns and factors:

$$\begin{cases} Y_t = \alpha + \boldsymbol{\beta}^\top \boldsymbol{X}_t + \varepsilon_t,\ \ \varepsilon_t = \eta_t\sigma_t,\ \ \sigma_t^2 = w + \sum_{i=1}^{p} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \\ X_{t,k} = \mu_k + \sum_{i=1}^{s_k} \phi_{i,k} X_{t-i,k} + \sum_{j=1}^{r_k} \psi_{j,k} \bar{\varepsilon}_{t-j,k} + \bar{\varepsilon}_{t,k},\ \ \bar{\varepsilon}_{t,k} = \bar{\eta}_{t,k}\bar{\sigma}_{t,k}, \\ \bar{\sigma}_{t,k}^2 = w_k + \sum_{i=1}^{p_k} a_{i,k} \bar{\varepsilon}_{t-i,k}^2 + \sum_{j=1}^{q_k} b_{j,k} \bar{\sigma}_{t-j,k}^2 \text{ for } k = 1, \dots, d, \end{cases} \tag{1}$$

where each of $\{\eta_t\}, \{\bar{\eta}_{t,1}\}, \dots, \{\bar{\eta}_{t,d}\}$ is a sequence of independent and identically distributed random variables with mean zero and variance one, and we will specify the dependence among sequences later. We model the conditional mean of each factor as an ARMA process because the daily market factor, SMB factor, and Momentum factor are significantly autocorrelated, with Q-statistics being 19.51, 14.17, and 71.86, respectively. An extension to generalized GARCH models is straightforward and is presented in a remark.

In the existing literature, researchers use α in (1) to measure fund managers' stock picking skills. Both parametric and nonparametric methods have been proposed to estimate fund managers' market timing ability. The parametric method adds a related volatility variable to the factor model and uses the corresponding coefficient to measure market timing. For example, given the market volatility variable $H(X_{t,1})$, researchers employ the coefficient of $\tilde{\gamma}$ in the following factor model:

[6] The test in Jiang (2003) does not suffer from the higher moment issue.

$$Y_t = \tilde{\alpha} + \tilde{\boldsymbol{\beta}}^\top \mathbf{X}_t + \tilde{\gamma} H(X_{t,1}) + \tilde{\varepsilon}_t, \tag{2}$$

which results in the so-called Treynor-Mazuy (Henriksson-Merton) market timing measure when $H(X_{t,1}) = X_{t,1}^2$ ($H(X_{t,1}) = \max(0, X_{t,1})$). Estimation and test for zero market timing ability use the least squares estimation and t-test, requiring enough finite moments. As this market timing measure $\tilde{\gamma}$ is defined as a parameter in a linear model, we call the measure, estimation, and test parametric ones. Recently, Jiang et al. (2022) develop a test for $H_0 : \tilde{\gamma} = 0$ by taking some stylized facts of daily data into account.

Alternatively, instead of adding a volatility variable into the factor model directly, we can also measure market timing ability nonparametrically by using the covariance between errors in the factor model (i.e., ε_t in (1)) and a volatility variable (i.e., $H(X_{t,1})$). An obvious advantage of studying nonparametric measures is the robustness against model misspecification. This leads to the nonparametric market timing measure

$$\gamma = E(\varepsilon_t H(X_{t,1})), \tag{3}$$

which is the nonparametric Treynor-Mazuy (Henriksson-Merton) market timing measure when we use $H(X_{t,1}) = X_{t,1}^2$ ($H(X_{t,1}) = \max(0, X_{t,1})$). Especially, the nonparametric Treynor-Mazuy market timing measure is the same as the coskewness studied in Back et al. (2018).

Put $\mu_H = EH(X_{t,1})$ and write Eq. (2) as

$$Y_t = \tilde{\alpha} + \tilde{\gamma}\mu_H + \tilde{\boldsymbol{\beta}}^\top \mathbf{X}_t + \varepsilon_t \text{ with } \varepsilon_t = \tilde{\varepsilon}_t + \tilde{\gamma}\{H(X_{t,1}) - \mu_H)\}.$$

Then, the nonparametric market timing measure defined in (3) becomes

$$\begin{aligned}\gamma &= E\{\tilde{\varepsilon}_t H(X_{t,1}) + \tilde{\gamma}(H(X_{t,1}) - \mu_H)H(X_{t,1})\} = \tilde{\gamma}\left\{EH^2(X_{t,1}) - \mu_H^2\right\} \\ &= \tilde{\gamma}E(H(X_{t,1}) - \mu_H)^2\end{aligned}$$

because of $E(\tilde{\varepsilon}_t H(X_{t,1})) = 0$ in (2). Therefore, the nonparametric market timing measure is proportional to the parametric market timing measure. Since $E(H(X_{t,1}) - \mu_H)^2$ is non-negative and only determined by the factors, the nonparametric measure will not change the cross-sectional ranking of funds under the parametric measure theoretically if all funds have the same time horizons. In practice, because of many funds with zero market timing ability, estimates may have a misclassified sign for many funds. Hence, it is crucial to test for zero market timing ability and separate funds by negative, zero, and positive market timing ability for the studies, such as the characteristics of the funds with different timing abilities and the association between stock picking ability and market timing ability.

The naive nonparametric estimator for (3) is

$$\hat{\gamma} = \frac{1}{n}\sum_{t=1}^{n}\{Y_t - \hat{\alpha} - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_t\}H(X_{t,1}), \tag{4}$$

where $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ are the least squares estimators of α and $\boldsymbol{\beta}$, minimizing the following least squares

$$\sum_{t=1}^{n}\{Y_t - \alpha - \boldsymbol{\beta}^\top \mathbf{X}_t\}^2.$$

To ensure that $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\gamma}$ have a joint normal limit, we need at least

$$E\{\varepsilon_t H(X_{t,1})\}^2 < \infty. \tag{5}$$

When ε_t and $X_{t,1}$ are independent, condition (5) is equivalent to

$$E(\varepsilon_t^2) < \infty \text{ and } E(H^2(X_{t,1})) < \infty.$$

However, when ε_t and $X_{t,1}$ are dependent, condition (5) may require that

$$E|\varepsilon_t|^{2p} < \infty \quad \text{and} \quad E|H(X_{t,1})|^{2q} < \infty \quad \text{for some} \quad p \geq 1, q \geq 1, p^{-1} + q^{-1} = 2 \tag{6}$$

by using Hölder's inequality.

It is known that many daily financial returns are heavy-tailed distributed. When $\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j = 1$, $\{\varepsilon_t\}$ is stationary but has an infinite variance. The nonparametric market timing measure in (3) is ill-defined when $E(|\varepsilon_t H(X_{t,1})|) = \infty$, and the asymptotic distribution of $\hat{\gamma}$ is nonnormal when $E(\varepsilon_t^2 H^2(X_{t,1})) = \infty$, which challenges the test for the absence of market timing ability. When ε_t and $X_{t,1}$ are dependent, and $H(X_{t,1}) = X_{t,1}^2$, Eq. (6) with $p = q = 2$ implies that the asymptotic normality of $\hat{\gamma}$ requires both $E\varepsilon_t^4 < \infty$ and $EX_{t,1}^8 < \infty$, which may be questionable to daily financial returns. To empirically investigate this problem for mutual fund daily returns, we estimate the tail index a (> 0) of $|\varepsilon_t H(X_{t,1})|$ for each of 3210 mutual funds from September 1, 1998 to December 31, 2018 by assuming that

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

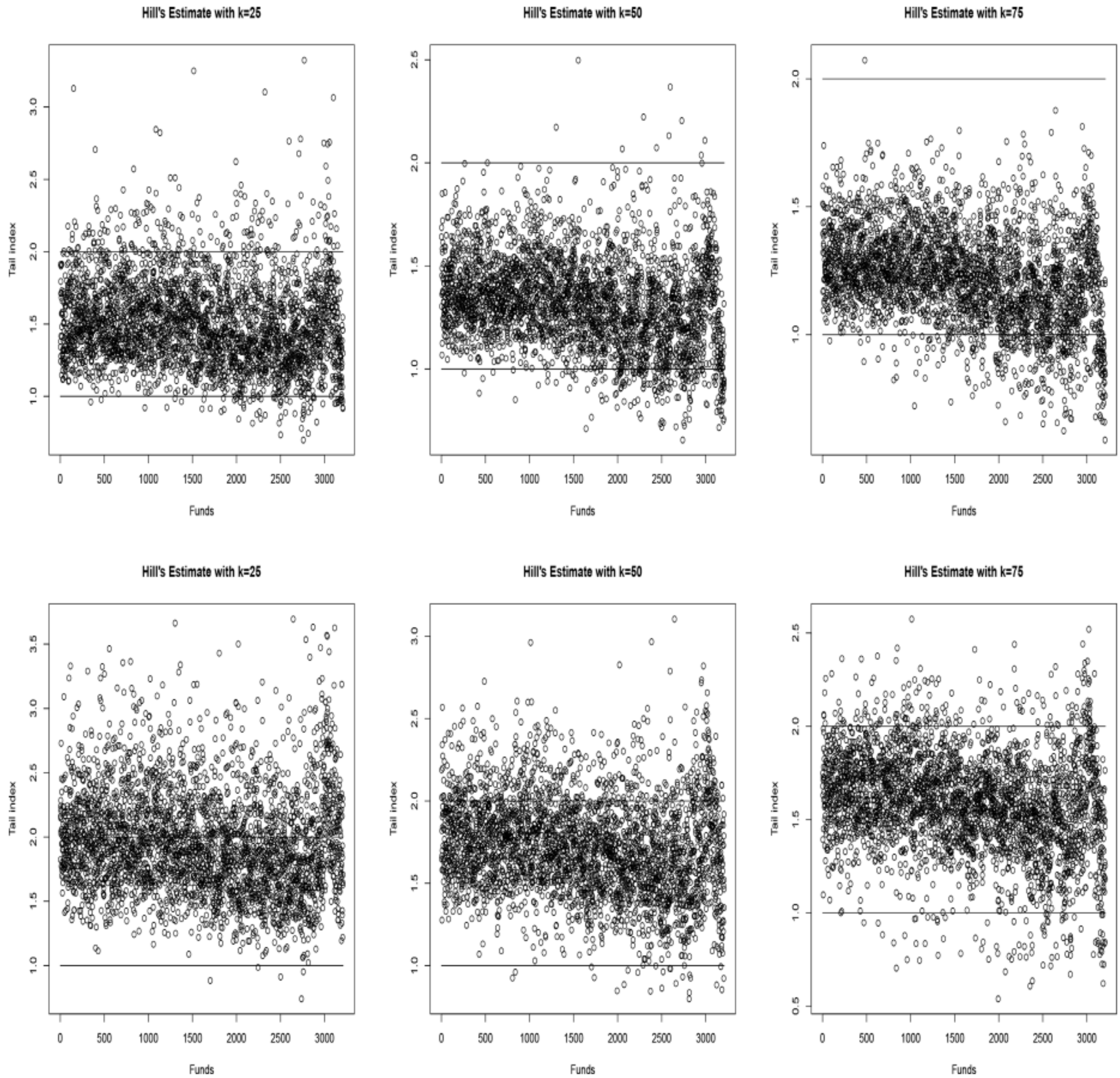

Fig. 1. *The Hill estimates for Model (1) with* $d = 4$. Using daily fund returns for each of 3210 mutual funds from September 1, 1998 to December 31, 2018, the figures plot the Hill estimates for $|\hat{\varepsilon}_t X_{t,1}^2|$ (upper panels) and $|\hat{\varepsilon}_t \max(0, X_{t,1})|$ (lower panels) with upper order statistics $k = 25, 50, 75$, respectively. Points below the horizontal lines $y = 2$ and $y = 1$ indicate the possible nonnormal limit of $\hat{\gamma}$ and ill-defined market timing measure, respectively.

Under this assumption, $\varepsilon_t H(X_{t,1})$ will have an infinite (a finite) second moment if $a < 2$ (> 2). A well-known estimator for a is the so-called Hill estimator described in Hill (1975), which is based on the k upper order statistics with $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. We refer to Hsing (1991) to estimate the tail index of dependent data. To motivate our weighted nonparametric market timing measure and test below, we compute the tail indexes of U.S. mutual funds analyzed in Section 4 with $k = 25, 50$, and 75 for the nonparametric Treynor-Mazuy measure and the nonparametric Henriksson-Merton measure. Fig. 1 plots these Hill estimates, which confirms that some funds have $E(|\varepsilon_t H(X_{t,1})|) = \infty$, many funds have $E(\varepsilon_t^2 H^2(X_{t,1})) = \infty$, and the number of funds with infinite moments for the nonparametric Treynor-Mazuy measure is larger than that for the nonparametric Henriksson-Merton measure. For example, take $k = 50$, we find that 0.65% and 6.48% of the total 3210 funds have $E(\varepsilon_t H(X_{t,1})) = \infty$ for the nonparametric Henriksson-Merton measure and nonparametric Treynor-Mazuy measure, respectively, and 84.14% and 99.60% of the total 3210 funds have $E(\varepsilon_t^2 H^2(X_{t,1})) = \infty$ for the nonparametric Henriksson-Merton measure and nonparametric Treynor-Mazuy measure, respectively. That is, the coskewness in Back et al. (2018) (i.e., the nonparametric Treynor-Mazuy timing measure in (3)) is ill-defined for some funds, and its naive nonparametric estimator $\hat{\gamma}$ has a nonnormal limit for many funds, which complicates the test for zero timing ability. Hence, it is useful to provide a nonparametric market timing measure and a test without requiring $E|\varepsilon_t H(X_{t,1})| < \infty$.

Because the least squares estimators for α and $\boldsymbol{\beta}$ do not have a normal limit when either ε_t or $\boldsymbol{X}_t$ has a very heavy tail, we use the weighted least squares estimation by minimizing

$$(\hat{\hat{\alpha}}, \hat{\hat{\boldsymbol{\beta}}}^\top)^\top = \arg\min_{\alpha,\boldsymbol{\beta}} \sum_{t=1}^{n} \{Y_t - \alpha - \boldsymbol{\beta}^\top \boldsymbol{X}_t\}^2 w_{t-1,\varepsilon}^{-1} w_{t-1,\boldsymbol{X}}^{-1}, \tag{7}$$

where $w_{t-1,\varepsilon}$ and $w_{t-1,\boldsymbol{X}}$ control σ_t in ε_t and $\bar{\sigma}_{t,i}$'s in $\boldsymbol{X}_t$, respectively. Therefore, the asymptotic normality of the above weighted least squares estimators holds when $E(\eta_t^2 \bar{\eta}_{t,i}^2) < \infty$ instead of $E(\varepsilon_t^2 X_{t,i}^2) < \infty$ for $i = 1, \ldots, d$. Appendix A provides details of the choice of weights. Furthermore, we study the weighted nonparametric market timing measure defined by

$$\bar{\gamma} = E\Big(\frac{\varepsilon_t H(X_{t,1})}{w_{t-1,\varepsilon} w_{t-1,\boldsymbol{X}} \sqrt{1 + H^2(X_{t,1})}}\Big), \tag{8}$$

where $\sqrt{1 + H^2(X_{t,1})}$ controls the moment effect of $H(X_{t,1})$. Therefore, $\bar{\gamma}$ is always well-defined (i.e., $|\bar{\gamma}| < \infty$) because $E(|\eta_t|) < \infty$. Note that we use $w_{t-1,\varepsilon} w_{t-1,\boldsymbol{X}}$ instead of $w_{t-1,\varepsilon}$ because $n^{-1} \sum_{t=1}^{n} \frac{\hat{\hat{\varepsilon}}_t}{w_{t-1,\varepsilon} w_{t-1,\boldsymbol{X}}}$ is much closer to zero than $n^{-1} \sum_{t=1}^{n} \frac{\hat{\hat{\varepsilon}}_t}{w_{t-1,\varepsilon}}$, where

$$\hat{\hat{\varepsilon}}_t = Y_t - \hat{\hat{\alpha}} - \hat{\hat{\boldsymbol{\beta}}}^\top \boldsymbol{X}_t, \; t = 1, \ldots, n.$$

We estimate the weighted nonparametric market timing ability by

$$\hat{\bar{\gamma}} = \frac{1}{n} \sum_{t=1}^{n} \frac{\hat{\hat{\varepsilon}} H(X_{t,1})}{w_{t-1,\varepsilon} w_{t-1,\boldsymbol{X}} \sqrt{1 + H^2(X_{t,1})}}. \tag{9}$$

Let $\mathcal{F}_t$ denote the σ-field generated by $\{\bar{\eta}_{s_1,1}, \bar{\eta}_{s_2,2}, \ldots, \bar{\eta}_{s_2,d}, \eta_{s_2} : s_1 \le t+1, \; s_2 \le t\}$. Note that $\mathcal{F}_t$ includes $\bar{\eta}_{t+1,1}$ for better reducing the moment effect of $H(X_{t,1})$. Theorem 1 below derives the asymptotic limit of $\hat{\bar{\gamma}}$ with the regularity conditions and proofs given in Appendices B and C, respectively.

Theorem 1. *Suppose model (1) holds with regularity conditions C1)–C4) in Appendix B. Then,*

$$\sqrt{n}\{\hat{\bar{\gamma}} - \frac{1}{n} \sum_{t=1}^{n} \frac{E(\eta_t|\mathcal{F}_{t-1}) \sigma_t H(X_{t,1})}{w_{t-1,\varepsilon} w_{t-1,\boldsymbol{X}} \sqrt{1 + H^2(X_{t,1})}}\} \xrightarrow{d} N(0, \bar{\sigma}^2)$$

as $n \to \infty$, where the formula of $\bar{\sigma}^2$ is given in Appendix C.

When

$$E(\eta_t|\mathcal{F}_{t-1}) = 0 \text{ for all } t \ge 1, \tag{10}$$

implying $\bar{\gamma} = 0$, we can use $\hat{\bar{\gamma}}$ to test $H_0 : \bar{\gamma} = 0$ against $H_a : \bar{\gamma} \ne 0$. An example satisfying (10) is

$$\eta_t = \eta_{t,0} G(\bar{\boldsymbol{\eta}}_t) \text{ or } \eta_t = \eta_{t,0} G(\bar{\boldsymbol{\eta}}_{t-1}),$$

for some function G, where $\{\eta_{t,0}\}$ is a sequence of independent and identically distributed random variables with zero means and is independent of $\{\bar{\boldsymbol{\eta}}_t = (\bar{\eta}_{t,1}, \ldots, \bar{\eta}_{t,d})^\top\}$.

To avoid estimating the complicated asymptotic variance $\bar{\sigma}^2$, we adopt the random weighted bootstrap method in Jin et al. (2001) and used by Zhu and Ling (2015), Zhu (2016) for heteroscedastic time series models. The residual-based bootstrap method is not applicable because we do not infer the GARCH models.

- Ai) Draw a random sample with size n from a distribution with mean one and variance one, such as the standard exponential distribution. Denote them by $\xi_1^b, \ldots, \xi_n^b$.
- Aii) Compute

$$\hat{\bar{\gamma}}^b = \sum_{t=1}^{n} \xi_t^b \frac{\hat{\hat{\varepsilon}}_t^b H(X_{t,1})}{w_{t-1,\varepsilon} w_{t-1,\boldsymbol{X}} \sqrt{1 + H^2(X_{t,1})}} / \sum_{t=1}^{n} \xi_t^b,$$

  where $\hat{\hat{\varepsilon}}_t^b = Y_t - \hat{\hat{\alpha}}^b - \hat{\hat{\boldsymbol{\beta}}}^{b\top} \boldsymbol{X}_t$ and

$$(\hat{\hat{\alpha}}^b, \hat{\hat{\boldsymbol{\beta}}}^{b\top})^\top = \arg\min_{\alpha,\boldsymbol{\beta}} \sum_{t=1}^{n} \xi_t^b \{Y_t - \alpha - \boldsymbol{\beta}^\top \boldsymbol{X}_t\}^2 w_{t-1,\varepsilon}^{-1} \varepsilon_{t-1,\boldsymbol{X}}^{-1}.$$

- Biii) Repeat the above two steps B times to get $\{\hat{\bar{\gamma}}^b\}_{b=1}^{B}$. Hence, the bootstrap variance estimator is $\hat{\bar{\sigma}}^2 = \frac{n}{B} \sum_{b=1}^{B} \{\hat{\bar{\gamma}}^b - \hat{\bar{\gamma}}\}^2$.

Theorem 2. *Under conditions of Theorem 1, we have*

$$\hat{\bar{\sigma}}/\bar{\sigma} \to 1 \text{ as } B \to \infty \text{ and } n \to \infty.$$

Based on Theorems 1 and 2, for testing $H_0 : \bar{\gamma} = 0$ against $H_0 : \bar{\gamma} \neq 0$, we reject the null hypothesis at level a if $n\hat{\bar{\gamma}}^2/\hat{\bar{\sigma}}^2 > \chi^2_{1,1-a}$. This test has an asymptotically correct size when model (1) holds with conditions in Theorem 2 and (10). Moreover, the power of this test approaches one when

$$\sqrt{n}|E(\frac{\varepsilon_t H(X_{t,1})}{w_{t-1,\varepsilon} w_{t-1,X}\sqrt{1+H^2(X_{t,1})}})| \to \infty \text{ as } n \to \infty.$$

Remark 1. The above test is still valid when we replaced the GARCH errors in (1) by the generalized GARCH model $\sigma_t^2 = f_t(\eta_{t-1}, \eta_{t-2}, \dots)$ and $\bar{\sigma}^2_{t,k} = \bar{f}_{t,k}(\bar{\eta}_{t-1,k}, \bar{\eta}_{t-2,k}, \dots)$ with f_t and $\bar{f}_{t,k}$ being measurable and satisfying that $C_0 \le f_t \le \sum_{i=0}^{\infty} \rho^i |\varepsilon_{t-1-i}|$ and $C_0 \le \bar{f}_{t,k} \le \sum_{i=0}^{\infty} \rho^i |\bar{\varepsilon}_{t-1-i,k}|$ for some $\rho \in (0, 1)$ and $C_0 > 0$. The proof is straightforward, and we skip details. The generalized GARCH model includes the GIJ model and nonlinear GARCH model in Section 5 of Zhu and Ling (2015), the absolute value GARCH model in Schwert (1989), the nonlinear GARCH model in Engle (1990), the volatility switching GARCH model in Fornari and Mele (1997), the threshold GARCH models in Zakoian (1994), and the generalized quadratic ARCH models in Sentana (1995).

An important application of the developed test above is to examine the association between stock picking skill and market timing ability. Back et al. (2018) compare the signs of $\hat{\gamma}$ and $\hat{\alpha}$ without studying the asymptotic properties of $\hat{\gamma}$. For heavy-tailed daily returns, $E(\varepsilon_t^2 H^2(X_{t,1}))$ may be pretty large or even infinite, and the asymptotic variance of $\hat{\gamma}$ could also be significant. A significant value of $|\hat{\gamma}|$ may indicate zero nonparametric market timing ability. Furthermore, Welch and Goyal (2008) argue that most previously documented predictors can not forecast stock market return out-of-sample. Thus, we expect most mutual fund managers to have zero market timing ability. Hence, excluding those noisy funds with insignificant market timing will improve our ability to study the association between stock picking skills and market timing ability. In our data analysis below, we first apply the developed tests for zero timing ability to each fund and then examine the tradeoff association by including and excluding funds with zero timing ability. Our sequential stepwise procedure is similar to Hansen (2005), Hansen et al. (2011), where the test power is increased by removing inferior alternatives. In our context, we remove the funds with insignificant timing abilities.

3. Simulation study

This section investigates the finite sample performance of the proposed test for zero weighted nonparametric Treynor-Mazuy market timing measure (i.e., $H(X_{t,1}) = X^2_{t,1}$) in terms of both size and power. Also, we compare it with the unweighted nonparametric measure, i.e., the coskewness in Back et al. (2018), and show that it is necessary to employ the weighted one due to heavy tails.

We draw 10,000 random samples with sample size $n = 500$, or 1000, or 2000, or 5000 from model (1) with $d = 1$, $s_1 = 1$, and $r_1 = 0$. Using the Statistical software R package ‘fGarch’, we fit the ARMA (1,0)-GARCH(1,1) model to the market excess returns in our real dataset from September 01, 1998 to December 31, 2018, which is $\mu_1 = 0.07726$, $\phi_{1,1} = -0.035865$, $w_1 = 0.01026$, $a_{1,1} = 0.09749$, and $b_{1,1} = 0.90001$. Similarly, using the single longest-tenured fund in our dataset, we fit the regression and a GARCH(1,1) model to ε_t's, which gives $\alpha = -0.00874$, $\beta = 0.96928$, $w = 0.00016$, $a_1 = 0.06851$, and $b_1 = 0.93084$. We use $B = 10000$ in the proposed random weighted bootstrap methods and employ the weight functions in (11) with $h = 0.2, 0.3, 0.4, 0.5$. For using $\hat{\gamma}$ in (4), we employ a similar random weighted bootstrap method to estimate its asymptotic variance. We use the following settings for η_t and $\eta_{t,1}$ to study the size, satisfying that

$$E(\eta_t) = 0,\ E(\eta_t^2) = 1,\ E(\eta_{t,1}) = 0,\ E(\eta^2_{t,1}) = 1, \text{ and } E(\eta_t|\mathcal{F}_{t-1}) = 0 \text{ (i.e., } \gamma = \bar{\gamma} = 0).$$

- *Case 1)* $U_t \sim t(\nu)$, $V_t \sim N(0, 1)$, $\eta_{t,1} = \frac{U_t}{\sqrt{\nu/(\nu-2)}}$, $\eta_t = V_t$, where $\nu = 4.5$, and U_t's and V_t's are independent.
- *Case 2)* $U_t \sim t(\nu)$, $V_t \sim N(0, 1)$, $\eta_{t,1} = \frac{U_t}{\sqrt{\nu/(\nu-2)}}$, and $\eta_t = \frac{U^2_{t-1}}{\sqrt{3\nu^2/(\nu-2)/(\nu-4)}} V_t$, where $\nu = 4.5$, and U_t's and V_t's are independent.
- *Case 3)* $U_t \sim t(\nu)$, $V_t \sim N(0, 1)$, $\bar{V}_t \sim N(0, 1)$, $\eta_{t,1} = \frac{U_t}{\sqrt{\nu/(\nu-2)}} V_t$, and $\eta_t = \frac{U^2_{t-1}}{\sqrt{3\nu^2/(\nu-2)/(\nu-4)}} \bar{V}_t$, where $\nu = 4.5$, and U_t's, V_t's, and $\bar{V}_t$'s are independent.
- *Case 4)* $U_t \sim t(\nu)$, $V_t \sim N(0, 1)$, $\bar{V}_t \sim N(0, 1)$, $\eta_{t,1} = \frac{U_t}{\sqrt{\nu/(\nu-2)}} V_t$, and $\eta_t = \frac{U_t}{\sqrt{\nu/(\nu-2)}} \bar{V}_t$, where $\nu = 4.5$, and U_t's, V_t's, and $\bar{V}_t$'s are independent.

Tables 1–4 report the empirical sizes for testing $H_0 : \gamma = 0$ against $H_a : \gamma \neq 0$ at the 10% and 5% levels by using the unweighted estimator $\hat{\gamma}$ in (4) and the weighted estimator $\hat{\bar{\gamma}}$ in (9). The weighted test of using $\hat{\bar{\gamma}}$ is not sensitive to the choice of h and has an accurate size for all cases except Case 4 with $n = 500$. In contrast, the unweighted test of using $\hat{\gamma}$ has a distorted size for Cases 2)–4), which have $E(\varepsilon_t^2 H^2(X_{t,1})) = \infty$, i.e., the asymptotic normality does not hold for $\hat{\gamma}$.

To study power, we consider the following setting for η_t and $\eta_{t,1}$ satisfying that

$$E(\eta_t) = 0,\ E(\eta_t^2) = 1,\ E(\eta_{t,1}) = 0,\ E(\eta^2_{t,1}) = 1,\ E(\eta_t \eta_{t,1}) = 0, \text{ and } E(\eta_t \eta^2_{t,1}) = 1.$$

Table 1
Empirical sizes of the tests using $\hat{\gamma}$ in (4) and $\hat{\hat{\gamma}}$ in (9) for Case 1.

n	Level	$\hat{\gamma}$	$\hat{\hat{\gamma}}, h = 0.2$	$\hat{\hat{\gamma}}, h = 0.3$	$\hat{\hat{\gamma}}, h = 0.4$	$\hat{\hat{\gamma}}, h = 0.5$
500	10%	0.1123	0.1102	0.1118	0.1135	0.1123
	5%	0.0527	0.0571	0.0578	0.0573	0.0569
1000	10%	0.1124	0.1034	0.1049	0.1051	0.1063
	5%	0.0522	0.0540	0.0541	0.0534	0.0537
2000	10%	0.1049	0.0982	0.0992	0.1021	0.1025
	5%	0.0468	0.0498	0.0490	0.0487	0.0480
5000	10%	0.1027	0.0981	0.0958	0.0961	0.0954
	5%	0.0453	0.0449	0.0445	0.0436	0.0444

Note: We compute the empirical sizes of the tests using $\hat{\gamma}$ in (4) and $\hat{\hat{\gamma}}$ in (9) with $h = 0.2, 0.3, 0.4, 0.5$ to test for zero nonparametric market timing ability for Case 1, where $U_t \sim \chi^2(\nu)$, $V_t \sim N(0, 1)$, $\eta_{t,1} = \frac{U_t}{\sqrt{\nu/(\nu-2)}}$, $\eta_t = V_t$, $\nu = 4.5$, and U_t's and V_t's are independent.

Table 2
Empirical sizes of the tests using $\hat{\gamma}$ in (4) and $\hat{\hat{\gamma}}$ in (9) for Case 2.

n	Level	$\hat{\gamma}$	$\hat{\hat{\gamma}}, h = 0.2$	$\hat{\hat{\gamma}}, h = 0.3$	$\hat{\hat{\gamma}}, h = 0.4$	$\hat{\hat{\gamma}}, h = 0.5$
500	10%	0.0766	0.1097	0.1095	0.1112	0.1131
	5%	0.0265	0.0516	0.0516	0.0528	0.0510
1000	10%	0.0815	0.1074	0.1054	0.1072	0.1060
	5%	0.0265	0.0516	0.0519	0.0509	0.0499
2000	10%	0.0760	0.1019	0.1024	0.1028	0.1010
	5%	0.0254	0.0509	0.0511	0.0519	0.0526
5000	10%	0.0756	0.0996	0.0994	0.0999	0.0999
	5%	0.0249	0.0520	0.0512	0.0491	0.0476

Note: We compute the empirical sizes of the tests using $\hat{\gamma}$ in (4) and $\hat{\hat{\gamma}}$ in (9) with $h = 0.2, 0.3, 0.4, 0.5$ to test for zero nonparametric market timing ability for Case 2, where $U_t \sim \chi^2(\nu)$, $V_t \sim N(0, 1)$, $\eta_{t,1} = \frac{U_t}{\sqrt{\nu/(\nu-2)}}$, and $\eta_t = \frac{U_{t-1}^2}{\sqrt{3\nu^2/(\nu-2)/(\nu-4)}}V_t$, $\nu = 4.5$, and U_t's and V_t's are independent.

Table 3
Empirical sizes of the tests using $\hat{\gamma}$ in (4) and $\hat{\hat{\gamma}}$ in (9) for Case 3.

n	Level	$\hat{\gamma}$	$\hat{\hat{\gamma}}, h = 0.2$	$\hat{\hat{\gamma}}, h = 0.3$	$\hat{\hat{\gamma}}, h = 0.4$	$\hat{\hat{\gamma}}, h = 0.5$
500	10%	0.0788	0.1209	0.1191	0.1178	0.1186
	5%	0.0241	0.0477	0.0489	0.0495	0.0461
1000	10%	0.0708	0.1105	0.1104	0.1100	0.1099
	5%	0.0227	0.0452	0.0452	0.0446	0.0434
2000	10%	0.0717	0.1066	0.1069	0.1062	0.1062
	5%	0.0225	0.0446	0.0445	0.0449	0.0442
5000	10%	0.0659	0.1041	0.1037	0.1020	0.1022
	5%	0.0212	0.0410	0.0429	0.0441	0.0448

Note: We compute the empirical sizes of the tests using $\hat{\gamma}$ in (4) and $\hat{\hat{\gamma}}$ in (9) with $h = 0.2, 0.3, 0.4, 0.5$ to test for zero nonparametric market timing ability for Case 3, where $U_t \sim \chi^2(\nu)$, $V_t \sim N(0, 1)$, $\bar{V}_t \sim N(0, 1)$, $\eta_{t,1} = \frac{U_t}{\sqrt{\nu/(\nu-2)}}V_t$, and $\eta_t = \frac{U_{t-1}^2}{\sqrt{3\nu^2/(\nu-2)/(\nu-4)}}\bar{V}_t$, $\nu = 4.5$, and U_t's, V_t's, and $\bar{V}_t$'s are independent.

Table 4
Empirical sizes of the tests using $\hat{\gamma}$ in (4) and $\hat{\hat{\gamma}}$ in (9) for Case 4.

n	Level	$\hat{\gamma}$	$\hat{\hat{\gamma}}, h = 0.2$	$\hat{\hat{\gamma}}, h = 0.3$	$\hat{\hat{\gamma}}, h = 0.4$	$\hat{\hat{\gamma}}, h = 0.5$
500	10%	0.1523	0.1265	0.1281	0.1316	0.1321
	5%	0.0594	0.0601	0.0605	0.0613	0.0633
1000	10%	0.1378	0.1201	0.1181	0.1170	0.1177
	5%	0.0494	0.0560	0.0574	0.0576	0.0575
2000	10%	0.1272	0.1069	0.1083	0.1092	0.1110
	5%	0.0494	0.0514	0.0517	0.0515	0.0516
5000	10%	0.1182	0.1077	0.1084	0.1074	0.1066
	5%	0.0471	0.0517	0.0518	0.0506	0.0498

Note: We compute the empirical sizes of the tests using $\hat{\gamma}$ in (4) and $\hat{\hat{\gamma}}$ in (9) with $h = 0.2, 0.3, 0.4, 0.5$ to test for zero nonparametric market timing ability for Case 4, where $U_t \sim \chi^2(\nu)$, $V_t \sim N(0, 1)$, $\bar{V}_t \sim N(0, 1)$, $\eta_{t,1} = \frac{U_t}{\sqrt{\nu/(\nu-2)}}V_t$, and $\eta_t = \frac{U_t}{\sqrt{\nu/(\nu-2)}}\bar{V}_t$, $\nu = 4.5$, and U_t's, V_t's, and $\bar{V}_t$'s are independent.

Table 5
Empirical powers of the tests in using $\hat{\gamma}$ in (4) and $\hat{\tilde{\gamma}}$ in (9) for Case 5.

n	Level	$\hat{\gamma}$	$\hat{\tilde{\gamma}}, h = 0.2$	$\hat{\tilde{\gamma}}, h = 0.3$	$\hat{\tilde{\gamma}}, h = 0.4$	$\hat{\tilde{\gamma}}, h = 0.5$
500	10%	0.9974	1	1	1	1
	5%	0.9963	1	1	1	1
1000	10%	0.9989	1	1	1	1
	5%	0.9982	1	1	1	1
2000	10%	0.9994	1	1	1	1
	5%	0.9986	1	1	1	1
5000	10%	0.9996	1	1	1	1
	5%	0.9989	1	1	1	1

Note: We compute the empirical powers of the tests in using $\hat{\gamma}$ in (4) and $\hat{\tilde{\gamma}}$ in (9) to test for zero nonparametric market timing ability for Case 5, where $\eta_t = V_t$, $\eta_{t,1} = \bar{V}_t(V_t + 1)/\sqrt{2}$, V_t's and $\bar{V}_t$'s are independent with the standard normal distribution.

- *Case 5)* $\eta_t = V_t$ and $\eta_{t,1} = \frac{(V_t+1)}{\sqrt{2}}\bar{V}_t$, where V_t's and $\bar{V}_t$'s are independent with the standard normal distribution.

Table 5 reports the empirical powers of the tests using $\hat{\gamma}$ in (4) and $\hat{\tilde{\gamma}}$ in (9), showing that they are powerful. Because we model ε_t and $X_{t,1}$ by a GARCH model and an ARMA-GARCH model, respectively, it becomes nontrivial to find a setting such that $E(\varepsilon_t X_{t,1}) = 0$ but $E(\varepsilon_t X_{t,1}^2)$ close to zero for investigating the local power of the proposed tests.

In summary, the proposed tests for the absence of market timing ability have an accurate size and are powerful. The proposed weighted nonparametric market timing ability measure and test effectively handle heteroscedasticity and the factor model's dependence between factors and errors.

4. An empirical study of U.S. equity funds

Given the availability of daily fund return data in the Center for Research in Security Prices (CRSP) Survivorship Bias Free Mutual Fund Database, we choose a sample period of September 01, 1998 to December 31, 2018. We consider all actively managed U.S. equity mutual funds and base our selection criteria on the objective codes from CRSP following Kacperczyk et al. (2008). We also drop ETFs, annuities, and index funds based on indicator variables or fund names from CRSP. We require 80% of the assets under management to be invested in common stocks since we focus on equity funds. In addition, We restrict our sample to funds with at least \$15 million in assets under management and use the fund ticker creation date to control for the incubation bias noted by Evans (2010).

To determine the stock holding characteristics for funds with different levels of timing ability, we use the portfolio holding share to calculate a dollar value for each stock held by a mutual fund. The share information is from the Thomson Reuters Mutual Fund Holdings (formerly CDA/Spectrum S12) database. As a further filter, we remove funds with investment objective codes 1, 5, 6, 7, and 8 in Thomson Reuters Mutual Fund Holdings, representing International, Municipal Bond, Bond and Preferred, Balanced, and Metals funds. Finally, we merge the CRSP Mutual Fund database and the Thomson Reuters Mutual Fund Holdings database using the MFLINKS tables provided by WRDS. We exclude funds with fewer than 200 valid observations of daily fund returns, resulting in 3210 funds in our final sample.

For each of these funds, we fit the one-factor model in Jensen (1968), the three-factor model in Fama and French (1996), and the four-factor model in Carhart (1997) to calculate alphas as the measure of stock picking skill. We compute the fund timing ability estimates $\hat{\gamma}$ in (4) and $\hat{\tilde{\gamma}}$ in (9) and the p-values of the related tests for zero timing ability. In this way, we estimate stock-picking skills and market timing ability in an integrated model. We use $B = 1000$ in the random weighted bootstrap methods and the weight functions in (11) with $h = 0.2$. Figure 2 plots the estimates of the unweighted ($\hat{\gamma}$) and weighted ($\hat{\tilde{\gamma}}$) nonparametric Treynor-Mazuy market timing ability ($H(X_{t,1}) = X_{t,1}^2$) and the corresponding p-values by using the one-factor, three-factor, and four-factor models. Similarly, Fig. 3 plots the estimates of the unweighted and weighted nonparametric Henriksson-Merton market timing ability ($H(X_{t,1}) = \max(0, X_{t,1})$) and the corresponding p-values.

Figures 2 and 3 indicate that the p-value for most of the funds in our sample is larger than 10%, i.e., many funds' timing ability is not significantly different from zero. Panel A in Tables 6 and 7 provides summary statistics for estimates and p-values of the related tests for zero market timing ability. Although the weights are fund-specific, limiting the cross-sectional fund comparison, we can still compare the existence of timing based on our method with the traditional way. Panel B of Tables 6 and 7 indicates that by applying either the Treynor-Mazuy or Henriksson-Merton measure, our method disagrees with the inference of traditional parametric test for many funds in terms of timing ability. For example, in Table 6, out of 534 positive timing funds based on the traditional parametric method, 344 of them have either zero or perverse timing ability using our weighted nonparametric method. The traditional parametric method identifies 1837 funds with zero Treynor-Mazuy timing ability. The weighted nonparametric method identifies 2426 funds with zero Treynor-Mazuy timing ability, and only 1522 funds overlap with those based on the parametric approach. The numbers of funds classified as negative timing funds, zero timing funds, and positive timing funds by both parametric and nonparametric methods are 258, 1522, and 190, respectively, about 61% of the funds in our sample. Therefore, compared with the traditional test, our method shows different inferences for about 39% of funds in terms of timing. In Table 7, we find similar results for the use of the

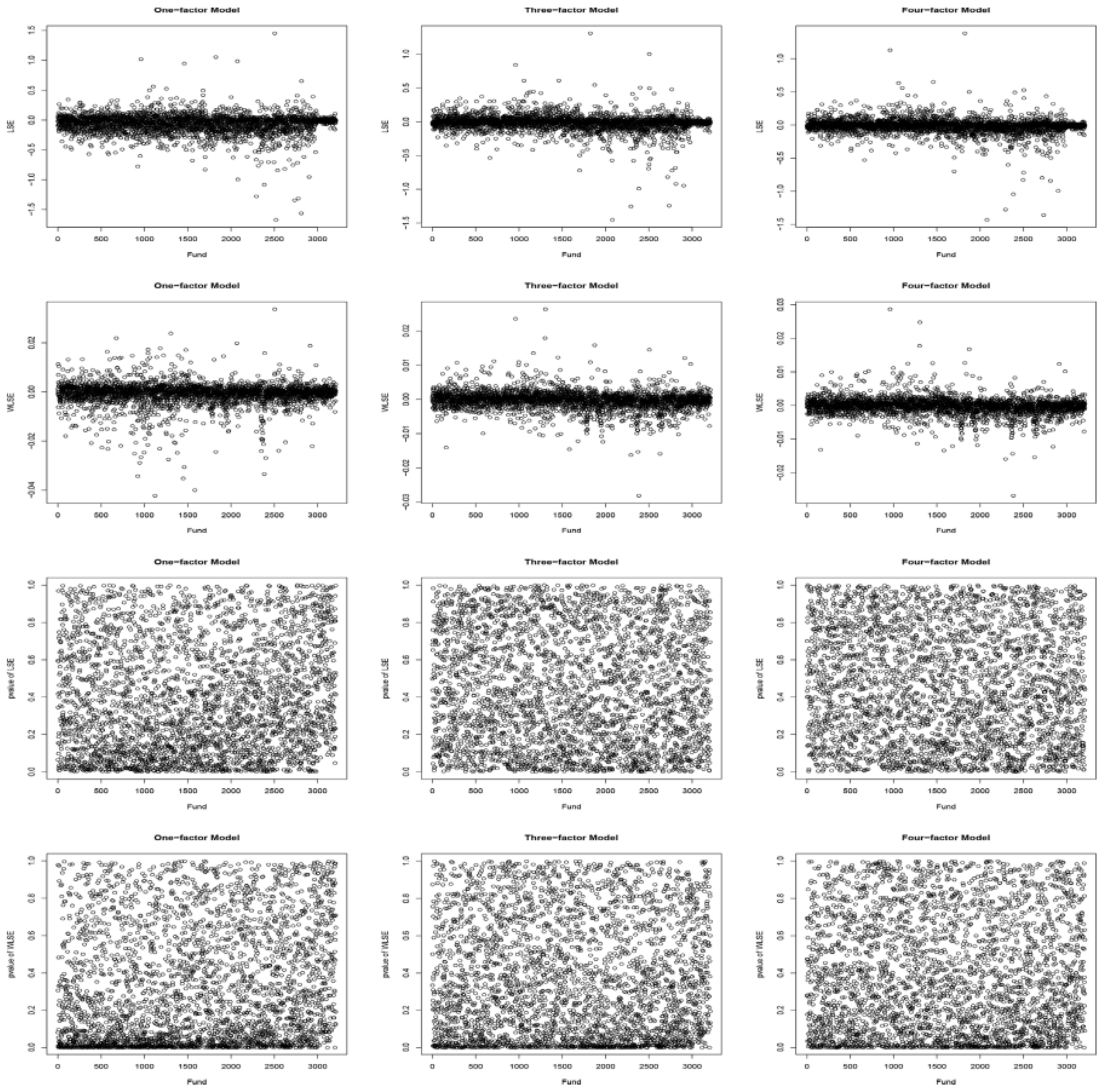

Fig. 2. *Nonparametric Treynor-Mazuy market timing measure* The figure shows nonparametric Treynor-Mazuy market timing ability $(H(X_{t,1}) = X_{t,1}^2)$ for 3210 funds using daily fund returns from September 1, 1998 to December 31, 2018 and the *p*-values. The first and third rows are the unweighted estimates and the corresponding *p*-values based on $\hat{\gamma}$ in (4). The second and fourth rows are the weighted estimates and the corresponding *p*-values based on $\hat{\hat{\gamma}}$ in (9) and weights in (11) with $h = 0.2$. They are the one-factor, three-factor, and four-factor models from left to right.

Henriksson-Merton measure.[7] In Panel C of both Tables 6 and 7, the construction of synthetic funds is based on actual funds' holding following Bollen and Busse (2001) to address spurious timing bias. By matching the actual funds' trading friction characteristics measured by firm size, we generate synthetic funds.[8] For each fund, we estimate the nonparametric Treynor-Mazuy measures (in Table 6) and Henriksson-Merton measure (in Table 7) of its corresponding 200 synthetic funds and use the mean of these estimations as a benchmark. For row "ControlSize" which is not subject to "spurious timing", we compare the estimated timing measure with distribution from 200 simulated funds of zero timing (as a benchmark).

[7] The sample used in our paper is dramatically different from Bollen and Busse (2001), resulting in a different classification of funds with timing ability based on traditional parametric measures.

[8] We first sort stocks into five groups by firm size. Then, for each fund, at the end of each month, we replace the stocks they hold with the stocks in the same quintile (sorted by firm size) at random, with a holding period of one month. Next, we take a value-weighted average of stock returns to get daily returns of synthetic funds and reduce the daily expense ratio to get a net synthetic return. We repeat this process 200 times, resulting in 200 sets of synthetic fund returns for each real fund. We remove funds with less than ten stocks in holdings. All funds are required to have at least 200 observations to be included in our sample. For each simulation, our final sample includes 2950 synthetic funds, each corresponding to one of the real funds. By construction, the synthetic funds should not exist (active) timing ability.

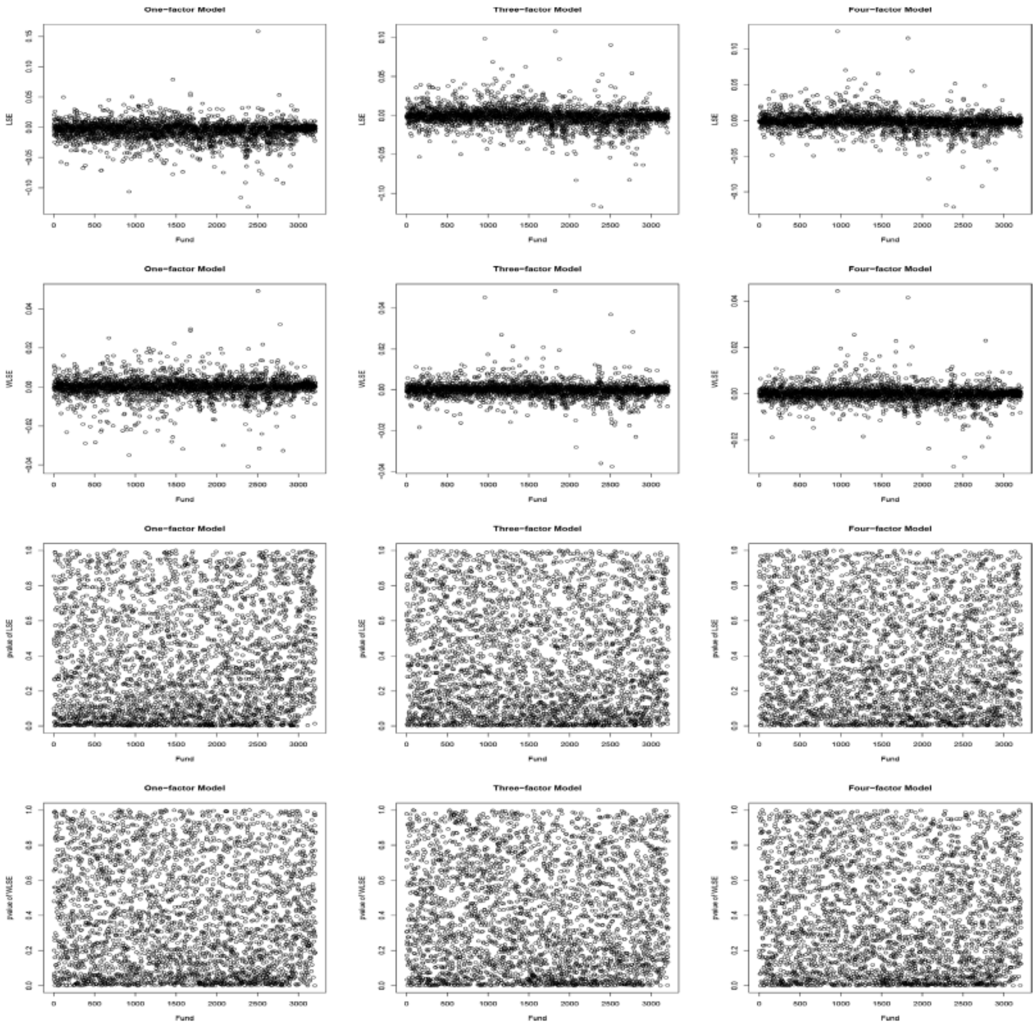

Fig. 3. *Nonparametric Henriksson-Merton market timing measure* The figure shows nonparametric Henriksson-Merson market timing ability ($H(X_{t,1}) = \max(0, X_{t,1})$) for 3210 funds using daily fund returns from September 1, 1998 to December 31, 2018, and the *p*-values. The first and third rows are the unweighted estimates and the corresponding *p*-values based on $\hat{\gamma}$ in (4). The second and fourth rows are the weighted estimates and the corresponding *p*-values based on $\hat{\hat{\gamma}}$ in (9) and weights in (11) with $h = 0.2$. They are the one-factor, three-factor, and four-factor models from left to right.

If the estimated timing measures based on our method lie below the 5th percentile, the funds are classified as significant negative timing funds. Similarly, if the estimated timing measures lie above the 95th percentile of zero timing distribution, the funds are classified as significant positive timing funds. We find that for the TM measure, there are more funds with negative timing than funds with positive timing while using HM, the number of funds with positive timing and negative timing are almost the same. For reference, we also report the row “Theorem 1” where we infer positive timing and negative timing using Theorem 1 of our paper. It indicates that the number of funds with positive and negative timing is almost the same (as expected). To conclude, “spurious timing” may affect the number of funds with positive or negative timing for the TM measure only. For HM measure, the results based on theorem one are almost the same as the results not subject to “spurious timing”.[9]

Based on our tests, we categorize funds into three groups: positive timing ability, no timing ability, and negative timing ability. We then examine the differences in fund holdings among these three groups. We estimate each fund’s weighted nonparametric Treynor-Mazuy timing measures based on the Carhart (1997) model and set the significance level to 10%. We report the time-series mean of the cross-sectional averages of monthly stock characteristics in each group and the difference between funds with positive timing and negative timing ability. The stock holding characteristics are computed based on

[9] In Table IA. 1 in the internet appendix, we calculate coskewness with firm size-sorted portfolios and find both positive and negative coskewness in the ten portfolios. Based on the standard deviation of those methods (the last two columns), none of the coskewness in the ten portfolios is statistically significant.

Table 6
Treynor-Mazuy Market Timing Measure.

Panel A: Summary Statistics of the Nonparametric Measure

One-factor	$\hat{\alpha}$	$\hat{\gamma}$	p-value (using $\hat{\gamma}$)	$\hat{\hat{\alpha}}$	$\hat{\hat{\gamma}}$	p-value (using $\hat{\hat{\gamma}}$)
10%	-0.0188	-0.2385	0.0406	-0.0193	-0.0052	0.0060
25%	-0.0092	-0.1384	0.1309	-0.0084	-0.0023	0.0450
50%	-0.002	-0.0356	0.3354	-0.0007	-0.0004	0.2290
75%	0.0058	0.0176	0.6225	0.0093	0.0012	0.5680
90%	0.0143	0.0794	0.8483	0.0198	0.0028	0.8235
Mean	-0.0023	-0.0636	0.3903	-0.0001	-0.0009	0.3254
SD	0.0179	0.1636	0.2925	0.0200	0.0047	0.3071
Three-factor						
10%	-0.0180	-0.1280	0.0591	-0.0186	-0.0031	0.0120
25%	-0.0096	-0.0554	0.1798	-0.0089	-0.0014	0.0820
50%	-0.0035	-0.0082	0.4198	-0.0029	-0.0001	0.3035
75%	0.0021	0.0283	0.7309	0.0027	0.0009	0.6330
90%	0.0079	0.0841	0.8905	0.0095	0.0022	0.8480
Mean	-0.0046	-0.0169	0.4537	-0.0041	-0.0003	0.3693
SD	0.0157	0.1228	0.303	0.0166	0.0027	0.3088
Four-factor						
10%	-0.0180	-0.1188	0.0671	-0.0179	-0.0029	0.0170
25%	-0.0096	-0.0535	0.1912	-0.0093	-0.0013	0.1030
50%	-0.0037	-0.0077	0.4430	-0.0032	-0.0001	0.3230
75%	0.0019	0.0279	0.7294	0.0025	0.0009	0.6380
90%	0.0081	0.0820	0.8945	0.0092	0.0021	0.8530
Mean	-0.0045	-0.0157	0.4635	-0.0041	-0.0002	0.3813
SD	0.0154	0.1183	0.2997	0.0161	0.0026	0.3063

Panel B: Number of Funds in Different Portfolios

		Weighted Nonparametric Measure			
		Negative	Zero	Positive	Total
Parametric Measure	Negative	258	564	17	839
	Zero	150	1522	165	1837
	Positive	4	340	190	534
	Total	412	2426	372	3210

Panel C: Theorem1 vs. ControlSize

		Negative	Zero	Positive	Total
Theorem1	Frequency	317	2267	366	2950
	Percentage	10.75%	76.85%	12.40%	100%
ControlSize	Frequency	835	1945	170	2950
	Percentage	28.31%	65.93%	5.76%	100%

*Notes:*Panel A reports the summary statistics for estimates and tests using $\hat{\gamma}$ and $\hat{\hat{\gamma}}$ with $h = 0.2$ based on the one-factor, three-factor, and four-factor models. In panel B, we compare our proposed estimation with the traditional parametric estimation of the timing coefficients in the Treynor-Mazuy approach based on the four-factor model, using daily data for 3210 funds from September 1, 1998 to December 31, 2018. We report the number of funds in different portfolios. Panel C reports the fund's timing sign based on Theorem 1 and synthetic funds using daily data for 2950 funds.

fund holdings and the characteristics of stocks held by each fund. We construct five stock characteristics in the category of "trading frictions" in Hou et al. (2015), including market capitalization and industry-adjusted market capitalization (*mve* and *mve_ia*), illiquidity (*ill*), volatility of liquidity based on dollar trading volume (*std_dolvol*), and the number of zero trading days (*zerotrade*). We calculate these quantities following Green et al. (2017). Mutual funds must disclose their holding every six months before May 2004 and every three months afterward. We compute t-statistics of the differences between the positive and negative timing funds using the Newey-West (1987) method with six lags.

As shown in Panel A of Table 8, the stocks held by funds with positive $\hat{\hat{\gamma}}$ and funds with negative $\hat{\hat{\gamma}}$ are dramatically different. Compared to funds with negative timing ability, funds with positive timing ability hold stocks with the larger market capitalization (*mve* and *mve_ia*), lower values of illiquidity (*ill*), lower volatility of liquidity based on dollar trading volume (*std_dolvol*), and a smaller number of zero trading days (*zerotrade*). To address spurious timing bias caused by holding stocks with option-like features (Jagannathan and Korajczyk, 1986), following Bollen and Busse (2001), we generate synthetic funds that match the actual funds' trading friction characteristics, such as firm size (*mve*). We find evidence of spurious timing bias in Panels C in Table 8. However, even after controlling for the firm size, we still find that good-timers tend to hold stocks with higher market capitalization than bad-timers. Funds holding stocks with lower trading frictions and higher liquidity can adjust their portfolios according to the market conditions more quickly and at a lower cost. In contrast, mutual funds holding stocks with higher trading frictions may be reluctant to time the market due to high trading costs. Another reason might be that those fund managers with timing ability intentionally hold stocks with low trading frictions because

Table 7
Henriksson-Merton Market Timing Measure.

Panel A: Summary Statistics of the Nonparametric Measure						
One-factor	$\hat{\alpha}$	$\hat{\gamma}$	*p*-value (using $\hat{\gamma}$)	$\hat{\hat{\alpha}}$	$\hat{\hat{\gamma}}$	*p*-value (using $\hat{\hat{\gamma}}$)
10%	-0.0188	-0.0219	0.0162	-0.0193	-0.0049	0.0260
25%	-0.0092	-0.0116	0.0871	-0.0084	-0.0016	0.1130
50%	-0.0020	-0.0033	0.3007	-0.0007	0.0002	0.3430
75%	0.0058	0.0019	0.6261	0.0093	0.0019	0.6550
90%	0.0143	0.0083	0.8530	0.0198	0.0044	0.8580
Mean	-0.0023	-0.0054	0.3708	-0.0001	-0.0001	0.3957
SD	0.0179	0.0148	0.3078	0.0200	0.0050	0.3067
Three-factor						
10%	-0.0180	-0.0139	0.0236	-0.0186	-0.0034	0.0250
25%	-0.0096	-0.0064	0.1046	-0.0089	-0.0013	0.1200
50%	-0.0035	-0.0011	0.3224	-0.0029	0.0000	0.3510
75%	0.0021	0.0030	0.6424	0.0027	0.0013	0.6560
90%	0.0079	0.0093	0.8573	0.0095	0.0029	0.8520
Mean	-0.0046	-0.0017	0.3882	-0.0041	-0.0001	0.3982
SD	0.0157	0.0122	0.3067	0.0166	0.0037	0.3027
Four-factor						
10%	-0.0180	-0.0129	0.0315	-0.0179	-0.0033	0.0205
25%	-0.0096	-0.0057	0.1249	-0.0093	-0.0013	0.1090
50%	-0.0037	-0.0012	0.3561	-0.0032	0.0000	0.3370
75%	0.0019	0.0029	0.6448	0.0025	0.0012	0.6480
90%	0.0081	0.0086	0.8534	0.0092	0.0028	0.8560
Mean	-0.0045	-0.0016	0.4006	-0.0041	-0.0001	0.3917
SD	0.0154	0.0116	0.2996	0.0161	0.0035	0.3056

Panel B: Number of Funds in Different Portfolios					
		Weighted Nonparametric Measure			
		Negative	Zero	Positive	Total
Parametric Measure	Negative	333	503	5	841
	Zero	92	1649	148	1889
	Positive	0	287	193	480
	Total	425	2439	346	3210

Panel C: Theorem1 vs. ControlSize					
		Negative	Zero	Positive	Total
Theorem1	Frequency	356	2257	337	2950
	Percentage	12.07%	76.51%	11.42%	100%
ControlSize	Frequency	354	2092	504	2950
	Percentage	12.00%	70.92%	17.08%	100%

*Notes:*Panel A reports the summary statistics for estimates and tests using $\hat{\gamma}$ and $\hat{\hat{\gamma}}$ with $h = 0.2$ based on the one-factor, three-factor, and four-factor models. In panel B, we compare our proposed estimation with the traditional parametric estimation of the timing coefficient in the Henriksson-Merton approach based on the four-factor model using daily data for 3210 funds from September 1, 1998 to December 31, 2018. We report the number of funds in different portfolios. Panel C reports the fund's timing sign based on Theorem 1 and synthetic funds using daily data for 2950 funds.

they expect to time the market in the future. Disentangling these two alternative hypotheses is an essential task for future research.

As robustness checks, we also estimate the weighted nonparametric Henriksson-Merton market timing measure for each fund based on the Carhart (1997) model. We report the differences in stock characteristics in Panel B of Table 8. The results are qualitatively similar. In Panel D, we find that even controlling for trading friction (measured by stock size), funds with positive timing ability and negative timing ability still hold stocks with bigger market capitalization which are more liquid, indicating that market timing is related to fund holdings.

Finally, based on our identification of funds with zero and nonzero timing ability, we further examine the association between market timing ability and stock picking skill after excluding funds with insignificant timing measures. The timing of insignificant funds may not be accurately measured, or those funds may not try to time the market. For each fund, we estimate alpha (i.e., stock picking skill) based on commonly used benchmarks, including the CAPM model, the Fama and French (1996) model, and the Carhart (1997) model. We exclude funds with insignificant market timing ability from the sample, following the discussions in Section 2, and report the percentages of each combination of signs in Panel A and B of Tables 9.

As seen in Panel A of Table 9, after removing funds with zero market timing ability, calculations based on nonparametric Treynor-Mazuy market timing ability suggest that the negative association between fund alpha and market timing ability dominates the positive association for the one-factor model. This confirms the finding in Back et al. (2018). The results based

Table 8
Mutual Fund Holding Characteristics – Trading Frictions.

Variable	Negative $\hat{\hat{\gamma}}$	Zero $\hat{\hat{\gamma}}$	Positive $\hat{\hat{\gamma}}$	Pos. - Neg.	*t*-stat
Panel A: Weighted nonparametric Treynor-Mazuy market timing measure					
mve	15.4999	16.0274	16.0973	0.597***	(18.69)
mve_ia	11.7030	18.5814	21.4318	9.729***	(13.69)
ill	0.0346	0.0113	0.0105	-0.024***	(-5.53)
std_dolvol	0.4276	0.3949	0.3901	-0.037***	(-16.20)
zerotrade	0.0084	0.0051	0.0054	-0.003***	(-5.03)
Panel B: Weighted nonparametric Henriksson-Merton market timing measure					
mve	15.3744	16.0071	16.4288	1.054***	(74.22)
mve_ia	10.0676	18.5401	24.5951	14.527***	(18.27)
ill	0.0340	0.0116	0.0049	-0.029***	(-5.38)
std_dolvol	0.4313	0.3960	0.3747	-0.057***	(-15.54)
zerotrade	0.0082	0.0055	0.0020	-0.006***	(-6.11)
Panel C: Weighted nonparametric TM measure after controlling *mve*					
mve	15.5388	16.1816	16.0817	0.543***	(15.83)
Panel D: Weighted nonparametric HM measure after controlling *mve*					
mve	15.3494	15.9903	16.2855	0.936***	(21.68)

Note: Using data for 3210 funds from September 1, 1998 to December 31, 2018, we compute the weighted nonparametric Treynor-Mazuy (Panel A) and Henriksson-Merton(Panel B) market timing measure based on the four-factor model and generate mutual fund portfolios with significantly positive, zero, and negative market timing ability. In Panels C and D, we control firm size (*mve*) to alleviate spurious timing bias and re-estimate the weighted nonparametric TM and HM measures. 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. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

on weighted nonparametric Treynor-Mazuy market timing ability support a negative association between market timing ability and stock picking skill across all three models. Therefore, removing funds with zero market timing ability generates a clearer picture of the tradeoff documented in Back et al. (2018). Panel B draws similar conclusions for the nonparametric Henriksson-Merton market timing measure. In Panels C and D, we control for spurious timing and re-investigate the tradeoff between alpha and gamma. Our main conclusions, such as the negative association between fund alpha and market timing ability dominates the positive association for the one-factor model, still hold.

We conjecture that funds with positive timing do not focus on stock picking because of their holding. Based on Table 8, funds with positive timing hold stocks with less trading friction, and those stock characteristics also mean low expected returns. Furthermore, in Table 10, we sort funds based on alpha (a measure of stock picking skill) and find that funds with stock picking skill hold smaller stocks. In Panel B, we control holding stock size and classify positive and negative timing funds based on the corresponding synthetic funds. After sorting funds based on alpha, we still find that funds with different stock picking skills hold stocks with significantly different trading frictions, and funds with positive alpha hold less liquid stocks.[10]

Given the tradeoff, we compare the difference between our method and the traditional parametric method from the perspective of net return. We divide funds into two groups – funds with significantly positive and negative timing ability – based on nonparametric or parametric timing measures. We assume that investors buy the value of one dollar of each portfolio in January 2000 and hold the portfolio until December 2018. We compare the cumulative returns of these two portfolios. Fig. 4 shows the return on $1.00 invested in mutual funds based on the Treynor-Mazuy measure based on the four-factor model.

Panel A shows the result based on the weighted nonparametric measures. As shown in Table 6, the positive-timing portfolio has 372 funds, and the negative-timing portfolio has 412 funds. During the twenty years, the good timers have higher net returns than the poor timers, with the value near $2.2 and $1.9, respectively, at the end of 2018. That is, investing in the funds with positive timing ability under our methodology gives higher net returns for investors than investing in funds with negative timing ability. Panel B illustrates the net return of the two portfolios based on the traditional parametric model. The trend of the changing values of the positive-timing portfolio (with 534 funds) and the negative-timing portfolio (with 839 funds) are almost the same. But good-timers' cumulative net return is even a bit lower than poor ones. Thus, in terms of net return, our weighted nonparametric method performs better in identifying funds with significantly positive and negative timing abilities.

[10] Busse et al. (2021) find that even adjusted for the risk factor, holding stock characteristics (such as stock size) is still significantly correlated with fund alpha.

Table 9
Tradeoffs Between Timing and Stock Picking.

Panel A: Treynor-Mazuy market timing measures						
One-factor	641	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	1,142	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	2.65%	53.83%	$\hat{\tilde{\alpha}} > 0$	16.37%	36.79%
	$\hat{\alpha} < 0$	19.03%	24.49%	$\hat{\tilde{\alpha}} < 0$	28.28%	18.56%
Three-factor	508	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	881	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	11.22%	19.29%	$\hat{\tilde{\alpha}} > 0$	15.21%	27.13%
	$\hat{\alpha} < 0$	28.74%	40.75%	$\hat{\tilde{\alpha}} < 0$	30.08%	27.58%
Four-factor	460	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	784	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	13.48%	17.61%	$\hat{\tilde{\alpha}} > 0$	17.73%	23.60%
	$\hat{\alpha} < 0$	27.83%	41.08%	$\hat{\tilde{\alpha}} < 0$	29.72%	28.95%
Panel B: Henriksson-Merton market timing measures						
One-factor	871	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	754	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	3.21%	51.67%	$\hat{\tilde{\alpha}} > 0$	18.97%	29.70%
	$\hat{\alpha} < 0$	19.29%	25.83%	$\hat{\tilde{\alpha}} < 0$	37.27%	14.06%
Three-factor	772	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	701	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	12.18%	25.12%	$\hat{\tilde{\alpha}} > 0$	15.69%	26.25%
	$\hat{\alpha} < 0$	23.32%	39.38%	$\hat{\tilde{\alpha}} < 0$	31.24%	24.82%
Four-factor	683	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	771	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	14.06%	22.11%	$\hat{\tilde{\alpha}} > 0$	16.73%	25.94%
	$\hat{\alpha} < 0$	24.74%	39.09%	$\hat{\tilde{\alpha}} < 0$	28.15%	29.18%
Panel C: Treynor-Mazuy market timing measures (ControlSize)						
One-factor	1,136	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	1,351	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	4.04%	48.42%	$\hat{\tilde{\alpha}} > 0$	11.25%	43.60%
	$\hat{\alpha} < 0$	12.50%	35.04%	$\hat{\tilde{\alpha}} < 0$	16.36%	28.79%
Three-factor	1,650	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	1,136	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	7.22%	26.48%	$\hat{\tilde{\alpha}} > 0$	6.16%	32.13%
	$\hat{\alpha} < 0$	15.82%	50.48%	$\hat{\tilde{\alpha}} < 0$	14.88%	46.83%
Four-factor	1,459	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	1,005	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	6.65%	26.18%	$\hat{\tilde{\alpha}} > 0$	7.27%	27.46%
	$\hat{\alpha} < 0$	13.43%	53.74%	$\hat{\tilde{\alpha}} < 0$	15.52%	49.75%
Panel D: Henriksson-Merton market timing measures (ControlSize)						
One-factor	1,378	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	934	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	3.05%	49.49%	$\hat{\tilde{\alpha}} > 0$	19.38%	33.19%
	$\hat{\alpha} < 0$	10.23%	37.23%	$\hat{\tilde{\alpha}} < 0$	28.37%	19.06%
Three-factor	1,775	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	828	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	7.15%	26.59%	$\hat{\tilde{\alpha}} > 0$	17.27%	19.57%
	$\hat{\alpha} < 0$	17.41%	48.85%	$\hat{\tilde{\alpha}} < 0$	32.00%	31.16%
Four-factor	1,590	$\hat{\gamma} > 0$	$\hat{\gamma} < 0$	858	$\hat{\tilde{\gamma}} > 0$	$\hat{\tilde{\gamma}} < 0$
	$\hat{\alpha} > 0$	6.86%	26.35%	$\hat{\tilde{\alpha}} > 0$	22.26%	13.54%
	$\hat{\alpha} < 0$	14.59%	52.20%	$\hat{\tilde{\alpha}} < 0$	35.43%	28.67%

Note: In Panels A and B, we compute the percentage of 3210 funds from September 1, 1998 to December 31, 2018 with each possible combination of signs on the estimates for α and γ or $\tilde{\gamma}$ using daily fund returns using only funds with nonzero market timing measures. In Panels C and D, we control firm size (*mve*) to alleviate spurious timing bias and compute the percentage of 2950 funds with each possible combination. The integers denote the total number of funds with nonzero nonparametric Treynor-Mazuy market timing measures based on the test of using $\hat{\tilde{\gamma}}$ with $h = 0.2$ at the 10% significance level.

Table 10
Mutual Fund Characteristics Sorted by Alpha.

Trading Frictions	Negative $\hat{\tilde{\alpha}}$	Positive $\hat{\tilde{\alpha}}$	Pos. - Neg.	t-stat
Panel A: Weighted nonparametric Measure				
mve	16.0317	15.8715	-0.160***	(-9.21)
Panel B: Weighted nonparametric measure after controlling *mve*				
mve	15.7314	14.0038	-1.728***	(-6.03)

Note: In Panel A, using data for 3210 funds from September 1, 1998 to December 31, 2018, we compute fund alpha based on the four-factor model and generate mutual fund portfolios with positive and negative stock picking ability. Panel B reports the results based on the inference with synthetic funds. We report the time-series averages of the monthly cross-sectional means in each portfolio and the differences in means between these two portfolios. 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. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

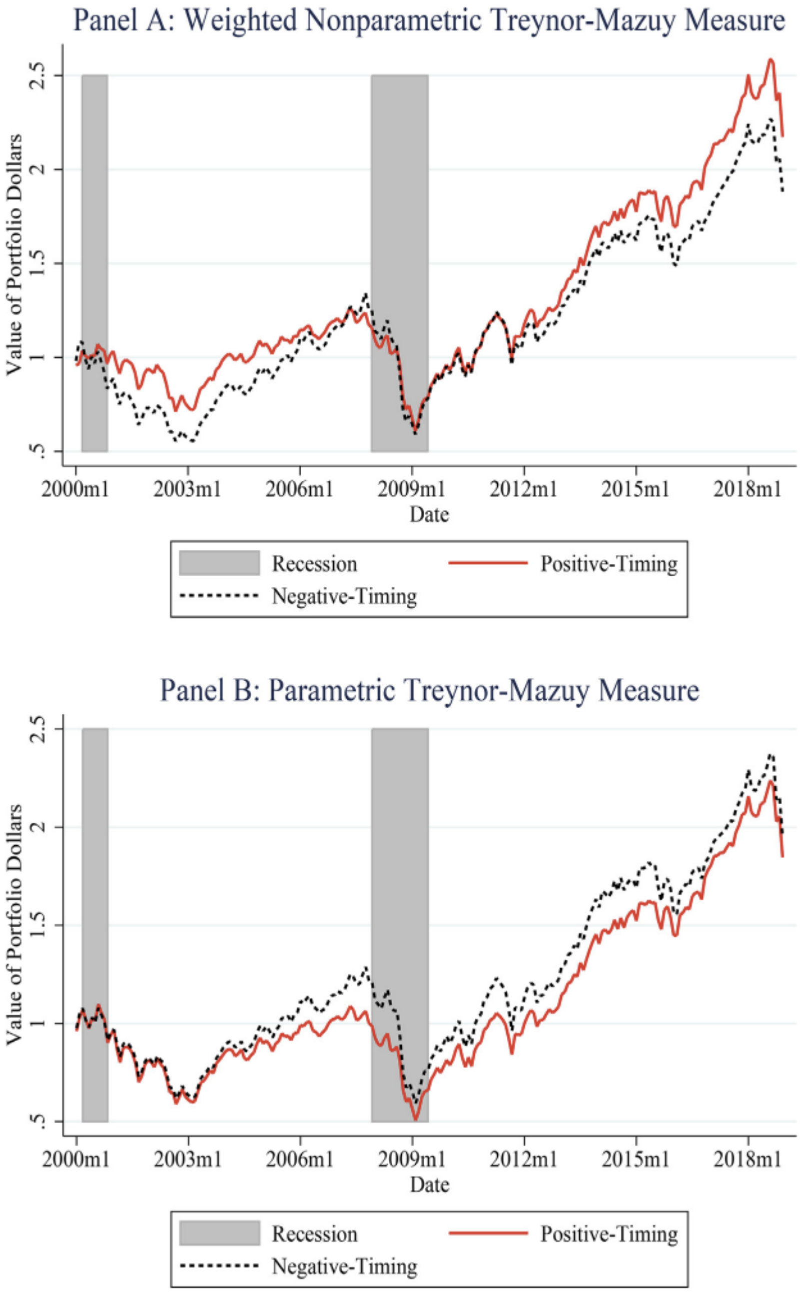

Fig. 4. *The Return on One Dollar Invested in Mutual Funds Based on Market Timing.* This figure shows the changing value of $1.00 invested in January 2000 through December 2018 in two equal-weighted portfolios of mutual funds. The portfolios are funds with significantly positive and negative timing abilities based on timing measures estimated by non-parametric (Panel A) and parametric (Panel B) Treynor-Mazuy approach based on the four-factor model. We measure recessions using the standard National Bureau of Economic Research (NBER) recession indicator.

5. Conclusions

Given the stylized facts of heavy tails and heteroscedasticity of daily fund returns, we propose robust methods for testing the absence of market timing ability for individual funds. We apply our tests to actively managed U.S. equity funds and find that our method disagrees with the inference of the traditional parametric approach for about 39% of funds in the sample in terms of timing ability. We also find that funds with positive and negative timing abilities hold stocks with dramatically different characteristics. Compared to funds with perverse timing ability, funds with positive timing ability hold stocks with lower trading frictions. The results can not be fully attributed to spurious timing bias. This association may be because only funds with liquid stocks time the market, or managers with positive timing ability tend to have liquid stocks. We also investigate the association between market timing ability and stock picking skill and find robust supporting evidence on the tradeoff between them, especially when excluding funds with zero timing ability. Finally, we find good timers provide higher net returns to investors than poor timers based on our weighted nonparametric measure.

Appendix A. The choice of weights

The weights in defining (7) and (8) are

$$\begin{cases} w_{t,\varepsilon} = w_{t,\varepsilon}(h) = w_{t,Y}(h) + w_{t,X}(h), \\ w_{t,Y} = w_{t,Y}(h) = \max\{1, \sum_{i=0}^{t} e^{\log(h)\log^2(i+1)} \frac{|Y_{t-i}|}{C_Y}\}, \\ w_{t,X} = w_{t,X}(h) = \max\{1, \sum_{i=0}^{t} e^{\log(h)\log^2(i+1)} \frac{\max(|X_{t-i,1}|,\ldots,|X_{t-i,d}|)}{C_X}\} \end{cases} \tag{11}$$

for some $h \in (0,1)$, C_Y and C_X are chosen as the 90% quantile of the distribution of $|Y_i|$ and $\max(|X_{t,1}|,\ldots,|X_{t,d}|)$, respectively to avoid overweight. Like Ling (2007) and Zhu and Ling (2011,2015), the key idea in choosing these weights is to bound

$$\sum_{i=0}^{\infty}\rho^i|\varepsilon_{t-i}|,\ \sum_{i=0}^{\infty}\rho^i|X_{t-i,1}|,\ldots,\ \sum_{i=0}^{\infty}\rho^i|X_{t-i,d}|$$

for some $\rho \in (0,1)$, depending on the underlying GARCH model. In practice, we replace C_Y and C_X by their sample quantiles, but we can show that the asymptotic limit remains the same by following the arguments in Ling (2007) or He et al. (2020). The simulation study in Section 3 shows that the proposed method is robust against $h = 0.2, 0.3, 0.4, 0.5$. Like kernel density estimation, choosing the optimal weight in terms of coverage probability is challenging, which requires the so-called Edgeworth expansion. We will not pursue it in this paper.

Appendix B. Regularity conditions

Recall that $\mathcal{F}_t$ is the σ-field generated by

$$\{\bar{\eta}_{s_1,1}, \bar{\eta}_{s_2,2},\ldots,\bar{\eta}_{s_2,d},\eta_{s_2} : s_1 \le t+1,\ s_2 \le t\}.$$

We list some regularity conditions for Theorems 1 and 2 in Section 2.

- C1) $\{\varepsilon_t\}$ and $\{\boldsymbol{X}_t\}$ are strictly stationary and ergodic; see conditions in Theorem 3.1 of Basrak et al. (2002).
- C2) Each of $\{\eta_t\}$ and $\{\bar{\boldsymbol{\eta}}_t = (\bar{\eta}_{t,1},\ldots,\bar{\eta}_{t,d})^\top\}$ is a sequence of independent and identically distributed random vectors with means zero and variances one. Also, $E(\eta_t\bar{\boldsymbol{\eta}}_t) = \boldsymbol{0}$, ensuring the consistency of the least squares estimation.
- C3) There exists $\delta > 0$ such that $E(|\varepsilon_t|^\delta) < \infty$, $E(|X_{t,k}|^\delta) < \infty$, $E(|\eta_t|^{2+\delta}) < \infty$, and $E(|\eta_t\bar{\eta}_{t,k}|^{2+\delta}) < \infty$ for $k = 1,\ldots,d$.
- C4) The covariance matrices of $(1,\boldsymbol{X}_t^\top)^\top/w_{t-1,X}$ and

$$\left(\frac{\{\varepsilon_t - E(\varepsilon_t|\mathcal{F}_{t-1})\}H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}},\ \frac{\varepsilon_t}{w_{t-1,\varepsilon}w_{t-1,X}},\ \frac{\varepsilon_t\boldsymbol{X}_t^\top}{w_{t-1,\varepsilon}w_{t-1,X}}\right)^\top$$

are positive definite.

Appendix C. Theoretical proofs

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

Proof of Theorem 1.. Define

$$\bar{\boldsymbol{X}}_t = (1,\boldsymbol{X}_t^\top)^\top \text{ for } t = 1,\ldots,n,\ \boldsymbol{\theta} = (\alpha,\boldsymbol{\beta}^\top)^\top,\ \hat{\boldsymbol{\theta}} = (\hat{\alpha},\hat{\boldsymbol{\beta}}^\top)^\top,$$

$$\boldsymbol{\Gamma}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\bar{\boldsymbol{X}}_t\bar{\boldsymbol{X}}_t^\top}{w_{t-1,\varepsilon}, w_{t-1,X}},\ \bar{\boldsymbol{\Gamma}}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\bar{\boldsymbol{X}}_t H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}},$$

$$\boldsymbol{W}_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t\bar{\boldsymbol{X}}_t}{w_{t-1,\varepsilon}w_{t-1,X}},\ Z_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\{\varepsilon_t - \sigma_t E(\eta_t|\mathcal{F}_{t-1})\}H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}}.$$

Then, we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \boldsymbol{\Gamma}_n^{-1}\boldsymbol{W}_n \tag{C.1}$$

and

$$\sqrt{n}\{\hat{\gamma} - \frac{1}{n}\sum_{t=1}^{n}\frac{\sigma_t E(\eta_t|\mathcal{F}_{t-1})H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}}\} = Z_n - \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top\bar{\boldsymbol{\Gamma}}_n. \tag{C.2}$$

Like Ling (2007), we can show that for any $\zeta > 0$

$$E(|\frac{\sigma_t}{w_{t-1,\varepsilon}}|^{\zeta}) < \infty, \ E(|\frac{\bar{\sigma}_{t,i}}{w_{t-1,X}}|^{\zeta}) < \infty \text{ for } i = 1, \ldots, d. \tag{C.3}$$

Using (C.3), the weak law of large numbers for a Martingale differences sequence in Hall and Heyde (1980), and the ergodicity of $\{\varepsilon_t\}$ and $\{\boldsymbol{X}_t\}$, as $n \to \infty$, we have

$$\begin{cases} \boldsymbol{\Gamma}_n \xrightarrow{p} \boldsymbol{\Gamma} := \lim_{t\to\infty} E(\frac{\bar{\boldsymbol{X}}_t\bar{\boldsymbol{X}}_t^\top}{w_{t-1,\varepsilon}w_{t-1,X}}), \\ \bar{\boldsymbol{\Gamma}}_n \xrightarrow{p} \bar{\boldsymbol{\Gamma}} := \lim_{t\to\infty} E(\frac{\bar{\boldsymbol{X}}_t H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}}), \\ \frac{1}{n}\sum_{t=1}^n \frac{\varepsilon_t^2\bar{\boldsymbol{X}}_t\bar{\boldsymbol{X}}_t^\top}{w_{t-1,\varepsilon}^2 w_{t-1,X}^2} \xrightarrow{p} \boldsymbol{\Sigma} := \lim_{t\to\infty} E\left(\frac{\varepsilon_t^2\bar{\boldsymbol{X}}_t\bar{\boldsymbol{X}}_t^\top}{w_{t-1,\varepsilon}^2 w_{t-1,X}^2}\right), \\ \frac{1}{n}\sum_{t=1}^n \frac{\{\varepsilon_t - \sigma_t E(\eta_t|\mathcal{F}_{t-1})\}^2 H^2(X_{t,1})}{w_{t-1,\varepsilon}^2 w_{t-1,X}^2\{1+H^2(X_{t,1})\}} \xrightarrow{p} \sigma^2 := \lim_{t\to\infty} E\left(\frac{\{\varepsilon_t - \sigma_t E(\eta_t|\mathcal{F}_{t-1})\}^2 H^2(X_{t,1})}{w_{t-1,\varepsilon}^2 w_{t-1,X}^2\{1+H^2(X_{t,1})\}}\right), \\ \frac{1}{n}\sum_{t=1}^n \frac{\varepsilon_t\{\varepsilon_t - \sigma_t E(\eta_t|\mathcal{F}_{t-1})\}\bar{\boldsymbol{X}}_t H(X_{t,1})}{w_{t-1,\varepsilon}^2 w_{t-1,X}^2\sqrt{1+H^2(X_{t,1})}} \xrightarrow{p} \boldsymbol{\xi} := \lim_{t\to\infty} E\left(\frac{\varepsilon_t\{\varepsilon_t - \sigma_t E(\eta_t|\mathcal{F}_{t-1})\}\bar{\boldsymbol{X}}_t H(X_{t,1})}{w_{t-1,\varepsilon}^2 w_{t-1,X}^2\sqrt{1+H^2(X_{t,1})}}\right). \end{cases} \tag{C.4}$$

Because

$$E(\frac{\varepsilon_t\bar{\boldsymbol{X}}_t}{w_{t-1,\varepsilon}w_{t-1,X}}|\mathcal{F}_{t-1}) = \boldsymbol{0}$$

and

$$E(\frac{\varepsilon_t H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}}|\mathcal{F}_{t-1}) = \frac{\sigma_t E(\eta_t|\mathcal{F}_{t-1})H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}},$$

using (C.3), (C.4), and the central limit theorem for a Martingale differences sequence in Hall and Heyde (1980), we have

$$(\boldsymbol{W}_n^\top, Z_n)^\top \xrightarrow{d} N(0, \bar{\boldsymbol{\Sigma}}) \text{ as } n \to \infty, \tag{C.5}$$

where $\bar{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\top & \sigma^2 \end{pmatrix}$. Therefore, it follows from (C.1), (C.2), and (C.5) that

$$\sqrt{n}\{\hat{\bar{\gamma}} - \frac{1}{n}\sum_{t=1}^n \frac{\sigma_t E(\eta_t|\mathcal{F}_{t-1})H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}}\} = Z_n - \boldsymbol{W}_n^\top(\boldsymbol{\Gamma}^{-1})^\top\bar{\boldsymbol{\Gamma}} + o_p(1) \xrightarrow{d} N(0, \bar{\sigma}^2)$$

as $n \to \infty$, where

$$\bar{\sigma}^2 = \sigma^2 - 2\boldsymbol{\xi}^\top(\boldsymbol{\Gamma}^{-1})^\top\bar{\boldsymbol{\Gamma}} + \mathrm{tr}((\boldsymbol{\Gamma}^{-1})^\top\bar{\boldsymbol{\Gamma}}\bar{\boldsymbol{\Gamma}}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma})$$

with $\mathrm{tr}(A)$ denoting the trace of matrix A. □

Proof of Theorem 2.. Define

$$\boldsymbol{\Gamma}_n^b = \frac{1}{n}\sum_{t=1}^n \frac{\delta_t^b\bar{\boldsymbol{X}}_t\bar{\boldsymbol{X}}_t^\top}{w_{t-1,\varepsilon}, w_{t-1,X}}, \ \bar{\boldsymbol{\Gamma}}_n^b = \frac{1}{n}\sum_{t=1}^n \frac{\delta_t^b\bar{\boldsymbol{X}}_t H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}},$$

$$\boldsymbol{W}_n^b = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\delta_t^b\varepsilon_t\bar{\boldsymbol{X}}_t}{w_{t-1,\varepsilon}w_{t-1,X}}, \ Z_n^b = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\delta_t^b\{\varepsilon_t - \sigma_t E(\eta_t|\mathcal{F}_{t-1})\}H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}}.$$

Then, we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^b - \boldsymbol{\theta}) = (\boldsymbol{\Gamma}_n^b)^{-1}\boldsymbol{W}_n^b$$

and

$$\sqrt{n}\{\hat{\bar{\gamma}}^b - \frac{1}{n}\sum_{t=1}^n \frac{\delta_t^b\sigma_t E(\eta_t|\mathcal{F}_{t-1})H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}}\} = Z_n^b - \sqrt{n}(\hat{\boldsymbol{\theta}}^b - \boldsymbol{\theta})^\top\bar{\boldsymbol{\Gamma}}_n^b.$$

Like the proof of Theorem 1, we have

$$\boldsymbol{\Gamma}_n^b \xrightarrow{p} \boldsymbol{\Gamma} \text{ and } \bar{\boldsymbol{\Gamma}}_n^b \xrightarrow{p} \bar{\boldsymbol{\Gamma}} \text{ as } n \to \infty.$$

Hence,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^b - \boldsymbol{\theta}) = \boldsymbol{\Gamma}^{-1}\boldsymbol{W}_n^b + o_p(1)$$

and

$$\sqrt{n}\{\hat{\bar{\gamma}}^b - \frac{1}{n}\sum_{t=1}^{n}\frac{\delta_t^b\sigma_t E(\eta_t|\mathcal{F}_{t-1})H(X_{t,1})}{w_{t-1,\varepsilon}w_{t-1,X}\sqrt{1+H^2(X_{t,1})}}\} = Z_n^b - \sqrt{n}(\hat{\boldsymbol{\theta}}^b - \boldsymbol{\theta})^\top\bar{\boldsymbol{\Gamma}} + o_p(1),$$

which imply that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}}) = \boldsymbol{\Gamma}^{-1}(\boldsymbol{W}_n^b - \boldsymbol{W}_n) + o_p(1)$$

and

$$\sqrt{n}\{\hat{\bar{\gamma}}^b - \hat{\bar{\gamma}}\} = (Z_n^b - Z_n) - \sqrt{n}(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}})^\top\bar{\boldsymbol{\Gamma}} + o_p(1).$$

Using the same arguments in proving Theorem 1, we can show that the joint limit of $Z_n^b - Z_n$ and $\boldsymbol{W}_n^b - \boldsymbol{W}_n$ is the same as that of Z_n and $\boldsymbol{W}_n$. Hence,

$$\sqrt{n}(\hat{\bar{\gamma}}^b - \hat{\bar{\gamma}}) \xrightarrow{d} N(0, \bar{\sigma}^2) \text{ as } n \to \infty.$$

Similarly, we can show

$$\frac{n}{B}\sum_{b=1}^{B}(\hat{\bar{\gamma}}^b - \hat{\bar{\gamma}})^2 = n\hat{\bar{\gamma}}^2 + o_p(1) = \bar{\sigma}^2 + o_p(1)$$

i.e., Theorem 2 holds. □

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jedc.2023.104635.

References

Back, K., Crane, A.D., Crotty, K., 2018. Skewness consequences of seeking alpha. Rev. Financ. Stud. 31, 4720–4761.
Basrak, B., Davis, R.A., Mikosch, T., 2002. Regular variation of GARCH processes. Stoch. Process. Their Appl. 99, 95–116.
Becker, C., Ferson, W., Myers, D.H., Schill, M.J., 1999. Conditional market timing with benchmark investors. J. Financ. Econ. 52, 119–148.
Bollen, N.P., Busse, J.A., 2001. On the timing ability of mutual fund managers. J. Finance 56, 1075–1094.
Bollen, N.P., Busse, J.A., 2005. Short-term persistence in mutual fund performance. Rev. Financ. Stud. 18, 569–597.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econom. 31, 307–327.
Breen, W., Jagannathan, R., Ofer, A., 1986. Correcting for heteroscedasticity in tests for market timing ability. J. Bus. 59, 585–598.
Busse, J., Jiang, L., Tang, Y., 2021. Double adjusted mutual fund performance. Rev. Asset Pricing Stud. 11, 169–208.
Busse, J.A., 1999. Volatility timing in mutual funds: evidence from daily returns. Rev. Financ. Stud. 12, 1009–1041.
Busse, J.A., Tong, L., Tong, Q., Zhang, Z., 2019. Trading regularity and fund performance. Rev. Financ. Stud. 32, 374–422.
Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52, 57–82.
Chance, D.M., Hemler, M.L., 2001. The performance of professional market timers: daily evidence from executed strategies. J. Financ. Econ. 62, 377–411.
Chang, E.C., Lewellen, W.G., 1984. Market timing and mutual fund investment performance. J. Bus. 57, 57–72.
Elton, E.J., Gruber, M.J., Blake, C.R., 2012. An examination of mutual fund timing ability using monthly holdings data. Rev. Finance 16, 619–645.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kindom inflation. Econometrica 50, 987–1007.
Engle, R.F., 1990. Discussion: stock market volatility and the crash of 1987. Rev. Financ. Stud. 3, 103–106.
Evans, R., 2010. Mutual fund incubation. J. Finance 65, 1581–1611.
Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. J. Finance 51, 55–84.
Fama, E.F., French, K.R., 2010. Luck versus skill in the cross-section of mutual fund returns. J. Finance 65, 1915–1947.
Fornari, F., Mele, A., 1997. Sign- and volatility-switching ARCH models: theory and applications to international stock markets. J. Appl. Econom. 12, 1779–1801.
Goetzmann, W.N., Ingersoll, J., Ivković, Z., 2000. Monthly measurement of daily timers. J. Financ. Quant. Anal. 35, 257–290.
Green, J., Hand, J.R., Zhang, X.F., 2017. The characteristics that provide independent information about average US monthly stock returns. Rev. Financ. Stud. 30, 4389–4436.
Grinblatt, M., Titman, S., 1988. The Evaluation of Mutual Fund Performance: An Analysis of Monthly Returns. Working Paper. John E. Anderson Graduate School of Management, University of California.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Applications. Academic Press, New York.
Hansen, P.R., 2005. A test for superior predictive ability. J. Bus. Econ. Stat. 23, 365–380.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.
He, Y., Hou, Y., Peng, L., Shen, H., 2020. Inference for conditional value-at-risk of a predictive regression. Ann. Stat. 48, 3442–3464.
Henriksson, R.D., 1984. Market timing and mutual fund performance: an empirical investigation. J. Bus. 57, 73–96.
Henriksson, R.D., Merton, R.C., 1981. On market timing and investment performance II. Statistical procedures for evaluating skills. J. Bus. 54, 513–533.
Hill, B., 1975. A simple general approach to inference about the tail of a distribution. Ann. Stat. 3, 1163–1174.
Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Rev. Financ. Stud. 28, 650–705.
Hsing, T., 1991. On tail index estimation using dependent data. Ann. Stat. 19, 1547–1569.
Jagannathan, R., Korajczyk, R., 1986. Assessing the market timing performance of managed portfolios. J. Bus. 59, 217–235.
Jensen, M.C., 1968. The performance of mutual funds in the period 1945–1964. J. Finance 23, 389–416.
Jiang, G.J., Yao, T., Yu, T., 2007. Do mutual funds time the market? Evidence from portfolio holdings. J. Financ. Econ. 86, 724–758.
Jiang, G.J., Zaynutdinova, G.R., Zhang, H., 2021. Stock-selection timing. J. Bank. Finance 125, 106089.

Jiang, L., Liu, W., Peng, L., 2022. Test for market timing using daily fund returns. J. Bus. Econ. Stat.. forthcoming
Jiang, W., 2003. A nonparametric test of market timing. J. Empir. Finance 10, 399–425.
Jin, Z., Ying, Z., Wei, L.J., 2001. A simple resampling method by perturbing the minimand. Biometrika 88, 381–390.
Kacperczyk, M., Nieuwerburgh, S.V., Veldkamp, L., 2014. Time-varying fund manager skill. J. Finance 69, 1455–1484.
Kacperczyk, M., Sialm, C., Zheng, L., 2008. Unobserved actions of mutual funds. Rev. Financ. Stud. 21, 2379–2416.
Ling, S., 2007. Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. J. Econom. 140, 849–873.
Mamaysky, H., Spiegel, M., Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. Rev. Financ. Stud. 21, 233–264.
Newey, W., West, K., 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Schwert, G.W., 1989. Why does stock market volatility change over time? J. Finance 45, 1129–1155.
Sentana, E., 1995. Quadratic ARCH models. Rev. Econ. Stud. 62, 639–661.
Treynor, J., Mazuy, K., 1966. Can mutual funds outguess the market? Harv. Bus. Rev. 44, 131–146.
Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21, 1455–1508.
Zakoian, J.M., 1994. Threshold heteroskedastic models. J. Econ. Dyn. Control 18, 931–955.
Zhu, K., 2016. Bootstrapping the portmanteau tests in weak autoregressive moving average models. J. R. Stat. Soc. Ser. B 78, 463–485.
Zhu, K., Ling, S., 2011. Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. Ann. Stat. 39, 2131–2163.
Zhu, K., Ling, S., 2015. LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises. J. Am. Stat. Assoc. 110, 784–794.