Statistics and Its Interface Volume 16 (2023) 337–347

Empirical likelihood-based portmanteau tests for autoregressive moving average models with possible infinite variance innovations*

Xiaohui Liu, Donghui Fan, Xu Zhang†, and Catherine Liu

It is an important task in the literature to check whether a fitted autoregressive moving average (ARMA) model is adequate, while the currently used tests may suffer from the size distortion problem when the underlying autoregressive models have low persistence. To fill this gap, this paper proposes two empirical likelihood-based portmanteau tests. The first one is naive but can serve as a benchmark, and the second is for the case with infinite variance innovations. The asymptotic distributions under the null hypothesis are derived under mild moment conditions, and their usefulness is demonstrated by simulation experiments and two real data examples.

AMS 2000 subject classifications: C12, C22, G1.
Keywords and phrases: ARMA model, GARCH process, Diagnostic checking, Empirical likelihood, Infinite variance.

1. INTRODUCTION

Consider the autoregressive moving average (ARMA) model with orders p and q, denoted by ARMA(p, q),

$$X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \varepsilon_t, \tag{1}$$

where $(\mu, \phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)$ contains unknown parameters, and $\{\varepsilon_t\}$ is a martingale difference series, and this model has been widely used in many fields such as finance and economics. It is an important task in time series analysis to check whether the fitted model is adequate, i.e. the orders p or/and q may not be correctly specified, and there is a huge literature for it. The early seminal works include the Box-Pierce statistic Q_m in [4] and Ljung-Box statistic $\tilde{Q}_m$ in [20], and they can be defined as

$$Q_m = n\hat{\boldsymbol{\rho}}^\top \hat{\boldsymbol{\rho}}, \quad \text{and} \quad \tilde{Q}_m = n\hat{\boldsymbol{\rho}}^\top W \hat{\boldsymbol{\rho}},$$

where the diagonal matrix $W = \text{diag}\{(n+2)/(n-1), (n+2)/(n-2), \cdots, (n+2)/(n-m)\}$, $\hat{\boldsymbol{\rho}} := (\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_m)^\top$, the residuals auto-correlation at lag k has the form of

$$\hat{\rho}_k = \sum_{t=k+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-k} \Big/ \sum_{t=1}^{n} \hat{\varepsilon}_t^2,$$

and $\{\hat{\varepsilon}_t\}$ are residuals from the fitted ARMA model at (1).

Note that $\tilde{Q}_m$ is a weighted version of Q_m, and it usually has a better performance especially when the sample size n is relatively small. Further improvements along this line include the weighted Ljund-Box test statistic in [11]. All these test statistics are easily implemented, and hence they have already been widely applied in practice. However, as pointed out by [33], the asymptotic properties of these test statistics are only valid under a strong condition that $\{\varepsilon_t\}$ are independent and identically distributed (*i.i.d.*) random variables. [33] developed an interesting random weighting (RW) technique to calculate the critical values of these test statistics, and hence these easy-to-implemented tests can be extended to the case that ε_t's are uncorrelated, but not necessarily independent.

In the meanwhile, when autoregressive (AR) models have low persistence, i.e., AR coefficients are relatively small, the RW method in [33] still suffers from significant size distortion; see simulation results in Section 3 for details. As a result, this paper revisits the literature of diagnostic checking for AMRA models, and a new test statistic is then proposed by the profile empirical likelihood (EL) method [24, 26]. It can be further shown that, under mild conditions, the proposed test statistic has the null chi-squared distribution, which is a desirable property for tests.

On the other hand, financial and economic data usually exhibit the phenomenon of volatility clustering, which can be interpreted by the conditional heteroscedasticity. [9] first suggested an autoregressive conditional heteroskedastic (ARCH) model for it. Moreover, by noting that the AR process usually needs a higher order than the ARMA process

arXiv: 2209.09704
*We thank the co-editors, co-guest editors and two anonymous reviewers for their valuable suggestions and comments, which have led to significant improvements of this paper. Xiaohui Liu's research is supported by NSF of China (Grant No.11971208), National Social Science Foundation of China (21&ZD152). Xu Zhang's research is supported by NSF of China (Grant No.12171167). Catherine Liu's research is partially supported by General Research Funding 15301519, RGC, HKSAR.
†Corresponding author.

in the actual modeling, [2] extended the ARCH model to a more flexible generalized autoregressive conditional heteroskedastic (GARCH) model, which not only reduces the number of parameters but also provides a better fit to the data; see, e.g., [23], [13], [25], [6], [18], [21], and references therein. The GARCH model has the form of

$$\varepsilon_t = \eta_t \sigma_t,\ \sigma_t^2 = \omega + \sum_{i=1}^{r} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} b_j \sigma_{t-j}^2, \tag{2}$$

where $\{\eta_t\}$ are *i.i.d.* random errors with means zero and variances one, $(\omega, a_1, \cdots, a_r, b_1, \cdots, b_s)$ contains unknown parameters, and ω, a_i's and b_j's are assumed to be positive. The GARCH process has the finite variance if $\sum_{i=1}^{r} a_i + \sum_{j=1}^{s} b_j < 1$, however, many financial data may exhibit an infinite variance of $\{\varepsilon_t\}$, i.e. $\sum_{i=1}^{r} a_i + \sum_{j=1}^{s} b_j$ may be very close to one.

When ε_t has an infinite variance, both the RW and EL test statistics perform poorly in terms of both sizes and powers since they are only valid for the case with finite variance innovations. As a result, this paper further proposes a weighted empirical likelihood (WeL) test statistic to check the adequacy of the fitted models at (1) with GARCH errors at (2), and the null distribution is also derived with the innovations being allowed to have infinite variance.

Both EL and WeL are developed based on the empirical likelihood methods in [26], and the original is attributed to [24]. Empirical likelihood is a popular nonparametric likelihood method and has wide and successful applications in many fields; see [28], [27] and among others. While it has attracted less attention in literature of time series. Empirical likelihood is first introduced by [5] to GARCH models to build likelihood ratio test statistics, and other applications include but are limited to constructing confidence intervals for the tail index and testing for zero median of errors; see [32], [21], etc. It is noteworthy to point out that the naive EL test only works for finite variance innovations, while the proposed WeL method is motivated by the self-weighting method to modify local quasi-maximum likelihood estimators in [18].

The remainder of the paper is organized as follows. Section 2 gives two tests, and their null distributions are also derived. Sections 3 and 4 provide simulation results and real analysis, respectively, and a quick summary is given in Section 5. The theoretical details are relegated to the Appendix.

2. METHODOLOGY AND MAIN RESULTS

Let $\boldsymbol{\gamma} := (\gamma_1, \gamma_2, \cdots, \gamma_m)^\top = (E(\varepsilon_t\varepsilon_{t-1}), E(\varepsilon_t\varepsilon_{t-2}), \cdots, E(\varepsilon_t\varepsilon_{t-m}))$ for some given $m \geq 1$, and $\boldsymbol{\theta} = (\mu, \phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)^\top$. The serial correlation hypotheses can be summarized into

$$\mathcal{H}_0 : \boldsymbol{\gamma} = 0 \quad \text{versus} \quad \mathcal{H}_1 : \boldsymbol{\gamma} \neq 0. \tag{3}$$

Assume that the observed time series $\{X_t\}_{t=1}^n$ are generated from model (1). Note that the definition of $\boldsymbol{\gamma}$ is related to the expectation. We propose to test (3) by using the empirical likelihood technique in [26]. We start with the case that ε_t has finite variance, and then extend the result to the weighted empirical likelihood test statistic for infinite variance innovations.

2.1 Finite variance innovations

For convenience, define $\varepsilon_t(\boldsymbol{\theta}) = X_t - \mu - \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \psi_j \varepsilon_{t-j}(\boldsymbol{\theta})$. Note that the least squares (LS) estimator $\hat{\boldsymbol{\theta}}$ minimizes

$$\sum_{t=1}^{n} \left(X_t - \mu - \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \psi_j \varepsilon_{t-j}(\boldsymbol{\theta}) \right)^2 .$$

That is, $\hat{\boldsymbol{\theta}}$ is the solution to

$$\sum_{t=1}^{n} \varepsilon_t(\boldsymbol{\theta}) \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0. \tag{4}$$

This motivates us to define the empirical likelihood function for testing $\mathcal{H}_0$ as follows:

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\gamma}) \; &= \sup\{\textstyle\prod_{t=m+1}^{n} (N p_t) : p_{m+1} \geq 0, \cdots, p_n \geq 0, \\ &\quad \textstyle\sum_{t=m+1}^{n} p_t = 1, \sum_{t=m+1}^{n} p_t \boldsymbol{Z}_t(\boldsymbol{\theta}, \boldsymbol{\gamma}) = 0\}, \end{aligned}$$

where $N = n - m$, and $\boldsymbol{Z}_t(\boldsymbol{\theta}, \boldsymbol{\gamma}) = (Z_{t,1}(\boldsymbol{\theta}, \boldsymbol{\gamma})^\top, Z_{t,p+q+1}(\boldsymbol{\theta}, \boldsymbol{\gamma}), \cdots, Z_{t,p+q+m}(\boldsymbol{\theta}, \boldsymbol{\gamma}))^\top$ with

$$\begin{cases} \boldsymbol{Z}_{t,1}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \varepsilon_t(\boldsymbol{\theta}) \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\ Z_{t,p+q+l}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \varepsilon_t(\boldsymbol{\theta}) \varepsilon_{t-l}(\boldsymbol{\theta}) - \gamma_l, \quad l = 1, 2, \cdots, m. \end{cases}$$

Throughout this paper, we compute $\partial \varepsilon_t(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ recursively by

$$\frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\tilde{X}_t - \sum_{j=1}^{q} \psi_j \frac{\partial \varepsilon_{t-j}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad t = 1, 2, \cdots, n,$$

where $\tilde{X}_t = (1, X_{t-1}, \cdots, X_{t-p}, \varepsilon_{t-1}(\boldsymbol{\theta}), \cdots, \varepsilon_{t-q}(\boldsymbol{\theta}))^\top$.

It follows from the Lagrange multiplier technique that

$$-2 \log L(\boldsymbol{\theta}, \boldsymbol{\gamma}) = -2 \sum_{t=m+1}^{n} \log\{1 + \boldsymbol{\lambda}^\top \boldsymbol{Z}_t(\boldsymbol{\theta}, \boldsymbol{\gamma})\},$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ satisfies

$$\sum_{t=m+1}^{n} \frac{\boldsymbol{Z}_t(\boldsymbol{\theta}, \boldsymbol{\gamma})}{1 + \boldsymbol{\lambda}^\top \boldsymbol{Z}_t(\boldsymbol{\theta}, \boldsymbol{\gamma})} = 0.$$

Since we are interested in testing $\boldsymbol{\gamma}$, we consider the log-profile empirical likelihood function as follows

$$\ell(\boldsymbol{\gamma}) = -2 \log\{\sup_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\gamma})\}.$$

Denote by Θ the parameters space, which is compact subset of $\mathbb{R}^{p+q+1}$. Suppose the following conditions hold, i.e.,

(C1) The true value, say $\boldsymbol{\theta}_0$, of $\boldsymbol{\theta}$ is an interior point in Θ, and for $\boldsymbol{\theta} \in \Theta$, $\phi(z) \neq 0$ and $\psi(z) \neq 0$ when $|z| < 1$, and $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ and $\psi(z) = 1 + \sum_{j=1}^{q} \psi_j z^j$ have no comment root with $\phi_p \neq 0$ or $\psi_q \neq 0$.

(C2) $E(|\varepsilon_t|^{4+\delta}) < \infty$ for some constant $\delta > 0$.

Based on the above assumptions, we have the following result.

Theorem 1. *Suppose that $\{\varepsilon_t\}$ is a martingale difference series, i.e. there is no serial correlation existing in $\{\varepsilon_t\}$. Then, under Conditions **(C1)-(C2)**, we have*

$$\ell(0) \xrightarrow{d} \chi_m^2, \quad \text{as } n \to \infty,$$

where '$\xrightarrow{d}$' denotes the convergence in distribution, and χ_m^2 denotes a chi-squared variable with m degrees of freedom.

Based on Theorem 1, we may reject the null hypothesis $\mathcal{H}_0$ if $\ell(0) \geq \chi_m^2(1-a)$ at the significance level $a \in (0,1)$, where $\chi_m^2(1-a)$ denotes the $(1-a)$-th quantile of the distribution of χ_m^2.

2.2 Infinite variance innovations

The ARMA models are usually used in analyzing the daily financial series, which may be heavy tailed. To account for this, we further consider the case in this part that the errors ε_t follow the GARCH process at (2) with possible infinite variance.

Note that the asymptotical validity of the empirical likelihood-based statistic depends on an assumption that $E(|\varepsilon_t|^{2+\nu}) < \infty$ for some positive $\nu > 0$, which is too strict. Hence, we propose to define the profile weighted empirical likelihood function to account for the infinite variance case for testing $\mathcal{H}_0$ as follows

$$\tilde{\ell}(\boldsymbol{\gamma}) = -2 \log\{\sup_{\boldsymbol{\theta}} \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\gamma})\},$$

where

$$\begin{aligned} \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\gamma}) &= \sup \Big\{ \textstyle\prod_{t=m+1}^{n} (N p_t) : p_{m+1} \geq 0, \cdots, p_n \geq 0, \\ &\qquad \textstyle\sum_{t=m+1}^{n} p_t = 1, \sum_{t=m+1}^{n} p_t \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, \boldsymbol{\gamma}) = 0 \Big\}, \end{aligned}$$

$\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, \boldsymbol{\gamma}) = (\tilde{Z}_{t,1}(\boldsymbol{\theta}, \boldsymbol{\gamma})^\top, \tilde{Z}_{t,p+q+1}(\boldsymbol{\theta}, \boldsymbol{\gamma}), \cdots, \tilde{Z}_{t,p+q+m}(\boldsymbol{\theta}, \boldsymbol{\gamma}))^\top$ with

$$\begin{cases} \tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = w_{t-1}^{-2} \varepsilon_t(\boldsymbol{\theta}) \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\ \tilde{Z}_{t,p+q+l}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = w_{t-1}^{-1} w_{t-1-l}^{-1} \varepsilon_t(\boldsymbol{\theta}) \varepsilon_{t-l}(\boldsymbol{\theta}) - \gamma_l, \end{cases}$$

for $l = 1, 2, \cdots, m$, and

$$w_t = \max\{M_X, \sum_{i=0}^{t} e^{-\log^2(i+1)} |X_{t-i}|\}. \quad (5)$$

In the sequel we take M_X to be the 90% sample quantile of $\{|X_t|\}$. A similar strategy can be found in [14].

For $\tilde{\ell}(\boldsymbol{\gamma})$, replace Condition **(C2)** with **(C3)** and further assume **(C4)** as follows:

(C3) $E(w_{t-1}^{-4} \xi_{\rho,t-1}^{4+\delta}) < \infty$ for any $\rho \in (0,1)$, where $\xi_{\rho,t} = 1 + \sum_{i=1}^{\infty} \rho^i |X_{t-i}|$ (we suggest to use $\rho = 0.95$ based on simulations), w_t is stationary and $\mathcal{F}_t$-measurable, and $\inf_t w_t > 0$. Hereafter, δ is an arbitrary small positive constant, and $\mathcal{F}_t$ denotes the sigma field generated by $\{\eta_s : s \leq t\}$, for $t = 1, 2, \cdots, n$.

(C4) $\nu^* < 0$, where ν^* is the Lyapunov exponent of the random matrix A_t, and

$$\nu^* = \inf \left\{ \frac{1}{n} E(\ln \|\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_n\|_{\max}) : n = 1, 2, \cdots \right\},$$

where $\|\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_n\|_{\max}$ means to maximize the norm of $\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_n$, and

$$\mathbf{A}_t = \begin{pmatrix} \tilde{a}_1^* & b_2 & \ldots & b_{s-1} & \beta_s & a_2 & a_3 & \ldots & a_r \\ 1 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & 0 & \ldots & 0 \\ \eta_t^2 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix},$$

with $\tilde{a}_1^* = a_1 \eta_t^2 + b_1$ and $\|\mathbf{A}_t\| = \sup_{|\boldsymbol{x}|=1} |\mathbf{A}_t \boldsymbol{x}|$. We can prove the following result.

Theorem 2. *Suppose that $\{\eta_t\}$ is a sequence of i.i.d. random variables with mean zero and variance one, indicating that there is no serial correlation existing in $\{\varepsilon_t\}$. Then, under Conditions **(C1)**, **(C3)**, and **(C4)**, we have*

$$\tilde{\ell}(0) \xrightarrow{d} \chi_m^2, \quad \text{as } n \to \infty.$$

Remark 1. *Conditions **(C1)-(C4)** commonly used in the literature. **(C1)** and **(C4)** are assumed to guarantee the stationarity of $\{X_t\}$ and $\{\sigma_t\}$, respectively; see, e.g., [18] and [21]. **(C3)** allows the weight to reduce the moment effect of σ_t. By [21], we have that the weigh defined in (5) satisfies Condition **(C3)**. Under **(C3)**, although σ_t may have infinite variance, the result of Theorem 2 still holds, fortunately.*

Remark 2. *By 'infinite variance' we mean that $E(\varepsilon_t^2|\mathcal{F}_{t-1})$ tends to infinite almost surely as $t \to \infty$, noting that $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$, while Theorem 1 depends on Condition **(C2)**, i.e., $E(|\varepsilon_t|^{4+\delta}) < \infty$ for some constant $\delta > 0$. Hereafter, $\mathcal{F}_t$ denotes the sigma field generated by $\{\eta_s : s \leq t\}$.*

Remark 3. *Compared with the unweighted empirical likelihood test, which requires at least finite 4th order moment on* *the data process $\{X_t\}$, the weighted empirical likelihood test needs no moment condition on $\{X_t\}$, but instead the condition $E(w_{t-1}^{-4}\xi_{\rho,t-1}^{4+\delta}) < \infty$ to guarantee the chi-squared limit distribution as indicated in Theorem 2.*

Theorem 2 indicates that through controlling the effect of the error variance, the weighted log-empirical likelihood ratio still has a standard limit distribution. Based on Theorem 2, we may similarly reject the null hypothesis $\mathcal{H}_0$ if $\tilde{\ell}(0) \geq \chi_m^2(1-a)$ at the significance level $a \in (0,1)$.

3. SIMULATION RESULTS

In this section, we carry out some simulation experiments to illustrate the finite sample properties of the proposed empirical likelihoods when the variance of ε_t is finite or infinite. For the sake of comparison, we also report the result of the $\tilde{Q}$ statistic in [33].

The simulated data $\{X_t\}_{t=1}^n$ are generated from:

$$\begin{cases} X_t = \mu + \phi X_{t-1} + \psi\varepsilon_{t-1} + \varepsilon_t, \\ \varepsilon_t = \eta_t\sigma_t,\ \sigma_t^2 = \omega + a\varepsilon_{t-1}^2 + b\sigma_{t-1}^2, \end{cases}$$

where ε_t follows a GARCH(1,1) process, $\eta_t = (\frac{c}{\sqrt{n}}e_{t-1} + e_t)/\sqrt{1+(\frac{c}{\sqrt{n}})^2}$, and $\{e_t\}$ is a sequence of *i.i.d.* random variables generated from the standard normal distribution. c is taken from $\{0,\ 5,\ 10,\ 15\}$ with $c = 0$ standing for the validity of $\mathcal{H}_0$, while $c = 5$ or $c = 10$ or $c = 15$ representing that the local alternative hypothesis of $\mathcal{H}_0$ holds. We set $\phi = 0.3$, $\psi = 0.4$, $\omega = 0.2$, and consider two different intercepts μ, i.e., 0, 0.5. For the GARCH process of ε_t, we choose $(a,\ b) = (0.1,\ 0.15)$ to represent the variance of ε_t being finite, while $(a,\ b) = (0.33,\ 0.66)$ to imply the infinite variance of ε_t approximately. Note that when $a+b$ is close to 1, we have $\sigma_t^2 \to \infty$ as $t \to \infty$.

For simplicity, we by 'EL' mean the naive empirical likelihood method, by 'WeL' the weighted empirical likelihood method, and by $\tilde{Q}$ the random weighted bootstrapping statistic given in [33], respectively. We investigate the performance of $\tilde{Q}$, EL, and WeL in testing whether the residuals are correlated at lags $m = 2$ or $m = 6$. Note that the diagonal matrix W^* for $\tilde{Q}$ is taken to be the identity matrix of order m, and the random weights are generated from the exponential distribution with parameter 1 ensuring that the weights have means one and variances one. The other settings for the random weighted bootstrap are the same as those in [33].

Tables 3.1-3.4 report the empirical ratios of rejecting $\mathcal{H}_0$ based on 2000 replications at significance levels $\tau = 0.1$ and 0.05. Three sample sizes, i.e., n = 400,800, and 1200, are considered, and there are four findings. (i) For the case of $(a,b) = (0.1, 0.15)$, the sizes of both EL and WeL are very close to the nominal levels, noting that EL is better than WeL. (ii) For the case of $(a,b) = (0.33, 0.66)$, as expected, WeL performs the best, but is slightly over-sized. Fortunately, its size decreases as n increases. Note that EL is highly over-sized and its size seems not to be convergent as the sample size increases. (iii) There is a size distortion for $\tilde{Q}$ in our reported cases compared to the proposed empirical likelihood methods. (iv) Both EL and WeL have nontrivial local powers, and their powers increase as the value of c increases.

Table 3.1. The finite variance case with $(a,b) = (0.1, 0.15)$, $(\phi,\psi) = (0.3, 0.4)$ and $m = 2$.

μ	n	c	$\tau = 0.1$			$\tau = 0.05$		
			$\tilde{Q}$	EL	WeL	$\tilde{Q}$	EL	WeL
0	400	0	0.017	0.094	0.104	0.006	0.049	0.053
		5	0.151	0.227	0.140	0.083	0.141	0.078
		10	0.616	0.607	0.314	0.502	0.488	0.213
		15	0.954	0.936	0.717	0.930	0.883	0.593
	800	0	0.027	0.104	0.108	0.012	0.055	0.057
		5	0.134	0.198	0.120	0.083	0.124	0.069
		10	0.517	0.541	0.226	0.405	0.408	0.133
		15	0.918	0.911	0.532	0.873	0.844	0.398
	1200	0	0.028	0.091	0.102	0.013	0.047	0.054
		5	0.118	0.177	0.126	0.068	0.101	0.062
		10	0.464	0.510	0.219	0.364	0.374	0.135
		15	0.883	0.871	0.459	0.812	0.791	0.341
0.5	400	0	0.017	0.103	0.099	0.005	0.052	0.048
		5	0.131	0.215	0.126	0.073	0.125	0.072
		10	0.597	0.597	0.374	0.496	0.473	0.255
		15	0.954	0.936	0.792	0.921	0.877	0.686
	800	0	0.027	0.108	0.106	0.013	0.056	0.055
		5	0.124	0.202	0.124	0.069	0.121	0.064
		10	0.504	0.529	0.270	0.394	0.399	0.183
		15	0.902	0.898	0.649	0.849	0.830	0.524
	1200	0	0.024	0.092	0.095	0.014	0.047	0.042
		5	0.113	0.178	0.111	0.065	0.102	0.053
		10	0.465	0.512	0.222	0.358	0.374	0.141
		15	0.881	0.875	0.566	0.813	0.790	0.434

It is noted that WeL suffers from a loss of power owing to the usage of the weighting technique compared to EL. Both EL and WeL are sightly over-sized for the finite variance case when $m = 6$, which indicates that the empirical likelihood-based testing methods, i.e., EL and WeL, are affected by the dimension of the auxiliary vectors. Similar phenomena have been observed in the literature. In practice, one may increase the precision of the chi-square approximation through adding proper pseudo-observations; see, e.g., [7] and [19] for details.

4. TWO APPLICATIONS

In this section, we conduct two real analyses based on modelling the monthly exchange rate on the stock market and the daily PM2.5 data in different cities by using the ARMA model discussed in this paper.

Table 3.2. The finite variance case with $(a,b) = (0.1, 0.15)$, $(\phi, \psi) = (0.3, 0.4)$ and $m = 6$.

μ	n	c	$\tau = 0.1$			$\tau = 0.05$		
			$\tilde{Q}$	EL	WeL	$\tilde{Q}$	EL	WeL
0	400	0	0.000	0.104	0.128	0.000	0.054	0.073
		5	0.000	0.184	0.178	0.000	0.115	0.106
		10	0.004	0.484	0.329	0.001	0.351	0.220
		15	0.033	0.855	0.617	0.007	0.766	0.489
	800	0	0.000	0.110	0.127	0.000	0.054	0.069
		5	0.000	0.170	0.144	0.000	0.095	0.091
		10	0.002	0.418	0.261	0.000	0.288	0.171
		15	0.029	0.793	0.519	0.005	0.691	0.388
	1200	0	0.000	0.097	0.107	0.000	0.049	0.058
		5	0.002	0.140	0.135	0.000	0.070	0.071
		10	0.003	0.364	0.234	0.001	0.244	0.138
		15	0.021	0.742	0.486	0.003	0.624	0.353
0.5	400	0	0.000	0.125	0.139	0.000	0.073	0.081
		5	0.000	0.177	0.177	0.000	0.109	0.110
		10	0.004	0.468	0.378	0.001	0.349	0.253
		15	0.032	0.845	0.691	0.007	0.763	0.566
	800	0	0.000	0.117	0.126	0.000	0.061	0.065
		5	0.000	0.164	0.156	0.000	0.091	0.088
		10	0.002	0.412	0.290	0.000	0.286	0.190
		15	0.029	0.790	0.603	0.005	0.687	0.472
	1200	0	0.000	0.096	0.107	0.000	0.050	0.055
		5	0.001	0.138	0.137	0.000	0.071	0.070
		10	0.003	0.364	0.273	0.001	0.241	0.178
		15	0.020	0.747	0.558	0.003	0.622	0.430

Table 3.3. The infinite variance case with $(a,b) = (0.33, 0.66)$, $(\phi, \psi) = (0.3, 0.4)$ and $m = 2$.

μ	n	c	$\tau = 0.1$			$\tau = 0.05$		
			$\tilde{Q}$	EL	WeL	$\tilde{Q}$	EL	WeL
0	400	0	0.012	0.211	0.106	0.006	0.148	0.057
		5	0.075	0.299	0.141	0.035	0.213	0.077
		10	0.295	0.471	0.335	0.206	0.359	0.233
		15	0.583	0.658	0.633	0.479	0.522	0.544
	800	0	0.020	0.294	0.114	0.013	0.234	0.064
		5	0.059	0.354	0.138	0.033	0.268	0.077
		10	0.201	0.487	0.257	0.133	0.385	0.185
		15	0.445	0.596	0.518	0.343	0.493	0.394
	1200	0	0.014	0.360	0.115	0.003	0.292	0.066
		5	0.043	0.401	0.144	0.020	0.325	0.080
		10	0.141	0.477	0.257	0.084	0.389	0.181
		15	0.328	0.588	0.487	0.234	0.480	0.377
0.5	400	0	0.021	0.217	0.110	0.010	0.148	0.064
		5	0.069	0.283	0.149	0.039	0.213	0.082
		10	0.290	0.478	0.351	0.192	0.371	0.246
		15	0.591	0.660	0.632	0.478	0.557	0.540
	800	0	0.016	0.298	0.111	0.009	0.229	0.064
		5	0.057	0.352	0.130	0.028	0.287	0.081
		10	0.201	0.465	0.285	0.133	0.369	0.187
		15	0.454	0.599	0.540	0.354	0.484	0.424
	1200	0	0.014	0.363	0.107	0.003	0.299	0.062
		5	0.043	0.400	0.141	0.021	0.325	0.081
		10	0.141	0.487	0.265	0.084	0.399	0.191
		15	0.328	0.601	0.485	0.233	0.496	0.378

4.1 The exchange rate on the stock market

We first collected the monthly exchange rate of eight countries including emerging and developed countries. The currencies of emerging countries that we use are: Indian rupee (INR), Malaysian ringgit (MYR), South Korea Won (KRW) and Thai baht (THB); the currencies in developed countries include: Canadian dollar (CAD), British sterling (GBP), Euro (EUR) and Japanese yen (JPY). The stock indices are: S&P/TSX (Canada), DAX (Germany), Nifty 50 (India), Nikkei 225 (Japan), FTSE KLCI (Malaysia), KOSPI Composite Index (South Korea), SET 50 (Thailand) and FTSE 100 (UK). All data are downloaded from *investing.com* and *Yahoo Finance*. We then transform all data by using $\log(\frac{P_t}{P_{t-1}})$, where P_t is the exchange rate at time t, so X_t represent the exchange rate return in our model.

The time spans of the data sets of these eight countries are summarized in Table 4.1. We check the ARCH effect of these data by using the Lagrange multiplier procedure suggested in [9], and found that the p-values are 0.0013, 0.0000, 0.0000, 0.0004, 0.027, 0.0000, 0.0249, 0.0117 for the monthly exchange rates of India, Malaysia, Korea, Thailand, Canada, UK, Germany, Japan, respectively. This shows the rationality of fitting these data by using the GARCH-type errors.

To ensure that we use the proper test, it is important to check if there is any heavy tail in residuals. In fact, as pointed out in [16], the heavy-tail feature is of key interest to risk managers, financial regulators, financial stability analysts, and policy makers. Several recent studies have suggested that many financial variables may be driven by infinite-variance innovations. For example, studies by [22], [3], [17], [1], [10], [16] provide evidence for infinite variance behavior in exchange rate return. We show their QQ-plots in Figure 1 with the standard normal distribution being compared. It seems that the distributions of these monthly data likely do *not* have infinite variances.

We fit the real data by using *auto.arima.R* contained in the R package 'forecast', and then test the possibility of existing serial correlation in the estimated residuals. All results of $\tilde{Q}$, EL and WeL are summarized in Table 4.1. The setting for $\tilde{Q}$ is the same as that in the simulations. From these results, we can see that the results of $\tilde{Q}$ indicate that no serial correlation exists in the residuals. It is not surprise by noting that $\tilde{Q}$ suffers from the undersized issue. On the other hand, both EL and WeL suggest rejecting some of the null hypotheses when $m = 2$, and EL suggests rejecting most of them when $m = 6$. Considering the good finite performance of EL as indicated in simulations, we may conclude that the results fitted by *auto.arima.R* sound good. Note that based on the testing results of $\tilde{Q}$, it seems difficult to obtain such a conclusion.

Table 3.4. The infinite variance case with $(a,b)=(0.33,0.66)$, $(\phi,\psi)=(0.3,0.4)$ and $m=6$.

μ	n	c	$\tau=0.1$			$\tau=0.05$		
			$\tilde{Q}$	EL	WeL	$\tilde{Q}$	EL	WeL
0	400	0	0.000	0.202	0.136	0.000	0.133	0.079
		5	0.000	0.291	0.185	0.000	0.209	0.111
		10	0.000	0.436	0.374	0.000	0.338	0.261
		15	0.003	0.608	0.560	0.000	0.504	0.471
	800	0	0.000	0.255	0.127	0.000	0.183	0.074
		5	0.000	0.314	0.181	0.000	0.232	0.111
		10	0.000	0.436	0.338	0.000	0.345	0.238
		15	0.003	0.556	0.559	0.000	0.472	0.453
	1200	0	0.000	0.292	0.122	0.000	0.223	0.073
		5	0.000	0.346	0.188	0.000	0.262	0.115
		10	0.001	0.416	0.323	0.000	0.330	0.235
		15	0.001	0.513	0.522	0.000	0.422	0.402
0.5	400	0	0.000	0.204	0.130	0.000	0.136	0.077
		5	0.000	0.278	0.196	0.000	0.200	0.112
		10	0.000	0.436	0.383	0.000	0.326	0.281
		15	0.003	0.604	0.565	0.000	0.495	0.491
	800	0	0.000	0.262	0.120	0.000	0.188	0.069
		5	0.000	0.312	0.174	0.000	0.233	0.110
		10	0.000	0.432	0.330	0.000	0.345	0.231
		15	0.003	0.550	0.531	0.000	0.469	0.428
	1200	0	0.000	0.312	0.115	0.000	0.238	0.066
		5	0.000	0.352	0.190	0.000	0.268	0.124
		10	0.000	0.426	0.331	0.000	0.337	0.236
		15	0.000	0.525	0.514	0.000	0.425	0.405

4.2 The PM2.5 in different cities

In our second application, we consider testing the possibility of existing serial correlation in residuals when using the ARMA model to fit the daily PM2.5 data. The PM2.5 data are taken from http://www.weather.com.cn/. Many researchers considered fitting these data by using the ARMA model; see, e.g., [8, 29, 31]. Some of them found that there may exist ARCH effect in the PM2.5 data [30]. Motivated by this, we also fit these datasets by using the ARMA-GARCH models based on $auto.arima.R$ and then test the possibility of existing serial correlation in the estimated residuals.

Since they are daily data, most of the related QQ-plots deviate from the diagonal line $y=x$, implying that their variances may possibly be quite large. Here, we do not present the QQ-plots for all these datasets in order to save space; see Figure 2 for details. The values of p, q are selected automatically by $auto.arima.R$. We then test $\mathcal{H}_0$ with three methods mentioned above. Their results are summarized in Table 4.2 for $m=2$. From these results, it is easy to check that the $\tilde{Q}$ statistic rarely rejects the null hypothesis, while the EL rejects the null hypothesis for almost all datasets. Compared to $\tilde{Q}$ and EL, WeL appears to have a relatively reasonable rejection, considering that $\tilde{Q}$ and EL suffer from a significant size distortion problem as indicated in the simulations.

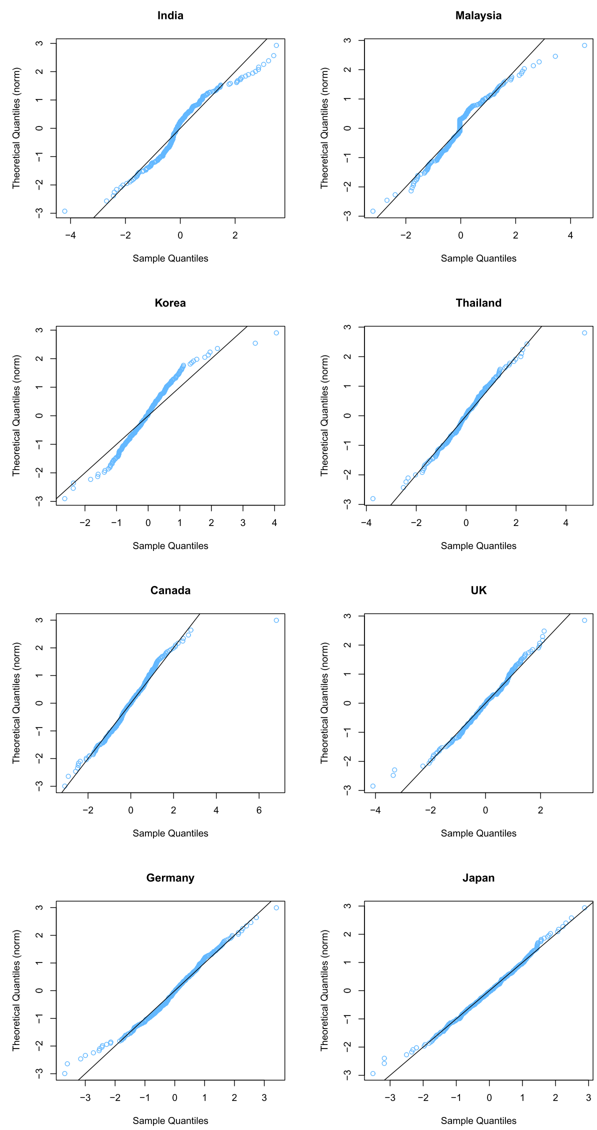

Figure 1. QQ-plots for residuals of the monthly exchange rate data from eight countries.

Note that the true conditional variances of the daily datasets may possibly tend to infinite, whereas when the true variance tends to infinite, the method in $auto.arima.R$ performs poorly in selecting the order of p, q owning to its lack of consideration of the effect of infinite variance [15]. In this sense, it is reasonable to consider that some of the residuals fitted by $auto.arima.R$ may show serial correlation because $auto.arima.R$ may select wrong p or/and p in some situations. It seems that this can not be reflected by the $\tilde{Q}$ and EL tests.

5. CONCLUDING DISCUSSIONS

In this paper, we considered the issue of diagnostic checking of AMAR models with a GARCH error by using the empirical likelihood. It turns out that the proposed log-

Table 4.1. The p-values of different tests with the monthly the stock market data, where EL(2) stands for the EL method with $m = 2$, and EL(6) is for the EL method with $m = 6$.

Country	Time	$\tilde{Q}(2)$	$\tilde{Q}(6)$	EL(2)	EL(6)	WeL(2)	WeL(6)
India	1996.01 – 2020.04	0.8111	0.7866	0.0759^{*}	0.3718	0.5537	0.0014^{***}
Malaysia	2002.07 – 2020.04	0.8080	0.8312	0.0008^{***}	0.0000^{***}	0.8823	0.1495
Korea	1997.08 – 2020.04	0.5966	0.7337	0.0813^{*}	0.0000^{***}	0.5556	0.0191^{**}
Thailand	2003.10 – 2020.04	0.7602	0.7953	0.3750	0.0681^{*}	1.0000	0.2412
Canada	1990.02 – 2020.04	0.7660	0.6682	0.1523	0.0227^{**}	0.3697	0.3164
UK	2001.03 – 2020.04	0.4467	0.5571	0.2771	0.0140^{**}	0.6344	0.0812^{*}
Germany	1990.02 – 2020.04	0.6805	0.6517	0.6493	0.0378^{**}	0.4956	0.1433
Japan	1995.02 – 2020.04	0.3849	0.7115	0.0003^{***}	0.0002^{***}	0.7796	0.1050

Significance levels: $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

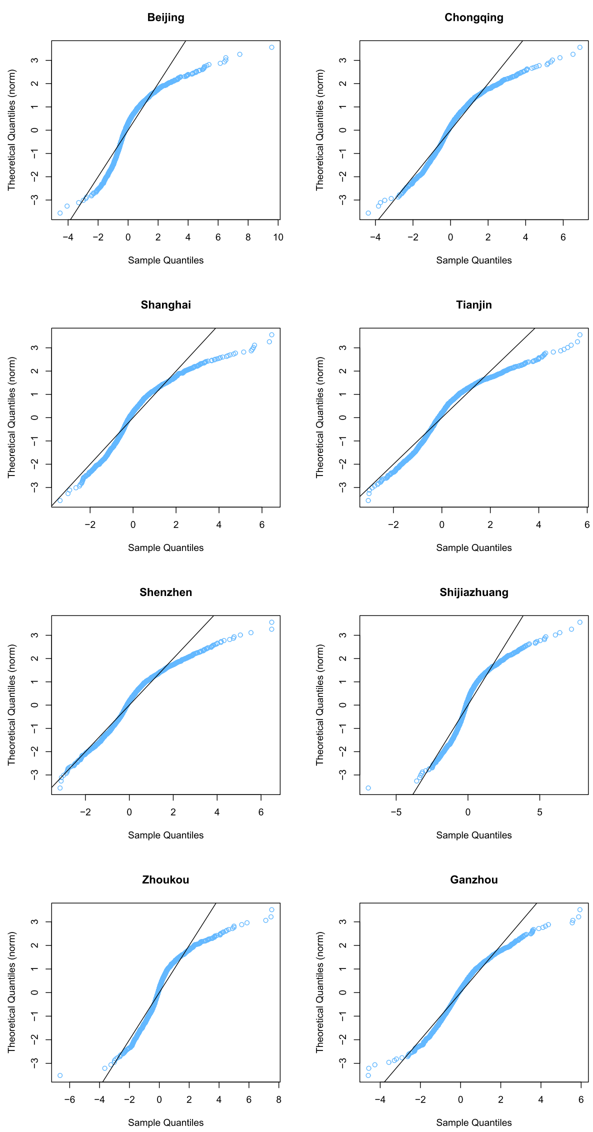

Figure 2. QQ-plots for residuals of the daily PM2.5 data of eight cities.

Table 4.2. The p-values of different tests with the daily PM2.5 data with $m = 2$.

Cities	Time	$\tilde{Q}$	EL	WeL
Chongqing	2013.10 – 2021.04	0.1474	0.0016^{***}	0.8223
Xiamen	2013.10 – 2021.04	0.4860	0.0000^{***}	0.0492^{**}
Suzhou	2015.01 – 2021.04	0.8091	0.0000^{***}	0.9975
Liuan	2015.01 – 2021.04	0.3916	0.0000^{***}	0.7435
Maanshan	2014.01 – 2021.04	0.4934	0.0000^{***}	0.8317
Tongling	2015.01 – 2021.04	0.1402	0.0000^{***}	0.9623
Hangzhou	2013.10 – 2021.04	0.0490^{*}	0.0005^{***}	0.0000^{***}
Anyang	2014.01 – 2021.04	0.3370	0.0000^{***}	0.9008
Hebi	2015.01 – 2021.04	0.3070	0.0000^{***}	0.8794
Jiaozuo	2014.01 – 2021.04	0.5782	0.0000^{***}	0.9810
Baoshan	2015.01 – 2021.04	0.8771	0.0000^{***}	0.4814
Ningbo	2013.10 – 2021.04	0.2391	0.0000^{***}	0.0000^{***}
Shaoxing	2013.10 – 2021.04	0.1159	0.0584^{*}	0.9701
Taizhou	2013.10 – 2021.04	0.1032	0.0001^{***}	0.0000^{***}
Wenzhou	2013.10 – 2021.04	0.1422	0.0000^{***}	0.0000^{***}
Yiwu	2014.01 – 2021.04	0.2324	0.0000^{***}	0.0000^{***}
Zhoushan	2013.10 – 2021.04	0.5299	0.0000^{***}	0.0118^{**}
Fuyang	2014.01 – 2021.04	0.0825^{*}	0.0000^{***}	0.9168
Aba	2015.01 – 2021.04	0.3172	0.8954	0.0000^{***}
Chengdu	2013.10 – 2021.04	0.9655	0.0000^{***}	0.0000^{***}

Significance levels: $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

empirical likelihood functions converge to a standard chi-squared distribution asymptotically. Since the empirical likelihood function does not involve the estimation of unknown variance, the new statistics do not need to estimate the GARCH parameters. We also compare the new method with the $\tilde{Q}$ statistic discussed in [33]. It turns out the empirical likelihood-based methods perform better than $\tilde{Q}$ especially when the model has low persistence, and are both computationally easy. Note that since a weighted technique is employed to reduce the moment effect of σ_t, the weighted empirical likelihood statistic suffers a little power loss when the underlying model variance is finite.

APPENDIX: PROOFS OF THE MAIN RESULTS

In this appendix, we provide the detailed proofs for the main results. Since the proof of Theorem 1 is like that of

Theorem 2. We only prove Theorem 2. Without confusion, denote $\boldsymbol{\theta}_0$ as the true value of $\boldsymbol{\theta}$, and $\mathcal{F}_t$ as the sigma field generated by $\{\eta_s : s \leq t\}$, and let

$$\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0) := \begin{pmatrix} \tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}, 0) \\ \tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}, 0) \end{pmatrix},$$

where $\tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}, 0) = w_{t-1}^{-2}\varepsilon_t(\boldsymbol{\theta})\partial\varepsilon_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, and $\tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}, 0) = (w_{t-1}^{-1}w_{t-2}^{-1}\varepsilon_t(\boldsymbol{\theta})\varepsilon_{t-1}(\boldsymbol{\theta}), \cdots, w_{t-1}^{-1}w_{t-m-1}^{-1}\varepsilon_t(\boldsymbol{\theta})\varepsilon_{t-m}(\boldsymbol{\theta}))^\top$, for $t = m+1, 2, \cdots, n$.

The following lemmas are useful in proving Theorem 2.

Lemma 1. *Suppose the same conditions of Theorem 2 holds. Then, there exist a constant $\rho \in (0,1)$, a constant $C > 0$, and a neighborhood Θ_0 such that*

$$\sup_{\boldsymbol{\theta}\in\Theta_0} |\varepsilon_t(\boldsymbol{\theta})| \leq C\xi_{\rho,t-1}, \quad \sup_{\boldsymbol{\theta}\in\Theta_0} \left\| \frac{\partial\varepsilon_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \right\| \leq C\xi_{\rho,t-1},$$

and

$$\sup_{\boldsymbol{\theta}\in\Theta_0} \left\| \frac{\partial^2\varepsilon_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \right\| \leq C\xi_{\rho,t-1},$$

*where $\xi_{\rho,t-1}$ is defined in Condition **(C3)**, and $\|A\|^2 = trace(A^\top A)$ for a given matrix A.*

Proof. This lemma is adopted from [18]. We omit the details. □

Lemma 2. *Let $\mathcal{B}_0 = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \frac{C}{\sqrt{n}}\}$ for some positive C. Then, under the same conditions of Theorem 2, as $n \to \infty$, we have uniformly for $\boldsymbol{\theta} \in \mathcal{B}_0$ that:*

(i). $\max\limits_{m+1\leq t\leq n} \sup_{\boldsymbol{\theta}\in\mathcal{B}_0} \|\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)\| = o_p(\sqrt{n})$;

(ii). $\frac{1}{N}\sum_{t=m+1}^n \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0) = \frac{1}{n}\sum_{t=m+1}^n \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}_0, 0) + O_p(\frac{1}{\sqrt{n}})$;

(iii). $\frac{1}{N}\sum_{t=m+1}^n \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)^\top = \tilde{\Sigma} + o_p(1)$, *where* $\tilde{\Sigma} = E(\tilde{\boldsymbol{Z}}_1(\boldsymbol{\theta}_0, 0)\tilde{\boldsymbol{Z}}_1(\boldsymbol{\theta}_0, 0)^\top)$.

Proof. We first prove Part (i). Note that

$$\|\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)\| \leq \|\tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}, 0)\| + \sum_{l=1}^m |\tilde{Z}_{t,p+q+l}(\boldsymbol{\theta}, 0)|.$$

By the proof of (i) in Lemma 2 of [21],

$$\max_{m+1\leq t\leq n} \sup_{\boldsymbol{\theta}\in\mathcal{B}_0} \|\tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}, 0)\| = o_p(\sqrt{n}),$$

For $\tilde{Z}_{t,p+q+l}(\boldsymbol{\theta}, 0)$, $l \in \{1, 2, \cdots, m\}$, note that

$$\begin{aligned} \sup_{\boldsymbol{\theta}\in\mathcal{B}_0} |\tilde{Z}_{t,p+q+l}(\boldsymbol{\theta}, 0)| &= \sup_{\boldsymbol{\theta}\in\mathcal{B}_0} |w_{t-1}^{-1}\varepsilon_t(\boldsymbol{\theta})w_{t-1-l}^{-1}\varepsilon_{t-l}(\boldsymbol{\theta})| \\ &\leq C^2 \underbrace{w_{t-1}^{-1}\xi_{\rho,t-1}}_{U_{t-1}} \underbrace{w_{t-1-l}^{-1}\xi_{\rho,t-1-l}}_{U_{t-1-l}}, \end{aligned}$$

by following Lemma 1. For any $\epsilon > 0$, by the Markov inequality and Cauchy-Schwarz inequality, it follows

$$\begin{aligned} & P\left(\max_{m+1\leq t\leq n} U_{t-1}U_{t-1-l} \geq \sqrt{n}\epsilon\right) \\ \leq & \sum_{t=m+1}^n P\left(U_{t-1}U_{t-1-l} \geq \sqrt{n}\epsilon\right) \\ \leq & \frac{1}{n\sqrt{n}^{\delta/2}\epsilon^{2+\delta/2}} \sum_{t=m+1}^n E\left(U_{t-1}^{2+\delta/2}U_{t-1-l}^{2+\delta/2}\right) \\ \leq & \frac{1}{\sqrt{n}^{\delta/2}\epsilon^{2+\delta/2}} \left\{\frac{1}{n}\sum_{t=m+1}^n \sqrt{E\left(U_{t-1}^{4+\delta}\right)E\left(U_{t-1-l}^{4+\delta}\right)}\right\} \\ \to & \ 0, \end{aligned}$$

as $n \to \infty$, based on Condition **(C3)**. This implies (i).

For (ii), since the proof of

$$\frac{1}{N}\sum_{t=m+1}^n \tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}, 0) = \frac{1}{N}\sum_{t=m+1}^n \tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}_0, 0) + O_p(\frac{1}{\sqrt{n}})$$

can be found in Lemma 2 of [21], we only need to show

$$\frac{1}{N}\sum_{t=m+1}^n \tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}, 0) = \frac{1}{N}\sum_{t=m+1}^n \tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}_0, 0) + O_p(\frac{1}{\sqrt{n}}).$$

Note that

$$\begin{aligned} & |w_{t-1}^{-1}\varepsilon_t(\boldsymbol{\theta})w_{t-1-l}^{-1}\varepsilon_{t-l}(\boldsymbol{\theta}) - w_{t-1}^{-1}\varepsilon_t(\boldsymbol{\theta}_0)w_{t-1-l}^{-1}\varepsilon_{t-l}(\boldsymbol{\theta}_0)| \\ \leq & \underbrace{|w_{t-1}^{-1}(\varepsilon_t(\boldsymbol{\theta}) - \varepsilon_t(\boldsymbol{\theta}_0))w_{t-1-l}^{-1}\varepsilon_{t-l}(\boldsymbol{\theta})|}_{V_{t,1}} \\ & - \underbrace{|w_{t-1}^{-1}\varepsilon_t(\boldsymbol{\theta}_0)w_{t-1-l}^{-1}(\varepsilon_{t-l}(\boldsymbol{\theta}) - \varepsilon_{t-l}(\boldsymbol{\theta}_0))|}_{V_{t,2}}. \end{aligned}$$

A simple derivation leads to that

$$\begin{aligned} & \sup_{\boldsymbol{\theta}\in\mathcal{B}_0} V_{t,1} \\ \leq & \sup_{\boldsymbol{\theta}\in\mathcal{B}_0} \left\{|w_{t-1}^{-1}w_{t-1-l}^{-1}\varepsilon_{t-l}(\boldsymbol{\theta})| \left\|\frac{\partial\varepsilon_{t-l}(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}}\right\| \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|\right\} \\ \leq & \frac{C^2C_0}{\sqrt{n}} w_{t-1}^{-1}w_{t-1-l}^{-1}\xi_{\rho,t-1}\xi_{\rho,t-1-l}, \end{aligned}$$

where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. This implies as $n \to \infty$ that

$$\frac{1}{N}\sum_{t=m+1}^n \sup_{\boldsymbol{\theta}\in\mathcal{B}_0} V_{t,1} = O_p\left(\frac{1}{\sqrt{n}}\right),$$

under Condition **(C3)**. Similarly, we can show

$$\frac{1}{N}\sum_{t=m+1}^n \sup_{\boldsymbol{\theta}\in\mathcal{B}_0} V_{t,2} = O_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{as } n \to \infty.$$

Hence, (ii) follows.

The proof of (iii) follows as similar fashion to that of (ii). We omit the details. □

Lemma 3. *Under the same conditions of Theorem 2, we have, as $n \to \infty$,*

$$\frac{1}{\sqrt{N}} \sum_{t=m+1}^{n} \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}_0, 0) \xrightarrow{d} N(0, \tilde{\Sigma}), \quad \text{and}$$
$$\frac{1}{N} \sum_{t=m+1}^{n} \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}_0, 0)\tilde{\boldsymbol{Z}}_t^\top(\boldsymbol{\theta}_0, 0) \xrightarrow{p} \tilde{\Sigma}.$$

Proof. It follows from the first part of Lemma 3 in [21] that $\frac{1}{\sqrt{N}}\sum_{t=m+1}^{n} \tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}_0, 0)$ is asymptotically normally distributed. Then, it suffices to show that, as $n \to \infty$,

$$(6) \quad \frac{1}{N} \sum_{t=m+1}^{n} E(\tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}_0, 0) w_{t-1}^{-1} w_{t-1-l}^{-1} \varepsilon_t(\boldsymbol{\theta}_0)\varepsilon_{t-l}(\boldsymbol{\theta}_0)|\mathcal{F}_{t-1}) \xrightarrow{p} \lim_{t\to\infty} E\left(\sigma_t^2 \frac{\partial \varepsilon_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} w_{t-1}^{-3} w_{t-1-l}^{-1} \varepsilon_{t-l}(\boldsymbol{\theta}_0)\right),$$

for $l = 1, 2, \cdots, m$, and

$$(7) \quad \frac{1}{\sqrt{N}} \sum_{t=m+1}^{n} \tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}_0, 0) \xrightarrow{d} N(0, E(\tilde{\boldsymbol{Z}}_{1,2}(\boldsymbol{\theta}_0, 0)\tilde{\boldsymbol{Z}}_{1,2}(\boldsymbol{\theta}_0, 0)^\top)).$$

Note that

$$\begin{aligned}
&E(\tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}_0, 0) w_{t-1}^{-1} w_{t-1-l}^{-1} \varepsilon_t(\boldsymbol{\theta}_0)\varepsilon_{t-l}(\boldsymbol{\theta}_0)|\mathcal{F}_{t-1}) \\
&= E(\varepsilon_t^2(\boldsymbol{\theta}_0) \frac{\partial \varepsilon_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} w_{t-1}^{-3} w_{t-1-l}^{-1} \varepsilon_{t-l}(\boldsymbol{\theta}_0)|\mathcal{F}_{t-1}) \\
&= \sigma_t^2 \frac{\partial \varepsilon_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} w_{t-1}^{-3} w_{t-1-l}^{-1} \varepsilon_{t-l}(\boldsymbol{\theta}_0).
\end{aligned}$$

We obtain (6) under Conditions **(C1)** and **(C3)** based on the weak law of large numbers for a martingale difference series given in [12] and the stationarity of $\{\sigma_t^2\}$, $\{X_t\}$, and $\{w_t\}$.

For (7), let $W_t = \boldsymbol{a}^\top \tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}_0, 0)$ with $\boldsymbol{a}$ being an any given m-dimensional nonzero vector. Then, it is easy to check that $E(W_t|\mathcal{F}_{t-1}) = 0$, for any $t = 1, 2, \cdots, n$. That is, $\{W_t\}$ is a martingale difference sequence.

Next, note that

$$(8) \quad \frac{1}{N} \sum_{t=m+1}^{n} E(W_t^2|\mathcal{F}_{t-1}) = \boldsymbol{a}^\top \frac{1}{N} \sum_{t=m+1}^{n} E\left(\tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}_0, 0)\tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}_0, 0)^\top|\mathcal{F}_{t-1}\right) \boldsymbol{a}.$$

For any $1 \le i, j \le m$, since by Condition **(C3)** and the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\left|\frac{1}{N} \sum_{t=m+1}^{n} w_{t-1}^{-2}\varepsilon_t^2(\boldsymbol{\theta}_0) w_{t-1-i}^{-1}\varepsilon_{t-i}(\boldsymbol{\theta}_0) w_{t-1-j}^{-1}\varepsilon_{t-j}(\boldsymbol{\theta}_0)\right| \\
\le\ & \frac{1}{N} \sum_{t=m+1}^{n} \Big(\frac{1}{2} w_{t-1}^{-4}\varepsilon_t^4(\boldsymbol{\theta}_0) + \frac{1}{4} w_{t-1-i}^{-4}\varepsilon_{t-i}^4(\boldsymbol{\theta}_0) + \frac{1}{4} w_{t-1-j}^{-4}\varepsilon_{t-j}^4(\boldsymbol{\theta}_0)\Big) \\
\le\ & \frac{1}{N} \sum_{t=m+1}^{n} \Big(\frac{1}{2} w_{t-1}^{-4}\xi_{\rho,t-1}^4 + \frac{1}{4} w_{t-1-i}^{-4}\xi_{\rho,t-1-i}^4 \ \frac{1}{4} w_{t-1-j}^{-4}\xi_{\rho,t-1-j}^4\Big) \\
\xrightarrow{p}\ & \lim_{t\to\infty} E(w_{t-1}^{-4}\xi_{\rho,t-1}^4),
\end{aligned}$$

as $n \to \infty$, where '$\xrightarrow{p}$' denotes the convergence in probability. Then, we may conclude that (8) converges by the dominated convergence theorem and the weak law of large numbers for a martingale difference series given in [12].

Furthermore, similar to the proof of (8), we can show that

$$\frac{1}{N} \sum_{t=m+1}^{n} E(W_t^2 I(|W_t| \ge \sqrt{n}\epsilon)|\mathcal{F}_{t-1}) \xrightarrow{p} 0, \text{ as } n \to \infty,$$

for any positive $\epsilon > 0$. Finally, we complete the proof of this lemma by using the central limit theorem of martingale differences [12]. This proves the first part.

The second part follows a similar fashion. We omit the details. □

Proof of Theorem 2. Based on Lemmas 2-3, the following proof is similar to that of Theorem 1 in [21].

Put $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \frac{\boldsymbol{u}}{\sqrt{n}}$ for some $(p+q+1)$-dimensional vector $\boldsymbol{u}$. Define

$$h(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{N} \sum_{t=m+1}^{N} \frac{\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)}{1 + \boldsymbol{\lambda}^\top \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)},$$

where $\boldsymbol{\lambda}$ is the solution to $h(\boldsymbol{\theta}, \boldsymbol{\lambda}) = 0$ for given $\boldsymbol{\lambda}$.

Write $\boldsymbol{\theta} = \rho \boldsymbol{v}$ with $\|\boldsymbol{v}\| = 1$. Note that

$$0 = \|h(\boldsymbol{\theta}, \boldsymbol{\lambda})\| \ge |\boldsymbol{v}^\top h(\boldsymbol{\theta}, \boldsymbol{\lambda})| = \left|\frac{1}{N} \sum_{t=m+1}^{N} \frac{\boldsymbol{v}^\top \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)}{1 + \rho \boldsymbol{v}^\top \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)}\right|.$$

Then, by a standard proof as that in [24] we can show that $\boldsymbol{\lambda} = O_p(\frac{1}{\sqrt{N}})$, and

$$\boldsymbol{\lambda} = T_n^{-1}(\boldsymbol{\theta}, 0)\left(\frac{1}{N} \sum_{t=m+1}^{N} \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)\right) + o_p\left(\frac{1}{\sqrt{N}}\right),$$

uniformly for $\boldsymbol{\theta} \in \mathcal{B}_0$ based on Lemma 2, where $T_n(\boldsymbol{\theta}, 0) = \frac{1}{N}\sum_{t=m+1}^{N} \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0)\tilde{\boldsymbol{Z}}_t^\top(\boldsymbol{\theta}, 0)$. Using this, we can further derive by the Taylor expansion and Lemma 2 that

$$\begin{aligned}
&-2\log(\tilde{L}(\boldsymbol{\theta}, 0)) \\
=\ & 2\log(1 + \boldsymbol{\lambda}^\top \tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta}, 0))
\end{aligned}$$

$$
\begin{aligned}
&= 2\boldsymbol{\lambda}^\top(\sum_{t=m+1}^{N}\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta},0)) - N\boldsymbol{\lambda}^\top T_n(\boldsymbol{\theta},0)\boldsymbol{\lambda} \\
&\quad + \frac{2}{3!}\sum_{t=m+1}^{N}\frac{1}{(1+\xi_t^*)^2}(\boldsymbol{\lambda}^\top\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta},0))^3 \\
&= S_n(\boldsymbol{\theta},0)^\top T_n^{-1}(\boldsymbol{\theta},0)S_n(\boldsymbol{\theta},0) + o_p(1) \\
&= S_n(\boldsymbol{\theta},0)^\top\tilde{\Sigma}^{-1}S_n(\boldsymbol{\theta},0) + o_p(1),
\end{aligned}
$$

uniformly for $\boldsymbol{\theta}\in\mathcal{B}_0$, where $|\xi_t^*| < |\boldsymbol{\lambda}^\top\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta},0)|$, $S_n(\boldsymbol{\theta},0) = \frac{1}{\sqrt{N}}\sum_{t=m+1}^{N}\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta},0)$. Note that

$$
\begin{aligned}
&\left|\sum_{t=m+1}^{N}\frac{1}{(1+\xi_t^*)^2}(\boldsymbol{\lambda}^\top\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta},0))^3\right| \\
&\leq C\sum_{t=m+1}^{N}\|\boldsymbol{\lambda}\|^3\|\tilde{\boldsymbol{Z}}_t(\boldsymbol{\theta},0)\|^3 = o_p(1),
\end{aligned}
$$

uniformly for $\boldsymbol{\theta}\in\mathcal{B}_0$ based on Lemma 2 as $n\to\infty$.

Furthermore, since $\boldsymbol{\theta}_0\in\mathcal{B}_0$, we have as $n\to\infty$

$$
-2\log(\tilde{L}(\boldsymbol{\theta}_0,0)) = S_n(\boldsymbol{\theta}_0,0)^\top\tilde{\Sigma}^{-1}S_n(\boldsymbol{\theta}_0,0) + o_p(1).
$$

That is,

$$
\begin{aligned}
&-2\log(\tilde{L}(\boldsymbol{\theta},0)) + 2\log(\tilde{L}(\boldsymbol{\theta}_0,0)) \\
&= S_n(\boldsymbol{\theta},0)^\top\tilde{\Sigma}^{-1}S_n(\boldsymbol{\theta},0) - \\
&\quad S_n(\boldsymbol{\theta}_0,0)^\top\tilde{\Sigma}^{-1}S_n(\boldsymbol{\theta}_0,0) + o_p(1).
\end{aligned} \tag{9}
$$

Note that for given $\boldsymbol{\theta}$, by the Taylor expansion and Lemmas 1-2, we have

$$
\begin{aligned}
&S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) \\
&= \frac{1}{\sqrt{N}}\sum_{t=m+1}^{N}\begin{pmatrix}\tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta},0) - \tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}_0,0)\\ \tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta},0) - \tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}_0,0)\end{pmatrix} \\
&= \left(\frac{1}{N}\sum_{t=m+1}^{N}\begin{pmatrix}\frac{\partial(\tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}_0,0))}{\partial\boldsymbol{\theta}^\top}\\ \frac{\partial(\tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}_0,0))}{\partial\boldsymbol{\theta}^\top}\end{pmatrix}\right)\sqrt{N}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o_p(1) \\
&= E\begin{pmatrix}\frac{\partial(\tilde{\boldsymbol{Z}}_{t,1}(\boldsymbol{\theta}_0,0))}{\partial\boldsymbol{\theta}^\top}\\ \frac{\partial(\tilde{\boldsymbol{Z}}_{t,2}(\boldsymbol{\theta}_0,0))}{\partial\boldsymbol{\theta}^\top}\end{pmatrix}\sqrt{N}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o_p(1) \\
&:= \tilde{\Gamma}\sqrt{N}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o_p(1),
\end{aligned}
$$

as $n\to\infty$. Hence, the minimizer, say $\hat{\boldsymbol{\theta}}$, of $-2\log(\tilde{L}(\boldsymbol{\theta},0))$ with respect to $\boldsymbol{\theta}$ satisfies that

$$
\begin{aligned}
0 &= \frac{-2\partial\log(\tilde{L}(\hat{\boldsymbol{\theta}},0))}{\partial\boldsymbol{\theta}} \\
&= 2\tilde{\Gamma}\tilde{\Sigma}^{-1}\tilde{\Gamma}^\top\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) + 2\tilde{\Gamma}\tilde{\Sigma}^{-1}S_n(\boldsymbol{\theta}_0) + o_p(1).
\end{aligned}
$$

For given $\boldsymbol{\theta}$, let $\varrho = \sqrt{N}\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|$, and $\boldsymbol{v} = \frac{\boldsymbol{\theta}-\boldsymbol{\theta}_0}{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|}$. Then, it is easy to check that

$$
\begin{aligned}
&\left\|\frac{-2\partial\log(\tilde{L}(\boldsymbol{\theta},0))}{\partial\boldsymbol{\theta}}\right\| \\
&\geq \left|\boldsymbol{v}^\top\frac{-2\partial\log(\tilde{L}(\boldsymbol{\theta},0))}{\partial\boldsymbol{\theta}}\right| \\
&\geq 2\varrho\boldsymbol{v}^\top\tilde{\Gamma}\tilde{\Sigma}^{-1}\tilde{\Gamma}^\top\boldsymbol{v} - 2|\boldsymbol{v}^\top\tilde{\Gamma}\tilde{\Sigma}^{-1}S_n(\boldsymbol{\theta}_0)| + o_p(1) \\
&\xrightarrow{p} \infty, \quad \text{as } \varrho\to\infty,
\end{aligned}
$$

by noting that $\boldsymbol{v}^\top\tilde{\Gamma}\tilde{\Sigma}^{-1}\tilde{\Gamma}^\top\boldsymbol{v} = O_p(1)$ and $|\boldsymbol{v}^\top\tilde{\Gamma}\tilde{\Sigma}^{-1}S_n(\boldsymbol{\theta}_0)| = O_p(1)$ as $n\to\infty$. This shows $\hat{\boldsymbol{\theta}}\in\mathcal{B}_0$. Further combining with (9), we obtain

$$
\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = -(\tilde{\Gamma}\tilde{\Sigma}^{-1}\tilde{\Gamma}^\top)^{-1}(\tilde{\Gamma}\tilde{\Sigma}^{-1}S_n(\boldsymbol{\theta}_0)) + o_p(1),
$$

as $n\to\infty$.

Finally, as in [26], we show that

$$
\begin{aligned}
&\inf\left\{-2\log(\tilde{L}(\boldsymbol{\theta},0))\right\} \\
&= -2\log(\tilde{L}(\hat{\boldsymbol{\theta}},0)) \\
&= S_n^\top(\boldsymbol{\theta}_0)(\tilde{\Sigma} - \tilde{\Gamma}^\top(\tilde{\Gamma}\tilde{\Sigma}^{-1}\tilde{\Gamma})^{-1}\tilde{\Gamma})S_n(\boldsymbol{\theta}_0) + o_p(1) \\
&\xrightarrow{d} \chi_m^2,
\end{aligned}
$$

as $n\to\infty$. This completes the proof of this theorem. □

Received 29 November 2021

REFERENCES

[1] Akgiray, V., Booth, G. G. and Seifert, B. (1988), 'Distribution properties of Latin American black market exchange rates', *Journal of International Money and Finance* **7**(1), 37–48.

[2] Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327. MR0853051

[3] Boothe, P. and Glassman, D. (1987), 'The statistical distribution of exchange rates: empirical evidence and economic implications', *Journal of International Economics* **22**(3-4), 297–319.

[4] Box, G. E. and Pierce, D. A. (1970), 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of the American Statistical Association* **65**(332), 1509–1526. MR0273762

[5] Chan, N. H. and Ling, S. (2006), 'Empirical likelihood for GARCH models', *Econometric Theory* **22**, 403–428. MR2256098

[6] Chan, N. H. and Zhang, R.-M. (2010), 'Inference for unit-root models with infinite variance GARCH errors', *Statistica Sinica* **20**(4), 1363–1393. MR2777329

[7] Chen, J., Variyath, A. M. and Abraham, B. (2008), 'Adjusted empirical likelihood and its properties', *Journal of Computational and Graphical Statistics* **17**(2), 426–443. MR2439967

[8] Cheng, Y., Zhang, H., Liu, Z., Chen, L. and Wang, P. (2019), 'Hybrid algorithm for short-term forecasting of PM2. 5 in China', *Atmospheric Environment* **200**, 264–279.

[9] Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica: Journal of the Econometric Society* **50**(4), 987–1007. MR0666121

[10] Falk, B. and Wang, C.-H. (2003), 'Testing long-run PPP with infinite-variance returns', *Journal of Applied Econometrics* **18**(4), 471–484.

[11] Fisher, T. J. and Gallagher, C. M. (2012), 'New weighted portmanteau statistics for time series goodness of fit testing', *Journal of the American Statistical Association* **107**(498), 777–787. MR2980084

[12] Hall, P. and Heyde, C. C. (2014), *Martingale Limit Theory and Its Application*, Academic Press. MR0624435
[13] Hall, P. and Yao, Q. (2003), 'Inference in ARCH and GARCH models with heavy–tailed errors', *Econometrica* **71**(1), 285–317. MR1956860
[14] He, Y., Hou, Y., Peng, L. and Shen, H. (2020), 'Inference for conditional value-at-risk of a predictive regression', *The Annals of Statistics* **48**(6), 3442–3464. MR4185815
[15] Hyndman, R. J. and Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**, 1–22.
[16] Ibragimov, M., Ibragimov, R. and Kattuman, P. (2013), 'Emerging markets and heavy tails', *Journal of Banking and Finance* **37**(7), 2546–2559.
[17] Koedijk, K. G. and Kool, C. J. (1992), 'Tail estimates of East European exchange rates', *Journal of Business and Economic Statistics* **10**(1), 83–96.
[18] Ling, S. (2007), 'Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models', *Journal of Econometrics* **140**(2), 849–873. MR2408929
[19] Liu, Y. and Chen, J. (2010), 'Adjusted empirical likelihood with high-order precision', *The Annals of Statistics* **38**(3), 1341–1362. MR2662345
[20] Ljung, G. M. and Box, G. E. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**(2), 297–303.
[21] Ma, Y., Zhou, M., Peng, L. and Zhang, R. (2021), 'Test for zero median of errors in an ARMA–GARCH model', *Econometric Theory* pp. 1–26. MR4438296
[22] Mandelbrot, B. B. (1963), 'The variation of certain speculative prices', *The Journal of Business* **36**(4), 394–419.
[23] Mikosch, T. and Starica, C. (2000), 'Limit theory for the sample autocorrelations and extremes of a GARCH(1, 1) process', *The Annals of Statistics* **28**(5), 1427–1451. MR1805791
[24] Owen, A. B. (2001), *Empirical Likelihood*, Chapman and Hall/CRC.
[25] Peng, L. and Yao, Q. (2003), 'Least absolute deviations estimation for ARCH and GARCH models', *Biometrika* **90**(4), 967–975. MR2024770
[26] Qin, J. and Lawless, J. (1994), 'Empirical likelihood and general estimating equations', *The Annals of Statistics* **22**(1), 300–325. MR1272085
[27] Shen, J., Yu, H., Yang, J. and Liu, C. (2019), 'Semiparametric bayesian analysis for longitudinal mixed effects models with non-normal AR(1) errors', *Statistics and Computing* **29**(3), 571–583. MR3969061
[28] Shen, J., Yuen, K. C. and Liu, C. (2016), 'Empirical likelihood confidence regions for one-or two-samples with doubly censored data', *Computational Statistics and Data Analysis* **93**, 285–293. MR3406212
[29] Wang, P., Zhang, H., Qin, Z. and Zhang, G. (2017), 'A novel hybrid-GARCH model based on ARIMA and SVM for PM2. 5 concentrations forecasting', *Atmospheric Pollution Research* **8**(5), 850–860.
[30] Yao, K.-C., Hsueh, H.-W., Huang, M.-H. and Wu, T.-C. (2022), 'The role of GARCH effect on the prediction of air pollution', *Sustainability* **14**(8), 4459.
[31] Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., Tan, H., Lin, D. and Wang, J. (2018), 'Trend analysis and forecast of PM2. 5 in Fuzhou, China using the ARIMA model', *Ecological Indicators* **95**, 702–710.
[32] Zhang, R., Li, C. and Peng, L. (2019), 'Inference for the tail index of a GARCH(1,1) model and an AR(1) model with ARCH(1) errors', *Econometric Reviews* **38**(2), 151–169. MR3942932
[33] Zhu, K. (2016), 'Bootstrapping the portmanteau tests in weak auto-regressive moving average models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**(2), 463–485. MR3454205

Xiaohui Liu
School of Statistics
Jiangxi University of Finance and Economics
Nanchang, Jiangxi, 330013
China
Key Laboratory of Data Science in Finance and Economics
Jiangxi University of Finance and Economics
Nanchang, Jiangxi, 330013
China
E-mail address: liuxiaohui@jxufe.edu.cn

Donghui Fan
School of Statistics
Jiangxi University of Finance and Economics
Nanchang, Jiangxi, 330013
China
E-mail address: fandonghui1114@foxmail.com

Xu Zhang
School of Mathematical Sciences
South China Normal University
Guangzhou, 510631
China
E-mail address: zhangx6690@gmail.com

Catherine Liu
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hong Kong SAR
E-mail address: macliu@polyu.edu.hk