LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises

Zhu, Ke and Ling, Shiqing

Chinese Academy of Science, Hong Kong University of Science and Technology

12 October 2014

LADE-based Inference for ARMA Models with Unspecified and Heavy-tailed Heteroscedastic Noises

BY KE ZHU

Institute of Applied Mathematics, Chinese Academy of Sciences, Haidian District, Zhongguancun, Beijing, China

AND SHIQING LING

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ABSTRACT

This paper develops a systematic procedure of statistical inference for the ARMA model with unspecified and heavy-tailed heteroscedastic noises. We first investigate the least absolute deviation estimator (LADE) and the self-weighted LADE for the model. Both estimators are shown to be strongly consistent and asymptotically normal when the noise has a finite variance and infinite variance, respectively. The rates of convergence of the LADE and the self-weighted LADE are $n^{-1/2}$ which is faster than those of LSE for the AR model when the tail index of GARCH noises is in $(0, 4]$, and thus they are more efficient in this case. Since their asymptotic covariance matrices can not be estimated directly from the sample, we develop the random weighting approach for statistical inference under this nonstandard case. We further propose a novel sign-based portmanteau test for model adequacy. Simulation study is carried out to assess the performance of our procedure and one real illustrating example is given.

Some key words: ARMA(p, q) models; Asymptotic normality; Heavy-tailed noises; G/ARCH noises; LADE; Random weighting approach; Self-weighted LADE; Sign-based portmanteau test; Strong consistency.

1. INTRODUCTION

It has been more or less accepted that the conditional volatilities depend on the past information and change from time to time in economics and financial industries since the G/ARCH models were proposed by Engle (1982) and Bollerslev (1986). A lot of alternative G/ARCH-type models have been proposed in the literature, see, e.g., Fan and Yao (2003) or Francq and Zakoïan (2010) for an overview. Examples are the absolute value GARCH model of Taylor (1986) and Schwert (1989), the GJR model of Glosten et al. (1992), the threshold GARCH model of Zakoïan (1994), and the volatility switching GARCH model of Fornari and Mele (1997) among others. The ARMA model with the G/ARCH-type noise has been extensively applied in practice. For instance, Bollerslev (1986) used an AR(4)-GARCH$(1,1)$ model to study the GNP series in U.S., Franses and Van Dijk (1996) studied several stock market indexes by AR(1)-GJR$(1,1)$ models, Zhu and Ling (2011) fitted a MA(3)-GARCH$(1,1)$ model to the world oil prices, see also Tsay (2005) for more empirical evidences.

This paper considers the following ARMA(p,q) model:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_t \text{ and } \varepsilon_t = \eta_t h_t, \tag{1.1}$$

where $\{\eta_t\}$ is a sequence of i.i.d. innovations, $h_t \in \mathcal{F}_{t-1}$ is positive almost surely (a.s.), and $\mathcal{F}_t \equiv \sigma(\varepsilon_s; s \leq t)$ is a σ-filed. We do not specify the form of h_t. It can be GARCH models, threshold GARCH models, log-GARCH models, and many others. It can also be a function of exogenous variables or other random noises as long as Assumption 2.2 in Section 2 is satisfied and it is independent of η_t. When ε_t is i.i.d. (i.e., h_t is a constant), model (1.1) has been well considered. For example, when $E\varepsilon_t^2 < \infty$, Brockwell and Davis (1991) studied the Gaussian maximum likelihood estimator (GMLE) in the frequency-domain and Yao and Brockwell (2006) studied the same estimator in the time-domain, see also Davis and Dunsmuir (1997) for the local LADE. When $E\varepsilon_t^2 = \infty$, Davis et al. (1992) and Mikosch et al. (1995) obtained the limiting distribution of the M-estimator and Whittle estimator, respectively; and Pan et al. (2007) and Zhu and Ling (2012) considered the self-weighted LADE (SLADE). When ε_t is a martingale difference with $E[\varepsilon_t^2|\mathcal{F}_{t-1}] = \sigma^2$ (a constant), Robinson (1977) obtained the consistency and asymptotic normality of the least squares estimator (LSE) for MA(1) models; and Chan and Wei (1988) and Tsay and Tiao (1990) established a complete theory for AR models and ARMA models. The model specification theory and methodology in this case have been well established in Tiao and Tsay (1989).

When ε_t is G/ARCH noise and its tail index, denoted by α, is less than 4, it has the heavy-tailed feature and its sample autocorrelation function is neither $\sqrt{n}$-consistent nor asymptotically normal, see Davis and Mikosch (1998). Recently, Lange (2011) studied the LSE for the AR model [i.e., model (1.1) with $q = 0$] with ε_t defined by (2.2) in Section 2. When the tail index α is in $(2, 4)$, i.e., $E\varepsilon_t^4 = \infty$ and $E\varepsilon_t^2 < \infty$, he showed that the LSE, denoted by $\hat{\theta}_{LSE}$, is $n^{1-2/\alpha}$-consistent. Furthermore, for the AR model with ε_t being G-GARCH$(1, 1)$ noise in He and Terasvirta (1999), Zhang and Ling (2014) showed that

$$\frac{\sqrt{n}}{\log n}(\hat{\theta}_{LSE} - \theta_0) \to_d \text{Normal, if } \alpha = 4 \ (i.e. \ E\varepsilon_t^4 = \infty), \tag{1.2}$$

$$n^{1-2/\alpha}(\hat{\theta}_{LSE} - \theta_0) \to_d \text{Stable, if } \alpha \in (2, 4) \ (i.e. \ E\varepsilon_t^2 < \infty \ and \ E\varepsilon_t^4 = \infty), \tag{1.3}$$

$$\log n(\hat{\theta}_{LSE} - \theta_0) \to_d \text{Stable, if } \alpha = 2 \ (i.e. \ E\varepsilon_t^2 = \infty), \tag{1.4}$$

$$\hat{\theta}_{LSE} - \theta_0 \to_d \text{Stable, if } \alpha \in (0, 2) \ (i.e. \ E\varepsilon_t^2 = \infty), \tag{1.5}$$

when $n \to \infty$, where n is sample size and $\to_d$ denotes the convergence in distribution. The LSE not only has a slower rate of convergence but also is not asymptotically normal when $\alpha \in (0, 4)$. Thus, the classical theory and methodology (e.g., T-test, Wald test and Ljung-Box test, among others) do not work in this case. This raises a problem how to do statistical inference and model selection for model (1.1) when the form of h_t is unknown and the tail index α of ε_t is in $(0, 4]$.

This paper is to build up a systematic procedure of statistical inference for model (1.1) with unspecified and heavy-tailed h_t. We first investigate the LADE and SLADE for this model. Both estimators are shown to be strongly consistent and asymptotically normal when the noise ε_t has a finite variance and infinite variance, respectively. The rates of convergence of the LADE and the SLADE are $n^{-1/2}$ which is faster than those of LSE in (1.2)-(1.5), and thus they are more efficient in this case. The LADE for regression models has been well studied in the literature, see, e.g., Koenker and Bassett (1978, 1982) and Knight (1987, 1998) for earlier works. In the time series context, Chan and Peng (2005) studied the local weighted LADE in the DAR(1) model and Li and Li (2008) studied the local LADE when $E\varepsilon_t^2 < \infty$, see also Wu and Davis (2010) and references therein. The technique we used for LADE and SLADE in this paper is from Zhu and Ling (2011, 2012, 2013). However, since we do not specify the form of h_t, the results are fully different from those in the previous papers and the scop of applications is much wider.

Since the asymptotic covariance matrices of LADE and SLADE can not be estimated directly from the sample, we develop the random weighting approach for statistical inference under this nonstandard case. This approach, as a variant of the traditional wild bootstrap in Wu (1986), was originally proposed by Jin et al. (2001) and provides a way to do statistical inference when the covariance matrix of the estimator can not be estimated by the conventional methods, see, e.g., Chen et al. (2008), Chen et al. (2010), and Zhu and Li (2013) for more discussions. Model checking is an important step in modeling. The classical approach is to use the Ljung-Box portmanteau test in Ljung and Box (1978) for model adequacy. Since this test requires that $E\varepsilon_t^4 < \infty$ and $h_t \equiv$ a constant, it is invalid when the tail index $\alpha \leq 4$ or $h_t \not\equiv$ a constant as in model (1.1). Li and Li (2008) considered two LADE-based portmanteau tests, which are only applicable when ε_t follows a GARCH model with $E\varepsilon_t^2 < \infty$. Since the form of h_t is not specified in model (1.1), the test in Li and Li (2008) is not fitted in our setting. There is not any formal test for checking the adequacy of model (1.1) up to now. In this paper, we propose a novel sign-based portmanteau test for model adequacy. Simulation study is carried out to assess the performance of our procedure and one real illustrating example is given.

This paper is organized as follows. Section 2 studies the LADE/SLADE for model (1.1). The random weighting approach is proposed in Section 3. A sign-based portmanteau test is given in Section 4. Simulation results are reported in Section 5. One real example is given in Section 6. The conclusion and some discussions are offered in Section 7. The entire proofs are presented in the Appendix.

2. LADE AND SELF-WEIGHTED LADE

We first denote the unknown parameter of model (1.1) by $\theta \equiv (\mu, \phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)'$. Let θ_0 be the true value of θ and the parameter space Θ be a compact subset of $\mathcal{R}^m$, where $\mathcal{R} = (-\infty, \infty)$ and $m = p + q + 1$. We make the following two assumptions:

Assumption 2.1. θ_0 is an interior point in Θ, and for each $\theta \in \Theta$, $\phi(z) \equiv 1 - \sum_{i=1}^{p} \phi_i z^i \neq 0$ and $\psi(z) \equiv 1 + \sum_{i=1}^{q} \psi_i z^i \neq 0$ when $|z| \leq 1$, and $\phi(z)$ and $\psi(z)$ have no common root with $\phi_p \neq 0$ or $\psi_q \neq 0$.

Assumption 2.2. ε_t is strictly stationary and ergodic.

Assumption 2.1 is a usual condition for the stationarity, invertibility and identifiability of model (1.1). Given the observations $\{y_1, \cdots, y_n\}$ and the initial values $Y_0 \equiv \{y_0, y_{-1}, \cdots\}$, which are generated by model (1.1), we can write the parametric model as

$$\varepsilon_t(\theta) = y_t - \mu - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \psi_i \varepsilon_{t-i}(\theta).$$

Here, $\varepsilon_t(\theta_0) = \varepsilon_t$. We consider the following objective function:

$$L_{sn}(\theta) = \frac{1}{n} \sum_{t=1}^{n} w_t |\varepsilon_t(\theta)|,$$

where $w_t = w(y_{t-1}, y_{t-2}, ...) > 0$ and w is a measurable and bounded function on $\mathcal{R}^{Z_0}$ with $Z_0 = \{0, 1, 2...\}$, and it satisfies the following condition:

Assumption 2.3. $E[w_t\xi_{\rho t-1}^2] < \infty$ for any $\rho \in (0,1)$, where $\xi_{\rho t} = 1 + \sum_{i=0}^{\infty} \rho^i |y_{t-i}|$.

Since the initial values Y_0 are unobservable, we modify $L_{sn}(\theta)$ as

$$\tilde{L}_{sn}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{w}_t|\tilde{\varepsilon}_t(\theta)|, \tag{2.1}$$

where $\tilde{\varepsilon}_t(\theta)$ and $\tilde{w}_t$ are defined in the same way as $\varepsilon_t(\theta)$ and w_t, respectively, with Y_0 being replaced by some constants. The minimizer, $\tilde{\theta}_{sn}$, of $\tilde{L}_{sn}(\theta)$ on Θ is called the SLADE of θ_0, i.e.,

$$\tilde{\theta}_{sn} \equiv \arg\min_{\theta\in\Theta} \tilde{L}_{sn}(\theta).$$

When $E\varepsilon_t^2 < \infty$, we take $w_t \equiv 1$, $\tilde{\theta}_{sn}$ reduces to the usual LADE, denoted by $\tilde{\theta}_n$. The weight function w_t is to down weight the large value of y_t. Without this weight, one cannot obtain the asymptotic normality of the estimated parameters when $E\varepsilon_t^2 = \infty$, see Mikosch et al. (1995) and Zhang and Ling (2014). To make the the initial values Y_0 ignorable, we need the following assumption:

Assumption 2.4. $E|\varepsilon_t|^{2\iota} < \infty$ and $E\,|w_t - \tilde{w}_t|^{\iota/4} = O(t^{-2})$ for some $\iota \in (0,1)$.

THEOREM 2.1. *Assume that η_t has a zero median and Assumptions 2.1-2.4 hold. Then, (a) $\tilde{\theta}_{sn} \to \theta_0$ a.s. as $n \to \infty$; (b) furthermore, if $E\varepsilon_t^2 < \infty$, then $\tilde{\theta}_n \to \theta_0$ a.s. as $n \to \infty$.*

To study the asymptotic normality of $\tilde{\theta}_{sn}$ and $\tilde{\theta}_n$, we need two more assumptions:

Assumption 2.5. $h_t \geq c_0$ (a.s.) for some positive constant c_0.

Assumption 2.6. η_t has a zero median with a continuous density function $g(x)$ satisfying $g(0) > 0$ and $\sup_{x\in\mathcal{R}} g(x) < \infty$.

Assumption 2.5 is a mild condition for heteroscedastic noises. Assumption 2.6 is a basic set-up for the LADE, see Zhu and Ling (2011). We now can state the asymptotic normality of $\tilde{\theta}_{sn}$ as follows:

THEOREM 2.2. *Assume that Assumptions* 2.1-2.6 *hold. Then, (a)*

$$\sqrt{n}(\tilde{\theta}_{sn} - \theta_0) \to_d N\left(0, [2g(0)]^{-2}\Sigma^{-1}\Omega\Sigma^{-1}\right) \quad \textit{as } n \to \infty;$$

(b) furthermore, if $E\varepsilon_t^2 < \infty$*, then*

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d N\left(0, [2g(0)]^{-2}\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}\right) \quad \textit{as } n \to \infty,$$

where

$$\Omega = E\left[w_t^2 \frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}\right] \quad \textit{and} \quad \Sigma = E\left[\frac{w_t}{h_t}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}\right],$$

and Ω_0 *and* Σ_0 *are defined in the same way as* Ω *and* Σ*, respectively, with* $w_t \equiv 1$.

Remark 2.1. When $E\varepsilon_t^2 < \infty$, the LADE is more efficient than the SLADE, see Zhu and Ling (2011). Thus, we do not need a weight in this case. However, if there is not a clear evidence to show that $E\varepsilon_t^2 < \infty$, we should use SLADE from the view of robustness. Both LADE and SLADE are robust if there is not a clear form of h_t, compared with the ARMA-GARCH specification in Francq and Zakoïan (2004) and Zhu and Ling (2011).

When ε_t is i.i.d. noise, the LSE, LADE and M-estimator are $\sqrt{n}$-consistent and asymptotically normal if $E\varepsilon_t^2 < \infty$, and are $L(n)n^{1/\alpha}$-consistent and converge to a stable random variable if $E\varepsilon_t^2 = \infty$, where $L(n)$ is a slowly varying function and $\alpha \in (0, 2)$, see Davis et al. (1992), Mikosch et al. (1995) and Davis (1996). Unlike the i.i.d case, when ε_t is G/ARCH-type noise with $\alpha \in [2, 4]$, the rates of convergence of LSE given in (1.2)-(1.4) are slower than $n^{-1/2}$ and hence the LADE may be expected to have a slower rate of convergence. However, Theorem 2.2 shows that the rate of convergence of LADE is $n^{-1/2}$, which is faster than that of LSE.

Furthermore, when $\alpha \in (0, 2)$, the SLADE is still $\sqrt{n}$-consistent even if the LSE is not consistent as given in (1.5), and hence the SLADE is much more efficient than the LSE.

The SLADE is used only when ε_t is heteroscedastic noise with $\alpha \in (0, 2]$. Up to now, we do not know the rate of convergence of LADE in this case, yet. To explore its possible rate of convergence, we consider the following speical case:

$$y_t = \theta y_{t-1} + \varepsilon_t,\ \ \varepsilon_t = \eta_t h_t \ \text{ and } h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1}^2, \tag{2.2}$$

where $\alpha_0 > 0, \alpha_1 > 0$, and $\beta > 0$. Using the objective function:

$$\bar{L}_n(\theta) = \sum_{t=1}^{n} \left[|\varepsilon_t(\theta)| - |\varepsilon_t(\theta_0)| \right],$$

we should have the following score-type and information-type quantities:

$$\bar{T}_n = \sum_{t=1}^{n} y_{t-1} \left[I(\eta_t > 0) - I(\eta_t < 0) \right] \ \text{ and } \ \bar{\Sigma}_n = 2g(0) \sum_{t=1}^{n} \frac{y_{t-1}^2}{h_t}.$$

Logically, we should have the following expansion:

$$\bar{L}_n(\theta) = (\theta - \theta_0)\bar{T}_n + (\theta - \theta_0)^2 \left[\bar{\Sigma}_n + (\theta - \theta_0) R_n(\theta) \right],$$

where $R_n(\theta)$ is a remainder term. If $E[\theta_0^2 / \sqrt{\alpha_1 \eta_t^2 + \beta}] < 1$, then we can show that $E[y_{t-1}^2 / h_t] < \infty$ when $\alpha \in (1, 2]$, and hence by the ergordic theorem, we have

$$\frac{1}{n}\bar{\Sigma}_n \to \bar{\Sigma} = [2g(0)] E \frac{y_{t-1}^2}{h_t} \text{ a.s. as } n \to \infty. \tag{2.3}$$

Furthermore, if we can show that $(\tilde{\theta}_n - \theta_0) R_n(\tilde{\theta}_n)/n = o_p(1)$, then we have the expansion:

$$\tilde{\theta}_n - \theta_0 = -\frac{1}{n} \left[\bar{\Sigma} + o_p(1) \right]^{-1} \bar{T}_n.$$

By Theorem 3.1 in Zhang et al. (2014), we can show that

$$a_n^{-1} \bar{T}_n \to_d \xi_\alpha \text{ as } n \to \infty,$$

where ξ_α is α-stable random variable and $a_n = c n^{1/\alpha}$ with a constant c. Thus, $\tilde{\theta}_n$ should have the rate of convergence $n^{1/\alpha - 1}$ which is slower than $n^{-1/2}$ if $\alpha \in (1, 2]$. Hence, when $\alpha \in (0, 2]$, we conjecture that the SLADE should be more efficient than that of the LADE for model (2.2). A simulation study in Section 5 confirms our conjecture. However, if h_t is other heteroscedastic process or includes exogenous variable, it is not clear the asymptotic behavior of $\bar{\Sigma}_n$ in (2.3) at all and hence we cannot work out the rate of convergence of LADE in this case.

It is clear that the limiting distribution of $\tilde{\theta}_{sn}$ depends on the way that we choose the weight function w_t. Note that the tail index of ε_t is the same as the one of y_t. As in Ling (2007), we choose w_t according to the tail index α of y_t. For instance, when $\alpha \in (1, \infty)$ (i.e., $E|y_t| < \infty$), we can choose the weight function as

$$w_t = \left(\max \left\{ 1, C^{-1} \sum_{k=1}^{\infty} \frac{1}{k^9} |y_{t-k}| I\left\{ |y_{t-k}| > C \right\} \right\} \right)^{-4}, \tag{2.4}$$

where $C > 0$ is a constant. In practice, it works well when we select C as the 90% or 95% quantile of data set $\{y_1, \cdots, y_n\}$. When $q = 0$, for any $\alpha > 0$, the weight can be selected as

$$w_t = \left(\max \left\{ 1, C^{-1} \sum_{k=1}^{p} \frac{1}{k^9} |y_{t-k}| I\left\{ |y_{t-k}| > C \right\} \right\} \right)^{-4}. \tag{2.5}$$

When $\alpha \in (0, 1]$ and $q > 0$, the weight function need to be modified as

$$w_t = \left(\max \left\{ 1, C^{-1} \sum_{k=1}^{\infty} \frac{1}{k^{1+8/\iota_0}} |y_{t-k}| I\left\{ |y_{t-k}| > C \right\} \right\} \right)^{-4}, \tag{2.6}$$

where ι_0 is any positive constant such that $2\iota_0 < \alpha$. Moreover, when $q = 0$, we can also use Huber's influence function to select the weight function as follows:

$$w_t = \begin{cases} 1 & \text{if } a_t = 0, \\ C^2/a_t^2 & \text{if } a_t \neq 0, \end{cases}$$

where $a_t = \sum_{i=1}^{p} |y_{t-i}| I(|y_{t-i}| > C)$. Obviously, these weight functions satisfy Assumptions 2.3-2.4, see also Ling (2005) and Pan et al. (2007) for more choices of w_t. Theoretically, how to choose the optimal weight function is still a challenging open question. In practice, we can estimate the tail index α by Hill's estimators and get some useful guidance to pick up our weight function, see Zhu and Ling (2011).

3. RANDOM WEIGHTING PROCEDURE

To do inference for model (1.1), we need to estimate the covariance matrix in Theorem 2.2. However, both $g(0)$ and Σ_0 can not be directly estimated from the sample since $\{h_t\}$ is unobservable. To solve this problem, we use the random weighting method to approximate the limiting distribution in Theorem 2.2. Let $w_1^* \cdots, w_n^*$ be a sequence of i.i.d. nonnegative random variables, with mean and variance both equal to 1. Define

$$\tilde{L}_{sn}^*(\theta) = \frac{1}{n}\sum_{t=1}^{n} w_t^* \tilde{w}_t |\tilde{\varepsilon}_t(\theta)|,$$

and $\tilde{\theta}_{sn}^* = \arg\min_{\Theta} \tilde{L}_{sn}^*(\theta)$. Based on Assumption 3.1 below, we can show that the distribution of $\sqrt{n}(\tilde{\theta}_{sn} - \theta_0)$ can be approximated by the resampling distribution of $\sqrt{n}(\tilde{\theta}_{sn}^* - \tilde{\theta}_{sn})$.

Assumption 3.1. (i) $E|w_t^*|^{2+\delta_0} < \infty$ for some $\delta_0 > 0$; (ii) $\{w_t^*\}$ and $\{y_t\}$ are independent.

THEOREM 3.1. *Assume that the conditions in Theorem* 2.2 *and Assumption* 3.1 *hold. Conditional on* $\{y_1, \cdots, y_n\}$, *then (a)*

$$\sqrt{n}(\tilde{\theta}_{sn}^* - \tilde{\theta}_{sn}) \to_d N\left(0, \Sigma^{-1}\Omega\Sigma^{-1}\right) \text{ in probability } \text{ as } n \to \infty;$$

(b) furthermore, if $E\varepsilon_t^2 < \infty$, *then*

$$\sqrt{n}(\tilde{\theta}_{n}^* - \tilde{\theta}_{n}) \to_d N\left(0, \Sigma_0^{-1}\Omega_0\Sigma_0^{-1}\right) \text{ in probability } \text{ as } n \to \infty,$$

where Σ, Ω, Ω_0 *and* Σ_0 *are defined as in Theorem* 2.2.

According to Theorems 2.2 and 3.1, we can approximate the asymptotic variance-covariance matrix of $\tilde{\theta}_{sn}$ via the resampling procedure as follows. First, we generate J replications of the i.i.d. random weights $\{w_1^*, \cdots, w_n^*\}$ from the standard exponential distribution, which has mean and variance both equal to one. For each replication, we compute $\tilde{\theta}_{sn}^*$. Denote them as $\{b_1, \cdots, b_J\}$. Then, the sample variance-covariance matrix of $\{b_1 - \tilde{\theta}_{sn}, \cdots, b_J - \tilde{\theta}_{sn}\}$, denoted by $\tilde{V}$, provides a good approximation for the asymptotic variance-covariance matrix of

$\tilde{\theta}_{sn}$ in Theorem 2.2 when J is large. Hence, we can construct a Wald test statistic

$$W_n = (\Gamma\tilde{\theta}_{sn} - r)' \left(\Gamma\tilde{V}\Gamma'\right)^{-1} (\Gamma\tilde{\theta}_{sn} - r) \tag{3.1}$$

to detect the following linear null hypothesis:

$$H_0 : \Gamma\theta_0 = r,$$

where Γ is a $s \times m$ constant matrix with rank s and r is a $s \times 1$ constant vector. If W_n is larger than the upper-tailed critical value of χ_s^2, then the null hypothesis H_0 is rejected. Otherwise, H_0 is not rejected. Similarly, we can construct a Wald test statistic based on $\tilde{\theta}_n$ when $E\varepsilon_t^2 < \infty$.

4. Sign-Based Ljung-Box Portmanteau Test

To construct a test for model checking of model (1.1), we first define a weighted sign-component as follows:

$$\zeta_t(\theta) = w_t[sgn(\varepsilon_t(\theta))] := w_t\xi_t(\theta).$$

If model (1.1) is correct, then the autocorrelation function of $\{\zeta_t(\theta_0)\}$:

$$\rho_k \equiv \frac{E\zeta_t(\theta_0)\zeta_{t-k}(\theta_0)}{E\zeta_t^2(\theta_0)} = 0,$$

for all $k \geq 1$. Let $\tilde{\zeta}_t(\theta)$ and $\tilde{\xi}_t(\theta)$ be defined in the same way as $\zeta_t(\theta)$ and $\xi_t(\theta)$, respectively, with $\varepsilon_t(\theta)$ and w_t being replaced by $\tilde{\varepsilon}_t(\theta)$ and $\tilde{w}_t$. Then, we can estimate ρ_k by its sample autocorrelation function defined by

$$\tilde{\rho}_k = \frac{\sum_{t=k+1}^{n}[\tilde{\zeta}_t(\tilde{\theta}_{sn}) - \bar{\zeta}_n][\tilde{\zeta}_{t-k}(\tilde{\theta}_{sn}) - \bar{\zeta}_n]}{\sum_{t=1}^{n}[\tilde{\zeta}_t(\tilde{\theta}_{sn}) - \bar{\zeta}_n]^2},$$

where $\bar{\zeta}_n = \sum_{t=1}^{n} \tilde{\zeta}_t(\tilde{\theta}_{sn})/n$. Denote $\xi_t := \xi_t(\theta_0)$ and $\zeta_t := w_t\xi_t$. We have the following result:

THEOREM 4.1. *Let $\tilde{\rho} = (\tilde{\rho}_1, \cdots, \tilde{\rho}_M)'$ for a given positive integer M. Then, under the conditions of Theorem 2.2, we have*

$$\sqrt{n}\tilde{\rho} \to_d N\left(0, (Ew_t^2)^{-2}A\right) \quad \textit{as } n \to \infty,$$

where $A = VZV'$ *with* $V = [I_M, -W\Sigma^{-1}]$, $Z = E[Z_t Z_t']$, $W = (W_1, \cdots, W_M)'$, *and*

$$W_k = E\left[\frac{w_t \zeta_{t-k}}{h_t} \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta}\right], \ Z_t = \left(\zeta_t \zeta_{t-1}, \cdots, \zeta_t \zeta_{t-M}, \zeta_t \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta'}\right)'.$$

As in Theorem 2.2, we cannot estimate A directly. We use the random weighting approach in Section 3 to estimate A. Define

$$\tilde{\rho}_{w^*k} = \frac{\sum_{t=k+1}^{n} w_t^* [\tilde{\zeta}_t(\tilde{\theta}_{sn}^*) - \bar{\zeta}_n^*][\tilde{\zeta}_{t-k}(\tilde{\theta}_{sn}^*) - \bar{\zeta}_n^*]}{\sum_{t=1}^{n} [\tilde{\zeta}_t(\tilde{\theta}_{sn}^*) - \bar{\zeta}_n^*]^2},$$

where $\bar{\zeta}_n^* = \sum_{t=1}^{n} \tilde{\zeta}_t(\tilde{\theta}_{sn}^*)/n$. Then, we are ready to give the following result:

THEOREM 4.2. *Let* $\tilde{\rho}_{w^*} = (\tilde{\rho}_{w^*1}, \cdots, \tilde{\rho}_{w^*M})'$ *for a given positive integer* M. *Then, under the conditions of Theorem* 3.1 *and conditional on* $\{y_1, \cdots, y_n\}$, *we have*

$$\sqrt{n}(\tilde{\rho}_{w^*} - \tilde{\rho}) \to_d N\left(0, (Ew_t^2)^{-2} A\right) \ \textit{in probability} \ \ \textit{as } n \to \infty,$$

where A *is defined as in Theorem* 4.1.

In view of Theorems 4.1-4.2, we can approximate the asymptotic variance-covariance matrix of $\tilde{\rho}$ via the resampling procedure as follows. First, we generate J replications of the i.i.d. random weights $\{w_1^*, \cdots, w_n^*\}$ from the standard exponential distribution. For each replication, we compute $\tilde{\rho}_{w^*}$. Denote them as $\{c_1, \cdots, c_J\}$. Then, the sample variance-covariance matrix of $\{c_1 - \tilde{\rho}, \cdots, c_J - \tilde{\rho}\}$, denoted by $\tilde{V}_*$, provides a good approximation for the asymptotic variance-covariance matrix of $\tilde{\rho}$ in Theorem 4.1 when J is large. Finally, we can construct the following portmanteau test statistic

$$S_M = \tilde{\rho}' \tilde{V}_*^{-1} \tilde{\rho}, \tag{4.1}$$

and compare it to the upper-tailed critical value of χ_M^2 at an appropriate level. If S_M is larger than the critical value, then the fitted model (1.1) is adequate. Otherwise, it is not adequate. Similarly, we can construct the sign-based Ljung-Box portmanteau test based on $\tilde{\theta}_n$ when $E\varepsilon_t^2 < \infty$.

Up to now, there is not any test that can be used to check the adequacy of model (1.1) when the form of h_t is unknown and the tail index of ε_t is in $(0, 4]$. Our sign-based test is the first try for this purpose in the literature, and we hope that more satisfactory tests can be built in the future.

5. SIMULATION

In this section, we first assess the performance of the LADE ($\tilde{\theta}_n$), the SLADE ($\tilde{\theta}_{sn}$), and the corresponding random weighting approach in the finite sample. We generate 10,000 replications of sample size $n =$ 200 and 400 from the following model:

$$y_t = 0.4y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t, \tag{5.1}$$

where ε_t is generated from the following model (5.2) (i.e., GJR$(1, 1)$ model) and model (5.3) (i.e., non-linear GARCH$(1, 1)$ model), respectively:

$$\varepsilon_t = \eta_t h_t,\ h_t^2 = \alpha_0 + \left\{\beta + [\alpha_1 + \omega I(\eta_{t-1} < 0)]\, \eta_{t-1}^2\right\} h_{t-1}^2, \tag{5.2}$$

$$\varepsilon_t = \eta_t h_t,\ h_t^2 = \alpha_0 + \left\{\beta + \alpha_1 \left[1 - 2\omega sgn(\eta_{t-1}) + \omega^2\right] \eta_{t-1}^2\right\} h_{t-1}^2. \tag{5.3}$$

Here, the innovation η_t in models (5.2) and (5.3) is chosen to be re-scaled Laplace$(0, 1)$, N$(0, 1)$, and t_3 distribution, respectively, such that $E\eta_t^2 = 1$. Denote $\lambda = (\alpha_0, \alpha_1, \beta, \omega)$ be the unknown parameter of model (5.2) or (5.3). For the case that $E\varepsilon_t^2 < \infty$, we take $\lambda = (0.1, 0.1, 0.8, 0.0)$ and $(0.1, 0.1, 0.8, 0.3)$ in model (5.2) and (5.3), respectively. For the case that $E\varepsilon_t^2 = \infty$ (i.e., the tail index of ε_t is in $(0, 2]$), we take $\lambda = (0.1, 0.2, 0.8, 0.1)$ and $(0.1, 0.1, 0.9, 0.2)$ in model (5.2) and (5.3), respectively. Tables 1-2 report the sample biases, the sample standard deviations (SD), the average estimated asymptotic standard deviations (AD), and the average bootstrapped sample standard deviations (BD) of $\tilde{\theta}_n$ and $\tilde{\theta}_{sn}$, respectively. The ADs are calculated from Theorem 2.2 via the plug-in by assuming both $g(0)$ and the true structure of h_t are known. The BDs are obtained by using the random weighting method in Theorem 3.1 with the bootstrap sample size $J = 500$. In all calculations (hereafter), we choose the weight function w_t as in (2.6) with

$\iota_0 = 1/3$ and C being the 90% quantile of data sample. From Tables 1-2, we first find that both $\tilde{\theta}_n$ and $\tilde{\theta}_{sn}$ have good accuracy, and their SDs and ADs are close to each other in all cases. Second, we can see that the disparity between BD and AD in each case is small, and hence our random weighting approach is reliable regardless of the structure of ε_t and distribution of η_t. Third, as we expected, all of the SDs, ADs and BDs become smaller as n increases from 200 to 400.

Table 1. *Bias and standard deviations of* $\tilde{\theta}_n$ *for model* (5.1) *with* $\theta_0 = (0.4, 0.5)$ *and* $E\varepsilon_t^2 < \infty$.

		$\varepsilon_t \sim$ model (5.2)				$\varepsilon_t \sim$ model (5.3)			
		$n = 200$		$n = 400$		$n = 200$		$n = 400$	
$\eta_t \sim$		$\tilde{\phi}_n$	$\tilde{\psi}_n$	$\tilde{\phi}_n$	$\tilde{\psi}_n$	$\tilde{\phi}_n$	$\tilde{\psi}_n$	$\tilde{\phi}_n$	$\tilde{\psi}_n$
Laplace(0, 1)	Bias	-0.0041	0.0005	-0.0028	0.0001	-0.0026	-0.0016	-0.0016	-0.0001
	SD	0.0800	0.0763	0.0553	0.0522	0.0777	0.0750	0.0534	0.0501
	AD	0.0864	0.0812	0.0592	0.0556	0.0838	0.0789	0.0570	0.0537
	BD	0.0883	0.0851	0.0591	0.0561	0.0856	0.0827	0.0571	0.0544
N(0, 1)	Bias	-0.0037	-0.0010	-0.0024	-0.0004	-0.0037	-0.0011	-0.0021	0.0000
	SD	0.1195	0.1172	0.0838	0.0802	0.1173	0.1150	0.0825	0.0788
	AD	0.1248	0.1170	0.0872	0.0819	0.1225	0.1148	0.0854	0.0803
	BD	0.1259	0.1220	0.0876	0.0844	0.1237	0.1198	0.0857	0.0822
t_3	Bias	-0.0041	-0.0004	-0.0013	-0.0010	-0.0043	-0.0003	-0.0017	-0.0003
	SD	0.0921	0.0872	0.0658	0.0625	0.0890	0.0842	0.0629	0.0598
	AD	0.0872	0.0817	0.0618	0.0581	0.0841	0.0800	0.0594	0.0559
	BD	0.0970	0.0936	0.0679	0.0648	0.0938	0.0905	0.0654	0.0624

Next, we assess the finite sample performance of the Wald test statistic W_n in (3.1) and the portmanteau test statistic S_M in (4.1). We generate 10,000 replications of sample size $n = 200$ and 400 from the following model:

$$y_t = 0.4y_{t-1} + \kappa y_{t-2} + \varepsilon_t, \tag{5.4}$$

where ε_t is chosen as in Tables 1-2, and $\kappa = 0.0$, 0.2 or 0.4. In the case of $E\varepsilon_t^2 < \infty$ and $E\varepsilon_t^2 = \infty$, we fit each replication by an AR(2) model with the LADE and the SLADE method, respectively, and then use W_n to detect the hypothesis that $\kappa = 0$ in model (5.4). Furthermore,

Table 2. *Bias and standard deviations of* $\tilde{\theta}_{sn}$ *for model* (5.1) *with* $\theta_0 = (0.4, 0.5)$ *and* $E\varepsilon_t^2 = \infty$.

		$\varepsilon_t \sim$ model (5.2)				$\varepsilon_t \sim$ model (5.3)			
		$n = 200$		$n = 400$		$n = 200$		$n = 400$	
$\eta_t \sim$		$\tilde{\phi}_{sn}$	$\tilde{\psi}_{sn}$	$\tilde{\phi}_{sn}$	$\tilde{\psi}_{sn}$	$\tilde{\phi}_{sn}$	$\tilde{\psi}_{sn}$	$\tilde{\phi}_{sn}$	$\tilde{\psi}_{sn}$
Laplace$(0, 1)$	Bias	0.0003	0.0007	0.0009	-0.0010	-0.0031	0.0016	-0.0009	0.0001
	SD	0.1181	0.1123	0.0860	0.0828	0.1007	0.0881	0.0680	0.0593
	AD	0.1227	0.1108	0.0976	0.0826	0.1070	0.0911	0.0722	0.0622
	BD	0.1296	0.1183	0.0918	0.0867	0.1111	0.0975	0.0726	0.0634
N$(0, 1)$	Bias	0.0060	-0.0014	0.0066	-0.0025	0.0003	-0.0005	0.0020	0.0002
	SD	0.1754	0.1644	0.1417	0.1374	0.1458	0.1266	0.0991	0.0864
	AD	0.1773	0.1594	0.1390	0.1295	0.1482	0.1256	0.1020	0.0874
	BD	0.1814	0.1643	0.1443	0.1333	0.1517	0.1321	0.1037	0.0901
t_3	Bias	-0.0018	0.0006	0.0015	-0.0014	-0.0027	0.0005	0.0006	-0.0007
	SD	0.1258	0.1167	0.0895	0.0850	0.1163	0.1012	0.0796	0.0705
	AD	0.1178	0.1045	0.0832	0.0767	0.1088	0.0925	0.0756	0.0653
	BD	0.1346	0.1227	0.0946	0.0880	0.1239	0.1089	0.0845	0.0743

we fit each replication by an AR(1) model with the LADE and the SLADE method, respectively, and then use S_M to check whether an AR(1) model is adequate for the data sample generated from model (5.4). In all cases, we set the significance level $\underline{\alpha} = 0.05$ and $M = 6$. The empirical sizes and power of both tests are reported in Tables 3-4. Their sizes correspond to the results for the case with $\kappa = 0.0$. From Tables 3-4, it is clear that the sizes of W_n and S_M are always close to their nominal ones, although the sizes of S_M are conservative when n is small. For the power of both tests, it is generally as expected. First, all the powers become large as n increases. Second, both tests become more powerful as κ becomes larger. Overall, the tests W_n and S_M based on the random weighting approach have a good performance especially when the sample size is large.

Finally, we compare the performance of $\tilde{\theta}_n$ and $\tilde{\theta}_{sn}$ via a small simulation when ε_t is heteroscedastic noise with $\alpha \in (0, 2]$. We generate 10,000 replications of sample size $n = 5,000$, $10,000$ and $20,000$ from model (2.2) with true value $\theta_0 = 0.5$, where the innovation η_t is chosen to be N$(0, 1)$ distribution, and $(\alpha_0, \alpha_1, \beta)$ are set to be $(0.002, 0.2, 0.8)$ and $(0.002, 0.225, 0.8)$

Table 3. *Size and Power of* W_n *and* S_M $(\times 100)$ *for model (5.4) with* $E\varepsilon_t^2 < \infty$.

		$\varepsilon_t \sim$ model (5.2)				$\varepsilon_t \sim$ model (5.3)			
		$n = 200$		$n = 400$		$n = 200$		$n = 400$	
$\eta_t \sim$	κ	W_n	S_M	W_n	S_M	W_n	S_M	W_n	S_M
Laplace(0,1)	0.0	4.0	2.6	4.8	3.5	4.5	2.6	4.6	3.5
	0.2	79.4	21.9	98.1	57.3	82.3	22.3	98.5	58.7
	0.4	99.9	82.5	100	99.7	100	84.0	100	99.7
N(0,1)	0.0	6.1	2.8	5.5	3.1	6.2	2.7	5.5	3.2
	0.2	50.5	11.6	79.6	30.7	51.8	12.0	81.4	31.1
	0.4	97.5	63.3	100	96.3	97.9	64.1	100	96.6
t_3	0.0	5.2	2.7	5.9	3.4	4.3	2.3	5.9	3.7
	0.2	71.2	18.1	94.2	48.1	82.5	23.2	95.7	49.8
	0.4	99.7	80.0	100	99.4	99.8	80.9	100	99.6

† Both tests are calculated based on $\tilde{\theta}_n$.

Table 4. *Size and Power of* W_n *and* S_M $(\times 100)$ *for model (5.4) with* $E\varepsilon_t^2 = \infty$.

		$\varepsilon_t \sim$ model (5.2)				$\varepsilon_t \sim$ model (5.3)			
		$n = 200$		$n = 400$		$n = 200$		$n = 400$	
$\eta_t \sim$	κ	W_n	S_M	W_n	S_M	W_n	S_M	W_n	S_M
Laplace(0,1)	0.0	5.0	2.6	5.0	3.1	4.2	2.7	4.3	3.2
	0.2	58.8	14.3	82.4	41.3	73.4	17.1	96.3	47.3
	0.4	97.3	66.6	99.6	97.2	99.6	71.4	100	98.2
N(0,1)	0.0	6.3	2.6	6.4	3.2	6.3	2.7	5.9	3.5
	0.2	35.7	8.2	50.1	21.0	47.0	10.0	75.6	25.9
	0.4	85.7	46.4	94.6	87.1	95.5	53.0	99.9	92.1
t_3	0.0	6.1	2.3	6.1	3.2	5.2	2.5	5.9	3.2
	0.2	52.1	12.4	75.6	33.0	63.8	13.1	90.4	36.3
	0.4	96.0	62.2	99.6	95.8	99.1	65.5	100	97.1

† Both tests are calculated based on $\tilde{\theta}_{sn}$.

corresponding to the cases that $\alpha = 2$ and $\alpha \in (0, 2)$, respectively. Here, w_t is chosen as in (2.5) to calculate $\tilde{\theta}_{sn}$. Table 5 reports the bias and the relative efficiency of $\tilde{\theta}_n$ and $\tilde{\theta}_{sn}$: $R(\tilde{\theta}_n, \tilde{\theta}_{sn}) \equiv SD(\tilde{\theta}_n)/SD(\tilde{\theta}_{sn})$. From Table 5, we can see that both $\tilde{\theta}_n$ and $\tilde{\theta}_{sn}$ have small bias, and $\tilde{\theta}_{sn}$ is more efficient than $\tilde{\theta}_n$. Moreover, $R(\tilde{\theta}_n, \tilde{\theta}_{sn})$ is increasing as n becomes large.

This indicates that, when ε_t is heteroscedastic noise with $E\varepsilon_t^2 = \infty$, $\tilde{\theta}_n$ should have a slower rate of convergence than $n^{-1/2}$, and hence it confirms our conjecture in Remark 2.1.

Table 5. *Bias and Relative efficiency of* $\tilde{\theta}_n$ *and* $\tilde{\theta}_{sn}$ *for model (2.2) with* $\theta_0 = 0.5$.

	$\alpha \in (0, 2)$			$\alpha = 2$		
n	Bias of $\tilde{\theta}_n$	Bias of $\tilde{\theta}_{sn}$	$R(\tilde{\theta}_n, \tilde{\theta}_{sn})$	Bias of $\tilde{\theta}_n$	Bias of $\tilde{\theta}_{sn}$	$R(\tilde{\theta}_n, \tilde{\theta}_{sn})$
$5,000$	-0.0011	0.0005	1.5521	-0.0004	-0.0003	1.1621
$10,000$	-0.0009	-0.0003	2.0411	-0.0003	-0.0002	1.2621
$20,000$	-0.0015	-0.0001	2.8710	-0.0003	-0.0005	1.3508

6. APPLICATION TO HKD/USD EXCHANGE RATE

In this section, we study the daily HKD/USD exchange rate from January 21, 1998 to July 6, 2000, which has in total 601 observations. Denote the log-return ($\times 100$) of this data sample by $\{y_t\}_{t=1}^{600}$. To begin with, we first estimate the tail index of $\{y_t\}$ by Hill's estimator $\hat{\alpha}_y(k)$, where

$$\hat{\alpha}_y(k) = \left[\frac{1}{k} \sum_{i=1}^{k} \log \frac{y_{(n-i)}}{y_{(n-k)}} \right]^{-1} \tag{6.1}$$

with $\{y_{(t)}\}_{t=1}^{n}$ being the ascending order statistics of $\{y_t\}_{t=1}^{n}$. The plot of $\{\hat{\alpha}_y(k)\}_{k=10}^{180}$ is given in Figure 1, from which we can see that the tail of y_t is most likely between 1 and 2, i.e., $Ey_t^2 = \infty$ but $E|y_t| < \infty$. Next, we use an ARMA$(4, 2)$ model to fit $\{y_t\}$:

$$\begin{aligned} y_t = \underset{(0.0007)}{0.0016} - \underset{(0.0858)}{0.1238} y_{t-1} + \underset{(0.1517)}{0.0071} y_{t-2} - \underset{(0.0508)}{0.0232} y_{t-3} - \underset{(0.0370)}{0.0197} y_{t-4} \\ - \underset{(0.0945)}{0.0827} \varepsilon_{t-1} + \underset{(0.1456)}{0.0273} \varepsilon_{t-2} + \varepsilon_t, \end{aligned} \tag{6.2}$$

where model (6.2) is estimated by using the SLADE method with the weight function chosen as in (2.4), and the standard errors in parentheses are calculated via the random weighting method in Section 3 with $J = 500$. The p-values of the sign-based portmanteau tests S_{12} and S_{24} are 0.057 and 0.248, respectively. Hence, model (6.2) is adequate to fit $\{y_t\}$ at the significance level 5%. Moreover, we use the Wald test W_n to detect the hypothesis $H_0 : \phi_2 = \phi_3 = \phi_4 = \psi_1 = \psi_2 = 0$. The p-value of W_n is 0.571, and it turns out that we can not reject H_0 at the significance level

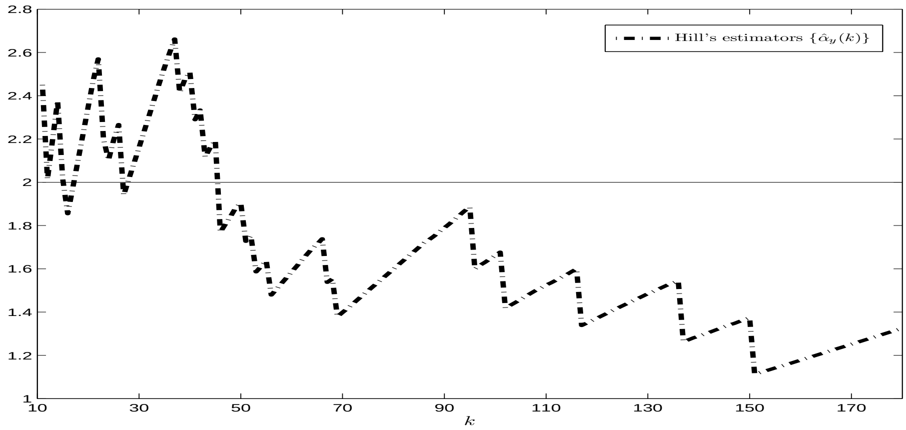

Fig. 1. Hill estimators $\{\hat{\alpha}_y(k)\}$ for $\{y_t\}$.

5%. Thus, we can further use an AR(1) model to fit $\{y_t\}$:

$$\begin{aligned} y_t = \underset{(0.0005)}{0.0016} - \underset{(0.0650)}{0.2492} y_{t-1} + \varepsilon_t, \end{aligned} \tag{6.3}$$

where model (6.3) is estimated in the same way as model (6.2), and the p-values of the sign-based portmanteau tests S_{12} and S_{24} are 0.065 and 0.255, respectively. This implies that model (6.3) is also adequate to fit $\{y_t\}$ at the significance level 5%. We should mention that the stationarity assumption plays an important role in the selected model (6.3). If we remove one or two "outliers", the model may be significantly changed.

Finally, we are interested in fitting the residuals $\{\tilde{\varepsilon}_t\}$ from model (6.3) by a GARCH$(1, 1)$ model, and the corresponding fitted model is as follows:

$$\tilde{\varepsilon}_t = \eta_t h_t \text{ and } h_t^2 = \underset{(0.0000)}{0.0000} + \underset{(0.0571)}{0.2127} \tilde{\varepsilon}_{t-1}^2 + \underset{(0.0632)}{0.5871} h_{t-1}^2, \tag{6.4}$$

where model (6.4) is estimated by using the Laplacian QMLE method in Berkes and Horváth (2004) with the standard errors in parentheses, and the estimated value of $E|\eta_t|$ is 0.997. A visual inspection of the Hill's estimators plot of η_t^2 (not displayed here) implies that $E\eta_t^2 < \infty$. Note that the portmanteau test $Q_a(M)$ in Li and Li (2008) is valid for pure GARCH model, when $E\varepsilon_t^2 = \infty$ and $E\eta_t^2 < \infty$. The p-values of $Q_a(12)$ and $Q_a(24)$ are 0.174 and 0.674, respectively, and hence model (6.4) is adequate at the significance level 5%. Particularly, the estimated value of $(E\eta_t^2)\alpha_1 + \beta$ is 1.1118. This implies that $Ey_t^2 = \infty$, and so our SLADE method used for model (6.3) is necessary in modelling the return series of HKD/USD exchange rate. Also, we have revisited the real example on world crude oil price in Zhu and Ling (2011), and the details can be found in one online supplementary material of this paper.

7. Conclusion and Discussion

In this paper, we propose the LADE and SLADE for the ARMA model with unspecified and heavy-tailed heteroscedastic noises. Both estimators are shown to be strongly consistent and asymptotically normal when the noise has a finite variance and infinite variance, respectively. Moreover, a Wald test based on the random weighting method is proposed to test the linear constraint in the true value, and a sign-based portmanteau test is investigated for model checking. Hence, a systematic procedure for statistical inference of ARMA model with unspecified and heavy-tailed heteroscedastic noises is feasible based on the LADE and SLADE methods, and its importance is further demonstrated by simulation studies and one real example.

The self-weighting approach can be applied for the M- and self-weighted M- estimators, see Huber (1977), He et al. (1990), and references therein for the classical M-estimation. Given the observations $\{y_1, \cdots, y_n\}$, as for (2.1), the objective function for M-estimation is:

$$\bar{W}_{sn}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{w}_t \tau(\tilde{\varepsilon}_t(\theta)),$$

where $\tau(\cdot)$ is a convex loss function, e.g., $\tau(x) = |x|^m$ for $m \in (1, 2]$ in Davis et al. (1992), Huber's pseudo-loss function $\tau(x) = \lambda^2[\sqrt{1 + (x/\lambda)^2} - 1]$ for $\lambda > 0$, and $\tau(x) = [(c+1)/2]\ln(x^2 + c)$ for $c > 0$ in Lucas (1995). The self-weighted M-estimator of θ_0 is $\bar{\theta}_{sn} \equiv \arg\min_{\theta\in\Theta} \bar{W}_{sn}(\theta)$. Under the regularity conditions given in Ling and McAleer (2010), $\bar{\theta}_{sn}$ is $\sqrt{n}$-consistent and asymptotically normal. However, its asymptotic covariance highly depends on the choice of $\tau(x)$ and it may be difficult to select a useful weight $\tilde{w}_t$ in practice. To see this, we look at AR(1) model in (2.2) and $\tau(x) = x^2$, i.e., self-weighted LSE. Then,

$$\bar{\theta}_{sn} - \theta_0 = \sum_{t=2}^{n} \tilde{w}_t y_{t-1}\varepsilon_t / \sum_{t=2}^{n} \tilde{w}_t y_{t-1}^2.$$

Since we do not know the form of h_t, it is not easy to choose $\tilde{w}_t$ such that $E(\tilde{w}_t y_{t-1}\varepsilon_t)^2 = E(\tilde{w}_t^2 y_{t-1}^2 h_t^2) < \infty$ even if $E\varepsilon_t^2 < \infty$. According to our preliminary research, it seems that the LADE or SLADE probably is the most useful approach for model (1.1).

ACKNOWLEDGEMENT

The authors greatly appreciate two anonymous referees, the Associate Editor and the Editors X. He and D. Ruppert for their very constructive suggestions and comments. This work is supported in part by Hong Kong Research Grants Commission Grants HKUST641912 and HKUST603413, National Natural Science Foundation of China (No.11201459), Key Lab of Random Complex Structures and Data Science of Chinese Academy of Sciences, and National Center for Mathematics and Interdisciplinary Sciences of Chinese Academy of Sciences.

APPENDIX: PROOFS

To facilitate the proofs, we first claim that there exist constants C and $\rho \in (0,1)$ such that the following holds uniformly in θ:

$$\sup_{\Theta} \|\varepsilon_{t-1}(\theta)\| \leq C\xi_{\rho t-1},$$
$$\sup_{\Theta} \left\| \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \right\| \leq C\xi_{\rho t-1},$$
$$\sup_{\Theta} \left\| \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq C\xi_{\rho t-1},$$

where $\xi_{\rho t}$ is defined as in Assumption 2.3. These three inequalities are used without mentioned, and their proofs are given in Ling (2007) based on Assumption 2.1.

Proof of Theorems 2.1. The proof of Theorem 2.1 follows the same one as for Theorem 2.1 and Theorem 2.3(i) in Zhu and Ling (2012), and hence the details are omitted. Q.E.D.

Proof of Theorems 2.2. We first re-parameterize the objective function (2.1) as $H_n(u) = n\tilde{L}_{sn}(\theta_0 + u) - n\tilde{L}_{sn}(\theta_0)$, where $u \in \Lambda \equiv \{u = (u_1', u_2')' : u + \theta_0 \in \Theta\}$. Let $\tilde{u}_n = \tilde{\theta}_{sn} - \theta_0$. Then, $\tilde{u}_n$ is the minimizer of $H_n(u)$ in Λ. Furthermore, we have

$$H_n(u) = \sum_{t=1}^{n} w_t \tilde{A}_t(u) + \tilde{\Pi}_{1n}(u), \tag{A1}$$

where

$$\tilde{A}_t(u) = |\tilde{\varepsilon}_t(\theta_0 + u)| - |\tilde{\varepsilon}_t(\theta_0)| \quad \text{and} \quad \tilde{\Pi}_{1n}(u) = \sum_{t=1}^{n} (\tilde{w}_t - w_t)\tilde{A}_t(u).$$

Let $I(\cdot)$ be the indicator function. Using the identity

$$|x - y| - |x| = -y[sgn(x)] + 2\int_0^y [I(x \le s) - I(x \le 0)]ds$$

for $x \neq 0$, we can show that

$$\tilde{A}_t(u) = \tilde{q}_t(u)[sgn(\tilde{\varepsilon}_t)] + 2\int_0^{-\tilde{q}_t(u)} \tilde{X}_t(s)ds, \tag{A2}$$

where $\tilde{\varepsilon}_t := \tilde{\varepsilon}_t(\theta_0)$, $\tilde{X}_t(s) = I(\tilde{\varepsilon}_t < s) - I(\tilde{\varepsilon}_t < 0)$,

$$\tilde{q}_t(u) = u' \frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial \theta} + \frac{1}{2} u' \frac{\partial^2 \tilde{\varepsilon}_t(\theta^\dagger)}{\partial \theta \partial \theta'} u,$$

and $\theta^\dagger$ lies between θ_0 and $\theta_0 + u$. Moreover, let

$$\tilde{\xi}_t(u) = 2w_t \int_0^{-u' \frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial \theta}} \tilde{X}_t(s)ds.$$

Then, from (A2), we have

$$\sum_{t=1}^{n} w_t \tilde{A}_t(u) = u'\tilde{T}_n + \tilde{\Pi}_{2n}(u) + \tilde{\Pi}_{3n}(u), \tag{A3}$$

where

$$\tilde{T}_n = \sum_{t=1}^{n} w_t \frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial \theta} [sgn(\tilde{\varepsilon}_t)], \qquad \tilde{\Pi}_{2n}(u) = \sum_{t=1}^{n} \tilde{\xi}_t(u),$$

$$\tilde{\Pi}_{3n}(u) = \frac{u'}{2} \sum_{t=1}^{n} w_t \frac{\partial^2 \tilde{\varepsilon}_t(\theta^\dagger)}{\partial \theta \partial \theta'} [sgn(\tilde{\varepsilon}_t)]u + 2\sum_{t=1}^{n} w_t \int_{-u' \frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial \theta}}^{-\tilde{q}_t(u)} \tilde{X}_t(s)ds.$$

By Assumptions 2.1-2.6, as for Lemma A.4 in Zhu and Ling (2012), we can show that

$$\frac{1}{\sqrt{n}}\left(\tilde{T}_n - T_n\right) = o_p(1), \tag{A4}$$

$$\tilde{\Pi}_{1n}(\tilde{u}_n) = o_p(\sqrt{n}\|\tilde{u}_n\|), \tag{A5}$$

$$\tilde{\Pi}_{2n}(\tilde{u}_n) - \Pi_{2n}(\tilde{u}_n) = o_p(\sqrt{n}\|\tilde{u}_n\|), \tag{A6}$$

$$\tilde{\Pi}_{3n}(\tilde{u}_n) - \Pi_{3n}(\tilde{u}_n) = o_p(n\|\tilde{u}_n\|^2), \tag{A7}$$

where T_n, $\Pi_{2n}(u)$ and $\Pi_{3n}(u)$ are defined in the same way as $\tilde{T}_n$, $\tilde{\Pi}_{2n}(u)$ and $\tilde{\Pi}_{3n}(u)$, respectively, with $\tilde{\varepsilon}_t(\theta)$ being replaced by $\varepsilon_t(\theta)$. Using the same arguments as for Lemmas 2.2 and 2.3 in Zhu and Ling (2012), we can show that

$$\Pi_{2n}(u_n) = (\sqrt{n}u_n)'[g(0)\Sigma](\sqrt{n}u_n) + o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2), \tag{A8}$$

$$\Pi_{3n}(u_n) = o_p(n\|u_n\|^2), \tag{A9}$$

where Σ is defined as in Theorem 2.2. Note that the conditional median of ε_t on $\mathcal{F}_{t-1}$ is zero. Directly by central limit theorem for martingale difference sequence, we have

$$\frac{1}{\sqrt{n}}T_n \to_d N(0, \Omega), \quad \text{as } n \to \infty, \tag{A10}$$

where Ω is defined as in Theorem 2.2. By (A1) and (A3)-(A10), exactly following the same procedure as for Theorem 2.2 in Zhu and Ling (2012), we can complete the proof and the details are omitted. Q.E.D.

Proof of Theorem 3.1. Let $T_n = \sum_{t=1}^n w_t(\partial\varepsilon_t(\theta_0)/\partial\theta)[sgn(\varepsilon_t)]$. According to the proof of Theorem 2.2, by Assumptions 2.1-2.6, we can show that

$$\sqrt{n}(\tilde{\theta}_{sn} - \theta_0) = -[2g(0)\Sigma]^{-1}T_n/\sqrt{n} + o_p(1). \tag{A11}$$

Similarly, since $Ew_t^{*2} < \infty$ and $\{w_t^*\}$ and $\{y_t\}$ and independent by Assumption 3.1, we have

$$\sqrt{n}(\tilde{\theta}^*_{sn} - \theta_0) = -[2g(0)\Sigma^*]^{-1}T^*_n/\sqrt{n} + o_p(1), \tag{A12}$$

where

$$\Sigma^* = E\left[\frac{w_t^* w_t}{h_t}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}\right] \quad \text{and} \quad T^*_n = \sum_{t=1}^n w_t^* w_t \frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}[sgn(\varepsilon_t)].$$

Then, by the independence of $\{w_t^*\}$ and $\{y_t\}$ and (A11)-(A12), we know that $\Sigma^* = \Sigma$ and

$$\sqrt{n}(\tilde{\theta}^*_{sn} - \tilde{\theta}_{sn}) = \frac{[2g(0)\Sigma]^{-1}}{\sqrt{n}}\sum_{t=1}^n(1 - w_t^*)w_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}[sgn(\varepsilon_t)] + o_p(1). \tag{A13}$$

Furthermore, by using the same argument as for Lemma A.4 in Zhu and Ling (2012), we can get

$$\sqrt{n}(\tilde{\theta}_{sn}^{*} - \tilde{\theta}_{sn}) = [2g(0)\Sigma]^{-1} \sum_{t=1}^{n} J_{tn} + o_p(1), \tag{A14}$$

where $J_{tn} = n^{-1/2}(1 - w_t^*)\tilde{w}_t(\partial\tilde{\varepsilon}_t(\theta_0)/\partial\theta)[sgn(\tilde{\varepsilon}_t)]$. Let E^* be the conditional expectation on $\{y_1, \cdots, y_n\}$ and $c \in \mathcal{R}^m$ be a constant vector. We now study the conditional distribution of $\sum_{t=1}^{n} c' J_{tn}$.

First, since w_t^* is independent to y_t with $Ew_t^* = 1$, we know that

$$E^*[c' J_{tn}] = 0. \tag{A15}$$

Next, since $Var(w_t^*) = 1$, by the independence of $\{w_t^*\}$ and $\{y_t\}$ and the same argument as for Lemma A.4 in Zhu and Ling (2012), we can show that

$$\begin{aligned}
\sum_{t=1}^{n} E^*[c' J_{tn} J_{tn}' c] &= c' \left\{ \frac{1}{n} \sum_{t=1}^{n} \tilde{w}_t^2 \frac{\partial\tilde{\varepsilon}_t(\theta_0)}{\partial\theta} \frac{\partial\tilde{\varepsilon}_t(\theta_0)}{\partial\theta'} [I(\tilde{\varepsilon}_t > 0) + I(\tilde{\varepsilon}_t < 0)] \right\} c \\
&= c' \left\{ \frac{1}{n} \sum_{t=1}^{n} w_t^2 \frac{\partial\varepsilon_t(\theta_0)}{\partial\theta} \frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'} [I(\varepsilon_t > 0) + I(\varepsilon_t < 0)] \right\} c + o_p(1) \\
&= c'\Omega c + o_p(1).
\end{aligned} \tag{A16}$$

Finally, we check the Lindeberg condition. Let C_0 be a positive generic constant. Since $E|w_t^*|^{2+\delta_0} < \infty$, by Hölder and Markov inequalities, for all $t = 1, \cdots, n$ and any given $\eta > 0$, we have

$$\begin{aligned}
E^*\left[(1 - w_t^*)^2 I(|c' J_{tn}| > \eta)\right] &\le \left\{E^*\left[|1 - w_t^*|^{2+\delta_0}\right]\right\}^{\frac{2}{2+\delta_0}} \left[E^* I\left(|c' J_{tn}| > \eta\right)\right]^{\frac{\delta_0}{2+\delta_0}} \\
&\le C_0 \left[\frac{E^*|c' J_{tn}|}{\eta}\right]^{\frac{\delta_0}{2+\delta_0}} \\
&\le C_0 \tilde{K}_n^{\frac{\delta_0}{2+\delta_0}},
\end{aligned}$$

where $\tilde{K}_n = n^{-1/2} \max_{1 \le t \le n} \|\tilde{w}_t \frac{\partial\tilde{\varepsilon}_t(\theta_0)}{\partial\theta}[sgn(\tilde{\varepsilon}_t)]\|$. Therefore, for any given $\eta > 0$, it follows that

$$\begin{aligned}
&\sum_{t=1}^{n} E^*[c' J_{tn} J_{tn}' c I(|c' J_{tn}| > \eta)] \\
&\quad = E^*\left[(1 - w_t^*)^2 I(|c' J_{tn}| > \eta)\right] \\
&\qquad \times c' \left\{ \frac{1}{n} \sum_{t=1}^{n} \tilde{w}_t^2 \frac{\partial\tilde{\varepsilon}_t(\theta_0)}{\partial\theta} \frac{\partial\tilde{\varepsilon}_t(\theta_0)}{\partial\theta'} [I(\tilde{\varepsilon}_t > 0) + I(\tilde{\varepsilon}_t < 0)] \right\} c \\
&\quad \le C_0 \tilde{K}_n^{\frac{\delta_0}{2+\delta_0}} c' \left\{ \frac{1}{n} \sum_{t=1}^{n} \tilde{w}_t^2 \frac{\partial\tilde{\varepsilon}_t(\theta_0)}{\partial\theta} \frac{\partial\tilde{\varepsilon}_t(\theta_0)}{\partial\theta'} [I(\tilde{\varepsilon}_t > 0) + I(\tilde{\varepsilon}_t < 0)] \right\} c \\
&\quad = C_0 [K_n + o_p(1)]^{\frac{\delta_0}{2+\delta_0}} \\
&\qquad \times c' \left\{ \frac{1}{n} \sum_{t=1}^{n} w_t^2 \frac{\partial\varepsilon_t(\theta_0)}{\partial\theta} \frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'} [I(\varepsilon_t > 0) + I(\varepsilon_t < 0)] + o_p(1) \right\} c,
\end{aligned}$$

where the last equality holds due to the same argument as for Lemma A.4 in Zhu and Ling (2012), and K_n is defined in the same way as $\tilde{K}_n$ with $\tilde{w}_t$ and $\tilde{\varepsilon}_t$ being replaced by w_t and ε_t, respectively. Note that $K_n = o_p(1)$ because $E\|w_t \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta}[sgn(\varepsilon_t)]\|^2 < \infty$. Hence, we know that

$$\sum_{t=1}^{n} E^*\left[c' J_{tn} J_{tn}' c I(|c' J_{tn}| > \eta)\right] = o_p(1). \tag{A17}$$

By (A15)-(A17), the Cramér-Wold device and central limit theorem in Pollard (1984, Theorem VIII.1) yield that conditional on $\{y_1, \cdots, y_n\}$,

$$\sum_{t=1}^{n} J_{tn} \to_d N(0, \Omega) \text{ in probability as } n \to \infty.$$

Now, the conclusion follows directly from (A14). Q.E.D.

Proof of Theorem 4.1. First, by Theorem 3.1 in Ling and McAleer (2003), the dominated convergence theorem and the same argument as for Lemma A.4 in Zhu and Ling (2012), we can show that

$$\begin{aligned}
&\bar{\zeta}_n = E(\zeta_t) + o_p(1) = o_p(1); \\
&\frac{1}{n}\sum_{t=1}^{n}\left[\tilde{\zeta}_t(\tilde{\theta}_{sn}) - \bar{\zeta}_n\right] = var(\zeta_t) + o_p(1) = Ew_t^2 + o_p(1);
\end{aligned}$$

and hence

$$\sqrt{n}\tilde{\rho}_k = \frac{1}{\sqrt{n}Ew_t^2}\sum_{t=k+1}^{n} \zeta_t(\tilde{\theta}_{sn})\zeta_{t-k}(\tilde{\theta}_{sn}) + o_p(1) := \frac{\sqrt{n}\hat{\rho}_k}{Ew_t^2} + o_p(1). \tag{A18}$$

Denote $\Theta_n = \{\theta \in \Theta : \sqrt{n}\|\theta - \theta_0\| \leq C_1\}$ for some constant $C_1 > 0$. Then, we want to show that

$$\sup_{\theta \in \Theta_n}\left\|\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left[\zeta_t(\theta)\zeta_{t-k}(\theta) - \zeta_t\zeta_{t-k}\right] - 2g(0)W_k'[\sqrt{n}(\theta - \theta_0)]\right\| = o_p(1). \tag{A19}$$

In order to prove (A19), we rewrite

$$\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left[\zeta_t(\theta)\zeta_{t-k}(\theta) - \zeta_t\zeta_{t-k}\right] = \frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left[d_{1t}(\theta) + d_{2t}(\theta)\right],$$

where $d_{1t}(\theta) = \zeta_{t-k}(\theta)[\zeta_t(\theta) - \zeta_t]$ and $d_{2t}(\theta) = \zeta_t[\zeta_{t-k}(\theta) - \zeta_{t-k}]$. By Taylor's expansion and Assumptions 2.3 and 2.5-2.6, a simple algebra gives us that

$$\begin{aligned}
E\sup_{\theta \in \Theta_n}|\zeta_t(\theta) - \zeta_t| &= 2E\sup_{\theta \in \Theta_n} w_t\left|I\left(\eta_t > -\frac{\theta - \theta_0}{h_t}\frac{\partial \varepsilon_t(\theta^\dagger)}{\partial \theta}\right) - I(\eta_t > 0)\right| \\
&\leq 2E\left[w_t I\left(-C_2 n^{-1/2}\xi_{\rho t-1} < \eta_t < C_2 n^{-1/2}\xi_{\rho t-1}\right)\right] = O(n^{-1/2}) \qquad \text{(A20)}
\end{aligned}$$

for some constant $C_2 > 0$, where $\theta^\dagger$ lies between θ and θ_0. By (A20) and some standard arguments, it is not hard to show that

$$\sup_{\theta\in\Theta_n}\left\|\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\{d_{1t}(\theta)-E\left[d_{1t}(\theta)|\mathcal{F}_{t-1}\right]\}\right\| = o_p(1), \tag{A21}$$

$$\sup_{\theta\in\Theta_n}\left\|\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}E\left[d_{1t}(\theta)|\mathcal{F}_{t-1}\right]-2g(0)W_k'[\sqrt{n}(\theta-\theta_0)]\right\| = o_p(1), \tag{A22}$$

$$\sup_{\theta\in\Theta_n}\left\|\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}d_{2t}(\theta)\right\| = o_p(1). \tag{A23}$$

Thus, by (A21)-(A23), it follows that (A19) holds. Denote $\xi_t(\tilde{\theta}_{sn})$ be $\hat{\xi}_t$. Let $\hat{\rho}_k$ be $\rho_k^\diamond$ when $\tilde{\theta}_{sn}$ is replaced by θ_0. Since $\sqrt{n}(\tilde{\theta}_{sn}-\theta_0)=O_p(1)$ by Theorem 2.2, from (A19), we have

$$\sqrt{n}\hat{\rho}_k-\sqrt{n}\rho_k^\diamond=\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}w_tw_{t-k}\left[\hat{\xi}_t\hat{\xi}_{t-k}-\xi_t\xi_{t-k}\right]=2g(0)W_k'[\sqrt{n}(\tilde{\theta}_{sn}-\theta_0)]+o_p(1),$$

which implies

$$\sqrt{n}\hat{\rho}-\sqrt{n}\rho^\diamond=2g(0)W[\sqrt{n}(\tilde{\theta}_{sn}-\theta_0)]+o_p(1), \tag{A24}$$

where $\hat{\rho}=(\hat{\rho}_1,\cdots,\hat{\rho}_M)'$ and $\rho^\diamond=(\rho_1^\diamond,\cdots,\rho_M^\diamond)'$. Furthermore, by (A11), (A18) and (A24), we have

$$Ew_t^2(\sqrt{n}\tilde{\rho})=VZ_n^\diamond+o_p(1),$$

where

$$Z_n^\diamond=\left(\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\zeta_t\zeta_{t-1},\cdots,\frac{1}{\sqrt{n}}\sum_{t=M+1}^{n}\zeta_t\zeta_{t-M},\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\zeta_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}\right)'.$$

Finally, the conclusion holds by central limit theorem for martingale difference sequence. Q.E.D.

Proof of Theorem 4.2. First, by the same argument as for (A18), we have

$$\sqrt{n}\tilde{\rho}_{w^*k}=\frac{1}{\sqrt{n}Ew_t^2}\sum_{t=k+1}^{n}w_t^*\zeta_t(\tilde{\theta}_{sn}^*)\zeta_{t-k}(\tilde{\theta}_{sn}^*)+o_p(1):=\frac{\sqrt{n}\hat{\rho}_{w^*k}}{Ew_t^2}+o_p(1),$$

and hence

$$Ew_t^2\left(\sqrt{n}\tilde{\rho}_{w^*k}-\sqrt{n}\tilde{\rho}_k\right)=\sqrt{n}\hat{\rho}_{w^*k}-\sqrt{n}\hat{\rho}_k+o_p(1). \tag{A25}$$

Let $\hat{\rho}_{w^*k}$ be $\rho_{w^*k}^\diamond$ when $\tilde{\theta}_{sn}^*$ is replaced by θ_0. In view of (A12) and $Ew_t^*=1$, by the same argument as for (A24), we can show that

$$\sqrt{n}\hat{\rho}_{w^*}-\sqrt{n}\rho_{w^*}^\diamond=2g(0)W[\sqrt{n}(\tilde{\theta}_{sn}^*-\theta_0)]+o_p(1), \tag{A26}$$

where $\hat{\rho}_{w^*} = (\hat{\rho}_{w^*1}, \cdots, \hat{\rho}_{w^*M})'$ and $\rho^{\diamond}_{w^*} = (\rho^{\diamond}_{w^*1}, \cdots, \rho^{\diamond}_{w^*M})'$. Therefore, by (A24)-(A26), we have

$$\begin{aligned} Ew_t^2(\sqrt{n}\tilde{\rho}_{w^*} - \sqrt{n}\tilde{\rho}) &= (\sqrt{n}\hat{\rho}_{w^*} - \sqrt{n}\rho^{\diamond}_{w^*}) - (\sqrt{n}\hat{\rho} - \sqrt{n}\rho^{\diamond}) \\ &\quad + (\sqrt{n}\rho^{\diamond}_{w^*} - \sqrt{n}\rho^{\diamond}) + o_p(1) \\ &= (\sqrt{n}\rho^{\diamond}_{w^*} - \sqrt{n}\rho^{\diamond}) + 2g(0)W[\sqrt{n}(\tilde{\theta}^*_{sn} - \tilde{\theta}_{sn})] + o_p(1). \end{aligned}$$

Now, by (A13), it follows that

$$Ew_t^2(\sqrt{n}\tilde{\rho}_{w^*} - \sqrt{n}\tilde{\rho}) = VZ_n^* + o_p(1),$$

where

$$Z_n^* = \left(\frac{1}{\sqrt{n}} \sum_{t=2}^{n} (w_t^* - 1)\zeta_t\zeta_{t-1}, \cdots, \frac{1}{\sqrt{n}} \sum_{t=M+1}^{n} (w_t^* - 1)\zeta_t\zeta_{t-M}, \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (w_t^* - 1)\zeta_t \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta'} \right)'.$$

Thus, by the same argument as for Theorem 3.1, we know that the conclusion holds. Q.E.D.

REFERENCES

[1] BERKES, I. & HORVÁTH, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics* **32**, 633-655.

[2] BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.

[3] BROCKWELL, P.J. and DAVIS, R.A. (1991). Time Series: Theory and Methods. 2nd ed. Springer-Verlag, New York.

[4] CHAN, N.H. and PENG, L. (2005). Weighted least absolute deviation estimation for an AR(1) process with ARCH(1) errors. *Biometrika* **92**, 477-484.

[5] CHAN, N.H. and WEI, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* **16**, 367-401.

[6] CHEN, K., GUO, S., LIN, Y., and YING, Z. (2010). Least absolute relative error estimation. *Journal of the American Statistical Association* **105**, 1104-1112.

[7] CHEN, K., YING, Z., ZHANG, H., and ZHAO, L. (2008). Analysis of least absolute deviation. *Biometrika* **95**, 107-122.

[8] DAVIS, R.A. (1996). Gauss-Newton and M-estimation for ARMA processes with infinite variance. *Stochastic Processes and Their Applications* **63**, 75-95.

[9] DAVIS, R.A. and DUNSMUIR, W.T.M. (1997). Least absolute deviation estimation for regression with ARMA errors. Dedicated to Murray Rosenblatt. *Journal of Theoretical Probability* **10**, 481-497.

[10] DAVIS, R.A., KNIGHT,K., and LIU, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Application* **40**, 145-180.

[11] DAVIS, R.A., and MIKOSCH, T. (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Annals of Statistics* **26** 2049-2080.

[12] ENGLE, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. *Econometrica* **50**, 987-1008.

[13] FAN, J. and YAO, Q. (2003). Nonlinear time series: Nonparametric and parametric methods. Springer, New York.

[14] FORNARI, F. and MELE, A. (1997). Sign- and volatility-switching ARCH models: theory and applications to international stock markets. *Journal of Applied Econometrics* **12**, 1779-1801.

[15] FRANCQ, C. and ZAKOÏAN, J.M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605-637.

[16] FRANCQ, C. and ZAKOÏAN, J.M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, Chichester, UK.

[17] FRANSES, P.H. and VAN DIJK, R. (1996). Forecasting stock market volatility using (non-linear) Garch models. *Journal of Forecasting* **15**, 229-235.

[18] JIN, Z., YING, Z., and WEI, L.J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381-390.

[19] GLOSTEN, L.R., JAGANNATHAN, R., and RUNKLE, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**, 1779-1801.

[20] HE, C. and TERASVIRTA, T. (1999). Properties of moments of a family of GARCH processes. *Journal of Econometrics* **92**, 173-192.

[21] HE, X., JURECKOVA, J., KOENKER, R. and PORTNOY, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* **58**, 1195-1214.

[22] HUBER, P.J. (1977). Robust Statistical Procedures. Philadelphia: Society for Industrial and Applied Mathematics.

[23] KNIGHT, K. (1987). Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregressions. *Journal of Time Series Analysis* **8**, 51-60.

[24] KNIGHT, K. (1998). Limiting distributions for L_1 regression estimators under general conditons. *Annals of Statistics* **26**, 755-770.

[25] KOENKER, R.W. and BASSETT, G.W. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

[26] KOENKER, R.W. and BASSETT, G.W. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Journal of Econometrics* **140** 849-873.

[27] LANGE, T. (2011). Tail behavior and OLS estimation in AR-GARCH models. *Statistica Sinica* **21**, 1191-1200.

[28] LI, G. and LI, W.K. (2008). Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika* **95**, 399-414.

[29] LING, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society* **B 67**, 381-393.

[30] LING, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* **140**, 849-873.

[31] LING, S. and MCALEER, M. (2003). Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* **19**, 280-310.

[32] LING, S. and MCALEER, M. (2010). A general asymptotic theory for time-series models. *Statistica Neerlandica* **64**, 97-111.

[33] LJUNG, G.M. and BOX, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297-303.

[34] LUCAS, A. (1995). Unit root tests based on M estimators. *Econometric Theory* **11**, 331-346

[35] MIKOSCH, T., GADRICH, T., KLÜPPELBERG, C., and ADLER, R.J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *Annals of Statistics* **23**, 305-326.

[36] PAN, J., WANG, H. and YAO, Q. (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory* **23**, 852-879.

[37] POLLARD, D. (1984). Convergence of stochastic processes. Springer, Berlin.

[38] ROBINSON, P.M. (1977). The estimation for a nonliear moving average model. *Stochastic Processes and their Applications* **5**, 81-90.

[39] SCHWERT, G.W. (1989). Why does stock market volatility change over time? *Journal of Finance* **45**, 1129-1155.

[40] TAYLOR, S. (1986). Modelling Financial Time Series. Wiley, New York.

[41] TIAO, G.C. and TSAY, R.S. (1989). Model specification in multivariate time series (with discussion). *Journal of the Royal Statistical Society* **B 51**, 157-213.

[42] TSAY, R.S. (1987). Conditional heteroscedastic time series models. *Journal of the American Statistical Association* **82**, 590-604.

[43] TSAY, R.S. (2005). Analysis of financial time series (2nd ed.). New York: John Wiley & Sons, Incorporated.

[44] TSAY, R.S. and TIAO, G.C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Annals of Statistics* **18**, 220-250.

[45] WU, C.F.J. (1986). Jacknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics* **14**, 1261-1350.

[46] WU, R. and DAVIS, R.A. (2010). Least absolute deviation estimation for general autoregressive moving average time series models. *Journal of Time Series Analysis* **31**, 98-112.

[47] YANG, M. and BEWLEY, R. (1995). Moving average conditional heteroskedastic processes. *Economics Letters* **49**, 367-372.

[48] YAO, Q. and BROCKWELL, P.J. (2006). Gaussian maximum likelihood estimation for ARMA models I: Time series. *Journal of Time Series Analysis* **27**, 857-875.

[49] ZAKOÏAN, J.M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18**, 931-955.

[50] ZHANG, R.M. and LING, S. (2014). Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory*. Forthcoming.

[51] ZHANG, R.M., SIN, C.-Y. and LING, S. (2014). On functional limits of short- and long-memory linear processes with GARCH(1,1) noises. Revised for *Stochastic Processes and their Applications*.

[52] ZHU, K. and LI, W.K. (2013). A bootstrapped spectral test for adequacy in weak ARMA models. Submitted.

[53] ZHU, K. and LING, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Annals of Statistics* **39**, 2131-2163.

[54] ZHU, K. and LING, S. (2012). The global weighted LAD estimators for finite/infinite variance ARMA(p, q) models. *Econometric Theory* **28**, 1065-1086.

[55] ZHU, K. and LING, S. (2013). Quasi-maximum exponential likelihood estimators for a double AR(p) model. *Statistica Sinica* **23**, 251-270.

LADE-based Inference for ARMA Models with Unspecified and Heavy-tailed Heteroscedastic Noises (Supplementary Material)

BY KE ZHU

Institute of Applied Mathematics, Chinese Academy of Sciences, Haidian District, Zhongguancun, Beijing, China

AND SHIQING LING

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

In this supplementary material, we revisit the real example on world crude oil price in Zhu and Ling (2011). The data sample they used is the weekly world crude oil price (dollars per barrel) from January 3, 1997 to August 6, 2010, which has in total 710 observations. Denote the log-return ($\times 100$) of this data sample by $\{y_t\}_{t=1}^{709}$. In Zhu and Ling (2011), a MA(3)-GARCH$(1,1)$ model is used to fit $\{y_t\}_{t=1}^{709}$, see model (5.2) in their paper. In the sequel, we will check whether a MA(3) model is adequate to fit $\{y_t\}$ in presence of conditional heteroskedasticity.

To begin with, we first estimate the tail index of $\{y_t\}$ by Hill's estimator $\hat{\alpha}_y(k)$ [see (6.1) in the paper]. The plot of $\{\hat{\alpha}_y(k)\}_{k=10}^{180}$ is given in Figure 1 below, from which we can see that the tail of y_t is most likely greater than 2, i.e., $Ey_t^2 < \infty$. This is consistent to the finding in Zhu and Ling (2011). Next, by looking at the first ten autocorrelation functions (ACFs) or partial autocorrelation functions (PACFs) of $\{y_t\}$ in Figure 2 below, we know that the 1st and 3rd ACF and the 1st, 3rd and 4th PACF exceed two asymptotic standard errors. Thus, it motives us to use

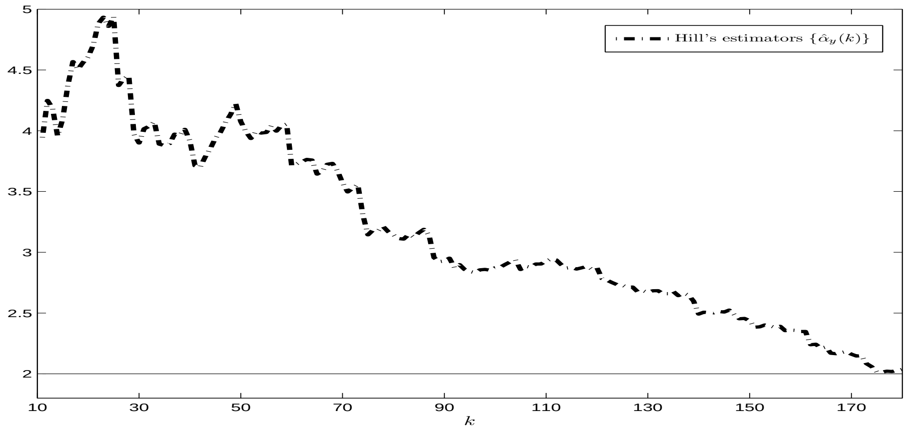

Fig. 1. Hill estimators $\{\hat{\alpha}_y(k)\}$ for $\{y_t\}$.

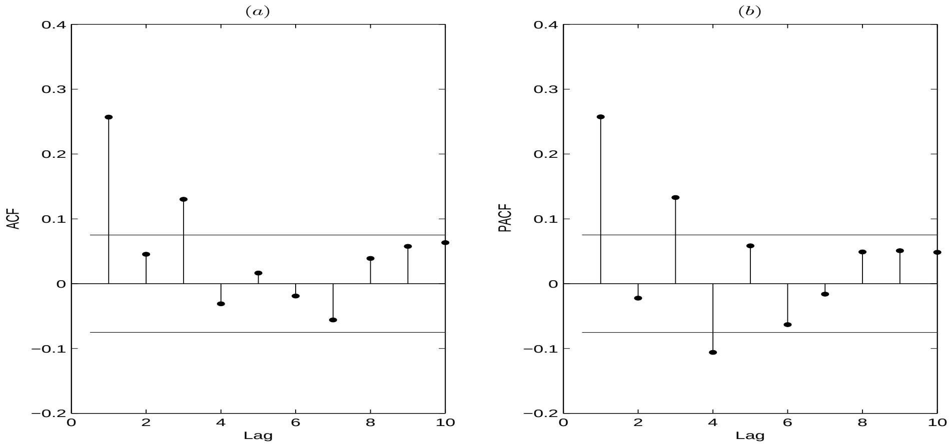

Fig. 2. (a) the autocorrelations for $\{y_t\}$; (b) the partial autocorrelations for $\{y_t\}$.

an ARMA$(4,3)$ model to fit $\{y_t\}$:

$$\begin{aligned} y_t = {} & 0.2047 - 0.1688 y_{t-1} + 0.3143 y_{t-2} + 0.3011 y_{t-3} - 0.1754 y_{t-4} \\ & (0.3008) \quad (0.3494) \qquad (0.3908) \qquad (0.2340) \qquad (0.0937) \\ & +0.4938\varepsilon_{t-1} - 0.1966\varepsilon_{t-2} - 0.2653\varepsilon_{t-3} + \varepsilon_t, \\ & (0.3459) \qquad (0.4419) \qquad (0.3316) \end{aligned} \tag{0.1}$$

where model (0.1) is estimated by using the LADE method, and the standard errors in parentheses are calculated via the random weighting method in Section 3 with $J = 500$. The p-values of the sign-based portmanteau tests S_{12} and S_{24} are 0.747 and 0.906, respectively. Hence, model (0.1) is adequate to fit $\{y_t\}$ at the significance level 5%. Moreover, we use the Wald test W_n to detect the hypothesis $H_0 : \mu = \phi_1 = \phi_2 = \phi_3 = \phi_4 = \psi_2 = 0$. The p-value of W_n is 0.123, and it turns out that we can not reject H_0 at the significance level 5%. Thus, we can further use a MA(3) model to fit $\{y_t\}$:

$$\begin{array}{c} y_t = 0.3085\varepsilon_{t-1} + 0.1334\varepsilon_{t-3} + \varepsilon_t, \\ \quad (0.0441) \qquad (0.0372) \end{array} \tag{0.2}$$

where model (0.2) is estimated in the same way as model (0.1), and the p-values of the sign-based portmanteau tests S_{12} and S_{24} are 0.499 and 0.633, respectively. This implies that model (0.2) is also adequate to fit $\{y_t\}$ at the significance level 5%. Hence, it is reasonable for us to use a MA(3)-GARCH$(1, 1)$ as in Zhu and Ling (2011) to fit $\{y_t\}$.