ZD-GARCH model: a new way to study heteroscedasticity

Li, Dong and Ling, Shiqing and Zhu, Ke

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

1 January 2016

ZD-GARCH model: a new way to study heteroscedasticity

BY DONG LI

Center for Statistical Science, Tsinghua University, Beijing 100084, China

malidong@tsinghua.edu.cn

SHIQING LING

Department of Mathematics, Hong Kong University of Science and Technology,

Clear Water Bay, Kowloon, Hong Kong

maling@ust.hk

AND KE ZHU

Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100190, China

kzhu@amss.ac.cn

ABSTRACT

This paper proposes a first-order zero-drift GARCH (ZD-GARCH$(1,1)$) model to study conditional heteroscedasticity and heteroscedasticity together. Unlike the classical GARCH model, ZD-GARCH$(1,1)$ model is always non-stationary regardless of the sign of the Lyapunov exponent γ_0, but interestingly when $\gamma_0 = 0$, it is *stable* with its sample path oscillating randomly between zero and infinity over time. Furthermore, this paper studies the generalized quasi-maximum likelihood estimator (GQMLE) of ZD-GARCH$(1,1)$ model, and establishes its strong consistency and asymptotic normality. Based on the GQMLE, an estimator for γ_0, a test for stability, and a portmanteau test for model checking are all constructed. Simulation studies are carried out to assess the finite sample performance of the proposed estimators and tests. Appli-

cations demonstrate that a stable ZD-GARCH$(1,1)$ model is more appropriate to capture heteroscedasticity than a non-stationary GARCH$(1,1)$ model, which suffers from an inconsistent QMLE of the drift term.

Some key words: Conditional heteroscedasticity; GARCH model; Generalized quasi-maximum likelihood estimator; Heteroscedasticity; Portmanteau test; Stability test; Top Lyapunov exponent; Zero-drift GARCH model.

1. INTRODUCTION

HETEROSCEDASTICITY is the often observed feature for economic and financial time series data. When the heteroscedastic error structure in regressions is correctly specified, one could gain substantial efficiency in using generalized least squares estimator (LSE), and more importantly, eliminate the ordinary LSE-based bias in standard errors resulting in valid inferences. Therefore, most of efforts made in the literature are to test heteroscedasticity by assuming a specified heteroscedastic error structure; see, e.g., Breusch and Pagan (1979) for earlier works and Greene (2002) and the references therein for more recent ones. In the last three decades, the conditionally heteroscedastic model has achieved a great success after the seminar work of Engle (1982) and Bollerslev (1986). However, less attempts have been made in the literature to capture conditional heteroscedasticity and heteroscedasticity together parametrically.

As one leading motivation, this paper provides a new parametric way to reach this goal. Let y_t be the error term in regressions. This paper proposes a first-order zero-drift generalized autoregressive conditional heteroscedasticity (ZD-GARCH$(1,1)$) model to capture y_t:

$$y_t = \eta_t \sqrt{h_t} \text{ and } h_t = \alpha_0 y_{t-1}^2 + \beta_0 h_{t-1}, \quad t = 1, ..., n, \tag{1.1}$$

with initial values $y_0 \in \mathcal{R}$ and $h_0 \geq 0$, where $\alpha_0 > 0$, $\beta_0 \geq 0$, $(y_0, h_0) \neq (0, 0)$, $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables, and η_t is independent of $\{y_j, j < t\}$. Particularly, model (1.1) nests the widely used exponentially weighted moving average (EWMA) model in RiskMetrics, from which the company J.P. Morgan calculates the daily volatility of many assets by this EWMA model; see Longerstaey and Zangari (1996). Clearly, when $E\eta_t^2 < \infty$, model (1.1) can capture the conditional heteroscedasticity of y_t, since $\text{var}(\eta_t)h_t$ designed as the conditional variance of y_t changes over time. Moreover, by letting $s_t^2 = \text{var}(y_t)$, we have $s_t^2 = [\alpha_0 \text{var}(\eta_t) + \beta_0]s_{t-1}^2$ in model (1.1) so that

$$s_t^2 = [\alpha_0 \text{var}(\eta_t) + \beta_0]^{t-1} s_1^2. \tag{1.2}$$

Therefore, y_t in model (1.1) is homoscedastic when $\alpha_0 \text{var}(\eta_t) + \beta_0 = 1$, and it is heteroscedastic with an exponentially decayed (or explosive) variance when $\alpha_0 \text{var}(\eta_t) + \beta_0 < 1$ (or > 1). Obviously, the heteroscedastic structure of y_t in (1.2) is different from the parametric ones presumed in Breusch and Pagan (1979) and White (1980) or the nonparametric ones studied in Dahlhaus (1997), Dahlhaus and Rubba Rao (2006), Engle and Rangel (2008) and many others. Thus, when $E\eta_t^2 < \infty$, model (1.1) provides us a new parametric way to study conditional heteroscedasticity and heteroscedasticity together.

Needless to say, model (1.1) is motivated by the classical GARCH$(1,1)$ model:

$$y_t = \eta_t \sqrt{h_t} \ \text{ and } \ h_t = \omega_0 + \alpha_0 y_{t-1}^2 + \beta_0 h_{t-1}, \quad t = 1, ..., n, \tag{1.3}$$

where all notations inherit from model (1.1) except for $\omega_0 > 0$. Model (1.3) initialized by Engle (1982) and Bollerslev (1986) has become the workhorse of financial applications, and it can be used to describe the volatility dynamics of almost any financial return series; see Engle (2004, p.408). Due to the importance of model (1.3), numerous works were devoted to its probabilistic structure and statistical inference (see, e.g., Francq and Zakoïan (2010) for a comprehensive review), but they all assume the positivity of ω_0. The case that $\omega_0 = 0$ (i.e., model (1.1)) would be meaningful but hardly touched, except for Hafner and Preminger (2015) who have studied an ARCH(1) model with intercept (i.e, model (1.1) with $\beta_0 = 0$). As the second motivation of this paper, it is of interest to fill in the gap from a theoretical viewpoint. Compared with Hafner and Preminger (2015), the technique developed here is much more involved due to the existence of β_0.

The third motivation of this paper comes from the invalidity of prediction in model (1.3) when $\gamma_0 \geq 0$, where γ_0 is the top Lyapunov exponent defined by

$$\gamma_0 = E \log(\beta_0 + \alpha_0 \eta_t^2). \tag{1.4}$$

Bougerol and Picard (1992a) showed that model (1.3) is stationary if and only if $\gamma_0 < 0$. Note that if $E\eta_t = 0$ and $E\eta_t^2 < \infty$, $s_t^2 = \omega_0(E\eta_t^2) + [\alpha_0(E\eta_t^2) + \beta_0]s_{t-1}^2$ in model (1.3). When $\gamma_0 \geq 0$, it implies that $\alpha_0(E\eta_t^2) + \beta_0 > 1$ and y_t in model (1.3) is heteroscedastic with an exponentially explosive variance. However, when $\gamma_0 \geq 0$, so far no consistent estimator is available for ω_0 as shown in Francq and Zakoïan (2012), and hence no prediction can be made in practice. Model (1.1) avoids this dilemma due to the absence of ω_0, and moreover, Section 2 below demonstrates that except for a different scale, its sample path has a similar shape as that of model (1.3) when $\gamma_0 \geq 0$. In view of this, model (1.1) could be more convenient than model (1.3) to study heteroscedasticity.

This paper gives an omnifaceted investigation of model (1.1). First, we obtain that after a suitable renormalization, the limit of the sample path of h_t or $|y_t|$ converges weakly to a geometric Brownian motion regardless of the sign of γ_0. This result makes a sharp difference from those for model (1.3) in Li, Li and Wu (2014) and Li and Wu (2015). From this result, we find that $|y_t|$ diverges to infinity or converges to zero almost surely (a.s.) at an exponential rate according to $\gamma_0 > 0$ or $\gamma_0 < 0$, while $|y_t|$ oscillates randomly between zero and infinity over time when $\gamma_0 = 0$. Following the terminology in Hafner and Preminger (2015), we call model (1.1) *stable* if $\gamma_0 = 0$ and *unstable* otherwise. Second, we study the generalized quasi-maximum likelihood estimator (GQMLE) of unknown parameter $\theta_0 \equiv (\alpha_0, \beta_0)'$. It is shown that the GQMLE is strongly consistent and asymptotically normal in a unified framework. Third, we consider the estimation for γ_0, and propose a t-test to check the model stability (i.e., $\gamma_0 = 0$). Fourth, we propose a portmanteau test for model checking. Simulation studies are carried out to assess the performance of all proposed estimators and tests in finite samples. Finally, applications to the KV-A stock return in Francq and Zakoïan (2012) and three major exchange rate returns during financial crisis in years 2007-2009 demonstrate that a stable ZD-GARCH$(1, 1)$ model is more appropriate to capture heteroscedasticity than a non-stationary GARCH$(1, 1)$ model.

The remainder of the paper is organized as follows. Section 2 investigates the limit of sample path of y_t in model (1.1). Section 3 studies the GQMLE with its asymptotics. Section 4 presents the estimation and test for γ_0. Section 5 proposes a portmanteau test for model checking. Simulation results are reported in Section 6. Empirical examples are given in Section 7. Concluding remarks and discussions are offered in Section 8. All of proofs are relegated to the Appendix.

2. SAMPLE PATH PROPERTIES

This section studies the sample path properties of renormalized h_t and $|y_t|$ in model (1.1). From model (1.1), we have $\log h_t = \log h_{t-1} + \log(\beta_0 + \alpha_0 \eta_{t-1}^2)$, and hence

$$\log h_t = \sum_{i=0}^{t-1} \log(\beta_0 + \alpha_0 \eta_i^2) + \log h_0.$$

Then, the theorem below follows directly from Donsker's Theorem in Billingsley (1999) (Theorem 8.2 on p.90).

THEOREM 2.1. *Suppose that* (i) $\{\eta_t\}$ *is a sequence of i.i.d. random variables with* $P(\eta_t = 0) = 0$ *and* $\sigma_{\gamma_0}^2 \equiv \mathrm{var}[\log(\beta_0 + \alpha_0 \eta_t^2)] \in (0, \infty)$*;* (ii) h_0 *is a positive random variable and independent of* $\{\eta_t : t \geq 1\}$*. Then, as* $n \to \infty$*,*

$$\frac{h_{[ns]}^{1/\sqrt{n}}}{\exp(s\gamma_0\sqrt{n})} \Longrightarrow \exp(\sigma_{\gamma_0}\mathbb{B}(s)) \quad \textit{in } \mathbb{D}[0, 1],$$

where γ_0 *is defined in* (1.4)*, '*$\Longrightarrow$*' stands for weak convergence,* $\mathbb{B}(s)$ *is a standard Brownian motion on* $[0,1]$*, and* $\mathbb{D}[0,1]$ *is the space of functions defined on* $[0,1]$*, which are right continuous and have left limits, endowed with the Skorokhod topology.*

Furthermore, it follows that as $n \to \infty$*,*

$$\frac{|y_{[ns]}|^{2/\sqrt{n}}}{\exp(s\gamma_0\sqrt{n})} \Longrightarrow \exp(\sigma_{\gamma_0}\mathbb{B}(s)) \quad \textit{in } \mathbb{D}[0,1].$$

Remark 2.1. Similar to Li and Wu (2015), the condition that η_t is i.i.d. in Theorem 2.1 can be relaxed to the one that η_t is strictly stationary and ergodic with $\{\log(\beta_0 + \alpha_0\eta_t^2)\}$ satisfying a suitable invariance principle.

Remark 2.2. For model (1.3), Li and Wu (2015) proved that as $n \to \infty$,

$$\frac{|y_{[ns]}|^{2/\sqrt{n}}}{\exp(s\gamma_0\sqrt{n})} \Longrightarrow \begin{cases} \infty & \text{in } \mathbb{D}[0,1], \text{ if } \gamma_0 < 0, \\ \exp\left(\sigma_{\gamma_0} \max\limits_{0\le\tau\le s} \mathbb{B}(\tau)\right) & \text{in } \mathbb{D}[0,1], \text{ if } \gamma_0 = 0, \\ \exp(\sigma_{\gamma_0}\mathbb{B}(s)) & \text{in } \mathbb{D}[0,1], \text{ if } \gamma_0 > 0. \end{cases} \tag{2.1}$$

Thus, the above limiting result varies according to the value of γ_0, and this is unlike the one in Theorem 2.1. Intuitively, when $\gamma_0 \ge 0$, the results in Theorem 2.1 and (2.1) are less helpful for us to make a useful formal test for hypotheses:

$$H_0 : \omega_0 = 0 \quad \text{v.s.} \quad H_1 : \omega_0 > 0. \tag{2.2}$$

Theorem 2.1 has two direct implications. First, it implies that y_t in model (1.1) is always non-stationary. This is not the case for model (1.3), which is non-stationary if and only if $\gamma_0 \ge 0$ (see, e.g., Bougerol and Picard (1992a)). Second, it indicates that the sample path property of y_t in model (1.1) depends on the sign of γ_0, and this is also the case for model (1.3) as shown in Nelson (1990), Francq and Zakoïan (2012), and Li, Li and Wu (2014). Precisely, we can see that $|y_t|$ in model (1.1) oscillates randomly between zero and infinity over time when $\gamma_0 = 0$, while $|y_t|$ either converges to zero or diverges to infinity a.s. as $t \to \infty$, according to the case that $\gamma_0 < 0$ or $\gamma_0 > 0$, respectively. In this sense, model (1.1) is *stable* if $\gamma_0 = 0$, and *unstable* otherwise; see also Hafner and Preminger (2015). To further illustrate it, Fig.1 depicts one sample path of $\{y_t\}_{t=1}^{200}$ from model (1.1) with $\eta_t \sim \mathcal{N}(0,1)$, $\beta_0 = 0.7$, and $\alpha_0 = 0.3$ (i.e., $\gamma_0 < 0$), 0.388 (i.e., $\gamma_0 = 0$), or 0.5 (i.e., $\gamma_0 > 0$), respectively. Under the same setting, the sample path of $\{y_t\}_{t=1}^{200}$ from model (1.3) with $\omega_0 = 0.001$ or 1 is also plotted as a comparison. Fig.1 confirms the conclusion drawn from Theorem 2.1 above and Theorems 2.1-2.2 in Li, Li and Wu (2014), and most importantly, it exhibits that when $\gamma_0 \geq 0$, apart from a larger scale, the sample path of y_t from model (1.3) has a similar shape as the one from model (1.1). This may suggest that when $\gamma_0 \geq 0$, it is difficult to examine hypotheses in (2.2), since ω_0 only reflects the scale of y_t in model (1.3). Hafner and Preminger (2015) suggested a bootstrap-assisted likelihood ratio (LR) test for this purpose, however our simulation study (not reported here but are available from us) shows that this LR test does not have satisfactory power when $\gamma_0 \geq 0$. Thus, how to test hypotheses in (2.2) remains a challenging open problem.

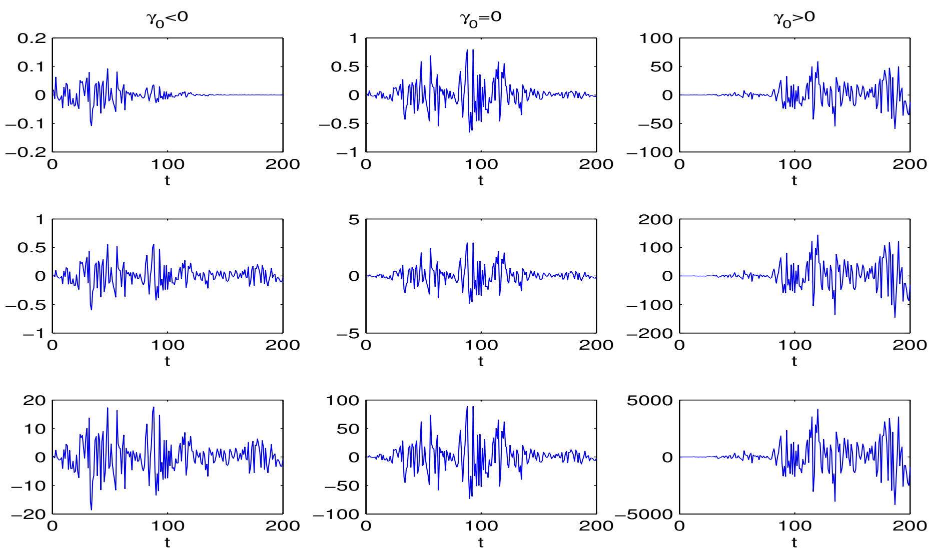

Fig. 1. One sample path $\{y_t\}_{t=1}^{200}$, where the columns from left to right correspond to the cases that $\gamma_0 < 0$, $\gamma_0 = 0$, and $\gamma_0 > 0$, respectively, and the rows from top to bottom correspond to model (1.1), model (1.3) with $\omega_0 = 0.001$, and model (1.3) with $\omega_0 = 1$, respectively.

3. THE GQMLE AND ITS ASYMPTOTICS

Let $\theta \equiv (\alpha, \beta)' \in \Theta$ be the unknown parameter of model (1.1) with the true parameter $\theta_0 = (\alpha_0, \beta_0)' \in \Theta$, where Θ is the parametric space. Assume the data sample $\{y_1, ..., y_n\}$ is from model (1.1). Then, as in Francq and Zakoïan (2013a), the generalized quasi-maximum likelihood estimator (GQMLE) of θ_0 is defined by

$$\widehat{\theta}_{n,r} = (\widehat{\alpha}_{n,r}, \widehat{\beta}_{n,r})' = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \ell_{t,r}(\theta),$$

where $r \geq 0$,

$$\ell_{t,r}(\theta) = \begin{cases} \log\{\sigma_t^r(\theta)\} + \frac{|y_t|^r}{\sigma_t^r(\theta)}, & \text{if } r > 0, \\ \{\log|y_t| - \log\sigma_t(\theta)\}^2, & \text{if } r = 0, \end{cases}$$

and $\sigma_t^2(\theta)$ is recursively defined by

$$\sigma_t^2(\theta) = \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\theta), \quad t = 1, ..., n,$$

with the initial values y_0 and $\sigma_0^2(\theta)$. Hereafter, we set $y_0 = y_0^*$ (a user-chosen non-zero constant) and $\sigma_0^2(\theta) = 0$ without loss of generality. In applications, we can always choose y_0^* be the first nonzero observation, and then do estimation for the remaining ones.

The non-negative user-chosen number r involved in $\widehat{\theta}_{n,r}$ indicates the used estimation method. Particularly, when $r = 2$, $\widehat{\theta}_{n,r}$ reduces to the Gaussian QMLE; and when $r = 1$, $\widehat{\theta}_{n,r}$ reduces to the Laplacian QMLE. Simulation studies in Section 6 imply that we shall choose a small (or large) value of r when η_t is heavy-tailed (or light-tailed).

To obtain the asymptotic property of $\widehat{\theta}_{n,r}$, we give two assumptions below. Assumption 3.1(i) is a regular condition for ARCH-type models, and Assumption 3.1(ii) is the identification condition for $\widehat{\theta}_{n,r}$; see, e.g., Francq and Zakoïan (2013a). Assumption 3.2 is commonly used in nonstationary ARCH-type models; see, e.g., Francq and Zakoïan (2012, 2013b).

Assumption 3.1. (i) η_t is i.i.d. and η_t^2 is non-degenerate with $P(\eta_t = 0) = 0$; (ii) $E|\eta_t|^r = 1$ when $r > 0$, and $E \log|\eta_t| = 0$ when $r = 0$.

Assumption 3.2. The parametric space $\Theta \equiv \{\theta : \alpha > 0,\ 0 \le \beta < e^{\gamma_0}\}$ is compact.

Let

$$\kappa_r = \begin{cases} \frac{4[E|\eta_t|^{2r}-1]}{r^2}, & \text{if } r > 0, \\ 4E[(\log|\eta_t|)^2], & \text{if } r = 0. \end{cases}$$

The following two theorems state the consistency and asymptotic normality of $\widehat{\theta}_{n,r}$, respectively.

THEOREM 3.1. *If Assumptions* 3.1-3.2 *hold, then* $\widehat{\theta}_{n,r} \to \theta_0$ *a.s. as* $n \to \infty$.

THEOREM 3.2. *If Assumptions* 3.1-3.2 *hold,* $\kappa_r < \infty$ *and* θ_0 *is an interior point of* Θ*, then*

$$\sqrt{n}(\widehat{\theta}_{n,r} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \kappa_r \mathcal{I}^{-1}) \quad \text{as } n \to \infty,$$

where $\xrightarrow{d}$ *stands for convergence in distribution, and*

$$\mathcal{I} = \begin{pmatrix} \frac{1}{\alpha_0^2} & \frac{\nu_1}{\alpha_0\beta_0(1-\nu_1)} \\ \\ \frac{\nu_1}{\alpha_0\beta_0(1-\nu_1)} & \frac{(1+\nu_1)\nu_2}{\beta_0^2(1-\nu_1)(1-\nu_2)} \end{pmatrix} \quad \text{with} \quad \nu_i = E\left(\frac{\beta_0}{\beta_0 + \alpha_0\eta_t^2}\right)^i \text{ for } i = 1, 2.$$

Remark 3.1. From two preceding theorems, we have a unified framework on the consistency and asymptotic normality of the GQMLE in model (1.1), regardless of the sign of γ_0. However, this is not the case for the Gaussian QMLE in model (1.3); see, e.g., Jensen and Rahbek (2004a,b) and Francq and Zakoïan (2012). Particularly, it is worth noting that when $\gamma_0 = 0$, an additional assumption on the distribution of η_t, which is not necessary for us, is needed for Francq and Zakoïan (2012) to establish the asymptotic normality of the Gaussian QMLE. However, this additional assumption is hard to be verified even for $\eta_t \sim \mathcal{N}(0, 1)$.

Remark 3.2. By Theorem 3.2, we can do statistical inference on θ_0 (e.g., the Wald test for the linear constraint of θ_0). To accomplish it, we need to estimate both κ_η and $\mathcal{I}$. Denote the residual $\widehat{\eta}_{t,r} \equiv y_t/\widehat{\sigma}_{t,r}$, where $\widehat{\sigma}_{t,r}(>0)$ is recursively calculated by

$$\widehat{\sigma}^2_{t,r} = \widehat{\alpha}_{n,r} y^2_{t-1} + \widehat{\beta}_{n,r}\widehat{\sigma}^2_{t-1,r}, \quad t = 1, ..., n,$$

with $y_0 = y^*_0$ and $\widehat{\sigma}^2_{0,r} = 0$. Then, we can consistently estimate κ_r and $\mathcal{I}$ by their sample counterparts based on the residuals $\{\widehat{\eta}_{t,r}\}$.

Remark 3.3. In Theorem 3.2, θ_0 is required to be an interior point of Θ. If θ_0 can be on the boundary of Θ (e.g., $\beta_0 = 0$), we need the condition of $E(\eta_t^{-4}) < \infty$ for Lemma A.3 in the Appendix, so that under the conditions of Theorem 3.2 and the mild condition that Θ contains a hypercube, the similar argument as Francq and Zakoïan (2007) shows that

$$\sqrt{n}(\widehat{\theta}_{n,r} - \theta_0) \xrightarrow{d} \lambda^\Lambda \quad \text{as } n \to \infty,$$

where $\lambda^\Lambda = \left(\mathcal{Z}_1 - \frac{\alpha_0\nu_1}{\beta_0(1-\nu_1)}\mathcal{Z}_2 I(\mathcal{Z}_2 < 0), \mathcal{Z}_2 - \mathcal{Z}_2 I(\mathcal{Z}_2 < 0)\right)'$ and $(\mathcal{Z}_1, \mathcal{Z}_2)' \sim \mathcal{N}(0, \kappa_r \mathcal{I}^{-1})$. However, the condition that $E(\eta_t^{-4}) < \infty$ fails even for $\eta_t \sim \mathcal{N}(0,1)$, and hence any statistical inference on θ_0, including the estimation, Wald test, and LR test, is hardly useful in this case.

Since Theorem 3.2 rules out the case that $\beta_0 = 0$, we shall consider this special case independently. When $\beta_0 = 0$, model (1.1) becomes

$$y_t = \eta_t\sqrt{h_t} \quad \text{and} \quad h_t = \alpha_0 y^2_{t-1}, \quad t = 1, ..., n. \tag{3.1}$$

This is the ARCH(1) model without intercept in Hafner and Preminger (2015). Denote $\Theta_\alpha \subset (0, \infty)$ be the parametric space of model (3.1). Then, the GQMLE of α_0 in model (3.1) is

$$\widetilde{\alpha}_{n,r} = \begin{cases} \arg\min_{\alpha\in\Theta_\alpha} \sum_{t=2}^n \left[\frac{r}{2}\log\{\alpha y^2_{t-1}\} + \frac{1}{\alpha^{r/2}}\frac{|y_t|^r}{|y_{t-1}|^r}\right], & \text{if } r > 0, \\ \arg\min_{\alpha\in\Theta_\alpha} \sum_{t=2}^n \left[\log|y_t| - \frac{1}{2}\log(\alpha y^2_{t-1})\right]^2, & \text{if } r = 0. \end{cases}$$

Without the compactness of Θ_α, the asymptotical normality of $\widetilde{\alpha}_{n,r}$ is given below:

THEOREM 3.3. *If Assumption* 3.1 *holds and* $\kappa_r < \infty$*, then*

$$\sqrt{n}(\widetilde{\alpha}_{n,r} - \alpha_0) \xrightarrow{d} \mathcal{N}(0, \kappa_r \alpha_0^2) \quad \textit{as } n \to \infty.$$

Remark 3.4. The proof of Theorem 3.3 is much simpler than that of Theorem 3.2, since an explicit expression of $\widetilde{\alpha}_{n,r}$ is available in this case. Particularly, when $r = 2$, Hafner and Preminger (2015) have obtained the same result but in an indirect way.

Remark 3.5. Besides the GQMLE, one may consider many other estimation methods for model (1.1); see, e.g., Peng and Yao (2003), Berkes and Horváth (2004), Fan, Qi and Xiu (2014), and Zhu and Li (2015a). Moreover, the condition that $\kappa_r < \infty$ is necessary for the asymptotic normality of the GQMLE, one may be of interest to study the GQMLE when $\kappa_r = \infty$; see, e.g., Hall and Yao (2003). These are two promising directions for the future study.

As shown in Remark 3.3, the Wald and LR tests are not suitable to detect whether $\beta_0 = 0$ in model (1.1). One may try the score test as in Engle (1982) for this purpose. However, this is not suitable as well. To see it clearly, we consider the limiting distribution of the score $\sqrt{n}\,\Pi_{n,r}(\widetilde{\theta}_{n,r})$ under the constraint that $\beta_0 = 0$, where

$$\Pi_{n,r}(\theta) = \frac{1}{n}\sum_{t=2}^{n} \frac{\partial \ell_{t,r}(\theta)}{\partial \beta} \quad \text{and} \quad \widetilde{\theta}_{n,r} = (\widetilde{\alpha}_{n,r}, 0).$$

A direct calculation shows that

$$\begin{aligned}\sqrt{n}\,\Pi_{n,r}(\widetilde{\theta}_{n,r}) = {} & \left(\frac{r}{2\alpha_0 n}\sum_{t=2}^{n}\frac{1}{\eta_{t-1}^2 \widetilde{\alpha}_{n,r}^{r/2}}\right)[\sqrt{n}(\widetilde{\alpha}_{n,r}^{r/2} - \alpha_0^{r/2})] \\ & + \left(\frac{r\alpha_0^{r/2}}{2\alpha_0 \widetilde{\alpha}_{n,r}^{r/2}}\right)\left[\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\frac{1 - |\eta_t|^r}{\eta_{t-1}^2}\right] \quad \text{for } r > 0.\end{aligned}$$

Hence, the limiting distribution of $\sqrt{n}\,\Pi_{n,r}(\widetilde{\theta}_{n,r})$ exists only when $E(\eta_t^{-4}) < \infty$, which fails even for $\eta_t \sim \mathcal{N}(0,1)$. Similarly, the conclusion holds when $r = 0$. In Section 5, a portmanteau test is available to detect whether $\beta_0 = 0$ in model (1.1).

4. Inference of the Lyapunov exponent

Generally, γ_0 plays a key role in determining stationarity or stability of nonlinear time series models. In model (1.3), there exists a strictly stationary solution if and only if $\gamma_0 < 0$; see Nelson (1990) and Bougerol and Picard (1992a,b). Similarly, γ_0 plays an equally important role in determining the stability of model (1.1). Thus, it is necessary to do statistical inference for γ_0.

From the definition of γ_0 in (1.4), a natural plug-in estimator of γ_0 is defined as

$$\widehat{\gamma}_{n,r} = \frac{1}{n}\sum_{t=1}^{n} \log(\widehat{\beta}_{n,r} + \widehat{\alpha}_{n,r}\widehat{\eta}_{t,r}^2).$$

Particularly, $\widehat{\gamma}_{n,r}$ admits a simple form for model (3.1):

$$\widehat{\gamma}_{n,r} = \frac{1}{n}\sum_{t=1}^{n} \left[\log(y_t^2) - \log(y_{t-1}^2)\right] = \frac{2}{n}\left(\log|y_n| - \log|y_0|\right).$$

Interestingly, the preceding definition of $\widehat{\gamma}_{n,r}$ is independent to the estimation method, and it has been used in Hafner and Preminger (2015). Furthermore, we have the following theorem:

THEOREM 4.1. *If the conditions in Theorem* 3.2 *are satisfied, then as* $n \to \infty$*,*

(i) $\widehat{\gamma}_{n,r} \to \gamma_0$ *in probability*;

(ii) $\sqrt{n}(\widehat{\gamma}_{n,r} - \gamma_0) \xrightarrow{d} \mathcal{N}(0, \sigma_{\gamma_0}^2)$,

where $\sigma_{\gamma_0}^2$ *is defined in Theorem* 2.1. *Moreover, if Assumption* 3.1*(i) holds and* $\sigma_{\gamma_0}^2 \in (0,\infty)$*, the same conclusion holds for model (*3.1*).*

Remark 4.1. Although $\widehat{\gamma}_{n,r}$ depends on r (i.e., the estimation method), its asymptotic variance is free of that. Intuitively, this suggests that the performance of $\widehat{\gamma}_{n,r}$ and its related stable test defined below is less affected by the estimation method. Simulation studies in Section 6 will confirm this statement.

Since model (1.1) is stable if and only if $\gamma_0 = 0$, it is of interest to consider hypotheses:

$$H_0 : \gamma_0 = 0 \quad \text{against} \quad H_1 : \gamma_0 \neq 0. \tag{4.1}$$

From Theorem 4.1, we propose a t-type test statistic $T_{n,r}$ to detect H_0 in (4.1), where

$$T_{n,r} = \sqrt{n}\frac{\widehat{\gamma}_{n,r}}{\widehat{\sigma}_{\eta,r}}$$

with $\widehat{\sigma}^2_{\eta,r} = \frac{1}{n}\sum_{t=1}^{n}\{\log(\widehat{\beta}_{n,r} + \widehat{\alpha}_{n,r}\widehat{\eta}^2_{t,r})\}^2 - \widehat{\gamma}^2_{n,r}$. Note that for model (3.1), $\widehat{\sigma}^2_{\eta,r}$ admits a simple form: $\widehat{\sigma}^2_{\eta,r} = \frac{4}{n}\sum_{t=1}^{n}\{\log|y_t| - \log|y_{t-1}|\}^2 - \frac{4}{n^2}\left(\log|y_n| - \log|y_0|\right)^2$, and hence $T_{n,r}$ is independent to the estimation method. Under H_0, it is not hard to see that $T_{n,r} \xrightarrow{d} \mathcal{N}(0,1)$ as $n \to \infty$. So, at the significance level $\underline{\alpha} \in (0,1)$, H_0 in (4.1) is rejected if $|T_{n,r}| > |\Phi^{-1}(\underline{\alpha}/2)|$, where $\Phi(\cdot)$ is the cdf of $\mathcal{N}(0,1)$; otherwise, it is not rejected.

5. Model diagnostic checking

This section proposes a portmanteau test to check the adequacy of model (1.1). We first define the lag-k autocorrelation function (ACF) of the s-th power of the absolute residuals $\{|\widehat{\eta}_{t,r}|^s\}$ as

$$\widehat{\rho}_{r,s}(k) = \frac{\sum_{t=k+1}^{n}(|\widehat{\eta}_{t,r}|^s - \widehat{a}_{r,s})(|\widehat{\eta}_{t-k,r}|^s - \widehat{a}_{r,s})}{\sum_{t=1}^{n}(|\widehat{\eta}_{t,r}|^s - \widehat{a}_{r,s})^2},$$

where $r \geq 0$, $s > 0$, k is a positive integer, and

$$\widehat{a}_{r,s} = \frac{1}{n}\sum_{t=1}^{n}|\widehat{\eta}_{t,r}|^s.$$

Next, we introduce the following notations:

$$a_s = E|\eta_t|^s, \quad b_s = \mathrm{var}(|\eta_t|^s), \quad p_s(k) = \left(0,\ \frac{\nu_1^{k-1}}{1-\nu_1}E\left[\frac{|\eta_t|^s - a_s}{\beta_0 + \alpha_0\eta_t^2}\right]\right),$$

$$V_{r,s} = \begin{cases} (p_s'(1), p_s'(2), \cdots, p_s'(m))'\left(\frac{2}{r}\mathcal{I}^{-1}\right), & \text{if } r > 0, \\ (p_s'(1), p_s'(2), \cdots, p_s'(m))'\left(2\mathcal{I}^{-1}\right), & \text{if } r = 0, \end{cases}$$

$$W_{r,s} = \begin{cases} (p_s'(1), p_s'(2), \cdots, p_s'(m))'E[(|\eta_t|^s - a_s)(|\eta_t|^r - 1)], & \text{if } r > 0, \\ (p_s'(1), p_s'(2), \cdots, p_s'(m))'E[(|\eta_t|^s - a_s)\log|\eta_t|], & \text{if } r = 0. \end{cases}$$

Let m be a given positive integer. The following theorem is crucial to derive the limiting distribution of our portmanteau test.

THEOREM 5.1. *Suppose that $b_s < \infty$, $b_r < \infty$, and $\|W_{r,s}\| < \infty$ for given $r \geq 0$ and $s > 0$. If model (1.1) is correctly specified and the conditions in Theorem 3.2 hold, then*

$$\sqrt{n}(\widehat{\rho}_{r,s}(1), ..., \widehat{\rho}_{r,s}(m))' \xrightarrow{d} \mathcal{N}(0, \Sigma_{r,s}(m)) \quad as\ n \to \infty,$$

where

$$\Sigma_{r,s}(m) = \frac{1}{b_s^2}(\mathbf{I}_m, -V_{r,s}) \begin{pmatrix} b_s^2\mathbf{I}_m, \, W_{r,s} \\ W'_{r,s} \quad b_r\mathcal{I} \end{pmatrix} (\mathbf{I}_m, -V_{r,s})'$$

and $\mathbf{I}_m$ is the $m \times m$ identity matrix. Moreover, if model (3.1) is correctly specified and Assumption 3.1(i) holds, then the same conclusion holds with $\Sigma_{r,s}(m) = \mathbf{I}_m$.

For model (1.1) with $\beta_0 > 0$, let $\widehat{\Sigma}_{r,s}(m)$ be the sample counterpart of $\Sigma_{r,s}(m)$, based on the residuals $\{\widehat{\eta}_{t,r}\}_{t=1}^n$; and for model (1.1) with $\beta_0 = 0$ (i.e., model (3.1)), let $\widehat{\Sigma}_{r,s}(m) = \mathbf{I}_m$. Then, our portmanteau test is defined by

$$Q_{r,s}(m) := n(n+2)\left(\frac{\widehat{\rho}_{r,s}(1)}{n-1}, ..., \frac{\widehat{\rho}_{r,s}(m)}{n-m}\right)[\widehat{\Sigma}_{r,s}(m)]^{-1}\left(\frac{\widehat{\rho}_{r,s}(1)}{n-1}, ..., \frac{\widehat{\rho}_{r,s}(m)}{n-m}\right)'.$$

Here, m is generally taken 6 or 12 in applications. When $r = s = 2$, $Q_{r,s}(m)$ is defined in the same way as the portmanteau test in Li and Mak (1994). When $r = 2$ (or 1) and $s = 1$, $Q_{r,s}(m)$ is analogous to the portmanteau test $Q^2(M)$ in Li and Li (2005) (or Q_r in Li and Li (2008)). We relax the choices of r and s so that $Q_{r,s}(m)$ with small (or large) r and s is expected to have a good performance when η_t is heavy-tailed (or light-tailed). Our portmanteau test $Q_{r,s}(m)$ is the first formal diagnostic checking tool for non-stationary ARCH-type models in the literature. For more discussions on the diagnostic checking of stationary ARCH-type models, we refer to Li (2004), Escanciano (2007, 2008), Ling and Tong (2011), and Chen and Zhu (2015).

By Theorem 5.1, we have $Q_{r,s}(m) \xrightarrow{d} \chi_m^2$ as $n \to \infty$. Thus, at the significance level $\underline{\alpha} \in (0, 1)$, we conclude that model (1.1) is not adequate if $Q_{r,s}(m) > \Psi_m^{-1}(1 - \underline{\alpha})$, where $\Psi_d(\cdot)$ is the cdf of χ_d^2; otherwise, it is adequate.

In the end, it is worth noting that the estimation effect does not affect the limiting distribution of $Q_{r,s}(m)$ for model (3.1), and this is different from most of portmanteau tests in times series analysis; see, e.g., Zhu and Li (2015b), Zhu (2016) and references therein for more discussions in this context. For model (1.1) with $\beta_0 > 0$, the estimation effect is involved into the limiting distribution of $Q_{r,s}(m)$, but interestingly, if $\alpha_0/\beta_0 \approx 0$ (as often observed in applications), it is not hard to see that $\widehat{\Sigma}_{r,s}(m) \approx \mathbf{I}_m$ due the the fact that $V_{r,s} \approx 0$, and hence the estimation effect is negligible in this case.

6. Simulation studies

In this section, we first assess the finite-sample performance of $\widehat{\theta}_{n,r}$, $\widehat{\gamma}_{n,r}$, and $T_{n,r}$. We generate 1000 replications from the following ZD-GARCH$(1,1)$ model:

$$y_t = \eta_t \sqrt{h_t}, \quad h_t = \alpha_0 y_{t-1}^2 + 0.9 h_{t-1}, \tag{6.1}$$

where η_t is taken as $\mathcal{N}(0,1)$, the standardized Student's t_5 (st_5) or the standardized Student's t_3 (st_3) such that $E\eta_t^2 = 1$. Here, we fix $\beta_0 = 0.9$, and choose α_0 as in Table 1, where the values of α_0 correspond to the cases of $\gamma_0 > 0$, $\gamma_0 = 0$, and $\gamma_0 < 0$, respectively. For the indicator r, we choose it to be 2, 1, 0.5, and 0. Since each GQMLE has a different identification condition, $\widehat{\theta}_{n,r}$ has to be re-scaled for θ_0 in model (6.1), and it is defined as

$$\widehat{\theta}_{n,r} = \left(\frac{\overline{\alpha}_{n,r}}{(E|\eta_t|^r)^{2/r}}, \overline{\beta}_{n,r} \right) \text{ for } r > 0 \quad \text{and} \quad \widehat{\theta}_{n,r} = \left(\frac{\overline{\alpha}_{n,r}}{\exp(2E \log |\eta_t|)}, \overline{\beta}_{n,r} \right) \text{ for } r = 0,$$

where $\overline{\theta}_{n,r}$ is the GQMLE calculated from the data sample, and the true values of $(E|\eta_t|^r)^{2/r}$ and $\exp(2E\log|\eta_t|)$ are used.

Tables 2-4 report the empirical bias, empirical standard deviation (SD), and the average of the asymptotic standard deviations (AD) of $\widehat{\theta}_{n,r}$ and $\widehat{\gamma}_{n,r}$ when the sample size $n = 500$ and 1000. The ADs of $\widehat{\theta}_{n,r}$ and $\widehat{\gamma}_{n,r}$ are evaluated from the asymptotic covariances in Theorems 3.2 and 4.1, respectively. From these tables, we find that (i) except $\widehat{\theta}_{n,2}$ in the case of $\eta_t \sim \mathrm{st}_3$, all GQMLEs

Table 1. *The values of the pair* (α_0, γ_0) *when* $\beta_0 = 0.9$.

$\eta_t \sim \mathcal{N}(0,1)$		$\eta_t \sim \mathrm{st}_5$		$\eta_t \sim \mathrm{st}_3$	
α_0	γ_0	α_0	γ_0	α_0	γ_0
0.1	-0.0082	0.1	-0.0152	0.1	-0.0300
0.1096508	0.0000	0.1201453	0.0000	0.1508275	0.0000
0.2	0.0706	0.2	0.0548	0.2	0.0263

Table 2. *Summary for* $\widehat{\theta}_{n,r}$ *and* $\widehat{\gamma}_{n,r}$ *when* $\gamma_0 < 0$.

			$r=2$			$r=1$			$r=0.5$			$r=0$		
η_t	n		$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$
$\mathcal{N}(0,1)$	500	Bias	-0.0051	0.0047	0.0001	-0.0047	0.0047	0.0001	-0.0028	0.0036	0.0000	-0.0021	0.0038	0.0001
		SD	0.0235	0.0199	0.0055	0.0246	0.0208	0.0055	0.0278	0.0233	0.0055	0.0373	0.0310	0.0056
		AD	**0.0227**	**0.0187**	0.0052	0.0244	0.0200	0.0052	0.0278	0.0226	0.0053	0.0362	0.0292	0.0053
	1000	Bias	-0.0032	0.0030	0.0001	-0.0028	0.0028	0.0001	-0.0011	0.0019	0.0001	-0.0007	0.0021	0.0001
		SD	0.0167	0.0139	0.0039	0.0176	0.0147	0.0039	0.0199	0.0165	0.0039	0.0262	0.0215	0.0039
		AD	**0.0163**	**0.0134**	0.0038	0.0175	0.0144	0.0038	0.0199	0.0162	0.0038	0.0261	0.0211	0.0038
st_5	500	Bias	-0.0045	0.0031	-0.0002	-0.0048	0.0044	-0.0002	-0.0025	0.0030	-0.0002	-0.0012	0.0022	-0.0002
		SD	0.0391	0.0268	0.0066	0.0268	0.0198	0.0066	0.0278	0.0205	0.0066	0.0345	0.0254	0.0066
		AD	0.0321	0.0229	0.0064	**0.0258**	**0.0186**	0.0064	0.0271	0.0193	0.0065	0.0336	0.0239	0.0065
	1000	Bias	-0.0026	0.0015	0.0002	-0.0018	0.0020	0.0002	-0.0005	0.0014	0.0002	-0.0007	0.0018	0.0002
		SD	0.0264	0.0189	0.0047	0.0191	0.0138	0.0047	0.0197	0.0141	0.0047	0.0247	0.0178	0.0047
		AD	0.0239	0.0172	0.0047	**0.0188**	**0.0135**	0.0047	0.0194	0.0138	0.0047	0.0239	0.0170	0.0047
st_3	500	Bias	2.2728	-0.0015	0.0001	-0.0054	0.0043	0.0000	-0.0045	0.0033	0.0000	-0.0019	0.0028	0.0000
		SD	71.572	0.0547	0.0080	0.0339	0.0193	0.0080	0.0279	0.0166	0.0080	0.0325	0.0189	0.0080
		AD	0.8192	0.0276	0.0073	0.0300	0.0169	0.0071	**0.0273**	**0.0155**	0.0072	0.0325	0.0181	0.0073
	1000	Bias	0.0078	-0.0022	-0.0001	-0.0030	0.0023	-0.0001	-0.0023	0.0015	-0.0001	0.0000	0.0011	-0.0001
		SD	0.1333	0.0382	0.0055	0.0245	0.0138	0.0055	0.0197	0.0114	0.0055	0.0225	0.0128	0.0055
		AD	0.0495	0.0236	0.0054	0.0220	0.0124	0.0052	**0.0197**	**0.0112**	0.0053	0.0233	0.0130	0.0053

†The smallest value of AD for $\widehat{\theta}_{n,r}$ is in boldface.

have a small bias, and their SDs and ADs are close to each other; (ii) when η_t is light-tailed (i.e., $\eta_t \sim \mathcal{N}(0,1)$), $\widehat{\theta}_{n,2}$ (or $\widehat{\theta}_{n,0}$) is the best (or worst) estimator in terms of the minimized value of

Table 3. *Summary for $\widehat{\theta}_{n,r}$ and $\widehat{\gamma}_{n,r}$ when $\gamma_0 = 0$.*

			$r=2$			$r=1$			$r=0.5$			$r=0$		
η_t	n		$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$
$\mathcal{N}(0,1)$	500	Bias	-0.0050	0.0045	-0.0001	-0.0049	0.0045	-0.0001	-0.0030	0.0035	-0.0001	-0.0021	0.0036	0.0000
		SD	0.0247	0.0205	0.0060	0.0257	0.0214	0.0060	0.0290	0.0240	0.0060	0.0400	0.0329	0.0061
		AD	**0.0242**	**0.0196**	0.0056	0.0260	0.0210	0.0056	0.0295	0.0237	0.0057	0.0385	0.0307	0.0057
	1000	Bias	-0.0029	0.0028	0.0003	-0.0027	0.0029	0.0003	-0.0010	0.0020	0.0003	-0.0007	0.0024	0.0003
		SD	0.0184	0.0152	0.0042	0.0194	0.0160	0.0042	0.0217	0.0176	0.0042	0.0284	0.0228	0.0042
		AD	**0.0174**	**0.0141**	0.0041	0.0187	0.0151	0.0041	0.0212	0.0170	0.0041	0.0278	0.0221	0.0041
st_5	500	Bias	-0.0056	0.0042	0.0004	-0.0054	0.0053	0.0004	-0.0030	0.0041	0.0004	-0.0011	0.0030	0.0004
		SD	0.0440	0.0300	0.0078	0.0316	0.0228	0.0078	0.0322	0.0230	0.0078	0.0425	0.0299	0.0078
		AD	0.0368	0.0253	0.0073	**0.0297**	**0.0206**	0.0073	0.0310	0.0213	0.0074	0.0385	0.0263	0.0075
	1000	Bias	-0.0032	0.0013	-0.0001	-0.0022	0.0021	-0.0001	-0.0004	0.0013	-0.0002	-0.0004	0.0015	-0.0001
		SD	0.0297	0.0207	0.0052	0.0215	0.0153	0.0052	0.0223	0.0157	0.0052	0.0275	0.0193	0.0052
		AD	0.0273	0.0190	0.0053	**0.0215**	**0.0149**	0.0053	0.0223	0.0153	0.0054	0.0275	0.0189	0.0054
st_3	500	Bias	0.0162	-0.0020	0.0008	-0.0045	0.0044	0.0008	-0.0041	0.0037	0.0007	-0.0002	0.0033	0.0007
		SD	0.1958	0.0543	0.0097	0.0472	0.0245	0.0097	0.0408	0.0217	0.0096	0.0491	0.0253	0.0096
		AD	0.0975	0.0364	0.0096	0.0431	0.0219	0.0093	**0.0386**	**0.0199**	0.0094	0.0457	0.0230	0.0095
	1000	Bias	0.3294	-0.0034	0.0003	-0.0029	0.0024	0.0003	-0.0026	0.0019	0.0002	0.0003	0.0019	0.0002
		SD	10.068	0.0526	0.0070	0.0324	0.0164	0.0070	0.0282	0.0147	0.0070	0.0340	0.0174	0.0070
		AD	0.3720	0.0301	0.0070	0.0311	0.0159	0.0067	**0.0276**	**0.0143**	0.0068	0.0326	0.0165	0.0068

†The smallest value of AD for $\widehat{\theta}_{n,r}$ is in boldface.

AD; (iii) when η_t is heavy-tailed (i.e., $\eta_t \sim \mathrm{st}_3$ or st_5), $\widehat{\theta}_{n,1}$ or $\widehat{\theta}_{n,0.5}$ has the smallest value of AD, while $\widehat{\theta}_{n,2}$ has a larger value of AD than $\widehat{\theta}_{n,0}$ if $\eta_t \sim \mathrm{st}_5$, and it is even not applicable if $\eta_t \sim \mathrm{st}_3$; (iv) the performance of $\widehat{\gamma}_{n,r}$ seems to be unchanged for all choices of r, even when $\widehat{\theta}_{n,2}$ is not asymptotically normal in the case of $\eta_t \sim \mathrm{st}_3$.

Fig. 2 plots the power of $T_{n,1}$ in terms of different values of α_0 with $n = 500$ and 1000, where the sizes of $T_{n,1}$ correspond to the choices of α_0 in Table 1 for the case of $\gamma_0 = 0$. Here, we shall mention that the power of $T_{n,r}$ for other choices of r is almost the same as the one of $T_{n,1}$, and

Table 4. *Summary for $\widehat{\theta}_{n,r}$ and $\widehat{\gamma}_{n,r}$ when $\gamma_0 > 0$.*

			$r=2$			$r=1$			$r=0.5$			$r=0$		
η_t	n		$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$	$\widehat{\alpha}_{n,r}$	$\widehat{\beta}_{n,r}$	$\widehat{\gamma}_{n,r}$
$\mathcal{N}(0,1)$	500	Bias	-0.0092	0.0069	-0.0007	-0.0098	0.0076	-0.0007	-0.0076	0.0068	-0.0007	-0.0059	0.0069	-0.0006
		SD	0.0375	0.0275	0.0096	0.0385	0.0283	0.0096	0.0429	0.0312	0.0096	0.0575	0.0413	0.0096
		AD	**0.0364**	**0.0264**	0.0089	0.0389	0.0282	0.0089	0.0441	0.0317	0.0089	0.0576	0.0411	0.0089
	1000	Bias	-0.0036	0.0032	0.0002	-0.0035	0.0034	0.0002	-0.0015	0.0027	0.0002	-0.0018	0.0037	0.0003
		SD	0.0257	0.0190	0.0063	0.0276	0.0205	0.0063	0.0316	0.0233	0.0063	0.0426	0.0313	0.0063
		AD	**0.0263**	**0.0190**	0.0065	0.0281	0.0203	0.0065	0.0319	0.0228	0.0065	0.0416	0.0297	0.0065
st_5	500	Bias	-0.0034	0.0016	0.0003	-0.0046	0.0046	0.0003	-0.0035	0.0046	0.0003	-0.0026	0.0044	0.0003
		SD	0.0614	0.0376	0.0108	0.0459	0.0288	0.0108	0.0476	0.0298	0.0108	0.0613	0.0381	0.0108
		AD	0.0548	0.0336	0.0106	**0.0445**	**0.0274**	0.0105	0.0460	0.0282	0.0105	0.0568	0.0347	0.0105
	1000	Bias	-0.0014	0.0001	0.0002	-0.0015	0.0021	0.0002	-0.0002	0.0019	0.0002	0.0002	0.0020	0.0002
		SD	0.0479	0.0284	0.0076	0.0326	0.0202	0.0076	0.0331	0.0205	0.0076	0.0402	0.0247	0.0076
		AD	0.0412	0.0252	0.0076	**0.0320**	**0.0197**	0.0075	0.0330	0.0201	0.0076	0.0407	0.0248	0.0076
st_3	500	Bias	0.0326	-0.0034	0.0004	-0.0054	0.0049	0.0004	-0.0053	0.0044	0.0004	0.0013	0.0034	0.0004
		SD	0.6593	0.0661	0.0118	0.0638	0.0292	0.0117	0.0514	0.0255	0.0117	0.0615	0.0296	0.0117
		AD	0.1203	0.0418	0.0113	0.0543	0.0256	0.0109	**0.0488**	**0.0233**	0.0110	0.0582	0.0271	0.0111
	1000	Bias	0.0087	-0.0028	-0.0004	-0.0045	0.0027	-0.0004	-0.0043	0.0024	-0.0004	-0.0005	0.0024	-0.0004
		SD	0.1959	0.0487	0.0082	0.0428	0.0202	0.0082	0.0369	0.0177	0.0082	0.0425	0.0201	0.0082
		AD	0.0843	0.0343	0.0080	0.0387	0.0185	0.0078	**0.0347**	**0.0167**	0.0079	0.0411	0.0194	0.0079

†The smallest value of AD for $\widehat{\theta}_{n,r}$ is in boldface.

hence it is not reported here. From Fig. 2, we can see that the size of $T_{n,1}$ is precise especially for large n, and the power of $T_{n,1}$ increasing with n is satisfactory for all choices of η_t. Overall, our proposed estimators ($\widehat{\theta}_{n,r}$ and $\widehat{\gamma}_{n,r}$) and test ($T_{n,r}$) have a good finite-sample performance.

Next, we assess the finite-sample performance of $Q_{r,s}(m)$. We generate 1000 replications from the following higher-order ZD-GARCH model:

$$y_t = \eta_t \sqrt{h_t}, \quad h_t = \alpha_0 y_{t-1}^2 + z_0 y_{t-2}^2 + 0.9 h_{t-1}, \tag{6.2}$$

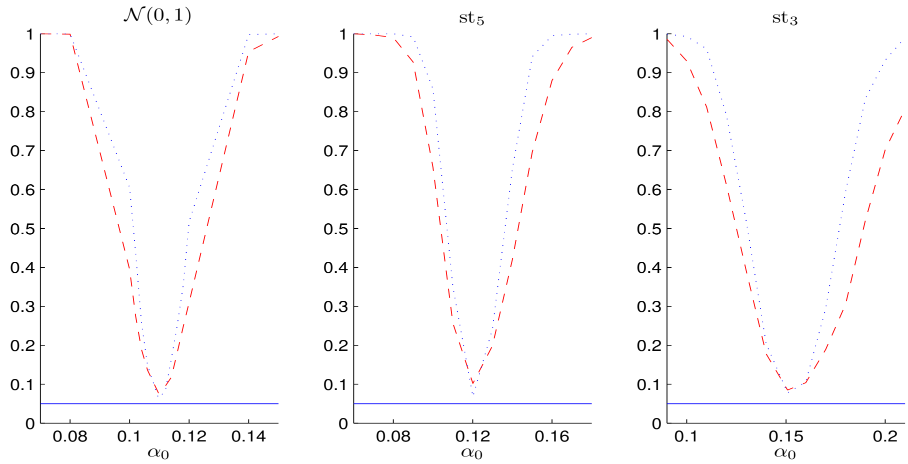

Fig. 2. The power of $T_{n,1}$ for $\eta_t \sim \mathcal{N}(0,1)$ (left panel), st$_5$ (middle panel) and st$_3$ (right panel) in terms of different values of α_0, where the sample size $n = 500$ (dash line) and 1000 (dot line), and the solid line stands for the significance level $\underline{\alpha} = 5\%$.

where η_t and α_0 are chosen as in model (6.1), and $z_0 = 0.0, 0.1, \cdots, 0.7$. For each replication, we use $Q_{r,s}(6)$ to detect whether model (6.1) is adequate to fit the generated data, where $(r,s) = (2,2), (1,2), (0,2), (1,1), (0,1)$ or $(0.5, 0.5)$. Fig. 3 plots the power of $Q_{r,s}(6)$ in terms of different choices of γ_0 (or α_0 equivalently) with $n = 500$ and 1000, and the sizes of $Q_{r,s}(6)$ correspond to the case that $z_0 = 0.0$. From this figure, we can see that (i) each $Q_{r,s}(6)$ has a precise size, although $Q_{r,s}(6)$ with $s = 2$ is not valid in the case of $\eta_t \sim$ st$_3$; (ii) when η_t is light-tailed (e.g., $\eta_t \sim \mathcal{N}(0,1)$), $Q_{r,s}(6)$ with all choices of (r,s) except that $(r,s) = (0.5, 0.5)$ has almost the same power performance; (iii) when η_t is heavy-tailed (e.g., $\eta_t \sim$ st$_5$ or st$_3$), $Q_{r,s}(6)$ with small values of s (e.g., $Q_{1,1}(6)$, $Q_{0,1}(6)$, and $Q_{0.5,0.5}(6)$) has a much better power performance than that with large values of s (e.g., $Q_{2,2}(6)$, $Q_{1,2}(6)$, and $Q_{0,2}(6)$). Thus, $Q_{r,s}(m)$ has the ability to detect the mis-specification of model (1.1) in the higher-order term.

Moreover, we are of interest to see whether $Q_{r,s}(m)$ can detect the mis-specification of model (1.1) in the drift term. We generate 1000 replications from the following GARCH$(1,1)$ model:

$$y_t = \eta_t \sqrt{h_t}, \quad h_t = z_0 + \alpha_0 y_{t-1}^2 + 0.9 h_{t-1}, \tag{6.3}$$

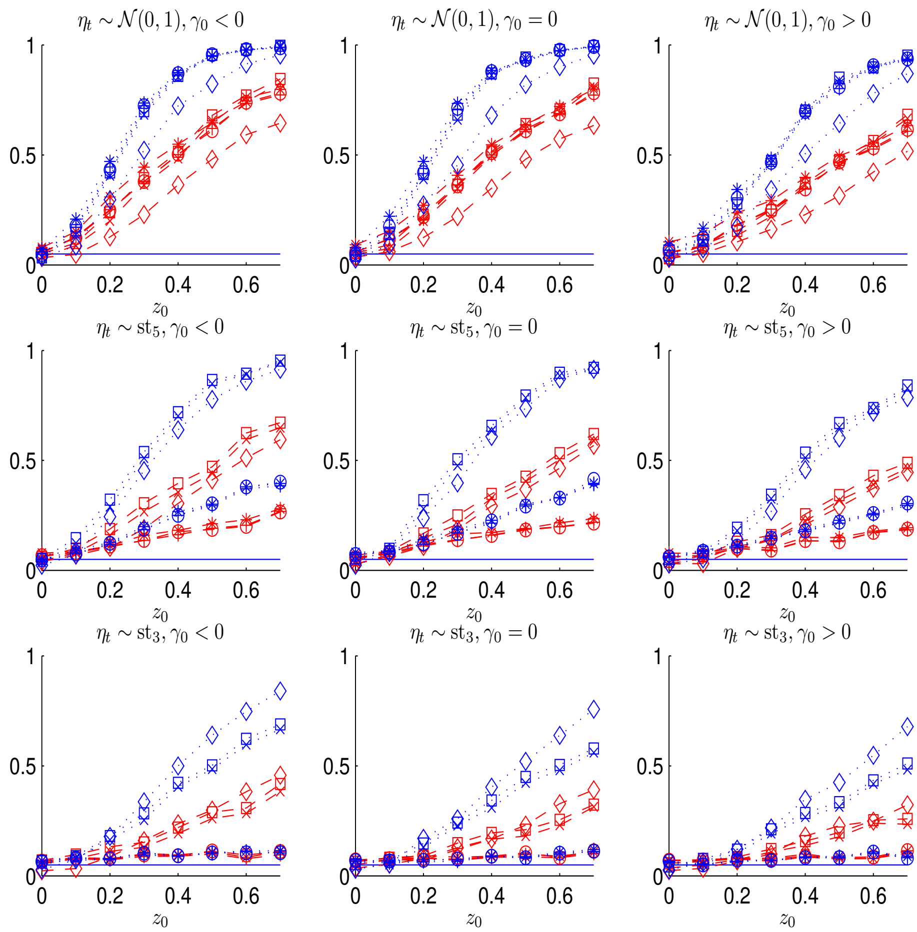

Fig. 3. The power of $Q_{2,2}(6)$ (circle line), $Q_{1,2}(6)$ (plus line), $Q_{0,2}(6)$ (star line), $Q_{1,1}(6)$ (cross line), $Q_{0,1}(6)$ (square line) and $Q_{0.5,0.5}(6)$ (diamond line) for $\eta_t \sim \mathcal{N}(0,1)$ (top panels), st_5 (middle panels), and st_3 (bottom panels) in terms of three different choices of γ_0, where the data sample is generated from model (6.2) with the sample size $n = 500$ (dash line) and 1000 (dot line), and the solid line stands for the significance level $\underline{\alpha} = 5\%$.

where η_t and α_0 are chosen as in model (6.1), and $z_0 = 0.0, 0.02, \cdots, 0.1$. For each replication, we use $Q_{r,s}(6)$ to detect whether model (6.1) is adequate to fit the generated data, where the values of (r, s) are chosen as before. Fig. 4 plots the power of $Q_{r,s}(6)$, and the sizes of $Q_{r,s}(6)$ correspond to the case that $z_0 = 0.0$. From this figure, we can see that (i) each $Q_{r,s}(6)$ has a precise size, although $Q_{r,s}(6)$ with $s = 2$ is not valid in the case of $\eta_t \sim \mathrm{st}_3$; (ii) when η_t is light-tailed (e.g., $\eta_t \sim \mathcal{N}(0,1)$), $Q_{r,s}(6)$ with large values of s (e.g., $s = 2$) has a much better

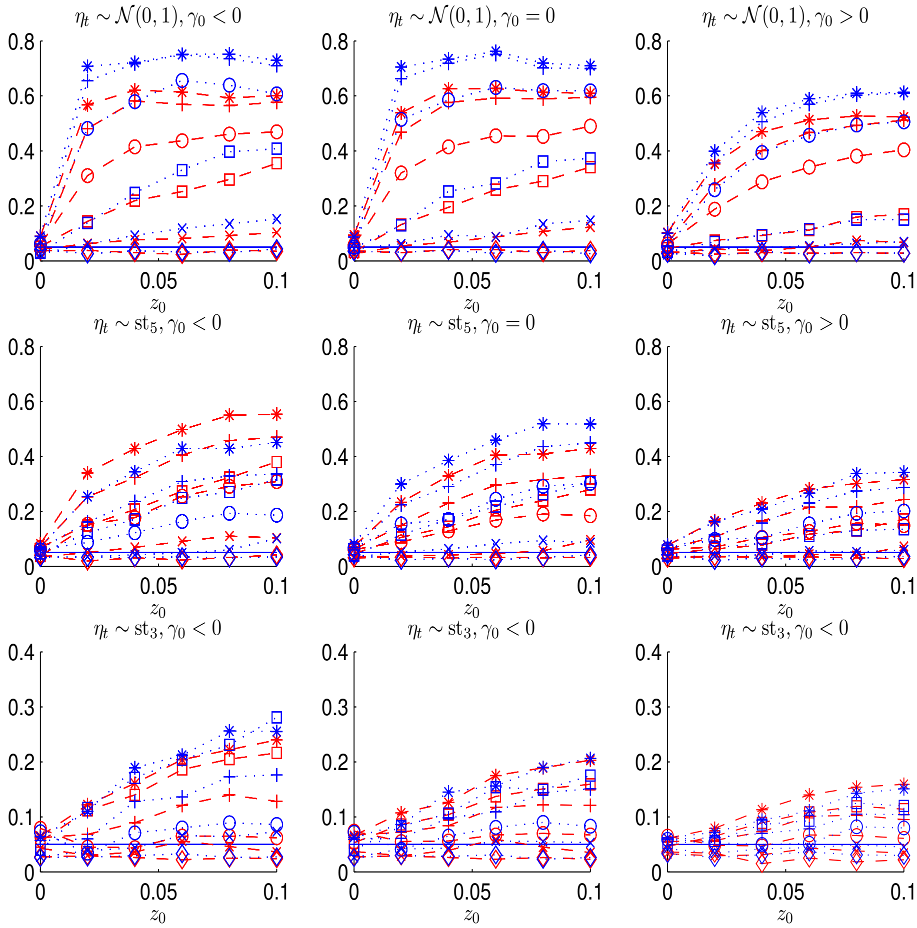

Fig. 4. The power of $Q_{2,2}(6)$ (circle line), $Q_{1,2}(6)$ (plus line), $Q_{0,2}(6)$ (star line), $Q_{1,1}(6)$ (cross line), $Q_{0,1}(6)$ (square line) and $Q_{0.5,0.5}(6)$ (diamond line) for $\eta_t \sim \mathcal{N}(0,1)$ (top panels), st_5 (middle panels), and st_3 (bottom panels) in terms of three different choices of γ_0, where the data sample is generated from model (6.3) with the sample size $n = 500$ (dash line) and 1000 (dot line), and the solid line stands for the significance level $\underline{\alpha} = 5\%$.

performance of that with small values of s (e.g., $s = 1$ or 0.5), and for a fixed choice of s, a smaller value of r will lead to a better power performance of $Q_{r,s}(6)$; (iii) when η_t is heavy-tailed (e.g., $\eta_t \sim \mathrm{st}_5$ or st_3), the power performance of $Q_{r,s}(6)$ for $s = 2$ or 1 becomes better when the value of r becomes smaller; (iv) $Q_{0.5,0.5}(6)$ has the worst power performance in all examined cases; (v) the performance of each portmanteau test becomes worse when η_t becomes more heavy-tailed, especially when $\gamma_0 > 0$; (vi) the power of each portmanteau test may not increase with the value of z_0, and this is probably because the positive z_0 only reflects the scale of y_t in model (6.3). In general, it is reasonable to conclude that $Q_{r,s}(m)$ has a desirable power to detect the mis-specification in the drift term especially for light-tailed η_t and the cases of $\gamma_0 \leq 0$.

7. APPLICATIONS

7·1. *Application to stock returns*

This subsection restudies the daily stock data of Monarch Community Bancorp (NasdaqCM: MCBF), KV Pharmaceutical (NYSE: KV-A), Community Bankers Trust (AMEX: BTC), and China MediaExpress (NasdaqGS: CCME) in Francq and Zakoïan (2012). The log-return ($\times 100$) of each stock data is non-stationary, and it was fitted by a non-stationary model (1.3) with the Gaussian QMLE in their paper. Fig. 5 plots the Hill's estimators $\{H_n(k)\}_{k=10}^{100}$ of the residuals from each fitted model (1.3) in Francq and Zakoïan (2012), where the Hill's estimator $H_n(k)$ of any sequence $\{z_t\}_{t=1}^n$ is defined by

$$H_n(k) = \left[\frac{1}{k}\sum_{i=1}^{k} \log \frac{z_{(n-i)}}{z_{(n-k)}}\right]^{-1}$$

with $\{z_{(t)}\}_{t=1}^n$ being the ascending order statistics of $\{z_t\}_{t=1}^n$, and k being a given positive integer. Clearly, Fig. 5 implies that η_t in each fitted model (1.3) has a finite second moment but an infinite fourth moment. Hence, the results in Francq and Zakoïan (2012) based on the Gaussian QMLE may not be reliable.

In this paper, we are of interest to see whether model (1.1) is able to fit these stock returns adequately. Fig. 6 plots the p-values of $Q_{1,1}(m)$ and $Q_{0.1,1}(m)$ for $m = 1, 2, \cdots, 20$. From this figure, we can see that model (1.1) is adequate to fit the KV-A return, and so we can fit this stock

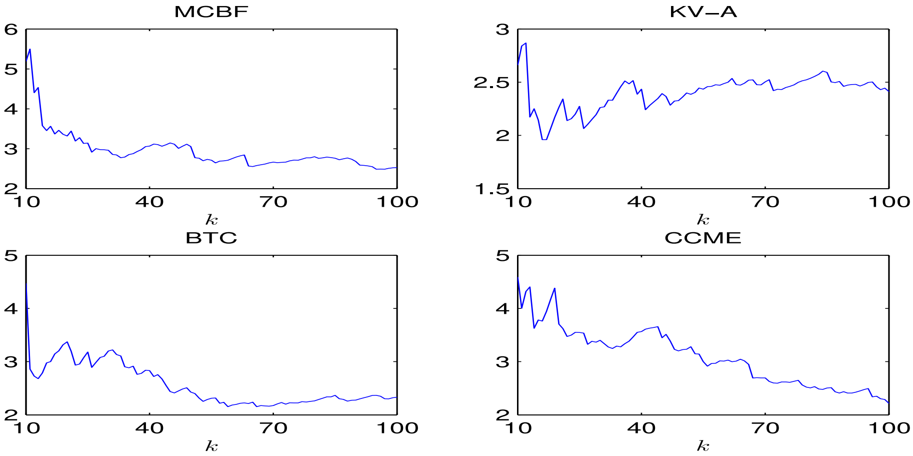

Fig. 5. Hill's estimators $\{H_n(k)\}_{k=10}^{100}$ for the residuals of each fitted model (1.3) in Francq and Zakoïan (2012).

return by

$$y_t = \eta_t \sqrt{h_t} \quad \text{and} \quad h_t = \underset{(0.0135)}{0.0588} y_{t-1}^2 + \underset{(0.0167)}{0.9081} h_{t-1}, \tag{7.1}$$

where model (7.1) is estimated by the GQMLE method with $r = 1$, the standard deviations of this estimator $\widehat{\theta}_{n,1}$ are in open parentheses, the estimate of $E|\eta_t|$ is 0.9998, and the value of Akaike Information Criterion (AIC) is 6433.1. Based on the residuals $\{\widehat{\eta}_{t,1}\}$, a plot of Hill's estimators $\{H_n(k)\}_{k=10}^{100}$ (not shown here for saving space) suggests that the tail index of η_t in model (7.1) lines between 2.2 and 2.5. Moreover, the value of stable test statistic $T_{n,1}$ is 0.934, and so there is no statistical evidence against the hypothesis that model (7.1) is stable. Also, the estimate of $\alpha_0 \text{var}(\eta_t) + \beta_0$ is 1.0958, and this implies that the KV-A return is heteroscedastic with an slightly exponentially explode variance. As a comparison, we also fit the KV-A return by model (1.3) with the same estimation method, and find that the fitted model (1.3) is non-stationary and its value of AIC is 6423.2, which is only 0.15% less than the one in model (7.1). Hence, in consideration of the inconsistency estimate of the drift term, a stable model (1.1) is more appropriate than a non-stationary model (1.3) to fit the KV-A return.

For the remaining three stock returns, Fig. 6 shows that model (1.1) can not fit them adequately, and we expect that model (1.3) with a robust estimation method can do it well.

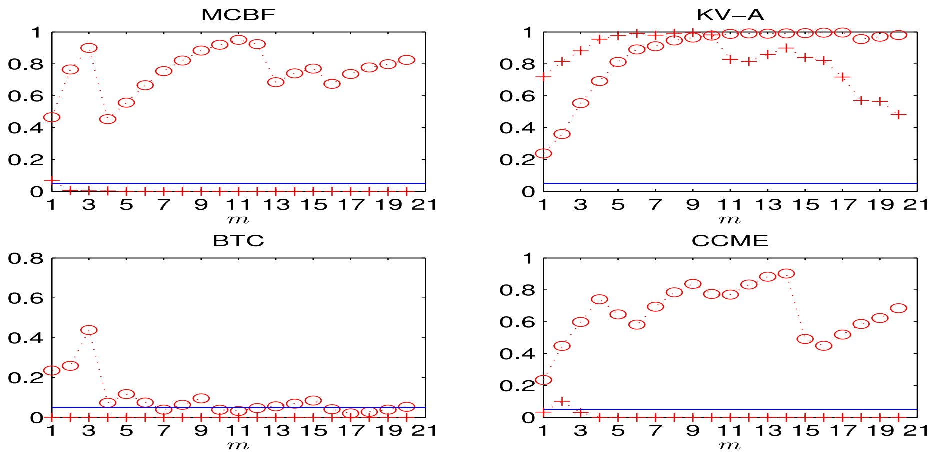

Fig. 6. The plot of p-values of $Q_{1,1}(m)$ (circle line) and $Q_{0.1,1}(m)$ (plus line) for each stock return, where the solid line stands for the significance level $\underline{\alpha} = 5\%$.

7·2. *Application to exchange rate returns*

This subsection studies the daily exchange rates of United States Dollars (USD) to Chinese Yuan (CNY), Euro (EUR), and British Pound (GBP) from January 2, 2007 to December 31, 2009, where each of data has in total 758 observations. We are of interest to see whether model (1.1) can fit the log-return ($\times 100$) of each exchange rate data. Here, since USD/CNY return exhibits some correlations in its conditional mean, it has been filtered by an ARMA$(2, 2)$ model with the LADE method in Zhu and Ling (2015). Fig. 7 plots the p-values of $Q_{1,2}(m)$, $Q_{0.1,2}(m)$, $Q_{1,1}(m)$, and $Q_{0.1,1}(m)$ for $m = 1, 2, \cdots, 20$. From this figure, we can see that model (1.1) can fit each exchange rate return adequately, although it is marginally inadequate to fit the USD/CNY return implied by the p-values of $Q_{1,2}(m)$ and $Q_{1,1}(m)$ at lags $m = 3$, 11, and 13.

Table 5 reports the estimation results based on $\widehat{\theta}_{n,1}$ for each exchange rate return. From this table, we find that each fitted model (1.1) is stable by looking at the values of $T_{n,1}$. Meanwhile, the estimated value of $\alpha_0 \text{var}(\eta_t) + \beta_0$ for each return is slightly larger than 1, and it implies that each

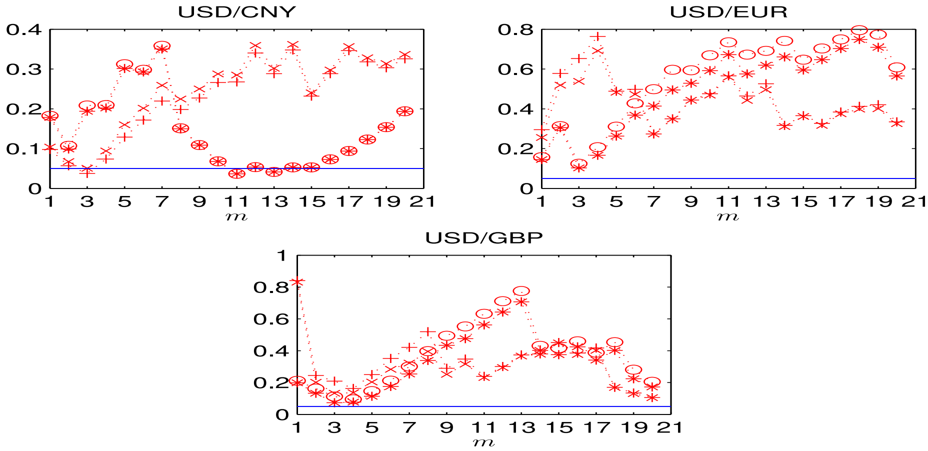

Fig. 7. The plot of p-values of $Q_{1,2}(m)$ (star line), $Q_{0.1,2}(m)$ (cross line), $Q_{1,1}(m)$ (circle line), and $Q_{0.1,1}(m)$ (plus line) for each exchange rate return, where the solid line stands for the significance level $\underline{\alpha} = 5\%$.

Table 5. *Estimation results based on* $\widehat{\theta}_{n,1}$ *for each exchange rate return*

	Log-return Series ($\times 100$)					
	USD/CNY		USD/EUR		USD/GBP	
	model (1.1)	model (1.3)	model (1.1)	model (1.3)	model (1.1)	model (1.3)
$\widehat{\omega}_{n,1}$		2.4e-5		0.0013		0.0013
		(1.4e-5)		(0.0008)		(0.0010)
$\widehat{\alpha}_{n,1}$	0.1011	0.1295	0.0364	0.0325	0.0390	0.0371
	(0.0216)	(0.0281)	(0.0093)	(0.0097)	(0.0095)	(0.0105)
$\widehat{\beta}_{n,1}$	0.8499	0.7990	0.9451	0.9428	0.9420	0.9379
	(0.0242)	(0.0315)	(0.0134)	(0.0158)	(0.0134)	(0.0164)
$\widehat{\gamma}_{n,1}$	0.0029	-0.0163	0.0036	-0.0045	0.0033	-0.0036
	(0.0106)	(0.4292)	(0.0034)	(0.1595)	(0.0037)	(0.1731)
$T_{n,1}$	0.2747	-1.0455	1.0429	-0.7722	0.9025	-0.5766
AIC	-1618.4	-1622.0	1440.8	1440.3	1582.5	1582.9
v_1	0.9999	1.0000	0.9995	1.0000	0.9999	1.0000
v_2	1.0844	1.1030	1.0088	1.0000	1.0095	1.0021

† The standard deviations of $\widehat{\theta}_{n,1}$ and $\widehat{\gamma}_{n,1}$ are in open parentheses, and v_1 and v_2 are the sample values of $E|\eta_t|$ and $\alpha_0 \mathrm{var}(\eta_t) + \beta_0$ based on the residuals, respectively. For model (1.3), $T_{n,1}$ is calculated analogously as in Francq and Zakoïan (2012). At the significance level 5%, model (1.1) is unstable if $|T_{n,1}| > 1.96$, and model (1.3) is stationary if $T_{n,1} < -1.65$.

return is heteroscedastic. This is in accordance with the visual evidence in Fig. 8, where along the sample path, the USD/CNY return has a seeming decreasing volatility, and the USD/EUR or USD/GBP return has a seeming increasing volatility. Moreover, a plot of Hill's estimators $\{H_n(k)\}_{k=10}^{100}$ in Fig. 9 suggests that η_t in fitted model (1.1) has a finite second moment but an infinite fourth moment for USD/CNY return, while it has a finite fourth moment for USD/EUR and USD/GBP returns.

As a comparison, we also fit each exchange rate return by model (1.3), and the related results are given in Table 5. From the values of $T_{n,1}$, we find that each fitted model (1.3) is non-stationary, and hence the values of $\widehat{\omega}_{n,1}$ and its standard deviation for fitted model (1.3) may be misleading, since $\widehat{\omega}_{n,1}$ is inconsistent according to a similar argument as Francq and Zakoïan (2012). Moreover, we find that model (1.1) and model (1.3) have very close values of AIC. In view of all of these, it is reasonable to conclude that model (1.1) is more appropriate than model (1.3) to fit each exchange rate return. Among years 2007-2009, the financial crisis happened so that most of exchange rate return data tend to be slightly heteroscedastic over time, and this might lead to the validity of model (1.1) in fitting each heteroscedastic return data.

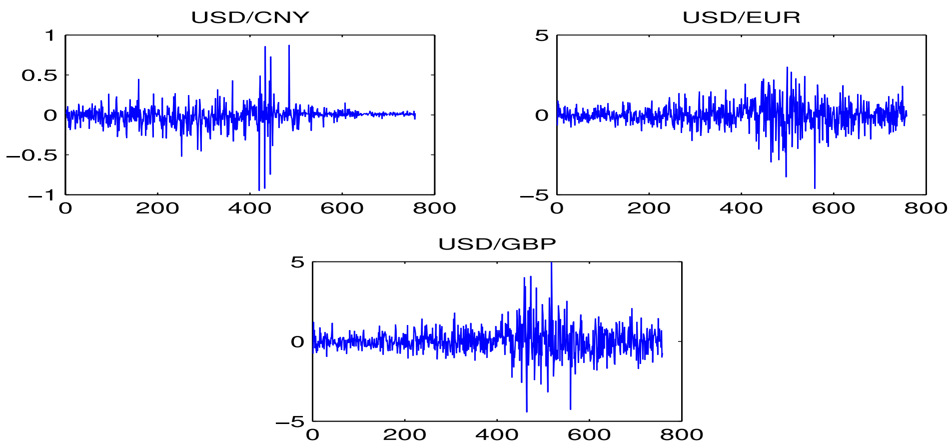

Fig. 8. The log-returns ($\times 100$) of three daily exchange rates from January 2, 2007 to December 31, 2009.

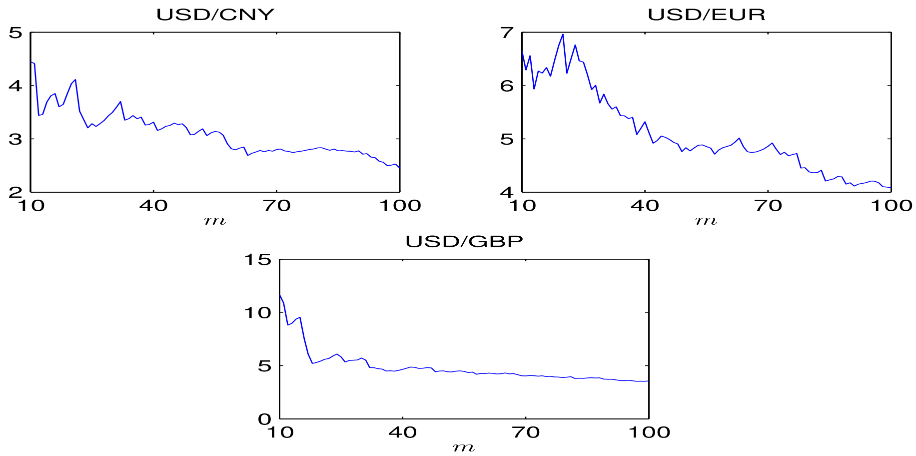

Fig. 9. Hill's estimators $\{H_n(k)\}_{k=10}^{100}$ for the residuals of each fitted model (1.1) in Table 5.

8. Concluding remarks and discussions

This paper proposes a ZD-GARCH$(1,1)$ model to study conditional heteroscedasticity and heteroscedasticity together. Unlike the classical GARCH$(1,1)$ model, ZD-GARCH$(1,1)$ model is always non-stationary, but interestingly when $\gamma_0 = 0$, it is stable with its sample path oscillating randomly between zero and infinity over time. Moreover, this paper studies the GQMLE of ZD-GARCH$(1,1)$ model, and establishes its strong consistency and asymptotic normality, regardless of the sign of γ_0. Based on the GQMLE, an estimator for γ_0, a test for stability, and a portmanteau test for model checking are all constructed. Simulation studies reveal that all proposed estimators and tests have a good finite sample performance. Applications demonstrate that a stable ZD-GARCH$(1,1)$ model is more appropriate than a non-stationary GARCH$(1,1)$ model to fit the KV-A stock return in Francq and Zakoïan (2012) and three major exchange rate returns during financial crisis in years 2007-2009.

It is worth noting that ZD-GARCH$(1,1)$ model is most likely stable in applications. This is not out of expectation, since only the stable ZD-GARCH$(1,1)$ model has a desirable sample path which is close to the often observed data track in the real world. Comparing with the non-stationary GARCH$(1,1)$ model, the stable ZD-GARCH$(1,1)$ model has the same ability to capture heteroscedasticity, and most importantly, it avoids the estimation for the drift-term, which is the troublesome for the non-stationary GARCH$(1,1)$ model.

The idea of setting drift term being zero can be easily applied to many other conditionally heteroscedastic models. However, the exploration of the corresponding properties of probabilistic structure and statistical inference is not trivial. Thus, considering the complexity of the extended heteroscedastic model, we will keep using ZD-GARCH$(1,1)$ model as a first step of introducing the phenomenon of "zero-drift". Although some readers might prefer to consider more comprehensive zero-drift conditionally heteroscedastic models, we hope that such readers will nonetheless find that our analysis in this paper is still helpful and stimulating.

APPENDIX: PROOFS OF THEOREMS

Define five $[0,\infty]$-valued processes

$$\begin{aligned}
v_t(\theta) &= \sum_{i=1}^{\infty} \frac{\alpha\eta_{t-i}^2}{\beta_0+\alpha_0\eta_{t-i}^2} \prod_{j=1}^{i-1} \frac{\beta}{\beta_0+\alpha_0\eta_{t-j}^2}, \\
d_t^{\alpha}(\theta) &= \sum_{i=1}^{\infty} \frac{\eta_{t-i}^2}{\beta_0+\alpha_0\eta_{t-i}^2} \prod_{j=1}^{i-1} \frac{\beta}{\beta_0+\alpha_0\eta_{t-j}^2}, \\
d_t^{\beta}(\theta) &= \sum_{i=2}^{\infty} \frac{(i-1)\alpha\eta_{t-i}^2}{\beta(\beta_0+\alpha_0\eta_{t-i}^2)} \prod_{j=1}^{i-1} \frac{\beta}{\beta_0+\alpha_0\eta_{t-j}^2}, \\
\nu_t^{\alpha\beta}(\theta) &= \sum_{i=2}^{\infty} \frac{(i-1)\eta_{t-i}^2}{\beta(\beta_0+\alpha_0\eta_{t-i}^2)} \prod_{j=1}^{i-1} \frac{\beta}{\beta_0+\alpha_0\eta_{t-j}^2}, \\
\nu_t^{\beta\beta}(\theta) &= \sum_{i=3}^{\infty} \frac{(i-1)(i-2)\alpha\eta_{t-i}^2}{\beta^2(\beta_0+\alpha_0\eta_{t-i}^2)} \prod_{j=1}^{i-1} \frac{\beta}{\beta_0+\alpha_0\eta_{t-j}^2}
\end{aligned}$$

with the convention $\prod_{k=1}^{j-1}=1$ when $j\leq 1$. Let Θ_0^* be any compact subset of Θ. Denote $\underline{\alpha}=\inf\{\alpha|\theta\in\Theta_0^*\}$, $\underline{\beta}=\inf\{\beta|\theta\in\Theta_0^*\}$, $\bar{\alpha}=\sup\{\alpha|\theta\in\Theta_0^*\}$, and $\bar{\beta}=\sup\{\beta|\theta\in\Theta_0^*\}$.

LEMMA A.1. Suppose that Assumptions 3.1(i) and 3.2 hold. Then, for any $\theta\in\Theta$, $v_t(\theta)$ is stationary and ergodic. Moreover, there exists a constant $c_0>0$ such that, as $t\to\infty$,

(i) $e^{c_0 t}\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t^2(\theta)}{h_t}-v_t(\theta)\right|\to 0$ a.s.;

(ii) $e^{c_0 t}\sup_{\theta\in\Theta_0^*}\left|\frac{h_t}{\sigma_t^2(\theta)}-\frac{1}{v_t(\theta)}\right|\to 0$ a.s.

Finally, for any $\theta \notin \Theta$, it holds that $\sigma_t^2(\theta)/h_t \to \infty$ a.s. as $t \to \infty$.

PROOF. For any $\theta \in \Theta$, $v_t(\theta)$ is finite (a.s.) by Assumption 3.1(i), Assumption 3.2 and the Cauchy root test. Since it is a measurable function of $\{\eta_j : j < t\}$, $v_t(\theta)$ is thus stationary and ergodic.

For (i), from (1.1), it follows that

$$\frac{h_{t-1}}{h_t} = \frac{1}{\beta_0 + \alpha_0 \eta_{t-1}^2}.$$

Note that

$$\begin{aligned}\frac{\sigma_t^2(\theta)}{h_t} &= \alpha \frac{h_{t-1}}{h_t}\frac{y_{t-1}^2}{h_{t-1}} + \beta \frac{h_{t-1}}{h_t}\frac{\sigma_{t-1}^2(\theta)}{h_{t-1}} \\ &= \frac{\alpha\eta_{t-1}^2}{\beta_0 + \alpha_0\eta_{t-1}^2} + \frac{\beta}{\beta_0 + \alpha_0\eta_{t-1}^2}\frac{\sigma_{t-1}^2(\theta)}{h_{t-1}} \\ &= \sum_{i=1}^{t} \frac{\alpha\eta_{t-i}^2}{\beta_0 + \alpha_0\eta_{t-i}^2}\prod_{j=1}^{i-1}\frac{\beta}{\beta_0 + \alpha_0\eta_{t-j}^2}.\end{aligned} \tag{A1}$$

Choose $c_0 = (\gamma_0 - \log \bar{\beta})/2$. Then, $c_0 > 0$ by Assumption 3.2, and

$$\begin{aligned}e^{c_0 t}\left|\frac{\sigma_t^2(\theta)}{h_t} - v_t(\theta)\right| &= e^{c_0 t}\sum_{i=t+1}^{\infty} \frac{\alpha\eta_{t-i}^2}{\beta_0 + \alpha_0\eta_{t-i}^2}\prod_{j=1}^{i-1}\frac{\beta}{\beta_0 + \alpha_0\eta_{t-j}^2} \\ &\le \sum_{i=t+1}^{\infty} \frac{\alpha\eta_{t-i}^2}{\beta_0 + \alpha_0\eta_{t-i}^2}\prod_{j=1}^{i-1}\frac{\beta e^{c_0}}{\beta_0 + \alpha_0\eta_{t-j}^2} \\ &\le \sum_{i=t+1}^{\infty} \frac{\bar{\alpha}\eta_{t-i}^2}{\beta_0 + \alpha_0\eta_{t-i}^2}\prod_{j=1}^{i-1}\frac{\bar{\beta} e^{c_0}}{\beta_0 + \alpha_0\eta_{t-j}^2}.\end{aligned}$$

Note that by the strong law of large numbers for stationary and ergodic sequences,

$$\frac{1}{i}\sum_{j=1}^{i-1}\log\left(\frac{\bar{\beta}e^{c_0}}{\beta_0 + \alpha_0\eta_{t-j}^2}\right) \to \log\bar{\beta} + c_0 - \gamma_0 = -c_0$$

as $i \to \infty$. By the Cauchy root test again, it follows that

$$e^{c_0 t}\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t^2(\theta)}{h_t} - v_t(\theta)\right| \to 0 \quad \text{a.s.}$$

as $t \to \infty$. Thus, it entails that (i) holds.

For (ii), a simple calculation yields that

$$\sup_{\theta\in\Theta_0^*}\left|\frac{h_t}{\sigma_t^2(\theta)} - \frac{1}{v_t(\theta)}\right| \le \frac{1}{v_t(\underline{\theta})\sigma_t^2(\underline{\theta})/h_t}\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t^2(\theta)}{h_t} - v_t(\theta)\right|,$$

where $\underline{\theta} = (\underline{\alpha}, \underline{\beta})$ and $\bar{\theta} = (\bar{\alpha}, \bar{\beta})$. Note that $\sigma_t^2(\underline{\theta})/h_t \to v_t(\underline{\theta})$ a.s. as $t \to \infty$ by (i) and

$$v_t(\underline{\theta}) > \frac{\underline{\alpha}\eta_{t-1}^2}{\beta_0 + \alpha_0\eta_{t-1}^2} > 0 \text{ a.s. by Assumption 3.1(i).}$$

By (i), it follows that (ii) holds.

Finally, for any $\theta \notin \Theta$, by (A1), it follows that $\sigma_t^2(\theta)/h_t \to \infty$ a.s. as $t \to \infty$ by the Cauchy root test when $\beta > e^{\gamma_0}$ and by the Chung-Fuchs theorem when $\beta = e^{\gamma_0}$. The proof is completed. $\square$

LEMMA A.2. Suppose that Assumptions 3.1(i) and 3.2 hold. Then, for any $\theta \in \Theta$, $d_t^{\alpha}(\theta)$, $d_t^{\beta}(\theta)$, $\nu_t^{\alpha\beta}(\theta)$ and $\nu_t^{\beta\beta}(\theta)$ are stationary and ergodic. Moreover, as $t \to \infty$,

(i) $$\sup_{\theta\in\Theta_0^*}\left\|\frac{1}{h_t}\frac{\partial\sigma_t^2(\theta)}{\partial\theta} - (d_t^{\alpha}(\theta), d_t^{\beta}(\theta))'\right\| \to 0 \quad \text{a.s.;}$$

(ii) $$\sup_{\theta\in\Theta_0^*}\left\|\frac{1}{h_t}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta\partial\theta'} - \begin{pmatrix} 0 & \nu_t^{\alpha\beta}(\theta) \\ & \\ \nu_t^{\alpha\beta}(\theta) & \nu_t^{\beta\beta}(\theta)\end{pmatrix}\right\| \to 0 \quad \text{a.s.}$$

PROOF. Since $\sigma_t^2(\theta) = \alpha\sum_{i=1}^t \beta^{i-1}y_{t-i}^2$, we have

$$\frac{\partial\sigma_t^2(\theta)}{\partial\alpha} = \sum_{i=1}^t \beta^{i-1}y_{t-i}^2, \qquad \frac{\partial\sigma_t^2(\theta)}{\partial\beta} = \alpha\sum_{i=2}^t (i-1)\beta^{i-2}y_{t-i}^2,$$
$$\frac{\partial^2\sigma_t^2(\theta)}{\partial\alpha^2} = 0, \qquad \frac{\partial^2\sigma_t^2(\theta)}{\partial\alpha\partial\beta} = \sum_{i=2}^t (i-1)\beta^{i-2}y_{t-i}^2,$$
$$\text{and} \quad \frac{\partial\sigma_t^2(\theta)}{\partial\beta^2} = \alpha\sum_{i=3}^t (i-1)(i-2)\beta^{i-3}y_{t-i}^2.$$

Then, the conclusion follows from the similar argument as for Lemma A.1. $\square$

LEMMA A.3. Suppose that the conditions in Theorem 3.2 hold. Then, as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial\ell_{t,r}(\theta_0)}{\partial\theta} \xrightarrow{d} \mathcal{N}(0, d_r\mathcal{I}),$$

where $d_r = r^4\kappa_r/16$ when $r > 0$, and $d_r = \kappa_r/4$ when $r = 0$.

PROOF. When $r > 0$, by a direct calculation, we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial\ell_{t,r}(\theta_0)}{\partial\theta} &= \frac{r}{2\sqrt{n}}\sum_{t=1}^n \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\left[1 - |\eta_t|^r\left(\frac{h_t}{\sigma_t^2(\theta_0)}\right)^{r/2}\right] \\ &= \frac{r}{2\sqrt{n}}\sum_{t=1}^n \frac{1}{h_t}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}(1 - |\eta_t|^r) + o_p(1),\end{aligned}$$

where the last equality holds by Lemma A.1(ii), Cauchy root test, and the fact that $v_t(\theta_0) = 1$ a.s. Similarly, when $r = 0$, we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial\ell_{t,r}(\theta_0)}{\partial\theta} &= -\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\left[\log|\eta_t| + \log\sqrt{\frac{h_t}{\sigma_t^2(\theta_0)}}\right] \\ &= -\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{1}{h_t}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\log|\eta_t| + o_p(1).\end{aligned}$$

Note that when θ_0 is an interior point of Θ, $d_t^\alpha(\theta_0)$ and $d_t^\beta(\theta_0)$ have moments of any order. Thus, by Lemma A.2(i), the conclusion follows from the same argument as Lemma A.4 in Francq and Zakoïan (2012). $\square$

LEMMA A.4. Let $d_t(\theta) = (d_t^\alpha(\theta), d_t^\beta(\theta))'$. Suppose that Assumptions 3.1-3.2 hold. Then, as $n \to \infty$,

$$\frac{1}{n}\sum_{t=1}^{n} \sup_{\theta\in\Theta_0^*} \left| \frac{\partial^2 \ell_{t,r}(\theta)}{\partial\theta\partial\theta'} - \Sigma_{t,r}(\theta) \right| \to 0 \quad \text{a.s.,}$$

where

$$\Sigma_{t,r}(\theta) = \frac{r}{2v_t(\theta)}\left[1 - \frac{|\eta_t|^r}{v_t^{r/2}(\theta)}\right]\begin{pmatrix} 0 & \nu_t^{\alpha\beta}(\theta) \\ & \\ \nu_t^{\alpha\beta}(\theta) & \nu_t^{\beta\beta}(\theta) \end{pmatrix} + \frac{r}{2v_t^2(\theta)}\left[\left(1 + \frac{r}{2}\right)\frac{|\eta_t|^r}{v_t^{r/2}(\theta)} - 1\right] d_t(\theta)d_t'(\theta)$$

when $r > 0$, and

$$\Sigma_{t,r}(\theta) = \left[\log\sqrt{v_t(\theta)} - \log|\eta_t|\right]\begin{pmatrix} 0 & \nu_t^{\alpha\beta}(\theta) \\ & \\ \nu_t^{\alpha\beta}(\theta) & \nu_t^{\beta\beta}(\theta) \end{pmatrix} + \left[\log|\eta_t| + \log\sqrt{\frac{1}{v_t(\theta)}} + \frac{1}{2}\right] d_t(\theta)d_t'(\theta)$$

when $r = 0$.

PROOF. Note that when $r > 0$,

$$\begin{aligned}\frac{\partial^2 \ell_{t,r}(\theta)}{\partial\theta\partial\theta'} &= \frac{r}{2}\left[1 - |\eta_t|^r\left(\frac{h_t}{\sigma_t^2(\theta)}\right)^{r/2}\right]\frac{1}{\sigma_t^2(\theta)}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta\partial\theta'} \\ &\quad + \frac{r}{2}\left[\left(1 + \frac{r}{2}\right)|\eta_t|^r\left(\frac{h_t}{\sigma_t^2(\theta)}\right)^{r/2} - 1\right]\frac{1}{\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'},\end{aligned}$$

and when $r = 0$,

$$\begin{aligned}\frac{\partial^2 \ell_{t,r}(\theta)}{\partial\theta\partial\theta'} &= \left[\log\sqrt{\frac{\sigma_t^2(\theta)}{h_t}} - \log|\eta_t|\right]\frac{1}{\sigma_t^2(\theta)}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta\partial\theta'} \\ &\quad + \left[\log|\eta_t| + \log\sqrt{\frac{h_t}{\sigma_t^2(\theta)}} + \frac{1}{2}\right]\frac{1}{\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}.\end{aligned}$$

Since the conclusion holds for the element-wise a.s. convergence, we only consider the convergence of $\partial^2 l_{t,2}(\theta)/\partial\beta^2$ for simplicity. By a direct calculation, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0^*}\left|\frac{\partial^2\ell_{t,2}(\theta)}{\partial\beta^2}-\Sigma_{t,2}^{\beta\beta}(\theta)\right| \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0^*}\left|\left\{1-\eta_t^2\frac{h_t}{\sigma_t^2(\theta)}\right\}\frac{1}{\sigma_t^2(\theta)}\frac{\partial^2\sigma_t^2(\theta)}{\partial\beta^2}-\left\{1-\frac{\eta_t^2}{v_t(\theta)}\right\}\frac{\nu_t^{\beta\beta}(\theta)}{v_t(\theta)}\right| \\
&+\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0^*}\left|\left\{2\eta_t^2\frac{h_t}{\sigma_t^2(\theta)}-1\right\}\frac{1}{\sigma_t^4(\theta)}\left[\frac{\partial\sigma_t^2(\theta)}{\partial\beta}\right]^2-\left\{\frac{2\eta_t^2}{v_t(\theta)}-1\right\}\frac{[d_t^\beta(\theta)]^2}{v_t^2(\theta)}\right| \\
&:= I_{1,n}+I_{2,n},
\end{aligned}$$

where $\Sigma_{t,2}^{\beta\beta}(\theta)$ is the last entry of $\Sigma_{t,2}(\theta)$. For $I_{1,n}$, we have

$$\begin{aligned}
I_{1,n} \leq\ & \frac{1}{n}\sum_{t=1}^{n}\eta_t^2\sup_{\theta\in\Theta_0^*}\left|\frac{h_t}{\sigma_t^2(\theta)}-\frac{1}{v_t(\theta)}\right|\frac{h_t}{\sigma_t^2(\theta)}\frac{1}{h_t}\frac{\partial^2\sigma_t^2(\theta)}{\partial\beta^2} \\
&+\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0^*}\left|1-\frac{\eta_t^2}{v_t(\theta)}\right|\left|\frac{h_t}{\sigma_t^2(\theta)}-\frac{1}{v_t(\theta)}\right|\frac{1}{h_t}\frac{\partial^2\sigma_t^2(\theta)}{\partial\beta^2} \\
&+\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0^*}\left|1-\frac{\eta_t^2}{v_t(\theta)}\right|\frac{1}{v_t(\theta)}\left|\frac{1}{h_t}\frac{\partial^2\sigma_t^2(\theta)}{\partial\beta^2}-\nu_t^{\beta\beta}(\theta)\right| \\
:=\ & I_{11,n}+I_{12,n}+I_{13,n}.
\end{aligned}$$

Now, we deal with $I_{13,n}$. By Lemma A.1(i) and Lemma A.2(ii), it follows that

$$I_{13,n}\leq\frac{1}{n}\sum_{t=1}^{n}\frac{1}{v_t(\underline{\theta})}\left[1+\frac{\eta_t^2}{v_t(\underline{\theta})}\right]\sup_{\theta\in\Theta_0^*}\left|\frac{1}{h_t}\frac{\partial^2\sigma_t^2(\theta)}{\partial\beta^2}-\nu_t^{\beta\beta}(\theta)\right|\to 0\quad\text{a.s.}$$

as $n\to\infty$. Similarly, by Lemma A.1(ii) and Lemma A.2, we can prove that $I_{11,n}\to 0$ and $I_{12,n}\to 0$ a.s. as $n\to\infty$. Thus, it follows that $I_{1,n}\to 0$ a.s. as $n\to\infty$. Using the same procedure, we can show that $I_{2,n}\to 0$ a.s. as $n\to\infty$. Therefore, as $n\to\infty$,

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0^*}\left|\frac{\partial^2\ell_{t,2}(\theta)}{\partial\beta^2}-\Sigma_{t,2}^{\beta\beta}(\theta)\right|\to 0\quad\text{a.s.}$$

and in turn the conclusion holds. □

PROOF OF THEOREM 3.1. We firs consider the case that $r>0$. Note that $\widehat{\theta}_{n,r}=\arg\min_{\theta\in\Theta}Q_{n,r}(\theta)$, where

$$Q_{n,r}(\theta)=\frac{1}{n}\sum_{t=1}^{n}\left[|\eta_t|^r\left\{\left(\frac{h_t}{\sigma_t^2(\theta)}\right)^{r/2}-1\right\}+\frac{r}{2}\log\frac{\sigma_t^2(\theta)}{h_t}\right]:=O_{n,r}(\theta)+R_{n,r}(\theta)$$

with

$$O_{n,r}(\theta)=\frac{1}{n}\sum_{t=1}^{n}\left[|\eta_t|^r\left\{\left(\frac{1}{v_t(\theta)}\right)^{r/2}-1\right\}+\frac{r}{2}\log v_t(\theta)\right]$$

and

$$R_{n,r}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left[|\eta_t|^r\left\{\left(\frac{h_t}{\sigma_t^2(\theta)}\right)^{r/2} - \left(\frac{1}{v_t(\theta)}\right)^{r/2}\right\} + \frac{r}{2}\log\frac{\sigma_t^2(\theta)}{h_t v_t(\theta)}\right].$$

Lemma A.1 implies that if $\theta \notin \Theta$, $Q_{n,r}(\theta) \to \infty$ a.s. as $n \to \infty$. Thus, it is sufficient to consider the case $\theta \in \Theta_0^*$, where Θ_0^* is an arbitrary compact subset of Θ. By the strong law of large numbers for stationary and ergodic sequences, we have

$$\lim_{n\to\infty} O_{n,r}(\theta) = E\left\{\left(\frac{1}{v_1(\theta)}\right)^{r/2} - 1 + \frac{r}{2}\log v_1(\theta)\right\} \geq 0 \quad \text{a.s.}$$

with the equality holding if and only if $v_1(\theta) = 1$ a.s. or equivalently, $\theta = \theta_0$ by Lemma A.2 in Francq and Zakoïan (2012).

Furthermore, since $v_t(\underline{\theta}) > 0$, Lemma A.1(i) and the mean value theorem entail that

$$\begin{aligned}
\lim_{n\to\infty}\sup_{\theta\in\Theta_0^*}\left|\frac{1}{n}\sum_{t=1}^{n}\log\frac{\sigma_t^2(\theta)}{h_t v_t(\theta)}\right| &\leq \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0^*}\left[\frac{h_t}{\sigma_t^2(\theta)} + \frac{1}{v_t(\theta)}\right]\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t^2(\theta)}{h_t} - v_t(\theta)\right| \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left[\frac{h_t}{\sigma_t^2(\underline{\theta})} - \frac{1}{v_t(\underline{\theta})}\right]\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t^2(\theta)}{h_t} - v_t(\theta)\right| \\
&\quad + \lim_{n\to\infty}\frac{2}{n}\sum_{t=1}^{n}\frac{1}{v_t(\underline{\theta})}\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t^2(\theta)}{h_t} - v_t(\theta)\right| \\
&= 0 \quad \text{a.s.}
\end{aligned} \tag{A2}$$

Meanwhile, by Lemma A.1(i), the mean value theorem, and the fact that $E|\eta_t|^r < \infty$, we can show that

$$\begin{aligned}
&\lim_{n\to\infty}\sup_{\theta\in\Theta_0^*}\left|\frac{1}{n}\sum_{t=1}^{n}|\eta_t|^r\left\{\left(\frac{h_t}{\sigma_t^2(\theta)}\right)^{r/2} - \left(\frac{1}{v_t(\theta)}\right)^{r/2}\right\}\right| \\
&\leq \lim_{n\to\infty}\frac{r}{2n}\sum_{t=1}^{n}|\eta_t|^r\sup_{\theta\in\Theta_0^*}\left[\left(\frac{h_t}{\sigma_t^2(\theta)}\right)^{r/2-1} + \left(\frac{1}{v_t(\theta)}\right)^{r/2-1}\right]\sup_{\theta\in\Theta_0^*}\left|\frac{h_t}{\sigma_t^2(\theta)} - \frac{1}{v_t(\theta)}\right| \\
&= 0 \quad \text{a.s.}
\end{aligned}$$

where the last equality holds as for (A2). Thus, it follows that

$$\lim_{n\to\infty}\sup_{\theta\in\Theta_0^*}|R_{n,r}(\theta)| = 0 \quad \text{a.s.} \tag{A3}$$

Then, since Θ is compact, the proof in the case of $r > 0$ is completed by standard arguments.

Next, we consider the case that $r = 0$. Note that $\widehat{\theta}_{n,0} = \arg\min_{\theta\in\Theta} Q_{n,0}(\theta)$, where

$$Q_{n,0}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left[\log|\eta_t|\log\left(\frac{h_t}{\sigma_t^2(\theta)}\right) + \left(\log\sqrt{\frac{h_t}{\sigma_t^2(\theta)}}\right)^2\right] := O_{n,0}(\theta) + R_{n,0}(\theta)$$

with

$$O_{n,0}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left[\log|\eta_t|\log\left(\frac{1}{v_t(\theta)}\right) + \left(\log\sqrt{\frac{1}{v_t(\theta)}}\right)^2\right]$$

and

$$R_{n,0}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left[\log|\eta_t|\log\left(\frac{h_t v_t(\theta)}{\sigma_t^2(\theta)}\right) + \left(\log\sqrt{\frac{h_t}{\sigma_t^2(\theta)}}\right)^2 - \left(\log\sqrt{\frac{1}{v_t(\theta)}}\right)^2\right].$$

By the strong law of large numbers for stationary and ergodic sequences, we have

$$\lim_{n\to\infty} O_{n,0}(\theta) = E\left[\left(\log\sqrt{\frac{1}{v_1(\theta)}}\right)^2\right] \geq 0 \quad \text{a.s.}$$

with the equality holding if and only if $v_1(\theta) = 1$ a.s. or equivalently, $\theta = \theta_0$ by Lemma A.2 in Francq and Zakoïan (2012). Meanwhile, as for (A3), it is not hard to show that

$$\lim_{n\to\infty}\sup_{\theta\in\Theta_0^*}|R_{n,0}(\theta)| = 0 \quad \text{a.s.}$$

Then, since Θ is compact, the proof in the case of $r = 0$ is completed by standard arguments. □

PROOF OF THEOREM 3.2. Define

$$\mathcal{I}_{n,r} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_{t,r}(\theta_0)}{\partial\theta\partial\theta'} \quad \text{and} \quad \mathcal{S}_{n,r} = -\mathcal{I}_{n,r}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_{t,r}(\theta_0)}{\partial\theta}.$$

Since $v_t(\theta_0) = 1$, $E\Sigma_t(\theta_0) = (r^2/4)\mathcal{I}$ when $r > 0$, and $E\Sigma_t(\theta_0) = \mathcal{I}/2$ when $r = 0$. Then, it is not hard to see that by Lemmas A.3 and A.4,

$$\mathcal{S}_{n,r} \xrightarrow{d} \mathcal{N}(0, \kappa_r\mathcal{I}^{-1}) \text{ as } n\to\infty.$$

By Taylor's expansion, Theorem 3.1, and Lemma A.4, standard arguments entail that

$$\sqrt{n}(\widehat{\theta}_{n,r} - \theta_0) = \mathcal{S}_{n,r} + o_p(1),$$

and hence the conclusion holds. This completes the proof. □

PROOF OF THEOREM 3.3. A direct calculation shows that $\widetilde{\alpha}_{n,r}$ has the following explicit expression:

$$\widetilde{\alpha}_{n,r}^{r/2} = \frac{1}{n-1}\sum_{t=2}^{n}\frac{y_t^2}{y_{t-1}^2} \text{ for } r > 0 \text{ and } \log\widetilde{\alpha}_{n,r} = \frac{2}{n-1}\sum_{t=2}^{n}\log\left|\frac{y_t}{y_{t-1}}\right| \text{ for } r = 0.$$

From this, by Assumption 3.1, it is straightforward to see that without the compactness of Θ_α, as $n\to\infty$,

$$\sqrt{n}(\widetilde{\alpha}_{n,r}^{r/2} - \alpha_0^{r/2}) = \frac{1}{\sqrt{n}}\sum_{t=2}^{n}(|\eta_t|^r - 1)\alpha_0^{r/2} + o_p(1) \xrightarrow{d} \mathcal{N}\Big(0, (r^2/4)\kappa_r\alpha_0^r\Big) \text{ for } r > 0,$$

and

$$\sqrt{n}(\log\widetilde{\alpha}_{n,r} - \log\alpha_0) = \frac{2}{\sqrt{n}}\sum_{t=2}^{n}\log|\eta_t| + o_p(1) \xrightarrow{d} \mathcal{N}(0, \kappa_r) \text{ for } r = 0.$$

By the delta method, it follows that as $n \to \infty$,

$$\sqrt{n}(\widetilde{\alpha}_{n,r} - \alpha_0) \xrightarrow{d} \mathcal{N}\Big(0, \kappa_r \alpha_0^2\Big) \text{ for } r \geq 0.$$

This completes the proof. □

PROOF OF THEOREM 4.1. For (i), the conclusion follows directly from Theorems 3.1-3.2 and the Taylor expansion of $\log(x)$.

For (ii), let $\gamma_n(\theta) = n^{-1}\sum_{t=1}^{n} \log(\beta + \alpha\eta_t^2(\theta))$ with $\eta_t(\theta) = y_t/\sigma_t(\theta)$. By Taylor's expansion, we have

$$\widehat{\gamma}_{n,r} := \gamma_n(\widehat{\theta}_{n,r}) = \gamma_n(\theta_0) + \frac{\partial \gamma_n(\theta_{n,r}^*)}{\partial \theta'}(\widehat{\theta}_{n,r} - \theta_0),$$

where $\theta_{n,r}^*$ satisfies $\|\theta_{n,r}^* - \theta_0\| \leq \|\widehat{\theta}_{n,r} - \theta_0\|$. Using the expression

$$\frac{\partial \eta_t^2(\theta)}{\partial \theta} = -\eta_t^2 \frac{h_t^2}{\sigma_t^4(\theta)} \frac{1}{h_t} \frac{\partial \sigma_t^2(\theta)}{\partial \theta}$$

and the same argument as for Lemmas A.1-A.2, we can show that

$$\sup_{\theta \in \Theta_0^*} \left| \frac{\partial \gamma_n(\theta)}{\partial \theta} - \frac{1}{n}\sum_{t=1}^{n} \Gamma_t(\theta) \right| \to 0 \quad \text{a.s. as } n \to \infty,$$

where

$$\Gamma_t(\theta) = \Big(\frac{\eta_t^2}{\beta v_t(\theta) + \alpha\eta_t^2}, \ \frac{v_t(\theta)}{\beta v_t(\theta) + \alpha\eta_t^2}\Big)' - \frac{\alpha\eta_t^2}{\beta v_t(\theta) + \alpha\eta_t^2}\frac{1}{v_t(\theta)}\Big(d_t^{\alpha}(\theta), \ d_t^{\beta}(\theta)\Big)'.$$

Note that $E\|\Gamma_t(\theta_0)\| \leq 2(1-\nu_1)/\alpha_0 + 2\nu_1/\beta_0 < \infty$, and $E\Gamma_t(\theta_0) = 0$ by the facts that $Ed_t^{\alpha}(\theta_0) = 1/\alpha_0$ and $Ed_t^{\beta}(\theta_0) = \nu_1/\{\beta_0(1-\nu_1)\}$, where $\nu_1 \in (0,1)$ is defined in Theorem 3.2. By Theorem 3.1 and the strong law of large numbers for stationary and ergodic sequences, it follows that

$$\frac{\partial \gamma_n(\theta_{n,r}^*)}{\partial \theta} = E\Gamma_t(\theta_0) + o(1) = o(1) \quad \text{a.s.}$$

Thus, since $\sqrt{n}(\widehat{\theta}_{n,r} - \theta_0) = O_p(1)$ by Theorem 3.2, we have

$$\begin{aligned}
\sqrt{n}(\widehat{\gamma}_{n,r} - \gamma_0) &= \sqrt{n}(\gamma_n(\theta_0) - \gamma_0) + \frac{\partial \gamma_n(\theta_{n,r}^*)}{\partial \theta'}\sqrt{n}(\widehat{\theta}_{n,r} - \theta_0) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\log(\beta_0 + \alpha_0\eta_t^2) - \gamma_0\} + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0, \sigma_{\gamma_0}^2) \text{ as } n \to \infty,
\end{aligned}$$

by the central limit theorem. This completes the proof of (ii).

Moreover, for model (3.1), it is straightforward to see that

$$\sqrt{n}(\widehat{\gamma}_{n,r} - \gamma_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\log(\alpha_0\eta_t^2) - E\log(\alpha_0\eta_t^2)\right].$$

Hence, parts (i) and (ii) hold by the central limit theorem. This completes all of the proofs. □

PROOF OF THEOREM 5.1. Recall that $d_t(\theta_0) = (d_t^\alpha(\theta_0), d_t^\beta(\theta_0))'$. By Taylor's expansion and the similar technique as for Lemma A.4, some calculations give us that

$$\begin{aligned}\sqrt{n}\widehat{\rho}_{r,s}(k) &= \frac{1}{b_s}\left\{J_s(k) - E[d_t'(\theta_0)\pi_{s,t}(k)][\sqrt{n}(\widehat{\theta}_{n,r} - \theta_0)]\right\} + o_p(1) \\ &= \frac{1}{b_s}\left\{J_s(k) - p_s(k)[\sqrt{n}(\widehat{\theta}_{n,r} - \theta_0)]\right\} + o_p(1), \end{aligned} \tag{A4}$$

where $\pi_{s,t}(k) = |\eta_{t-k}|^s - a_s$ and

$$J_s(k) = \frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}(|\eta_t|^s - a_s)(|\eta_{t-k}|^s - a_s).$$

Note that

$$\sqrt{n}(\widehat{\theta}_{n,r} - \theta_0) = \begin{cases} \frac{2}{r\sqrt{n}}\mathcal{I}^{-1}\sum_{t=1}^n d_t(\theta_0)(|\eta_t|^r - 1) + o_p(1), & \text{if } r > 0, \\ \frac{2}{\sqrt{n}}\mathcal{I}^{-1}\sum_{t=1}^n d_t(\theta_0)\log|\eta_t| + o_p(1), & \text{if } r = 0. \end{cases}$$

By (A4), it follows that

$$\sqrt{n}(\widehat{\rho}_{r,s}(1), \cdots, \widehat{\rho}_{r,s}(m)) = \frac{1}{b_s\sqrt{n}}\sum_{t=m+1}^{n}(\mathbf{I}_m, -V_{r,s})q_{r,s} + o_p(1),$$

where

$$q_{r,s} = \begin{cases} \big((|\eta_t|^s - a_s)(|\eta_{t-1}|^s - a_s), \cdots, (|\eta_t|^s - a_s)(|\eta_{t-m}|^s - a_s), \\ \qquad d_t'(\theta_0)(|\eta_t|^r - 1)\big)', \text{ if } r > 0, \\ \big((|\eta_t|^s - a_s)(|\eta_{t-1}|^s - a_s), \cdots, (|\eta_t|^s - a_s)(|\eta_{t-m}|^s - a_s), \\ \qquad d_t'(\theta_0)\log|\eta_t|\big)', \qquad \text{if } r = 0. \end{cases}$$

Particularly, for model (3.1), it is straightforward to see that

$$\sqrt{n}(\widehat{\rho}_{r,s}(1), \cdots, \widehat{\rho}_{r,s}(m)) = \frac{1}{b_s\sqrt{n}}\sum_{t=m+1}^{n}(\mathbf{I}_m, 0)q_{r,s} + o_p(1).$$

Thus, the conclusion holds by the central limit theorem for martingale difference sequence. □

REFERENCES

Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.* **32**, 633–655.

Billingsley, P. (1999) *Convergence of Probability Measures*. Wiley.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *J. Eonometrics.* **31**, 307–327.

Bougerol, P. and Picard, N. (1992a). Stationarity of GARCH processes and of some nonnegative time series. *J. Eonometrics.* **52**, 115–127.

Bougerol, P. and Picard, N. (1992b). Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730.

Breusch, T.S. and Pagan, A.R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **47**, 1287–1294.

Chen, M. and Zhu, K. (2015). Sign-based portmanteau test for ARCH-type models with heavy-tailed innovations. *J. Econometrics* **189**, 313–320.

Dahlhaus R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1–37.

Dahlhaus R. and Rubba Rao S. (2006). Statistical inference for time-varing ARCH processes. *Ann. Statist.* **34**, 1075–1114.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

Engle, R.F. (2004). Risk and Volatility: Econometric Models and Financial Practice. *American Economic Review* **94**, 405–420.

Engle, R.F. and Rangel, J.G. (2008). The spline GARCH model for unconditional volatility and its global macroeconomic causes. *Rev. Financ. Stud.* **21**, 1187–1222.

Escanciano, J.C. (2007). Model checks using residual marked emprirical processes. *Statist. Sinica* **17**, 115–138.

Escanciano, J.C. (2008) Joint and marginal specification tests for conditional mean and varnance models. *J. Econometrics* **143**, 74–87.

Fan, J., Qi, L. and Xiu, D. (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods (with discussion). *J. Bus. Econom. Statist.* **32**, 178–205.

Francq, C. and Zakoïan, J.-M. (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Process. Appl.* **117**, 1265–1284.

Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley.

Francq, C. and Zakoïan, J.-M. (2012). Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica* **80**, 821–861.

Francq, C. and Zakoïan, J.-M. (2013a). Optimal predictions of powers of conditionally heteroscedastic processes. *J. R. Stat. Soc. Series B* **75**, 345–367.

Francq, C. and Zakoïan, J.-M. (2013b). Inference in nonstationary asymmetric GARCH models. *Ann. Statist.* **41**, 1970–1998.

Greene, W. (2002). Econometric Analysis. Englewood Cliffs, NJ: Prentice-Hall.

Hafner, C. M. and Preminger, A. (2015). An ARCH model without intercept. *Econom. Lett.* **129**, 13–17.

Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**, 285–317.

Jensen, S.T. and Rahbek, A. (2004a). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* **72**, 641–646.

Jensen, S.T. and Rahbek, A. (2004b). Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**, 1203–1226.

Li, W.K. (2004). Diagnostic checks in time series. Chapman & Hall/CRC.

Li, D., Li, M. and Wu, W. (2014). On dynamics of volatilities in nonstationary GARCH models. *Statist. Probab. Lett.* **94**, 86–90.

Li, D. and Wu, W. (2015). Renorming volatilities in a family of GARCH models. *Under revision for Econometric Theory*.

Li, G. and Li, W.K. (2005) Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika* **92**, 691–701.

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.

Li, W.K. and Mak, T.K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroscedasticity, *J. Time Series Anal.* **15**, 627–636.

Ling, S. and Tong, H. (2011) Score based goodness-of-fit tests for time series. *Statist. Sinica* **21**, 1807–1829.

Longerstaey, J. and Zangari, A. (1996) RiskMetrics™-Technical Document, 4th ed. Morgan Guaranty Trust Company of New York, New York.

Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318–334.

Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967–975.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* **48**, 817–838.

Zhu, K. (2016). Bootstrapping the portmanteau tests in weak auto-regressive moving average models. To appear in *J. R. Stat. Soc. Series B*.

Zhu, K. and Li, W.K. (2015a). A new Pearson-type QMLE for conditionally heteroscedastic models. *J. Bus. Econom. Statist.* **33**, 552–565.

Zhu, K. and Li, W.K. (2015b). A bootstrapped spectral test for adequacy in weak ARMA models. *J. Econometrics* **187**, 113–130.

Zhu, K. and Ling, S. (2015). LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises. *J. Amer. Statist. Assoc.* **110**, 784–794.