J. Korean Math. Soc. **52** (2015), No. 3, pp. 503–522
http://dx.doi.org/10.4134/JKMS.2015.52.3.503

PARAMETER CHANGE TEST FOR NONLINEAR TIME SERIES MODELS WITH GARCH TYPE ERRORS

Jiyeon Lee and Sangyeol Lee

Abstract. In this paper, we consider the problem of testing for a parameter change in nonlinear time series models with GARCH type errors. We introduce two types of cumulative sum (CUSUM) tests: estimates-based and residual-based tests. It is shown that under regularity conditions, their limiting null distributions are the sup of independent Brownian bridges. A simulation study is conducted for illustration.

1. Introduction

Nonlinear time series models have been popular in modeling time series over decades and various nonlinear time series models have been proposed by many researchers. For a review of classical nonlinear time series models, we refer to [21]: see also [7], [19], and [20] for nonlinear GARCH models. Further, [2], [4], [12] and [13] studied the stability and asymptotic properties of nonlinear autoregressive models with pure GARCH errors. Among nonlinear autoregressive models, smooth transition autoregressive (STAR) models (cf. [3] and [15]) have attracted much attention from practitioners since they are designed to cope with smoothly varying changes in underlying models of time series: STAR models can be viewed as a continuous version of threshold models with abrupt regime changes. Later, to enhance the practicality of STAR models in the financial time series analysis, [1] and [14] designated STAR-GARCH models. Recently, [16] studied the asymptotic properties of nonlinear autoregressive models with first-order nonlinear GARCH errors that include various nonlinear models with conditional volatility equations such as pure AR-GARCH, asymmetric AR-GARCH and smooth transition GARCH models.

The parameter change test in time series models has long been a popular issue among researchers since time series often experience parameter changes due to critical events and policy changes. For a review of early works, we refer to [8] and the references therein: see also [10], [11] and [18]. The main objective

Received July 25, 2014; Revised November 22, 2014.
2010 *Mathematics Subject Classification.* 62M10.
Key words and phrases. nonlinear time series models with GARCH type errors, parameter change, CUSUM test, weak convergence to a Brownian bridge.

of this study is to establish a theoretical foundation on the CUSUM test in general nonlinear autoregressive models with first-order nonlinear GARCH errors introduced in [16]. To this task, we consider the estimates- and residual-based CUSUM tests and demonstrate that under regularity conditions, their limiting null distributions are the sup of independent Brownian bridges, which is a key result to perform the proposed CUSUM tests. Theoretically, the former has an advantage over the latter because it can detect a change of all model parameters, while the latter actually detects a change of a functional of parameters. In practice, however, the estimates-based CUSUM test has a defect to perform poorly in GARCH type models since the parameter estimates for small lags in the CUSUM tests are heavily biased and eventually produce severe size distortions: this phenomenon is escalated when the model complexity increases. In comparison, the residual-based CUSUM test does not suffer from size distortions since it eliminates the heteroscedasticity of time series as seen in Lee et al. [11] and Lee and Lee [9]: further, it is reasonably robust against model misspecification: see de Pooter and van Dijk [5]. In our simulation study, we consider the asymmetric GARCH (AGARCH) model and the logistic smooth transition AR-smooth transition GARCH (STAR-STGARCH) model. For the latter, we focus on the residual-based CUSUM test since the estimates-based CUSUM test performs poorly due to the difficulty that arises in parameter estimation.

The rest of this paper is organized as follows. In Section 2, we introduce the nonlinear autoregressive models with nonlinear GARCH models and establish the asymptotic properties of quasi-MLE (QMLE). In Section 3, we study the estimates- and residual-based CUSUM tests and derive their limiting null distributions. In Section 4, we perform a simulation study. In Section 5, concluding remarks are provided. All the proofs are given in the Appendix.

2. Nonlinear autoregressive model

Let us consider the model:

$$y_t = f(y_{t-1}, \ldots, y_{t-p}; \mu_0) + \sigma_t \epsilon_t, \tag{1}$$

$$\sigma_t^2 = g(u_{0t-1}, \sigma_{t-1}^2; \theta_0), \tag{2}$$

where $u_{0t} = y_t - f(y_{t-1}, \ldots, y_{t-p}; \mu_0)$ and ϵ_t are i.i.d. random variables with $E(\epsilon_t)=0$ and $E(\epsilon_t^2)=1$, independent of $\{y_s : s < t\}$. The term $f(y_{t-1}, \ldots, y_{t-p}; \mu_0)$ is the conditional mean of y_t, a function of p lagged past observations and the m-dimensional true parameter vector μ_0, while the term $g(u_{0t-1}, \sigma_{t-1}^2; \theta_0)$ is the conditional variance of y_t, a function of u_{0t-1}, σ_{t-1}^2 and $\theta_0 = (\mu_0^T, \lambda_0^T)^T$, where λ_0 denotes the l-dimensional true parameter vector associated with conditional variance. We set $\theta = (\mu^T, \lambda^T)^T \in \Theta = M \times \Lambda \subset \mathbb{R}^{m+l}$, where μ and λ do not have common elements and M and Λ are compact subsets of $\mathbb{R}^m$ and $\mathbb{R}^l$, respectively.

In what follows, we assume that the data is generated by a stationary and ergodic process with finite moments of some order as follows:

(DGP) (y_t, σ_t^2) in (1) and (2) is stationary and ergodic with $E|y_t|^{2r} < \infty$ and $E|\sigma_t^2|^r < \infty$ for some $r > 0$.

Sufficient conditions to ensure (DGP) can be found in [16].

Suppose that $y_1, \ldots, y_n$ are observed and one wishes to test the following hypotheses:

$$\begin{aligned} (3) \qquad & H_0: \text{ The true parameter } \theta_0 \text{ does not change over } y_1, \ldots, y_n. \text{ vs.} \\ & H_1: \text{ not } H_0. \end{aligned}$$

To perform a test, we estimate the true parameter θ_0 based on the quasi log-likelihood estimator (QMLE) of θ_0 as in [17], which is defined as the minimizer of the following objective function $L_n(\theta)$, that is,

$$\begin{aligned} \hat{\theta}_n &= \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n}\left(log(h_t(\theta)) + \frac{u_t^2(\mu)}{h_t(\theta)}\right) \\ &= \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) = \arg\min_{\theta\in\Theta} L_n(\theta), \end{aligned} \tag{4}$$

where $u_t(\mu)$ and $h_t(\theta)$ are defined recursively by

$$\begin{aligned} u_t(\mu) &= y_t - f(y_{t-1}, \ldots, y_{t-p}; \mu), \\ h_t(\theta) &= g(u_{t-1}(\mu), h_{t-1}(\theta); \theta), \end{aligned}$$

and the initial values are assumed to be given properly (cf. [17], page 1243).

Owing to Proposition 1 of [17], $h_t(\theta)$ approximates the stationary and ergodic solution $h_t^*(\theta)$ which coincides with true conditional variance σ_t^2 a.s. when $\theta = \theta_0$. We also define $L_n^*(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t^*(\theta)$, where $l_t^*(\theta) = \log(h_t^*(\theta)) + \frac{u_t^2(\mu)}{h_t^*(\theta)}$.

3. Cusum test

To test the hypotheses in (3), we introduce the two cusum tests based on the parameter estimates and residuals.

3.1. Cusum test based on the estimates of parameters

Let $\hat{\theta}_k$ be the QMLE from the observations up to time k. Then, the test statistic is given by

$$T_n^E = \max_{1\le k\le n} T_{n,k}^E = \max_{1\le k\le n} \frac{k^2}{n}\left(\hat{\theta}_k - \hat{\theta}_n\right)^T \hat{\Sigma}_n \left(\hat{\theta}_k - \hat{\theta}_n\right),$$

where $\hat{\Sigma}_n$ is a consistent estimator of $\Sigma = \mathcal{J}(\theta_0)\mathcal{I}(\theta_0)^{-1}\mathcal{J}(\theta_0)$ with the positive definite matrices

$$\mathcal{J}(\theta_0) = -E\left[\frac{\partial^2 l_t^*(\theta_0)}{\partial\theta\partial\theta^T}\right] \text{ and } \mathcal{I}(\theta_0) = E\left[\frac{\partial l_t^*(\theta_0)}{\partial\theta}\frac{\partial l_t^*(\theta_0)}{\partial\theta^T}\right].$$

Provided that $\hat{\theta}_n$ and $\hat{\Sigma}_n$ satisfy the following conditions under H_0:

(T1) $\hat{\theta}_n$ is a strongly consistent estimator of θ_0;

(T2) For $0 \le s \le 1$, there exists a positive definite matrix Σ such that

$$\frac{[ns]}{\sqrt{n}}\left(\hat{\theta}_{[ns]} - \hat{\theta}_n\right) \xrightarrow{w} \Sigma^{-1/2} W_d^o(s) \quad \text{as } n \to \infty,$$

where $d = m + l$ and W_d^o denotes a d-dimensional Brownian bridge;

(T3) $\hat{\Sigma}_n$ is a consistent estimator of Σ in (T2),

it can be seen that

$$T_n^E \xrightarrow{w} \sup_{0 \le s \le 1} |W_d^o(s)|^2, \ n \to \infty$$

(see Theorems 1 and 2 below), where $|\cdot|$ denotes the Euclidian norm for any scalars, vectors or matrices; the L_p-norm for random variables is denoted by $\|X\| = (E[X]^p)^{1/p}$.

Further, to ensure (T1)-(T3), one needs to assume following regularity conditions:

(C1) θ_0 lies in the interior of the compact set Θ.

(C2) (i) $g(u, x; \theta)$ is continuous in $(u, x, \theta) \in \mathbb{R} \times \mathbb{R}_+ \times \Theta$.
(ii) There exist $0 < \rho < 1$ and $C < \infty$ such that for all $u \in \mathbb{R}$, $x \in \mathbb{R}_+$ and $\theta \in \Theta$,

$$g(u, x; \theta) \le \rho x + C(1 + u^2).$$

(iii) There exists $0 < \kappa < 1$ such that for all $u \in \mathbb{R}$, $x_1, x_2 \in \mathbb{R}_+$ and $\theta \in \Theta$,

$$|g(u, x_1; \theta) - g(u, x_2; \theta)| \le \kappa |x_1 - x_2|.$$

(C3) (i) $f(y_1, \ldots, y_p; \mu)$ is continuous in $(y_1, \ldots, y_p) \in \mathbb{R}^p$ and is Borel-measurable in μ.
(ii) There exists $C < \infty$ such that for all $(y_1, \ldots, y_p) \in \mathbb{R}^p$ and $\mu \in M$,

$$|f(y_1, \ldots, y_p; \mu)| \le C(1 + \sum_{i=1}^{p} |y_i|).$$

(C4) For some $\underline{g} > 0$, $\inf_{(u,x,\theta) \in \mathbb{R} \times \mathbb{R}_+ \times \Theta} g(u, x; \theta) = \underline{g}$.

(C5) (i) $f(y_1, \ldots, y_p; \mu) = f(y_1, \ldots, y_p; \mu_0)$ a.s. implies $\mu = \mu_0$.
(ii) $h_t^*(\mu_0, \lambda) = \sigma_t^2$ a.s. implies $\mu = \mu_0$.

Below, we use the notation:
$\partial_\mu f(y_1, \ldots, y_p; \mu) = \frac{\partial f(y_1, \ldots, y_p; \mu)}{\partial \mu}$ and $\partial_{\mu\mu} f(y_1, \ldots, y_p; \mu) = \frac{\partial^2 f(y_1, \ldots, y_p; \mu)}{\partial \mu \partial \mu^T}$; $\partial_{v_1} g(u, h; \theta) = \frac{\partial g(u,h;\theta)}{\partial v_1}$ and $\partial_{v_1 v_2} g(u, h; \theta) = \frac{\partial^2 g(u,h;\theta)}{\partial v_1 \partial v_2^T}$, where v_1 and v_2 can be any of u, h and θ; $\partial_\theta h_t(\theta) = \frac{\partial h_t(\theta)}{\partial \theta}$ and $\partial_{\theta\theta} h_t(\theta) = \frac{\partial^2 h_t(\theta)}{\partial \theta \partial \theta^T}$; $\partial_\theta h_t^*(\theta)$ and $\partial_{\theta\theta} h_t^*(\theta)$ are similarly defined.

It is well known that Assumptions (C1)-(C5) guarantee (T1) (cf. [17]). Further, to deal with (T2) and (T3), we assume that θ_0 in an interior point of a

compact convex subset $\Theta_0 = M_0 \times \Lambda_0$ of Θ, and furthermore, the following conditions are fulfilled:

(N1) (i) $f(y_1, \ldots, y_p; \mu)$ is twice continuously differentiable with respect to μ on M_0 for all $(y_1, \ldots, y_p) \in \mathbb{R}^p$.
(ii) $g(u, x; \theta)$ is twice continuously differentiable with respect to u, x and θ on $\mathbb{R} \times \mathbb{R}_+ \times \Theta_0$.

(N2) (i) For any $(y_1, \ldots, y_p) \in \mathbb{R}^p$ and $\mu \in M_0$, $|\partial_\mu f(y_1, \ldots, y_p; \mu)|$ and $|\partial_{\mu\mu} f(y_1, \ldots, y_p; \mu)|$ are bounded by $C(1 + \sum_{i=1}^p |y_i|)$ for some $C > 0$.
(ii) For any $u \in \mathbb{R}$, $x \in \mathbb{R}_+$ and $\theta \in \Theta_0$, $|\partial_u g(u, h; \theta)|$, $|\partial_{\theta\theta} g(u, h; \theta)|$, $|\partial_{uu} g(u, h; \theta)|$, $|\partial_{\theta u} g(u, h; \theta)|$ and $|g_{u\theta}(u, h; \theta)|$ are bounded by $C(1 + u^2 + x)$ for some $C > 0$.
(iii) For any $u \in \mathbb{R}$, $x_1, x_2 \in \mathbb{R}_+$ and $\theta \in \Theta_0$,

$$|\partial_{v_1} g(u, x_1; \theta) - \partial_{v_1} g(u, x_2; \theta)| \le \kappa' |x_1 - x_2|,$$

$$|\partial_{v_1 v_2} g_(u, x_1; \theta) - \partial_{v_1 v_2} g(u, x_2; \theta)| \le \kappa' |x_1 - x_2|$$

for some $0 < \kappa' < 1$, where v_1 and v_2 can be any of u, h and θ.

(N3) Assumption (DGP) holds with $r = 2$ and $E(\epsilon_t^8) < \infty$.

(N4) $\| \sup_{\theta \in \Theta_0} \frac{|\partial_\theta h_t^*(\theta)|}{h_t^*(\theta)} \|_4 < \infty$ and $\| \sup_{\theta \in \Theta_0} \frac{|\partial_{\theta\theta} h_t^*(\theta)|}{h_t^*(\theta)} \|_2 < \infty$.

(N5) (i) The distribution of ϵ_t is not concentrated at two points.
(ii) $\nu_1^T \partial_\mu f(y_1, \ldots, y_p; \mu_0) = 0$ a.s. for some $\nu_1 \in \mathbb{R}^m$ implies $\nu_1 = 0$.
(iii) $\nu_2^T \partial_\lambda g(u_{0,t}, \sigma_t^2; \theta_0) = 0$ a.s. for some $\nu_2 \in \mathbb{R}^l$ implies $\nu_2 = 0$.

Particularly, (N5) implies that $\mathcal{J}(\theta_0)$ and $\mathcal{I}(\theta_0)$ are positive definite (cf. [17]).

To obtain the null distribution of T_n^E, we should check if (T2) holds. To task this, we apply a functional central limit theorem to $\hat{\theta}_{[ns]}$ for $0 \le s \le 1$. Since $L_{[ns]}(\theta)$ is twice continuously differentiable with respect to θ and has a maximum at $\theta = \hat{\theta}_{[ns]}$, by Taylor's theorem, we can express

$$0 = \partial_\theta L_{[ns]}(\hat{\theta}_{[ns]}) = \partial_\theta L_{[ns]}(\theta_0) + \partial_{\theta\theta}^2 L_{[ns]}(\bar{\theta}_{[ns]})(\hat{\theta}_{[ns]} - \theta_0),$$

where $\bar{\theta}_{[ns]}$ is an appropriate intermediate point between $\hat{\theta}_{[ns]}$ and θ_0. Thus, we have

$$\begin{aligned} \mathcal{J}(\theta_0)\left(\hat{\theta}_{[ns]} - \theta_0\right) &= \partial_\theta L_{[ns]}(\theta_0) + \left(\partial_{\theta\theta}^2 L_{[ns]}(\bar{\theta}_{[ns]}) + \mathcal{J}(\theta_0)\right)(\hat{\theta}_{[ns]} - \theta_0) \\ &= \partial_\theta L_{[ns]}^*(\theta_0) + \Delta_{1[ns]} + \Delta_{2[ns]}, \end{aligned} \tag{5}$$

where

$$\Delta_{1[ns]} = \left(\partial_\theta L_{[ns]}(\theta_0) - \partial_\theta L_{[ns]}^*(\theta_0)\right) \quad \text{and}$$
$$\Delta_{2[ns]} = \left(\partial_{\theta\theta}^2 L_{[ns]}(\bar{\theta}_{[ns]}) + \mathcal{J}(\theta_0)\right)(\hat{\theta}_{[ns]} - \theta_0),$$

and subsequently,

$$\mathcal{J}(\theta_0)\frac{[ns]}{\sqrt{n}}\left(\hat{\theta}_{[ns]} - \theta_0\right) = \frac{[ns]}{\sqrt{n}}\partial_\theta L_{[ns]}^*(\theta_0) + \frac{[ns]}{\sqrt{n}}\Delta_{1[ns]} + \frac{[ns]}{\sqrt{n}}\Delta_{2[ns]}. \tag{6}$$

Note that

$$\partial_\theta l_t^*(\theta) = \frac{1}{h_t^*(\theta)}\left(1 - \frac{u_t^2(\mu)}{h_t^*(\theta)}\right)\partial_\theta h_t^*(\theta) - 2\frac{u_t(\mu)}{h_t^*(\theta)}\partial_\mu f(y_{t-1},\ldots,y_{t-p};\mu),$$

and thus,

$$\partial_\theta l_t^*(\theta_0) = \frac{1}{\sigma_t^2}\left(1-\epsilon_t^2\right)\partial_\theta h_t^*(\theta_0) - 2\frac{\epsilon_t}{\sigma_t}\partial_\mu f(y_{t-1},\ldots,y_{t-p};\mu_0).$$

Then, it can be shown that $E\{\partial_\theta l_t^*(\theta_0)\,|\,\mathcal{F}_{t-1}\} = 0$ and $\mathcal{I}(\theta_0)$ is finite. Hence, $\{\mathcal{I}(\theta_0)^{-1/2}\partial_\theta l_t^*(\theta_0), \mathcal{F}_t\}$ forms a stationary and ergodic martingale difference sequence, and by using the functional central limit theorem for martingale difference arrays and the Wold-Craér device, it can be shown that

$$\frac{[ns]}{\sqrt{n}}\partial_\theta L_{[ns]}^*(\theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\partial_\theta l_t^*(\theta_0) \xrightarrow{w} \mathcal{I}(\theta_0)^{1/2}W_d(s) \tag{7}$$

in the $\mathbb{D}^d[0,1]$ space (cf. [6]), where W_d is a d-dimensional Brownian motion. Then, using (7), Lemmas 1 and 2 in Appendix, we obtain the following result, the proof of which is provided in the Appendix.

Theorem 1. *Suppose that assumptions* (DGP), (C1)-(C5) *and* (N1)-(N5) *hold. Then, under* H_0*, we have*

$$T_{0n}^E = \max_{1\le k\le n}\frac{k^2}{n}\left(\hat\theta_k - \hat\theta_n\right)^T\Sigma\left(\hat\theta_k - \hat\theta_n\right) \xrightarrow{w} \sup_{0\le s\le 1}\left|W_d^o(s)\right|^2, \quad n\to\infty,$$

where $\Sigma = \mathcal{J}(\theta_0)\mathcal{I}(\theta_0)^{-1}\mathcal{J}(\theta_0)$ *and* W_d^o *denotes a* d*-dimensional Brownian bridge.*

To apply the CUSUM test in real practice, we replace $\mathcal{J}(\theta_0)$ and $\mathcal{I}(\theta_0)$ by their consistent estimators $\mathcal{J}_n$ and $\mathcal{I}_n$. For instance, one can employ the estimators in Theorem 2 of [17]. To ensure their consistency, namely (T3), in addition to the assumptions in Theorem 1, one has to assume (DGP) with $r = 4$ to fulfill (N3).

Theorem 2. *Suppose that* (DGP) *with* $r = 4$*,* (C1)-(C5) *and* (N1)-(N5) *hold. Then, under* H_0*, we have*

$$T_n^E = \max_{1\le k\le n} T_{n,k}^E = \max_{1\le k\le n}\frac{k^2}{n}\left(\hat\theta_k - \hat\theta_n\right)^T\hat\Sigma_n\left(\hat\theta_k - \hat\theta_n\right) \xrightarrow{w} \sup_{0\le s\le 1}\left|W_d^o(s)\right|^2,$$

where $\Sigma_n = \mathcal{J}_n\mathcal{I}_n^{-1}\mathcal{J}_n$ *and* W_d^o *denotes a* d*-dimensional Brownian bridge.*

We reject H_0 if $T_n^E \ge C_\alpha$ at the nominal level α, where C_α is the $100(1-\alpha)$ quantile values of $\sup_{0\le s\le 1}|W_d^o(s)|^2$. The critical values for $\alpha = 0.01, 0.05, 0.10$ are provided in Table 1 in [8]. As mentioned in Introduction, because $\hat\theta_k$ for small k's can be severely biased from θ_0, to implement T_n^E in practice, one may need a fairly large sample size and a modification of the test, say, $\max_{k_n\le k\le n} T_{n,k}^E$, where k_n is a sequence of positive integers diverging to ∞

with $k_n/n \to 0$ as $n \to \infty$. Nevertheless, this remedy still may not work so well in many situations dealing with GARCH type models, and therefore, the residual-based CUSUM test below is taken into consideration as an alternative.

3.2. Cusum test based on the residuals

The residual-based CUSUM test has the form of

$$\frac{1}{\sqrt{nVar(\epsilon_1^2)}} \max_{1\le k\le n} \left| \sum_{t=1}^{k} \epsilon_t^2 - \left(\frac{k}{n}\right) \sum_{t=1}^{n} \epsilon_t^2 \right|.$$

Since ϵ_t^2 are not observable, we replace ϵ_t^2 by the residuals $\hat{\epsilon}_t^2 = \frac{u_t(\hat{\mu}_n)}{\sqrt{h_t(\hat{\theta}_n)}}$, where $\hat{\theta}_n = (\hat{\mu}_n^T, \hat{\lambda}_n^T)^T$ is the QMLE of θ_0 in (4).

Since

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} \hat{\epsilon}_t^2 - \left(\frac{k}{n}\right) \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right| \\
&\le \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} \epsilon_t^2 - \left(\frac{k}{n}\right) \sum_{t=1}^{n} \epsilon_t^2 \right| + \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} (\hat{\epsilon}_t^2 - \epsilon_t^2) - \left(\frac{k}{n}\right) \sum_{t=1}^{n} (\hat{\epsilon}_t^2 - \epsilon_t^2) \right|
\end{aligned}$$

and

$$\frac{1}{\sqrt{nVar(\epsilon_1^2)}} \left| \sum_{t=1}^{k} \epsilon_t^2 - \left(\frac{k}{n}\right) \sum_{t=1}^{n} \epsilon_t^2 \right| \xrightarrow{w} |W_1^o(s)| \quad \text{as} \quad n \to \infty,$$

provided that the following conditions are satisfied under H_0:

(R1) $\frac{1}{\sqrt{n}} \max_{1\le k\le n} \left| \sum_{t=1}^{k} (\hat{\epsilon}_t^2 - \epsilon_t^2) - \left(\frac{k}{n}\right) \sum_{t=1}^{n} (\hat{\epsilon}_t^2 - \epsilon_t^2) \right| = o_P(1)$;

(R2) $\hat{\tau}_n^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^4 - (\frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2)^2 \xrightarrow{P} Var(\epsilon_1^2)$,

it can be seen that

$$\frac{1}{\sqrt{n}\hat{\tau}_n} \max_{1\le k\le n} \left| \sum_{t=1}^{k} \hat{\epsilon}_t^2 - \left(\frac{k}{n}\right) \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right| \xrightarrow{w} \sup_{0\le s\le 1} |W_1^o(s)|, \ n \to \infty.$$

To verify (R1) and (R2), we express the residuals as the sum of true errors and additional terms as follows:

$$\begin{aligned}
(8) \quad \hat{\epsilon}_t^2 &= \frac{(y_t - f(y_{t-1},\ldots,y_{t-p};\hat{\mu}_n))^2}{h_t(\hat{\theta}_n)} \\
&= \epsilon_t^2 + \frac{\epsilon_t^2\left(\sigma_t^2 - h_t(\hat{\theta}_n)\right)}{\sigma_t^2} + \frac{\epsilon_t^2\left(\sigma_t^2 - h_t(\hat{\theta}_n)\right)^2}{\sigma_t^2 h_t(\hat{\theta}_n)} + \frac{2u_{0t}(\mu_0 - \hat{\mu}_n)^T \partial_\mu f(y_{t-1},\ldots,y_{t-p};\bar{\mu}_n)}{h_t(\hat{\theta}_n)} \\
&\quad + \frac{\left(u_{ot}(\mu_0 - \hat{\mu}_n)^T \partial_\mu f(y_{t-1},\ldots,y_{t-p};\bar{\mu}_n)\right)^2}{h_t(\hat{\theta}_n)} \\
&= \epsilon_t^2 + I_{1,t} + I_{2,t} + I_{3,t} + I_{4,t},
\end{aligned}$$

where $\bar{\mu}_n$ is an appropriate intermediate point between $\hat{\mu}_n$ and μ_0. Then, using Lemmas 3 and 4 in the Appendix, we obtain the following result.

Theorem 3. *Suppose that assumptions* (DGP) *with* $r = 4$, (C1)-(C5) *and* (N1)-(N5) *hold. Then, under* H_0*, we have*

$$\begin{aligned} T_n^R &= \max_{1\le k\le n} T_{n,k}^R \\ &= \frac{1}{\sqrt{n}\hat{\tau}_n} \max_{1\le k\le n} \left| \sum_{t=1}^{k} \hat{\epsilon}_t^2 - \left(\frac{k}{n}\right) \sum_{t=1}^{n} \hat{\epsilon}_t^2 \right| \xrightarrow{w} \sup_{0\le s\le 1} |W_1^o(s)|, \ n \to \infty. \end{aligned}$$

4. Simulation results

In this section, we evaluate the performance of the residual-based CUSUM test T_n^R proposed in Section 3. Among the nonlinear models with GARCH errors, we consider the AGARCH model and the logistic smooth transition AR-smooth transition GARCH (STAR-STGARCH) model. For these models, the conditional mean and the conditional variance in (1) and (2) are given as follows:

- AGARCH(1,1) model:

$$\begin{aligned} &f(y_{t-1}, \ldots, y_{t-p}; \mu_0) = 0, \\ &g(u_{0t-1}, \sigma_{t-1}^2; \theta_0) = \omega + \alpha(|u_{0t-1}| - \gamma u_{0t-1})^2 + \beta\sigma_{t-1}^2. \end{aligned}$$

- Logistic STAR(p)-STGARCH(1,1) model:

$$f(y_{t-1}, \ldots, y_{t-p}; \mu_0) = \phi_0 + \psi_0 F(y_{t-1}; \varphi_1, \varphi_2) + \sum_{j=1}^{p} (\phi_j + \psi_j F(y_{t-1}; \varphi_1, \varphi_2)) y_{t-j},$$

$$g(u_{0t-1}, \sigma_{t-1}^2; \theta_0) = \omega_0 + (\alpha_1 + \alpha_2 G(u_{0t-1}; \gamma_1, \gamma_2)) u_{0t-1}^2 + \beta\sigma_{t-1}^2,$$

where

$$\begin{aligned} F(y; \varphi_1, \varphi_2) &= [1 + \exp(-\varphi_2(y - \varphi_1))]^{-1}, \\ G(u; \gamma_1, \gamma_2) &= [1 + \exp(-\gamma_2(u - \gamma_1))]^{-1}. \end{aligned}$$

Some sufficient conditions to ensure the assumptions in Theorems 2 and 3 for the above models are given in Section 6 of [17].

For the null hypothesis, we consider the following two cases:

- Case I: AGARCH(1,1) model with the parameter vector $\theta_0^T = (\omega, \alpha, \gamma, \beta) = (0.5, 0.2, 0.2, 0.3)$.
- Case II: Logistic STAR(1)-STGARCH(1,1) model with the parameter vectors $\theta_0^T = (\mu_0^T, \lambda_0^T)$, $\mu_0^T = (\phi_0, \phi_1, \psi_0, \psi_1, \varphi_1, \varphi_2) = (-0.3, -0.5, 0.4, 1.0, 0.0, 1.0)$ and $\lambda_0^T = (\omega, \alpha_1, \alpha_2, \beta, \gamma_1, \gamma_2) = (0.5, 0.1, 0.2, 0.3, 1.0, 1.0)$.

To examine the power, we consider the alternative hypothesis:

$$H_1 : \theta_0 \text{ changes to } \theta_0' \text{ occurs at } t = [n/2].$$

For each case, sets of $n = 500$, 1000 and 2000 observations are generated from the model with $\epsilon_t \sim_{iid} N(0,1)$ in (1). The empirical sizes and powers are calculated at the nominal levels 0.01, 0.05 and 0.10 and summarized in Tables 1 and 3. The figures in each table stand for the proportion of the number of rejections of the null hypothesis "H_0: No changes occur in θ_0", out of 1000 repetitions.

Tables 1 and 2 report the empirical sizes and powers in Case I. Here, we compare the residual-based CUSUM test with the estimates-based CUSUM test. The figures in parentheses are for the latter. Table 1 shows that T_n^R and T_n^E produce no severe size distortions and the empirical size gets closer to the nominal levels as n increases. Table 2 also shows that both the tests produce good powers in most cases. As anticipated, the power increases remarkably as n increases and the parameters experience changes more significantly. However, it is noteworthy that our CUSUM test is not suitable to detect the change of the parameter γ, and a change of γ does not affect the performance of the CUSUM test. Although the result on T_n^E appear to be similar to that on T_n^R, T_n^R has merit over T_n^E in terms of convenience and the computation speed since the parameter estimation in T_n^E should be implemented for all k's.

In Case II, the estimation of STAR-STGARCH parameters can be problematic because it is highly sensitive to the choice of optimization algorithms and initial values. It is well known that the threshold values φ_1, γ_1 and the transition rates φ_2, γ_2 are difficult to estimate, especially with small samples (cf. [1]), which seriously damages the estimates-based CUSUM test. Thus, in Case II, we only focus on the residual-based CUSUM test. Table 3 illustrates that T_n^R has no severe size distortions and the empirical size gets closer to the nominal levels as n increases. It could be reasoned that the stability of the residual-based test is owing to its robustness property against model misspecification, but a careful analysis is needed to confirm this conjecture. Table 4 shows that the powers of T_n^R increases to 1 when n increases and more than two parameters change. Particularly, it can be seen that the powers for the single parameter change cases are rather low. However, this result is due to the fact that the magnitude of the changes is not large enough: it is because all the parameters must lie in a region to satisfy the stationarity assumption. Meanwhile, it turns out that T_n^R does not detect well the change of threshold value γ_1 and transition rate γ_2: in other words, the change of threshold value and transition rate does not much contribute to improving the performance of the test. This indicates that the residual-based CUSUM test has a limitation in the application to STAR-STGARCH models and a more refined study is required to overcome this shortcoming.

5. Concluding remarks

Thus far, we have studied the CUSUM test for nonlinear autoregressive models with nonlinear GARCH errors. To establish a theoretical foundation on the

TABLE 1. Empirical sizes for case I

n	0.01	0.05	0.10
500	0.004	0.039	0.096
	(0.021)	(0.061)	(0.113)
1000	0.010	0.045	0.097
	(0.017)	(0.058)	(0.109)
2000	0.014	0.053	0.096
	(0.011)	(0.055)	(0.108)

CUSUM test in this class of models, we obtained the limiting null distribution of the estimates- and residual-based CUSUM tests. Our simulation study confirms that the CUSUM test is a functional tool to detect a parameter change. The STAR-STGARCH model is an important example in practice and has its own merit since the model itself can accommodate parameter changes. However, as seen in [1], there are non-trivial difficulties such that the estimates are quite sensitive to the choice of optimization algorithms and initial values, which easily leads to a false conclusion with a high possibility. Also, it turned out that some model parameters are not well detected by the CUSUM method. Hence, we did not pursue a further empirical study in STAR-STGARCH models. We leave this issue as a task of our future study.

Appendix

In this section, we provide the proofs for the theorems presented in the previous section.

Lemma 1. *Suppose that the assumptions in Theorem 2 hold. Then, under H_0,*

$$\max_{1\le k\le n} \frac{k}{\sqrt{n}}|\Delta_{1k}| = o_P(1).$$

Proof. We can express

$$\begin{aligned}\max_{1\le k\le n} \frac{1}{\sqrt{n}}\left|\sum_{t=1}^{k}\partial_\theta l_t(\theta_0) - \sum_{t=1}^{k}\partial_\theta l_t^*(\theta_0)\right| &\le \frac{1}{\sqrt{n}}\sum_{t=1}^{n}|\partial_\theta l_t(\theta_0) - \partial_\theta l_t^*(\theta_0)| \\ &\le \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}|\partial_\theta l_t(\theta_0) - \partial_\theta l_t^*(\theta_0)|\,.\end{aligned}$$

Hence, using Lemma D.5 of [17], we have $\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}|\partial_\theta l_t(\theta_0) - \partial_\theta l_t^*(\theta_0)| < \infty$. Thus, the right hand side of (27) is $o_P(1)$ and the lemma is validated. □

Lemma 2. *Suppose that the assumptions in Theorem 2 hold. Then, under H_0,*

$$\max_{1\le k\le n} \frac{k}{\sqrt{n}}|\Delta_{2k}| = o_P(1).$$

TABLE 2. Empirical powers for case I

$\theta_0 \to \theta_0'$	n	0.01	0.05	0.10
	500	0.470	0.748	0.838
		(0.522)	(0.802)	(0.867)
$\omega : 0.5 \to 0.3$	1000	0.934	0.984	0.993
		(0.945)	(0.985)	(1.000)
	2000	0.999	0.999	1.000
		(1.000)	(1.000)	(1.000)
	500	0.063	0.208	0.315
		(0.111)	(0.259)	(0.350)
$\alpha : 0.2 \to 0.4$	1000	0.170	0.396	0.532
		(0.181)	(0.413)	(0.561)
	2000	0.495	0.743	0.828
		(0.510)	(0.755)	(0.881)
	500	0.194	0.446	0.592
		(0.231)	(0.510)	(0.622)
$\beta : 0.3 \to 0.5$	1000	0.592	0.812	0.899
		(0.612)	(0.819)	(0.905)
	2000	0.956	0.991	0.991
		(0.950)	(0.999)	(1.000)
	500	0.007	0.041	0.095
		(0.019)	(0.059)	(0.101)
$\gamma : 0.2 \to 0.4$	1000	0.009	0.048	0.097
		(0.013)	(0.050)	(0.105)
	2000	0.019	0.059	0.109
		(0.014)	(0.059)	(0.110)

TABLE 3. Empirical sizes for case II

n	0.01	0.05	0.10
500	0.004	0.040	0.090
1000	0.013	0.044	0.093
2000	0.011	0.045	0.093

Proof. Let $m(n)$ be a sequence of positive integer satisfying $m(n) \to \infty$ and $m(n)/\sqrt{n} \to 0$ as $n \to \infty$. We express

$$\max_{1 \le k \le n} \frac{k}{\sqrt{n}} |\Delta_{2k}| \le \max_{1 \le k \le m(n)} \frac{k}{\sqrt{n}} |\Delta_{2k}| + \max_{m(n) < k \le n} \frac{k}{\sqrt{n}} |\Delta_{2k}| . \tag{9}$$

TABLE 4. Empirical powers for case II

$\theta_0 \to \theta_0'$	n	0.01	0.05	0.10
	500	0.407	0.707	0.815
$\omega : 0.5 \to 0.3$	1000	0.901	0.972	0.983
	2000	1.000	1.000	1.000
	500	0.032	0.116	0.158
$\alpha_1 : 0.1 \to 0.2$	1000	0.073	0.208	0.281
	2000	0.079	0.378	0.473
	500	0.021	0.108	0.129
$\alpha_2 : 0.2 \to 0.3$	1000	0.032	0.153	0.187
	2000	0.043	0.189	0.291
$\alpha_1 : 0.1 \to 0.3$	500	0.153	0.408	0.530
$\alpha_2 : 0.2 \to 0.1$	1000	0.454	0.677	0.778
	2000	0.884	0.901	0.953
	500	0.205	0.545	0.670
$\beta : 0.3 \to 0.5$	1000	0.736	0.868	0.890
	2000	0.989	1.000	1.000
	500	0.031	0.093	0.134
$\gamma_1 : 1.0 \to 0$	1000	0.021	0.082	0.154
	2000	0.040	0.081	0.163
	500	0.011	0.042	0.116
$\gamma_2 : 1.0 \to 2.0$	1000	0.042	0.063	0.126
	2000	0.031	0.068	0.123

Since $\Delta_{2k} = \mathcal{J}(\theta_0)\left(\hat{\theta}_k - \theta_0\right) - \partial_\theta L_k^*(\theta_0) - \Delta_{1k}$, we have

(10)
$$\max_{1\le k\le m(n)} \frac{k}{\sqrt{n}}|\Delta_{2k}| \le |\mathcal{J}(\theta_0)|\frac{m(n)}{\sqrt{n}} \sup_{\theta\in\Theta_0} |\theta-\theta_0| + \frac{m(n)}{\sqrt{n}}\frac{1}{m(n)}\sum_{t=1}^{m(n)} |\partial_\theta l_t^*(\theta_0)| + \max_{1\le k\le n} \frac{k}{\sqrt{n}}|\Delta_{1k}|.$$

Since $\partial_\theta l_t^*(\theta_0)$ is stationary and $E|\partial_\theta l_t^*(\theta_0)| \le |\mathcal{I}(\theta_0)|^{1/2} < \infty$, we have

$$\frac{1}{m(n)}\sum_{t=1}^{m(n)} |\partial_\theta l_t^*(\theta_0)| = O_P(1).$$

This together with Lemma 1 implies the right hand side of (10) is $o_P(1)$.

Meanwhile, if $\partial^2_{\theta\theta} L_k(\bar{\theta}_k)$ is invertible, Δ_{2k} can be rewritten as

$$\Delta_{2k} = \left(\partial^2_{\theta\theta} L_k(\bar{\theta}_k) + \mathcal{J}(\theta_0)\right)\left(-\partial_\theta L_k(\theta_0)\right)\left(\partial^2_{\theta\theta} L_k(\bar{\theta}_k)\right)^{-1}.$$

Note that since $\mathcal{J}(\theta_0)$ is invertible, if $|M-\mathcal{J}(\theta_0)| \le (2|\mathcal{J}(\theta_0)^{-1}|)^{-1}$, the matrix M is invertible and $|M^{-1}|+|\mathcal{J}(\theta_0)^{-1}| \le C$, where C is a constant independent of M, and thus, if $\max_{m(n)<k\le n} | -\partial^2_{\theta\theta}L_k(\bar\theta_k) - \mathcal{J}(\theta_0)| \le (2|\mathcal{J}(\theta_0)^{-1}|)^{-1}$, it must hold that $\max_{m(n)<k\le n} |\partial^2_{\theta\theta}L_k(\bar\theta_k)^{-1}| \le C$.

Hence, since $\max_{m(n)<k\le n} |\partial^2_{\theta\theta}L_k(\bar\theta_k) + \mathcal{J}(\theta_0)| = o_P(1)$ and, due to (7),

$$\max_{1\le k\le n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{k}\partial_\theta l_t^*(\theta_0)\right| \xrightarrow{w} \sup_{0\le s\le 1}|\mathcal{I}(\theta_0)^{1/2}W_d(s)|,$$

we have

$$\begin{aligned}(11)\qquad &\max_{m(n)<k\le n}\frac{k}{\sqrt{n}}|\Delta_{2k}| \\ &\le C\max_{m(n)<k\le n}\left|\partial^2_{\theta\theta}L_k(\bar\theta_k)+\mathcal{J}(\theta_0)\right|\max_{1\le k\le n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{k}\partial_\theta l_t^*(\theta_0)\right| = o_P(1).\end{aligned}$$

Then, the lemma is asserted by (10) and (11). □

Proof of Theorem 1. In view of (6), we can express

$$\begin{aligned}\mathcal{J}(\theta_0)\frac{[ns]}{\sqrt{n}}(\hat\theta_{[ns]}-\hat\theta_n) &= \mathcal{J}(\theta_0)\frac{[ns]}{\sqrt{n}}\left(\hat\theta_{[ns]}-\theta_0-\hat\theta_n+\theta_0\right) \\ &= \frac{[ns]}{\sqrt{n}}\partial_\theta L^*_{[ns]}(\theta_0) - \frac{[ns]}{\sqrt{n}}\partial_\theta L^*_n(\theta_0) + \frac{[ns]}{\sqrt{n}}\left(\Delta_{1[ns]}-\Delta_{1n}\right) \\ &\quad + \frac{[ns]}{\sqrt{n}}\left(\Delta_{2[ns]}-\Delta_{2n}\right).\end{aligned}$$

Hence, we have

$$\begin{aligned}&\sup_{0\le s\le 1}\left|\mathcal{J}(\theta_0)\frac{[ns]}{\sqrt{n}}(\hat\theta_{[ns]}-\hat\theta_n) - \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\partial_\theta l_t^*(\theta_0) - \frac{[ns]}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\partial_\theta l_t^*(\theta_0)\right)\right| \\ &\le 2\sup_{0\le s\le 1}\frac{[ns]}{\sqrt{n}}|\Delta_{1[ns]}| + 2\sup_{0\le s\le 1}\frac{[ns]}{\sqrt{n}}|\Delta_{2[ns]}|.\end{aligned}$$

Further, by (7), we obtain

$$\begin{aligned}(12)\qquad &\frac{[ns]}{\sqrt{n}}\partial_\theta L^*_{[ns]}(\theta_0) - \frac{[ns]}{\sqrt{n}}\partial_\theta L^*_n(\theta_0) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\partial_\theta l_t^*(\theta_0) - \frac{[ns]}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\partial_\theta l_t^*(\theta_0) \\ &\xrightarrow{w} \mathcal{I}(\theta_0)^{1/2}(W_d(s)-sW_d(1)) = \mathcal{I}(\theta_0)^{1/2}W_d^o(s).\end{aligned}$$

Combining these and the results of Lemmas 1 and 2, we establish the theorem. □

Lemma 3. *Let $I_{i,t}$, $i = 1, 2, 3, 4$, be those in* (8) *and suppose that the assumptions in Theorem 3 hold. Then, under H_0,*

$$\frac{1}{\sqrt{n}} \max_{1\le k\le n} \left| \sum_{t=1}^{k} I_{i,t} - \left(\frac{k}{n}\right) \sum_{t=1}^{n} I_{i,t} \right| = o_P(1) \quad \textit{for } i = 1, 2, 3, 4.$$

Proof. First, we deal with $I_{1,t}$. By Taylor's theorem, we have

$$\begin{aligned} (13) \quad & \sigma_t^2 - h_t(\hat{\theta}_n) \\ &= \sigma_t^2 - h_t(\theta_0) + (\theta_0 - \hat{\theta}_n)^T \partial_\theta h_t^*(\bar{\theta}_n) + h_t(\theta_0) - h_t^*(\theta_0) + h_t^*(\hat{\theta}_n) - h_t(\hat{\theta}_n) \\ &:= J_{1,t} + J_{2,t} + J_{3,t} + J_{4,t}, \end{aligned}$$

where $\bar{\theta}_n$ is an appropriate intermediate point between $\hat{\theta}_n$ and θ_0. Further, by (N2)(iii), we have

$$\left|\sigma_t^2 - h_t(\theta_0)\right| \le \kappa^{t-1} \left|\sigma_1^2 - h_1(\theta_0)\right|. \tag{14}$$

Therefore, by using (C4), (N3) and the hölder's inequality, we have

$$\begin{aligned} & E\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\left|\sigma_t^2 - h_t(\theta_0)\right|}{\sigma_t^2} \epsilon_t^2 \right) \\ \le & \frac{1}{\underline{g}} \left(E\left|\sigma_1^2 - h_1(\theta_0)\right|^2 \right)^{1/2} \left(E(\epsilon_1^2)\right)^{1/2} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \kappa^{t-1} \xrightarrow{a.s.} 0, \end{aligned}$$

which in turn implies

$$\frac{1}{\sqrt{n}} \max_{1\le k\le n} \sum_{t=1}^{k} \left| J_{1,t} \frac{\epsilon_t^2}{\sigma_t^2} \right| \le \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| J_{1,t} \frac{\epsilon_t^2}{\sigma_t^2} \right| = o_P(1). \tag{15}$$

Meanwhile, concerning $J_{2,t}$, it suffices to show that

$$\frac{1}{\sqrt{n}} \max_{1\le k\le n} \left| \sum_{t=1}^{k} \left(J_{2,t} \frac{\epsilon_t^2}{\sigma_t^2} - E J_{2,t} \frac{\epsilon_t^2}{\sigma_t^2} \right) \right| = o_P(1), \tag{16}$$

since $\partial_\theta h_t^*(\bar{\theta}_n)$ and σ_t^2 are stationary and ergodic from Proposition 2 of [17] and (DGP). Owing to the invariance principle for stationary processes, we have

$$\frac{1}{\sqrt{n}} \max_{1\le k\le n} \left| \sum_{t=1}^{k} \left(\frac{\partial_\theta h_t^*(\bar{\theta}_n)\epsilon_t^2}{\sigma_t^2} - E \frac{\partial_\theta h_t^*(\bar{\theta}_n)\epsilon_t^2}{\sigma_t^2} \right) \right| = O_P(1).$$

Hence, since $\hat{\theta}_n \to \theta_0$ a.s., we obtain (16).

For $J_{3,t}$ and $J_{4,t}$, by using the fact that ϵ_t is stationary and $E(\epsilon_t^4) < \infty$, it holds that

$$\frac{\max_{1\le k\le n} \epsilon_t^2}{\sqrt{n}} = o_P(1). \tag{17}$$

Further, from Proposition 1 and Lemma A.2 of [17], we get

$$\sum_{t=1}^{n} \sup_{\theta\in\Theta} |h_t(\theta) - h_t^*(\theta)| < \infty.$$

Thus, by (C4), we have

(18)

$$\frac{1}{\sqrt{n}} \max_{1\le k\le n} \sum_{i=3}^{4} \sum_{t=1}^{k} \left| J_{i,t} \frac{\epsilon_t^2}{\sigma_t^2} \right| \le \frac{2}{\underline{g}} \frac{\max_{1\le k\le n} \epsilon_t^2}{\sqrt{n}} \sum_{t=1}^{n} \sup_{\theta\in\Theta} |h_t(\theta) - h_t^*(\theta)| = o_P(1).$$

Then, combining (15)-(18), we get

$$\frac{1}{\sqrt{n}} \max_{1\le k\le n} \left| \sum_{t=1}^{k} I_{1,t} - \left(\frac{k}{n}\right) \sum_{t=1}^{n} I_{1,t} \right| = o_P(1).$$

Next, we deal with $I_{2,t}$. Since $\sigma_t^2 \ge \underline{g} > 0$ and $h_t(\hat{\theta}_n) \ge \underline{g} > 0$ by (C4), to show

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |I_{2,t}| = o_P(1), \tag{19}$$

it suffices to verify $\frac{1}{\sqrt{n}} \sum_{i=1}^{4} \sum_{t=1}^{n} J_{i,t}^2 \epsilon_t^2 = o_P(1)$. Similarly to proof of (15), we can easily show that $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} J_{1,t}^2 \epsilon_t^2 = o_P(1)$. Further, due to Proposition 2 of [17], we can have $E(\sup_{\theta\in\Theta} |\partial_\theta h_t^*(\theta)|^2) < \infty$. This together with the fact that $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1)$ and (17) implies

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} J_{2,t}^2 \epsilon_t^2 \le n(\hat{\theta}_n - \theta_0)^T (\hat{\theta}_n - \theta_0) \frac{\max_{1\le k\le n} \epsilon_t^2}{\sqrt{n}} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta\in\Theta} |\partial_\theta h_t^*(\theta)|^2 = o_P(1).$$

Further, following essentially the same proof, we can easily obtain

$$\frac{1}{\sqrt{n}} \sum_{i=3}^{4} \sum_{t=1}^{n} J_{i,t}^2 \epsilon_t^2 = o_P(1).$$

Hence, we have (19) and this implies

$$\frac{1}{\sqrt{n}} \max_{1\le k\le n} \left| \sum_{t=1}^{k} I_{2,t} - \left(\frac{k}{n}\right) \sum_{t=1}^{n} I_{2,t} \right| = o_P(1).$$

Concerning $I_{3,t}$, we express

$$\begin{aligned} I_{3,t} = {} & \frac{2(\mu_0 - \hat{\mu}_n)^T \partial_\mu f(y_{t-1}, \ldots, y_{t-p}; \mu_0) \epsilon_t}{\sigma_t} + \frac{2\big(\sigma_t^2 - h_t(\hat{\theta}_n)\big)(\mu_0 - \hat{\mu}_n)^T \partial_\mu f(y_{t-1}, \ldots, y_{t-p}; \bar{\mu}_n) \epsilon_t}{\sigma_t h_t(\hat{\theta}_n)} \\ & + \frac{2(\mu_0 - \hat{\mu}_n)^T \partial_{\mu\mu} f(y_{t-1}, \ldots, y_{t-p}; \tilde{\mu}_n)(\mu_0 - \hat{\mu}_n) \epsilon_t}{\sigma_t} := K_{1,t} + K_{2,t} + K_{3,t}, \end{aligned}$$

where $\tilde{\mu}_n$ is an appropriate intermediate point between $\bar{\mu}_n$ and μ_0.

Owing to (DGP) and the invariance principle for stationary processes, we have

$$\frac{1}{\sqrt{n}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}\frac{\partial_\mu f(y_{t-1},\ldots,y_{t-p};\mu_0)\epsilon_t}{\sigma_t}-E\frac{\partial_\mu f(y_{t-1},\ldots,y_{t-p};\mu_0)\epsilon_t}{\sigma_t}\right|=O_P(1).$$

Since $\hat{\theta}_n$ converges to θ_0, it holds that

$$\frac{1}{\sqrt{n}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}K_{1,t}-\left(\frac{k}{n}\right)\sum_{t=1}^{n}K_{1,t}\right|=o_P(1).$$

Moreover, for $\epsilon>0$, by (DGP), the Markov's inequality and the Minkowski's inequality,

(20)
$$P\left(\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\left(1+\sum_{i=1}^{p}|y_{t-i}|\right)^2>\epsilon\right)\le\frac{\sum_{t=1}^{n}\left(1+\sum_{i=1}^{p}E^{1/2}|y_{t-i}|^2\right)^2}{\epsilon n\sqrt{n}}\xrightarrow{\text{a.s.}}0.$$

This together with (C4), (N2)(i), (19) and the fact that $\sqrt{n}\,(\mu_0-\hat{\mu}_n)=O_P(1)$ yields that

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|K_{2,t}|\\ &\le\frac{2C}{\sqrt{\underline{g}}}\sqrt{n}\,|\hat{\mu}_n-\mu_0|\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\epsilon_t^2\left(\sigma_t^2-h_t(\hat{\theta}_n)\right)^2}{\sigma_t^2h_t(\hat{\theta}_n)}\right)^{1/2}\left(\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\left(1+\sum_{i=1}^{p}|y_{t-i}|\right)^2\right)^{1/2}\\ &=o_P(1).\end{aligned}$$

Thus, $\frac{1}{\sqrt{n}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}K_{2,t}-\left(\frac{k}{n}\right)\sum_{t=1}^{n}K_{2,t}\right|=o_P(1)$.

For $K_{3,t}$, by (DGP), (N3), (C4), (N2)(i) and (20), we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|K_{3,t}|&\le\frac{2C}{\sqrt{\underline{g}}}n\,|\mu_0-\hat{\mu}_n|^2\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2\right)^{1/2}\left(\frac{1}{n^2}\sum_{t=1}^{n}\left(1+\sum_{i=1}^{p}|y_{t-i}|\right)^2\right)^{1/2}\\ &=o_P(1),\end{aligned}$$

which implies that $\frac{1}{\sqrt{n}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}I_{3,t}-\left(\frac{k}{n}\right)\sum_{t=1}^{n}I_{3,t}\right|=o_P(1)$.

Finally, we verify that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|I_{4,t}|=o_P(1)$. Similarly to (20), by (DGP), (C4) and (N2)(i), we also have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|I_{4,t}|\le\frac{C^2n\,|\hat{\mu}_n-\mu_0|^2}{\underline{g}}\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\left(1+\sum_{i=1}^{p}|y_{t-i}|\right)^2=o_P(1).\tag{21}$$

This completes the proof. □

Lemma 4. *Suppose that the assumptions in Theorem 3 hold. Then, under H_0,*

$$\hat{\tau}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^4 - (\frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^2)^2 \xrightarrow{P} Var(\epsilon_1^2).$$

Proof. By recalling the relationships in (8), we first verify that

$$\left|\frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^2 - \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2\right| \le \frac{1}{n}\sum_{i=1}^{4}\sum_{t=1}^{n}|I_{i,t}| = o_P(1). \tag{22}$$

Due to (C4), (N3), (19) and the Hölder's inequality, we can have

$$\frac{1}{n}\sum_{t=1}^{n}|I_{1,t}| \le \frac{1}{\sqrt{\underline{g}}}\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\epsilon_t^2\left(\sigma_t^2 - h_t(\hat{\theta}_n)\right)^2}{\sigma_t^2}\right)^{1/2} = o_P(1). \tag{23}$$

On the other hand, by (T1), (C4), (DGP), (N3) and (N2)(i), we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}|I_{3,t}| &\le \frac{2C^2\,|\hat{\mu}_n - \mu_0|}{\underline{g}}\left(\frac{1}{n}\sum_{t=1}^{n}u_{0t}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}\left(1+\sum_{i=1}^{p}|y_{t-i}|\right)^2\right)^{1/2}\\ &= o_P(1).\end{aligned} \tag{24}$$

Further, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|I_{2,t}| = o_P(1)$ and $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|I_{4,t}| = o_P(1)$ as we have in Lemma 3. Hence, combining this with (23) and (24), we obtain (22), which in turn implies $\frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^2 \xrightarrow{P} E(\epsilon_1^2)$.

Next, we verify that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\epsilon}_t^2 - \epsilon_t^2\right)^2 \le \frac{4}{n}\sum_{i=1}^{4}\sum_{t=1}^{n}|I_{i,t}|^2 = o_P(1). \tag{25}$$

First, for $|I_{1,t}|^2$, by using the results of proof of Lemma 3, we can easily get

$$\frac{1}{n}\sum_{t=1}^{n}|I_{1,t}|^2 \le \frac{1}{\underline{g}^2}\left(\frac{1}{\sqrt{n}}\max_{1\le t\le n}\epsilon_t^2\right)\frac{4}{\sqrt{n}}\sum_{i=1}^{4}\sum_{t=1}^{n}J_{i,t}^2\epsilon_t^2 = o_P(1). \tag{26}$$

Second, to deal with $|I_{2,t}|^2$, it suffices to verify

$$\frac{1}{n}\sum_{i=1}^{4}\sum_{t=1}^{n}J_{i,t}^4\epsilon_t^4 = o_P(1), \tag{27}$$

where we have used (C4) and (13). Further, since ϵ_t is stationary and $E\left(\epsilon_t^8\right) < \infty$ and (14), we have

$$\frac{1}{n}\sum_{t=1}^{n}J_{1,t}^4\epsilon_t^4 \le \left(\frac{1}{\sqrt{n}}\max_{1\le t\le n}\epsilon_t^4\right)\left(\sigma_1^2 - h_1(\theta_0)\right)^4\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\kappa^{4(t-1)} = o_P(1).$$

Moreover, due to Proposition 2, 3 and Lemma A.2 of [17], we also have

$$\frac{1}{n}\sum_{t=1}^{n} J_{2,t}^4\epsilon_t^4 \le \left(\sqrt{n}\left|\hat{\theta}_n-\theta_0\right|\frac{1}{\sqrt{n}}\max_{1\le t\le n}\left|\partial_\theta h_t^*(\theta_0)\right|\right)^4 \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^4 = o_P(1),$$

and

$$\frac{1}{n}\sum_{i=3}^{4}\sum_{t=1}^{n} J_{i,t}^4\epsilon_t^4 \le 2\left(\frac{\max_{1\le k\le n}\epsilon_t^2}{\sqrt{n}}\right)^2 \sum_{t=1}^{n}\sup_{\theta\in\Theta}\left|h_t(\theta)-h_t^*(\theta)\right|^4 = o_P(1).$$

Thus, the right hand side of (27) is $o_P(1)$.

Third, similarly to (24) and (21), it can be seen that

(28)

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}|I_{3,t}|^2 &\le \frac{2C^4\left|\hat{\mu}_n-\mu_0\right|^2}{\underline{g}^2}\left(\frac{1}{n}\sum_{t=1}^{n}u_{0t}^4\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}\left(1+\sum_{i=1}^{p}|y_{t-i}|\right)^4\right)^{1/2}\\ &= o_P(1),\end{aligned}$$

and

$$\frac{1}{n}\sum_{t=1}^{n}|I_{4,t}|^2 \le \frac{C^4\left(n\left|\mu_0-\hat{\mu}_n\right|^2\right)^2}{\underline{g}^2}\frac{1}{n}\sum_{t=1}^{n}\left(1+\sum_{i=1}^{p}|y_{t-i}|\right)^4 = o_P(1). \tag{29}$$

Thus, combining (26)–(29), we obtain (25).

Now, note that due to (25),

$$\begin{aligned}\left|\frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^4-\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^4\right| &\le \left(\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\epsilon}_t^2-\epsilon_t^2\right)^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\epsilon}_t^2+\epsilon_t^2\right)^2\right)^{1/2}\\ &\le \left(\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\epsilon}_t^2-\epsilon_t^2\right)^2\right)^{1/2}\left(\frac{2}{n}\sum_{t=1}^{n}\left(\hat{\epsilon}_t^2-\epsilon_t^2\right)^2+\frac{8}{n}\sum_{t=1}^{n}\epsilon_t^4\right)^{1/2}\\ &= o_P(1)(o_P(1)+O_P(1)) = o_P(1),\end{aligned}$$

and thus, $\frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^4 \xrightarrow{P} E(\epsilon_1^4)$. This together with (22) validates the lemma. □

Acknowledgements. We thank the referee for his/her helpful suggestions. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012R1A2A2A01046092).

References

[1] F. Chan and M. McAleer, *Maximum likelihood estimation of STAR and STAR-GARCH models: theory and Monte Carlo evidence*, J. Appl. Econometrics **17** (2002), 509–534.

[2] F. Chan, M. McAleer, and M. C. Medeiros, *Structure and asymptotic theory for nonlinear models with GARCH errors*, KIER Working Papers **754**, Kyoto University, Institute of Economic Research, 2010.

[3] K. S. Chan and H. Tong, *On estimating thresholds in autoregressive models*, J. Time Ser. Anal. **7** (1986), no. 3, 179–190.

[4] D. B. H. Cline, *Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models*, Adv. Appl. Probab. **39** (2007), 462–491.

[5] M. D. de Poorter and D. van Dijk, *Testing for changes in volatility in heteroscadastic time series – a further examination*, (No. EI 2004-38), Report / Econometric Institute, Erasmus University Rotterdam, 2004.

[6] P. Gaenssler and E. Haeusler, *On martingale central limit theory*, In E. Eberlein and M. S. Taqqu, (Eds.), Dependence in probability and statistics: A survey of recent results, pp. 303–334, Boston: Birkhäuser, 1986.

[7] D. Kristensen and A. Rahbek, *Asymptotics of the QMLE for a class of ARCH(q) models*, Econometric Theory **21** (2005), no. 5, 946–961.

[8] S. Lee, J. Ha, O. Na, and S. Na, *The cusum test for parameter change in time series models*, Scand. J. Statist. **30** (2003), no. 4, 781–796.

[9] S. Lee and J. Lee, *Residual based cusum test for parameter change in AR-GARCH models*, In Modeling Dependence in Econometrics, Advances in Intelligent Systems and Computing **251** (2014), 101–111.

[10] S. Lee and J. Song, *Test for parameter change in ARMA models with GARCH innovations*, Statist. Probab. Lett. **78** (2008), no. 13, 1990–1998.

[11] S. Lee, Y. Tokutsu, and K. Maekawa, *The cusum test for parameter change in regression models with ARCH errors*, J. Japan Statist. Soc. **34** (2004), 173–188.

[12] S. Ling, *On probability properties of a double threshold ARMA conditional heteroskedasticity model*, J. Appl. Probab. **36** (1999), 688–705.

[13] J. Liu, W. K. Li, and C. W. Li, *On a threshold autoregression with conditional heteroscedastic variances*, J. Statist. Plann. Inference **62** (1997), no. 2, 279–300.

[14] S. Lundbergh and T. Teräsvirta, *Modeling economic high frequency time series with STAR-STGARCH models*, SSE/EFI Working Paper Series in Economics and Finnce **291**, 1999.

[15] R. Luukkonen, P. Saikkonen, and T. Teräsvirta, *Testing linearity in univariate time series models*, Scand. J. Statist. **15** (1988), no. 3, 161–175.

[16] M. Meitz and P. Saikkonen, *Stability of nonlinear AR-GARCH models*, J. Time Ser. Anal. **29** (2008), no. 3, 453–475.

[17] ———, *Parameter estimation in nonlinear AR-GARCH models*, Econometric Theory **27** (2011), no. 6, 1236–1278.

[18] O. Na, J. Lee, and S. Lee, *Change point detection in copula ARMA-GARCH Models*, J. Time Ser. Anal. **33** (2012), no. 4, 554–569.

[19] J. Z. Pan, H. Wang, and H. Tong, *Estimation and tests for power-transformed and threshold GARCH models*, J. Econometrics **142** (2008), no. 1, 352–378.

[20] D. Straumann and T. Mikosch, *Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach*, Ann. Statist. **34** (2006), no. 5, 2449–2495.

[21] H. Tong, *Nonlinear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford, 1990.

Jiyeon Lee
Department of Statistics
Seoul National University
Seoul 151-747, Korea
E-mail address: leejy715@hanmail.net

Sangyeol Lee
Department of Statistics
Seoul National University
Seoul 151-747, Korea
E-mail address: sylee@stats.snu.ac.kr