JOURNAL OF TIME SERIES ANALYSIS
J. Time. Ser. Anal. **36**: 127–153 (2015)
Published online 14 October 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12098

ORIGINAL ARTICLE

TESTS FOR VOLATILITY SHIFTS IN GARCH AGAINST LONG-RANGE DEPENDENCE

TAEWOOK LEE,[a] MOOSUP KIM[b] AND CHANGRYONG BAEK[c] *
[a] *Department of Statistics, Hankuk University of Foreign Studies, Yongin, Kyunggi-do, Korea*
[b] *Department of Statistics, Seoul National University, Seoul, Korea*
[c] *Department of Statistics, Sungkyunkwan University, Seoul, Korea*

Many empirical findings show that volatility in financial time series exhibits high persistence. Some researchers argue that such persistency is due to volatility shifts in the market, while others believe that this is a natural fluctuation explained by stationary long-range dependence models. These two approaches confuse many practitioners, and forecasts for future volatility are dramatically different depending on which models to use. In this article, therefore, we consider a statistical testing procedure to distinguish volatility shifts in generalized AR conditional heteroscedasticity (GARCH) model against long-range dependence. Our testing procedure is based on the residual-based cumulative sum test, which is designed to correct the size distortion observed for GARCH models. We examine the validity of our method by providing asymptotic distributions of test statistic. Also, Monte Carlo simulations study shows that our proposed method achieves a good size while providing a reasonable power against long-range dependence. It is also observed that our test is robust to the misspecified GARCH models.

Received 13 August 2013; Revised 10 May 2014; Accepted 13 August 2014

Keywords: Volatility shifts; long-range dependence; GARCH models; change point analysis.
JEL. 62M10; 62G10; 60G18.

1. INTRODUCTION

Since the seminal articles of Engle (1982) and Bollerslev (1986), generalized AR conditional heteroscedasticity (GARCH, in short) models have been a popular tool to analyse the stylized features on the volatility of financial time series. In the applications of GARCH models, however, the parameter estimation frequently indicates high persistence. That is, the sum of the estimated coefficients of the squared lagged returns and the lagged conditional variance terms in GARCH models becomes closer to 1 for larger sample sizes. To incorporate such observations into a model, the integrated GARCH (IGARCH, in short) model was introduced by Engle and Bollerslev (1986). Unfortunately, the practitioners and researchers have been doubtful that IGARCH models have a practical use for modelling and long-term forecasting of volatility in financial applications. This is mainly because IGARCH modelling of financial time series has been reported to be spurious when volatility shifts or model misspecifications are present in GARCH models. See, for example, Mikosch and Stărică (2004), Berkes *et al.* (2005), Hillebrand (2005), Jensen and Lange (2010) and references therein.

An alternative way to incorporate high persistence is to use long-range dependent modelling of time series for suitably transformed data, for example, absolute return or squared return. Among many other literatures, for instance, Ding *et al.* (1993) and Baillie *et al.* (1996) document that long-range dependence (LRD, in short) successfully captures the stylized facts of financial time series volatility. Here, LRD for series $X = \{X_n\}_{n\in\mathbb{Z}}$ refers to a second-order stationary time series with a slowly decaying autocovariance function for large lags as

* Correspondence to: Changryong Baek, Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul, Korea 110-745. E-mail: crbaek@skku.edu

$$\gamma_X(h) = \mathrm{Cov}(X_0, X_h) \sim C h^{2d-1}, \quad \text{as} \quad h \to \infty, \tag{1}$$

where $C > 0$ is a constant and $d \in (0, 1/2)$ is the LRD parameter. Observe that under condition (1), autocovariances of LRD series are not absolutely summable, $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| = +\infty$. On the other hand, when the sum of autocovariances is absolutely summable, that is, $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$, it is referred to as short-range dependence (SRD, in short). However, it has a long history and debates over the use of LRD modelling because of the non-stationary-like features of LRD. In short, LRD series has slowly decaying positive autocovariance functions, so a stationary LRD series exhibits aperiodic local shifts for finite samples. It means that volatility shifts can produce spurious LRD (e.g. Klemeš, 1974; Teverovsky, 1999; Diebold and Inoue, 2001; Mikosch and Stărică, 2004 and references therein).

Therefore, prior to modelling the volatility of financial time series, it is crucial to make a decision on the selection of the true data generating process among IGARCH, volatility shifts and LRD. For volatility shifts, in particular, we only consider GARCH models on subsamples, also known as changing-parameter GARCH models for practical modelling perspectives. To be more specific, our test considers two hypotheses:

$$\mathrm{H}_0^{(R)} : \mathrm{VS-R\ model} \quad \text{against} \quad \mathrm{H}_1 : \mathrm{LRDmodels},$$

where VS-R represents that volatility shifts have occurred at *unknown* R locations; hence, each GARCH model is fitted for $(R+1)$ regimes. Also, note that the VS-R model is non-stationary, while LRD is second-order stationary; hence, its implications on the long-term forecasting of volatility could be dramatically different. Therefore, their discrimination is an interesting and important problem attracting many researchers for the last few decades.

Since the locations of volatility shifts are unknown, they need to be estimated. We focus on the popular cumulative sum (CUSUM) type of test statistic since an explicit test statistic is available and very easy to calculate. For example, Kokoszka and Leipus (2000) first considered the CUSUM test of squared asset returns for a single volatility shift by adapting the idea of the CUSUM statistic in detecting mean changes with i.i.d. observations. However, Andreou and Ghysels (2002) reported that the CUSUM test of Kokoszka and Leipus (2000) suffers from size distortions for finite samples due to strong dependence on observations. One may correct such a size distortion by considering the so-called Bartlett long-run variance estimator in the CUSUM test proposed by Berkes *et al.* (2006) and Zhang *et al.* (2007). However, the Bartlett long-run variance estimator is very sensitive to the choice of kernel bandwidth and contributes to low power against LRD as pointed out by Teverovsky *et al.* (1999) and Baek and Pipiras (2012).

To overcome such shortcomings, we adapt the so-called residual-based CUSUM test studied earlier. The residual-based CUSUM test uses the standardized residuals from an estimated GARCH model. Such residuals mimic innovation series in GARCH model; hence, dependence amongst observations can be diminished in the residual-based CUSUM statistic. For example, Lee *et al.* (2004) considered a residual-based CUSUM test of parameter changes in regression models with AR(∞) errors. Their statistic, however, needed tuning parameters for truncation and considered test for no parameter change only. Kulperger and Yu (2005) studied many interesting theoretical properties on the high-moment partial sum of GARCH residuals and proposed residual-based CUSUM test free from tuning parameter selection for truncation. However, their work was still confined to the test for no volatility shifts and always truncated the first observation. de Pooter and van Dijk (2004) conducted extensive simulations for volatility shifts and showed that the residual-based CUSUM test performs reasonably well amongst other methods. In our article, we study theoretical properties of the residual-based CUSUM test under the multiple-volatility-shifts context. Note also that our proposed method does not truncate any observations. The performance of our proposed method is numerically evaluated for various types of volatility models including misspecified models.

The organization of this article is as follows. In Section 2, we introduce our testing procedure for no volatility shifts against LRD. At this stage, the IGARCH model can also be dismissed from consideration. Then, the testing procedure is further examined for a single volatility shift in Section 3, and it leads to natural generalization to known R number of volatility shifts. The performance of our proposed method in the simulation study is reported

  J. Time. Ser. Anal. **36**: 127–153 (2015)
DOI: 10.1111/jtsa.12098

in Section 4. In Section 5, we apply our proposed method to several real financial time series. Conclusions and further discussions can be found in Section 6, and proofs of theorems are provided in Section 7.

2. TEST FOR NO VOLATILITY SHIFTS AGAINST LRD

Here, we consider the testing procedure of no volatility shifts on the observed volatility series $\{r_t\}_{t\in\mathbb{Z}}$ against LRD. In many financial time series, volatility is measured by log-return defined as

$$r_t = \log P_t - \log P_{t-1}, \tag{2}$$

where P_t represents the asset price or stock index at time t. Then, the celebrated GARCH(p,q) model for $\{r_t\}_{t\in\mathbb{Z}}$ satisfies the following relationship

$$r_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega^\circ + \sum_{i=1}^{q} \alpha_i^\circ r_{t-i}^2 + \sum_{j=1}^{p} \beta_j^\circ \sigma_{t-j}^2, \tag{3}$$

where the innovation $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of standard, that is, $\mathbb{E}\epsilon_t = 0$ and $\mathbb{E}\epsilon_t^2 = 1$, i.i.d. random variables. It is also assumed that $\omega^\circ > 0, \alpha_i^\circ \geq 0$ for all $i = 1, \ldots, q$ and $\beta_j^\circ \geq 0$ for all $j = 1, \ldots, p$. The true parameter vector is denoted by $\theta^\circ = (\omega^\circ, \alpha_1^\circ, \ldots, \alpha_q^\circ, \beta_1^\circ, \ldots, \beta_p^\circ)'$. Hereafter, the GARCH model in (3) is also denoted by VS-0 model for shortness's sake.

For the LRD model, it is widely reported that the return series $\{r_t\}_{t\in\mathbb{Z}}$ itself exhibits little autocorrelations, while power transformation of return data such as $\{r_t^2\}_{t\in\mathbb{Z}}$ exhibits pronounced autocorrelations. Therefore, we assume that $\{r_t^2\}_{t\in\mathbb{Z}}$ is an LRD series satisfying (1) for alternative hypothesis. Popular models of such types include the long-memory ARCH (LM-ARCH) model of Ding and Granger (1996), the closely related fractionally integrated GARCH (FIGARCH) model of Baillie *et al.* (1996) (although theoretical justification on long-memory property and the existence of stationary solution is not fully answered, see Giraitis *et al.*, 2000a) and the long-memory linear ARCH (LM-LARCH) model of Giraitis *et al.* (2000b). We can further consider the long-memory stochastic volatility model since it is well known that, under mild assumptions, the power-transformed series $\{|r_t|^\alpha\}_{t\in\mathbb{Z}}$ for any $\alpha > 0$ is again LRD in the sense of (1) with LRD parameter d. We will refer to Surgailis and Viano (2002), Robinson (2001) and references therein for more interesting discussions and generalizations.

Now, we are interested in testing the following hypotheses:

$H_0^{(0)}$: The observed data $\{r_t\}_{t\in\mathbb{Z}}$ follow the VS-0 model.
H_1: The observed data $\{r_t^2\}_{t\in\mathbb{Z}}$ follow the LRD model.

Note that GARCH(p,q) models are SRD because of the geometric ergodicity of random difference equations (e.g. Basrak *et al.*, 2002). In fact, Giraitis *et al.* (2000a) showed that all the ARCH(∞) models, which encapsule the class of GARCH(p,q) models and even hyperbolically decaying coefficients, are SRD under mild assumptions. It is deduced that $\{|r_t|^\alpha\}_{t\in\mathbb{Z}}, \alpha > 0$ is also SRD, to be more precise, geometrically ergodic, since $f(x) = |x|^\alpha$ is a continuous transformation. Therefore, the aforementioned test essentially distinguishes SRD and LRD models. Observe also that the VS-0 model encapsules the IGARCH model, that is, $\sum_{i=1}^q \alpha_i^\circ + \sum_{j=1}^p \beta_j^\circ = 1$, since our model (3) only assumed that parameters $\{\alpha_i^\circ\}$ and $\{\beta_j^\circ\}$ are non-negative. It indicates that our test can be used to rule out the IGARCH model when testing for $H_0^{(0)}$: VS-0.

For the test, one may consider the popular CUSUM test statistic based on $\{r_t\}_{t\in\mathbb{Z}}$, defined as

$$T_n = \frac{1}{\sqrt{n} s_n} \max_{1\leq k\leq n} \left| \sum_{t=1}^{k} r_t^2 - \frac{k}{n} \sum_{t=1}^{n} r_t^2 \right|,$$

DOI: 10.1111/jtsa.12098

where s_n^2 is a suitable estimator of the long-run variance $\sigma^2 = \sum_{h=-\infty}^{\infty} \mathrm{Cov}(r_0^2, r_h^2)$. However, the CUSUM test for GARCH models is reported to suffer from size distortions and low powers in finite samples (e.g. Lee *et al.*, 2004; de Pooter and van Dijk, 2004). This is because of the widely observed fact that long-run variance estimators such as heteroscedasticity and autocorrelation-consistent estimators are sensitive to the choice of kernel bandwidth when correlations between observations are strong as in the case of GARCH(p, q).

Instead, we bypass such deficit by taking uncorrelated observations into account. One such natural candidate is to consider the CUSUM test based on the innovation series $\{\epsilon_t\}$ in (3). That is,

$$\hat{T}_n = \frac{1}{\sqrt{n}\tau} \max_{1 \le k \le n} \left| \sum_{t=1}^{k} \epsilon_t^2 - \frac{k}{n} \sum_{t=1}^{n} \epsilon_t^2 \right|,$$

so the long-run variance is simply given by $\mathrm{Var}(\epsilon_1^2)$ since $\{\epsilon_t\}$ are an i.i.d. sequence of random variables. However, note that $\{\epsilon_t\}$ is unobservable and hence needs to be estimated. First, consider a consistent estimator of θ° such that

$$\sqrt{n} \left\| \hat{\theta} - \theta^\circ \right\| = O_P(1), \tag{4}$$

where $\hat{\theta} = (\hat{\omega}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_q, \hat{\beta}_1, \ldots, \hat{\beta}_p)'$. Then, the innovation series are estimated from (3) by

$$\tilde{\epsilon}_t := \frac{r_t}{\tilde{\sigma}_t\left(\hat{\theta}\right)}, \quad t = 1, \ldots, n, \tag{5}$$

where $\tilde{\sigma}_t^2(\theta)$ is calculated recursively from

$$\tilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2(\theta), \tag{6}$$

where $\theta = (\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$ is the parameter vector and $\{r_t\}$ represents the data at hand, which is a realization of the stationary solution in (3). We will also refer to (5) as the residuals in the context of Kulperger and Yu (2005). Finally, the CUSUM statistic based on the estimated innovation series $\{\tilde{\epsilon}_t\}$ is defined as

$$\tilde{T}_n = \frac{1}{\sqrt{n}\hat{\tau}} \max_{1 \le k \le n} \left| \sum_{t=1}^{k} \tilde{\epsilon}_t^2 - \frac{k}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^2 \right|, \tag{7}$$

where

$$\hat{\tau}^2 = \frac{1}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^4 - \left(\frac{1}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^2 \right)^2$$

is a method-of-moment estimator of $\mathrm{Var}(\epsilon_1^2)$.

It is worth noting that any $\sqrt{n}$-consistent estimator can be used to define the residuals (5). In particular, we employ the Gaussian quasi-maximum likelihood estimate (QMLE) based on the standard normal density, which is shown to be asymptotically normally distributed by Francq and Zakoïan (2004). Also note that, for given GARCH(p, q) orders, recursion (6) requires initial values for $r_0^2, \ldots, r_{1-q}^2$ and $\tilde{\sigma}_0^2, \ldots, \tilde{\sigma}_{1-p}^2$. Following Frank and Zakoïan (2004, p. 612), initial values could be any constant values taken to be fixed, neither random nor

DOI: 10.1111/jtsa.12098

functions of the parameters. In a computational perspective, it is already available in the `fGarch` package in R with all initial values taken to be the sample average of $r_1^2, \ldots, r_n^2$.

Remark 1. Berkes *et al.* (2003) introduced a different way of constructing residuals from an infinite-order moving average representation of GARCH(p, q). In their method, however, it is not possible to obtain residuals up to first $\max(p, q)$ terms. Because (5) is computationally more efficient and simulation results are only little different between two methods, we will use residuals defined as in (5) and (6).

So as to obtain the asymptotic properties, we need the following regularity conditions. Assume that the parameter vector $\theta = (\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$ belongs to the parameter space $\Theta \subset [c_1, c_2]^{p+q+1}$ for some $0 < c_1 < c_2 < \infty$.

(A1) θ° is an interior point of Θ, and Θ is compact.
(A2) The process $\{r_t\}_{t\in\mathbb{Z}}$ is strictly stationary and

$$\sup_{\theta\in\Theta} \sum_{j=1}^{p} \beta_j < 1.$$

Then, under the null hypothesis, we have the following convergence result.

Theorem 1. Assume that (4) and conditions **(A1)** and **(A2)** hold. Then, under $\mathrm{H}_0^{(0)}$, as $n \to \infty$,

$$\tilde{T}_n \xrightarrow{d} \sup_{0\le u\le 1} \left|\mathcal{B}^\circ(u)\right|, \tag{8}$$

where $\mathcal{B}^\circ(u)$ is a standard Brownian bridge.

Remark 2. For GARCH$(1, 1)$ model, **(A2)** holds if a top Lyapunov exponent is negative, or equivalently $E \log(\alpha_1\epsilon_t^2 + \beta_1) < 0$. For a formal definition of a negative top Lyapunov exponent in GARCH(p, q) model and detailed discussion, readers are referred to (3) of Francq and Zakoïan (2004), Giraitis *et al.* (2007) and references therein.

Remark 3. To verify condition (4), the following three additional conditions are assumed by Francq and Zakoïan (2004) together with **(A1)** and **(A2)**. Let $\mathcal{A}_\theta(z) = \sum_{i=1}^{q} \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$. Conventionally, if $q = 0, \mathcal{A}_\theta(z) = 0$, and if $p = 0, \mathcal{B}_\theta(z) = 1$.

(1) ϵ_t^2 is a non-degenerate random variable.
(2) If $p > 0, \mathcal{A}_{\theta^\circ}(z)$ and $\mathcal{B}_{\theta^\circ}(z)$ have no common root, $\mathcal{A}_{\theta^\circ}(1) \neq 0$, and $\alpha_q^\circ + \beta_p^\circ \neq 0$.
(3) $\mathbb{E}\epsilon_t^4 < \infty$.

We will assume that conditions (1)–(3) hold whenever (4) is used in this article. It is also noteworthy that according to Francq and Zakoïan (2007), (4) still holds when the true parameter value is on the boundary of the parameter space, and thus, Theorem 1 could be validated without the first part of Assumption (A1).

3. TEST FOR A SINGLE VOLATILITY SHIFT AGAINST LRD

Once the hypothesis test for no volatility shifts against LRD is rejected, it is natural to consider the testing procedure on volatility shifts for *known* R number of times, although their locations are unknown, against LRD. Hence, the interest of this section is in testing the following hypotheses:

$\mathrm{H}_0^{(R)}$: The observed data $\{r_t\}_{t\in\mathbb{Z}}$ follow the VS-R model.
H_1: The observed data $\{r_t^2\}_{t\in\mathbb{Z}}$ follow the LRD model.

Since the testing procedure can be easily extended to multiple volatility shifts, we will only consider $R = 1$, a single volatility shift, for a clear discussion of our method. Discussions on an *unknown* number of volatility shifts can be found in Section 6.

We first introduce the VS-1 model such that one volatility shift occurs at an unknown location k^* by following the framework used by Kokoszka and Leipus (2000). Consider two GARCH(p,q) models $\{r_{1,t}\}_{t\in\mathbb{Z}}$ and $\{r_{2,t}\}_{t\in\mathbb{Z}}$ from the same innovations, namely,

$$\begin{cases} r_{1,t} = \sigma_{1,t}\epsilon_t,\ \sigma_{1,t}^2 = \omega_1^\circ + \sum_{i=1}^{q}\alpha_{1,i}^\circ r_{1,t-i}^2 + \sum_{j=1}^{p}\beta_{1,j}^\circ \sigma_{1,t-j}^2, \\ r_{2,t} = \sigma_{2,t}\epsilon_t,\ \sigma_{2,t}^2 = \omega_2^\circ + \sum_{i=1}^{q}\alpha_{2,i}^\circ r_{2,t-i}^2 + \sum_{j=1}^{p}\beta_{2,j}^\circ \sigma_{2,t-j}^2, \end{cases} \tag{9}$$

where the innovations $\{\epsilon_t\}_{t\in\mathbb{Z}}$ are standard i.i.d. random variables. Assume that $\omega_m^\circ > 0, \alpha_{m,i}^\circ \geq 0, i = 1,\ldots,q$ and $\beta_{m,j}^\circ \geq 0, j = 1,\ldots,p$ for $m = 1,2$. The true parameter vectors are represented by $\theta_m^\circ = (\omega_m^\circ, \alpha_{m,1}^\circ,\ldots,\alpha_{m,q}^\circ,\beta_{m,1}^\circ,\ldots,\beta_{m,p}^\circ), m = 1,2$, for $\{r_{1,t}\}$ and $\{r_{2,t}\}$ respectively. Then, the VS-1 model refers to the model defined as

$$r_t = \begin{cases} r_{1,t}, & \text{if } 1 \leq t \leq k^*, \\ r_{2,t}, & \text{if } k^* < t \leq n, \end{cases} \tag{10}$$

where $k^* = [n\lambda_0], \lambda_0 \in (0,1)$, is an unknown change point.

So as to construct the test statistic, consider the change-point estimator

$$\hat{k} = \operatorname*{argmax}_{1\leq k\leq n} \left| \frac{k(n-k)}{n^2} \left(\frac{1}{k}\sum_{t=1}^{k} r_t^2 - \frac{1}{n-k}\sum_{t=k+1}^{n} r_t^2 \right) \right| \tag{11}$$

proposed by Kokoszka and Leipus (2000). Then, for a sequence of residuals $\{\tilde{\epsilon}_t, t = 1,\ldots,n\}$ explained later, the test statistic is given by

$$M_n = \max\left\{\tilde{T}_{n,1}, \tilde{T}_{n,2}\right\}, \tag{12}$$

where

$$\tilde{T}_{n,1} = \frac{1}{\hat{\tau}_1}\hat{k}^{-1/2} \max_{1\leq k\leq \hat{k}} \left| \sum_{t=1}^{k}\tilde{\epsilon}_t^2 - \frac{k}{\hat{k}}\sum_{t=1}^{\hat{k}}\tilde{\epsilon}_t^2 \right|, \quad \tilde{T}_{n,2} = \frac{1}{\hat{\tau}_2}\left(n-\hat{k}\right)^{-1/2} \max_{\hat{k}<k\leq n} \left| \sum_{t=\hat{k}+1}^{k}\tilde{\epsilon}_t^2 - \frac{k-\hat{k}}{n-\hat{k}}\sum_{t=\hat{k}+1}^{n}\tilde{\epsilon}_t^2 \right|,$$

and

$$\hat{\tau}_1^2 = \frac{1}{\hat{k}}\sum_{t=1}^{\hat{k}}\tilde{\epsilon}_t^4 - \left(\frac{1}{\hat{k}}\sum_{t=1}^{\hat{k}}\tilde{\epsilon}_t^2\right)^2, \quad \hat{\tau}_2^2 = \frac{1}{n-\hat{k}}\sum_{t=\hat{k}+1}^{n}\tilde{\epsilon}_t^4 - \left(\frac{1}{n-\hat{k}}\sum_{t=\hat{k}+1}^{n}\tilde{\epsilon}_t^2\right)^2.$$

DOI: 10.1111/jtsa.12098

The residuals $\{\tilde{\epsilon}_t, t = 1, \ldots, n\}$ are obtained basically by repeating the procedure in Section 2 to two subsamples segmented before and after the estimated change point $\hat{k}$. To be more specific, let $\tilde{r}_{1,t} = r_t$, $t = 1, \ldots, \hat{k}$, and $\tilde{r}_{2,t} = r_t, t = \hat{k} + 1, \ldots, n$. Define the recursion

$$\tilde{\sigma}^2_{1,t}(\theta_1) = \omega_1 + \sum_{i=1}^{q} \alpha_{1,i} \tilde{r}^2_{1,t-i} + \sum_{j=1}^{p} \beta_{1,j} \tilde{\sigma}^2_{1,t-j}(\theta_1), \quad t = 1, \ldots, \hat{k},$$
$$\tilde{\sigma}^2_{2,t}(\theta_2) = \omega_2 + \sum_{i=1}^{q} \alpha_{2,i} \tilde{r}^2_{2,t-i} + \sum_{j=1}^{p} \beta_{2,j} \tilde{\sigma}^2_{2,t-j}(\theta_2), \quad t = \hat{k} + 1, \ldots, n$$

for given fixed constant initial vectors $\left(\tilde{r}^2_{1,0}, \ldots, \tilde{r}^2_{1,1-q}\right), \left(\tilde{r}^2_{2,\hat{k}}, \ldots, \tilde{r}^2_{2,\hat{k}-q+1}\right), \left(\tilde{\sigma}^2_{1,0}, \ldots, \tilde{\sigma}^2_{1,1-p}\right)$, and $\left(\tilde{\sigma}^2_{2,\hat{k}}, \ldots, \tilde{\sigma}^2_{2,\hat{k}-p+1}\right)$ with parameter vectors denoted by $\theta_m = (\omega_m, \alpha_{m,1}, \ldots, \alpha_{m,q}, \beta_{m,1}, \ldots, \beta_{m,p})', m = 1, 2$. Then, the Gaussian QMLEs for θ_1 and θ_2 are given by

$$\hat{\theta}_1 = \underset{\theta_1 \in \Theta}{\operatorname{argmin}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\theta_1), \quad \hat{\theta}_2 = \underset{\theta_2 \in \Theta}{\operatorname{argmin}} \sum_{t=\hat{k}+1}^{n} \tilde{l}_t(\theta_2), \tag{13}$$

where Θ is a parameter space belonging to $[c_1, c_2]^{p+q+1}$ for some $0 < c_1 < c_2 < \infty$ and

$$\tilde{l}_t(\theta_m) = \frac{r_t^2}{\tilde{\sigma}_t^2(\theta_m)} + \log \tilde{\sigma}_t^2(\theta_m), \quad \tilde{\sigma}_t^2(\theta_m) = \tilde{\sigma}^2_{1,t}(\theta_m) I\left(t \le \hat{k}\right) + \tilde{\sigma}^2_{2,t}(\theta_m) I\left(t > \hat{k}\right), \quad m = 1, 2. \tag{14}$$

Finally, the residuals $\{\tilde{\epsilon}_t, t = 1, \ldots, n\}$ are obtained from the observations $\{r_t\}$ by calculating

$$\tilde{\epsilon}_t = \begin{cases} \dfrac{r_t}{\tilde{\sigma}_{1,t}\left(\hat{\theta}_1\right)}, & t = 1, \ldots, \hat{k}, \\ \dfrac{r_t}{\tilde{\sigma}_{2,t}\left(\hat{\theta}_2\right)}, & t = \hat{k} + 1, \ldots, n. \end{cases} \tag{15}$$

Remark 4. For the change-point estimator, we have used that of Kokoszka and Leipus (2000). In a theoretical perspective, any change-point estimator of the VS-1 model satisfying

$$\hat{k} - k^* = O_P(1) \tag{16}$$

will be sufficient. Theorem 6 of Kokoszka and Leipus (2000) verifies (16) for the change-point estimator in (11).

For the asymptotics of the test statistic (12) under the VS-1 model, namely $H_0^{(1)}$, we assume the following regularity conditions.

(B1) θ_1° and θ_2° are interior points of Θ, which is compact. Moreover, there exist positive constants c_1 and c_2 such that $\Theta \subset [c_1, c_2]^{p+q+1}$.
(B2) For $m = 1, 2$, the processes $\{r_{m,t}\}_{t \in \mathbb{Z}}$ are strictly stationary and

$$\sup_{\theta_m \in \Theta} \sum_{j=1}^{p} \beta_{m,j} < 1.$$

DOI: 10.1111/jtsa.12098

(B3) Suppose that

$$\Delta = \frac{\omega_1^\circ}{1 - \sum_{i=1}^{q} \alpha_{1,i}^\circ - \sum_{j=1}^{p} \beta_{1,j}^\circ} - \frac{\omega_2^\circ}{1 - \sum_{i=1}^{q} \alpha_{2,i}^\circ - \sum_{j=1}^{p} \beta_{2,j}^\circ} \neq 0.$$

We first describe that the parameter estimators in (13) are, in fact, $\sqrt{n}$-consistent estimators for the VS-1 model.

Theorem 2. Suppose that conditions **(B1)–(B3)** hold. Then, for a change-point estimator satisfying (16), the Gaussian QMLEs for θ_1 and θ_2 in (13) are $\sqrt{n}$-consistent estimators

$$\sqrt{n} \left\| \hat{\theta}_m - \theta_m^\circ \right\| = O_P(1), \quad m = 1, 2. \tag{17}$$

Remark 5. We also assume that conditions (1)–(3) in Remark 3 hold for θ_1° and θ_2°. In fact, we will always assume those conditions to hold whenever (17) is used in the proofs of other lemmas and theorems in this article.

For the VS-1 model, we have the following asymptotic result. Note that volatility shifts imply parameter changes, but parameter changes may not necessarily imply volatility shifts. The condition **(B3)** rules out such a case; thus, our test is for volatility shifts due to parameter changes.

Theorem 3. Suppose that the conditions of Theorem 2 hold. Then, for the VS-1 model (10) under $\mathrm{H}_0^{(1)}$, as $n \to \infty$,

$$M_n \xrightarrow{d} \max \left\{ \sup_{0 \le u \le 1} \left| \mathcal{B}_1^\circ(u) \right|, \sup_{0 \le u \le 1} \left| \mathcal{B}_2^\circ(u) \right| \right\}, \tag{18}$$

where $\mathcal{B}_1^\circ$ and $\mathcal{B}_2^\circ$ are independent standard Brownian bridges.

4. SIMULATION STUDY

Here, we discuss the finite-sample performance of our proposed method through Monte Carlo simulations. For volatility shifts, we consider GARCH(1, 1) models with non-negative coefficients satisfying $\alpha + \beta < 1$ for the second-order stationarity. For LRD models, we have considered LM-ARCH, FIGARCH, LM-LARCH and AR fractionally integrated moving average (FARIMA) models with the following specifications.

- LM-ARCH: The conditional variance of LM-ARCH model is given by

$$\sigma_t^2 = \sum_{i=1}^{N} w_i \sigma_{it}^2, \quad \sum_{i=1}^{N} w_i = 1,$$
$$\sigma_{it}^2 = \sigma^2(1 - \alpha_i - \beta_i) + \alpha_i r_{t-1}^2 + \beta_i \sigma_{it-1}^2$$

  for some weights w_i. We set $N = 20$ with equal weights, $\sigma^2 = 1, \beta_i$ follows Beta(5, d) distribution and $\alpha_i = 0.99(1 - \beta_i)$ in simulations.
- FIGARCH$(1, d, 1)$: FIGARCH models conditional variance as

$$(1 - \beta L)\sigma_t^2 = \omega + \left((1 - \beta L) + (1 - \phi L)(1 - L)^d\right) r_t^2,$$

DOI: 10.1111/jtsa.12098

where L is a backshift operator. We set $\omega = 0.6, \beta = 0.1$ and $\phi = 0.2$ in simulations.

- LM-LARCH: Instead of conditional variance, LM-LARCH models conditional standard deviation by

$$\sigma_t = a + \sum_{j=1}^{\infty} b_j r_{t-j},$$

and we considered $a = 0.3, b_1 = (1+d)/8, b_{j+1} = (j+d)/(j+1)b_j$ so that the coefficients $\{b_j\}$ decay similar to FARIMA$(0, d, 0)$ coefficients.

- FARIMA$(1, d, 0)$: We also considered LRD modelling of the return series $\{r_t\}_{t\in\mathbb{Z}}$ by generating Gaussian FARIMA$(1, d', 0)$ models. Under Gaussianity assumptions on $\{r_t\}_{t\in\mathbb{Z}}$, it follows from Taqqu (1979) that the squared return $\{r_t^2\}_{t\in\mathbb{Z}}$ follows the LRD series with parameter $d = 2d' - 1/2$.

All results are based on 1000 replications of sizes 1000, 2000 and 5000 with 5% significance level. Innovation series $\{\epsilon_t\}_{t\in\mathbb{Z}}$ are generated from N(0, 1) otherwise specified.

First, empirical sizes and powers of test for $H_0^{(0)}$: VS-0 against H_1: LRD are presented in Tables I and II. For empirical sizes reported in Table I, we have considered GARCH(1, 1) models with two types of innovations.

Table I. Empirical size of test for no volatility shifts

Innovations	$N(0,1)$			$t(5)$		
(ω, α, β)	$n = 1000$	$n = 2000$	$n = 5000$	$n = 1000$	$n = 2000$	$n = 5000$
$(0.1, 0.1, 0.8)$	0.040	0.038	0.050	0.027	0.042	0.033
$(0.1, 0.1, 0.6)$	0.031	0.058	0.043	0.023	0.025	0.037
$(0.1, 0.1, 0.4)$	0.035	0.032	0.030	0.023	0.030	0.053
$(0.1, 0.2, 0.6)$	0.039	0.044	0.047	0.020	0.034	0.036
$(0.3, 0.1, 0.8)$	0.021	0.031	0.050	0.021	0.027	0.045
$(0.3, 0.1, 0.89)$	0.019	0.021	0.032	0.010	0.026	0.028

GARCH(1, 1) models are considered. GARCH, generalized AR conditional heteroscedasticity.

Table II. Empirical power of testing $H_0^{(0)}$: VS-0 against H_1 : r_t^2 is LRD

d	$n = 1000$	$n = 2000$	$n = 5000$
	LM-ARCH		
0.25	0.333	0.479	0.684
0.35	0.344	0.480	0.615
0.45	0.351	0.468	0.683
	LM-LARCH		
0.25	0.455	0.613	0.699
0.35	0.576	0.698	0.770
0.45	0.561	0.690	0.691
	FIGARCH$(1, d, 1)$		
0.25	0.376	0.547	0.760
0.35	0.369	0.559	0.795
0.45	0.392	0.559	0.736
	FARIMA$(1, d, 0)$		
0.25	0.616	0.707	0.816
0.35	0.804	0.868	0.919
0.45	0.895	0.962	0.973

FARIMA, AR fractionally integrated moving average; FIGARCH, fractionally integrated generalized AR conditional heteroscedasticity; LM-ARCH, long-memory AR conditional heteroscedasticity; LM-LARCH, long-memory linear AR conditional heteroscedasticity. LRD, long-range dependence.

DOI: 10.1111/jtsa.12098

They are i.i.d. N(0,1) and t-distribution with a degree of freedom of 5, denoted by $t(5)$, which is standardized to have zero mean and unit variance. It can be seen that our method achieves excellent empirical sizes over the range of sample sizes and parameters. Even for the case when $\alpha + \beta = 0.1 + 0.89 = 0.99 \approx 1$, no size distortion is observed. It is also observed that empirical sizes are getting closer to the nominal 5% significance level as sample size increases. Even with heavier tails realized by $t(5)$ distributions, there is no size distortion in our proposed method. The empirical power of testing $\mathrm{H}_0^{(0)}$: VS-0 against H_1: LRD is reported in Table II. It is observed that our proposed method provides an acceptable power in all cases considered. It is also observed that the power approaches 1 as sample size increases and as dependence are getting stronger.

Now, we turn our attention to test M_n in (18) for a single volatility shift. Table III shows the empirical size of test for $\mathrm{H}_0^{(1)}$: VS-1, and the empirical power for the LRD alternative is presented in Table IV. Results are consistent with the previous cases of test $\tilde{T}_n$. The empirical size is slightly conservative than the VS-0 testing procedure, but approaches the nominal significance level as sample size increases. The empirical power against LRD is still powerful at this stage for moderate sample sizes. Therefore, it can be deduced that our proposed method successfully distinguishes volatility shifts against LRD.

Finally, we discuss the robustness of our tests to model misspecifications. We first consider the performance of our tests when GARCH orders are misspecified. Table V presents empirical size of test for $\mathrm{H}_0^{(1)}$: VS-1 when GARCH(2, 1) models are generated, but GARCH(1, 1) models are fitted to obtain residuals. It can be observed that all tests achieve a nominal 5% significance level in general even for model order misspecification.

Table III. Empirical size of test for a single volatility shift

$(\omega_1, \alpha_1, \beta_1)$	$n = 1000$	$n = 2000$	$n = 5000$
$(0.1, 0.1, 0.8)$ to $(0.1, 0.1, 0.6)$	0.019	0.042	0.040
$(0.1, 0.1, 0.8)$ to $(0.1, 0.1, 0.4)$	0.016	0.034	0.049
$(0.1, 0.1, 0.8)$ to $(0.1, 0.2, 0.6)$	0.030	0.029	0.040
$(0.1, 0.1, 0.8)$ to $(0.3, 0.1, 0.8)$	0.006	0.018	0.043

Parameters change at midpoint.

Table IV. Empirical power of testing for $\mathrm{H}_0^{(1)}$: VS-1 against H_1 : r_t^2 is LRD

d	$n = 1000$	$n = 2000$	$n = 5000$
		LM-ARCH	
0.25	0.265	0.389	0.593
0.35	0.284	0.407	0.610
0.45	0.311	0.405	0.604
		LM-LARCH	
0.25	0.313	0.486	0.703
0.35	0.457	0.596	0.842
0.45	0.513	0.645	0.863
		FIGARCH$(1, d, 1)$	
0.25	0.295	0.419	0.657
0.35	0.320	0.463	0.759
0.45	0.275	0.465	0.724
		FARIMA$(1, d, 0)$	
0.25	0.671	0.756	0.898
0.35	0.833	0.901	0.968
0.45	0.931	0.983	0.992

FARIMA, AR fractionally integrated moving average; FIGARCH, fractionally integrated generalized AR conditional heteroscedasticity; LM-ARCH, long-memory AR conditional heteroscedasticity; LM-LARCH, long-memory linear AR conditional heteroscedasticity. LRD, long-range dependence.

DOI: 10.1111/jtsa.12098

Table V. Empirical size of test for a single volatility shift when generalized AR conditional heteroscedasticity orders are misspecified

$(\omega_1, \alpha_{11}, \alpha_{12}, \beta_1)$ to $(\omega_2, \alpha_{21}, \alpha_{22}, \beta_2)$	$n = 1000$	$n = 2000$	$n = 5000$
$(0.1, 0.1, 0.1, 0.7)$ to $(0.1, 0.1, 0.1, 0.5)$	0.020	0.047	0.060
$(0.1, 0.1, 0.1, 0.7)$ to $(0.1, 0.1, 0.1, 0.3)$	0.034	0.072	0.078
$(0.1, 0.1, 0.1, 0.7)$ to $(0.1, 0.2, 0.2, 0.4)$	0.026	0.059	0.073
$(0.1, 0.1, 0.1, 0.7)$ to $(0.3, 0.1, 0.1, 0.7)$	0.009	0.029	0.036

Table VI. Empirical size of test for a single volatility shift when asymmetric power AR conditional heteroscedasticity models are misspecified

$(\omega_1, \alpha_1, \gamma_1, \beta_1, \delta_1)$ to $(\omega_2, \alpha_2, \gamma_2, \beta_2, \delta_2)$	$n = 1000$	$n = 2000$	$n = 5000$
$(0.035, 0.2, -0.2, 0.7, 1.6)$ to $(0.035, 0.2, -0.2, 0.5, 1.6)$	0.016	0.033	0.032
$(0.035, 0.2, -0.2, 0.7, 1.6)$ to $(0.035, 0.2, -0.2, 0.3, 1.6)$	0.016	0.028	0.040
$(0.035, 0.2, -0.2, 0.7, 1.6)$ to $(0.035, 0.1, -0.1, 0.7, 1.6)$	0.013	0.027	0.046
$(0.035, 0.2, -0.2, 0.7, 1.6)$ to $(0.035, 0.1, -0.3, 0.7, 1.6)$	0.010	0.029	0.031

Next, we consider the asymmetric power ARCH(1, 1) [A-PARCH(1, 1)] model proposed by Ding *et al.* (1993), which is different from the conditional variance in (3) of GARCH models, as follows:

$$\sigma_t^\delta = \omega + \alpha(|\epsilon_{t-1}| - \gamma\epsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta$$

with $\delta > 0$ and $-1 < \gamma < 1$. This model is well known to reflect the asymmetry of asset returns on volatility (e.g. Paolella, 2001; Giot and Laurent, 2004; Hartz *et al*., 2006). Tables V–VI report the empirical sizes of testing for a single volatility shift when the data are generated from GARCH(2, 1) and A-PARCH(1, 1) models with a parameter change in the midpoint, but GARCH(1, 1) model is fitted. Again, the empirical sizes are close to the nominal 5% significance level as sample size increases. We have observed similar conclusions for over-specified model in GARCH models, but not reported here for brevity. In summary, our simulation study strongly demonstrates the validity of our tests in various types of model misspecifications.

5. REAL DATA APPLICATIONS

We illustrate here how to apply our proposed method to analyse the volatility of a stock index. The volatility is measured by daily log-returns defined in (2), where P_t is the daily stock index of interest. Four different daily stock indices, namely KOSPI200, KOSDAQ, S&P500 and Nikkei 225 from 4 January 1999 to 31 August 2012, are considered. The data are plotted in Figure 1, and sample autocorrelation functions of squared log-returns are depicted in Figure 2. We employ the GARCH(1, 1) model for the conditional volatility model, and parameter estimates are provided in Table VII.

First, we observe very high persistence in volatility on all stock indices from Table VII. Also, Figure 1 reveals possible volatility shifts, for example, around the 1000th observation for KOSDAQ indices. On the other hand, sample autocorrelation functions of squared log-returns in Figure 2 decay slowly in a hyperbolic fashion for all stock indices considered except Nikkei 225. In turn, it substantiates LRD modelling for squared log-returns. This exactly shows that we need to distinguish between volatility shifts and LRD models for a better explanation of observations.

We first apply the test for no volatility shifts, and the results are shown in Table VIII. Under the 5% significance level, we do not reject the null hypothesis of $H_0^{(0)}$: VS-0 for KOSPI200, S&P500 and NIKKEI225 indices.

DOI: 10.1111/jtsa.12098

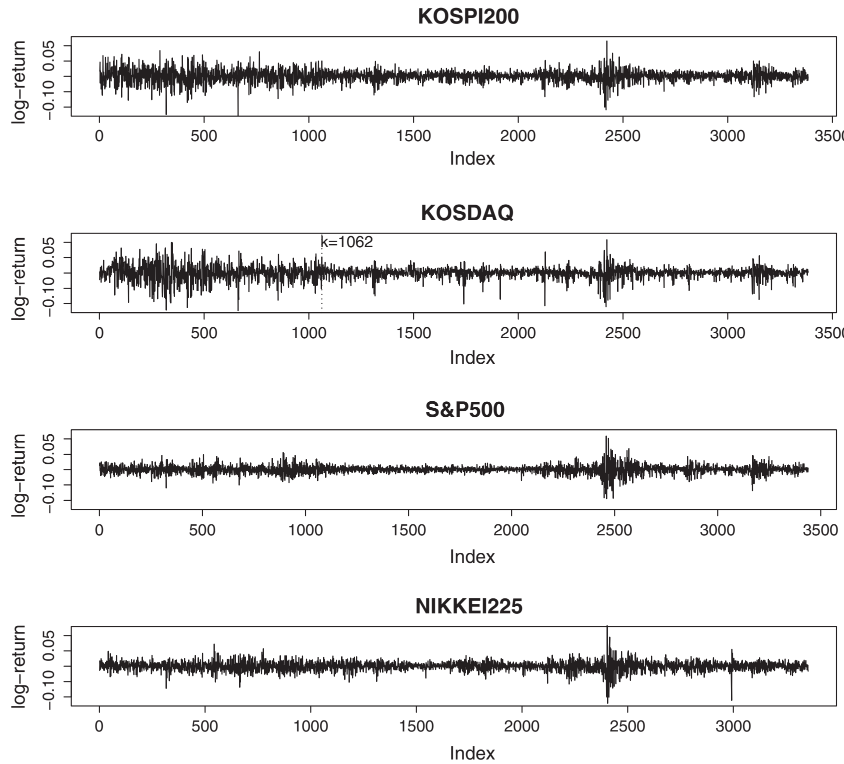

Figure 1. Time-series plot of log-returns of stock indices from 2 January 1999 to 31 August 2012

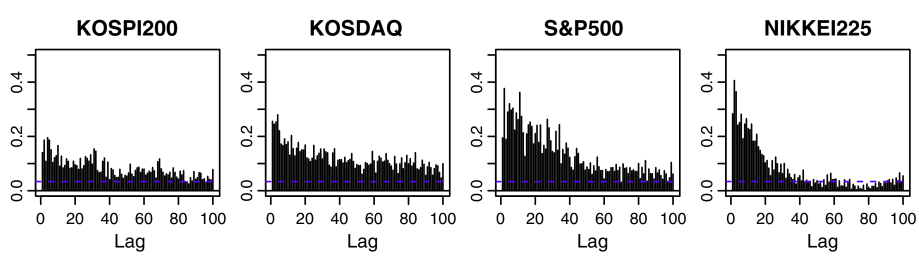

Figure 2. Sample autocorrelation functions for the squared S&P500 log-returns from 2 January 2002 to 30 December 2011

Table VII. Estimated parameters of GARCH(1, 1) models for stock indices and their persistence

	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
KOSPI200	2.04×10^{-6}	0.0751	0.9213	0.9964
KOSDAQ	8.20×10^{-6}	0.1845	0.8129	0.9974
S&P500	1.49×10^{-6}	0.0843	0.9071	0.9914
NIKKEI225	4.65×10^{-6}	0.1064	0.8759	0.9823

GARCH, generalized AR conditional heteroscedasticity.

J. Time. Ser. Anal. **36**: 127–153 (2015)
DOI: 10.1111/jtsa.12098

Table VIII. Test for no volatility shifts and estimated change point $\hat{k}$

	T_n	P-value	$\hat{k}$	Date of change point
KOSPI200	1.2417	0.0916	—	—
KOSDAQ	1.4647	0.0274	1062	28 April 2003
S&P500	0.9433	0.3358	—	—
NIKKEI225	0.8997	0.3931	—	—

Table IX. Estimated parameters and standard deviations from GARCH(1, 1) models together with sample standard deviations of two subsamples divided by change-point $\hat{k} = 1062$ for KOSDAQ

	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha} + \hat{\beta}$	Estimated SD	Sample SD
Subsample 1	4.69×10^{-5}	0.1233	0.8245	0.9478	0.0300	0.0295
Subsample 2	1.60×10^{-5}	0.2515	0.7031	0.9546	0.0188	0.0158

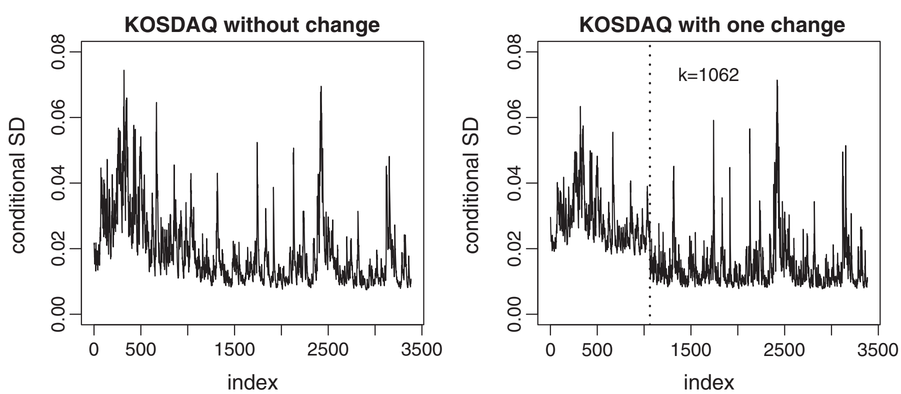

Figure 3. Estimated volatilities, that is, the estimated value of σ_t in the fitted GARCH(1, 1) models of KOSDAQ indices with and without structural changes

Therefore, we conclude that there is no strong evidence against volatility shifts or LRD models for KOSPI200, S&P500 and NIKKEI225 indices. However, we reject the null hypothesis of no volatility shifts for the KOSDAQ index, so we conduct the testing of $H_0^{(1)}$:VS-1 consequently. The change point is estimated by $\hat{k} = 1062$ (28 April 2003) from Kokoszka and Leipus (2000) with test statistics $M_n = 1.2648$, and the corresponding p-value is 0.1565. Hence, under the 5% significance level, we conclude that the VS-1 GARCH(1, 1) model is preferred to LRD models.

The parameter estimates of the final VS-1 GARCH(1, 1) model are presented in Table IX. Observe that the persistence in two subsamples is weakened compared with the estimated VS-0 GARCH(1, 1) model in Table VII. Furthermore, the estimated and sample standard deviations indicate that the volatility shifts from high to low periods for the KOSDAQ index. Figure 3 compares conditional volatility estimated from VS-0 GARCH(1, 1) and VS-1 GARCH(1, 1) with a change point at $\hat{k} = 1062$. It is observed that the volatility from the estimated VS-0 GARCH(1,1) model tends to underestimate high volatility period in VS-1 GARCH(1, 1) and slightly overestimate after the change-point.

DOI: 10.1111/jtsa.12098

6. CONCLUSIONS AND DISCUSSIONS

In this article, we have introduced a simple but effective testing procedure on distinguishing volatility shifts in GARCH against LRD. Our testing procedure is based on the CUSUM statistic from residual series obtained after fitting GARCH models. Since correlations are significantly removed in our test, our test performs better than the usual CUSUM test based on original observations. Simulations study confirms that our testing procedure improves the usual CUSUM test by achieving good size and providing reasonable power against LRD alternatives. In addition, our method is robust to various types of model misspecifications such as GARCH orders, heavier tails and asymmetry.

We have assumed that the number of volatility shifts is known *in priori*. Here, we briefly discuss about estimating the number of change-points by applying the so-called binary segmentation (BS) method. BS method provides a consistent estimate on the number of change-points while computationally very efficient. See, for example, Vostrikova (1981) and Bai (1997). The algorithm is very simple due to its recursive structure. First, start with the whole sample $\{r_1, \ldots, r_n\}$, and perform the testing of $\mathrm{H}_0^{(0)}$: VS-0. The test statistic is given in (7) and we reject the null hypothesis based on (8) for a given level of significance α. Once the hypotheses is rejected, we next test $\mathrm{H}_0^{(1)}$: VS-1. To this end, we estimate the first change-point $\hat{k}_1$ as in (11) and calculate test statistic in (12). If it is not rejected according to asymptotics in (18), we set $\hat{R} = 1$. Otherwise, we split the whole sample into two subsamples, $\left\{r_1, \ldots, r_{\hat{k}_1}\right\}$ and $\left\{r_{\hat{k}_1+1}, \ldots, n\right\}$. Then, we can repeat exactly the same testing procedure to detect a single change-point for each subsample. For example, if the testing on a single change-point is rejected on $\left\{r_1, \ldots, r_{\hat{k}_1}\right\}$, we can split it further into two subsamples and continue testing till no further change-points were found. The final estimated number of change-points is given as the number of subsamples minus 1. Since the change-point estimator of Kokoszka and Leipus (2000) in (11) satisfies (16), similar to Proposition 11 of Bai (1997) where changes in mean model has been considered, under suitable assumptions, we have the following consistency result:

$$P\left(\hat{R} = R\right) \to 1, \quad \text{as}\, n \to \infty, \tag{19}$$

where the size of the test $\alpha = \alpha(n)$ is such that $\alpha(n) \to 0$ but $\liminf_{n\to\infty} n\alpha(n) > 0$.

One may consider the volatility shifts in the sense of the following model

$$r_t = \begin{cases} \sigma_{1,t}\epsilon_t, & \text{if} \quad 1 \le t \le k^* \\ \sigma_{2,t}\epsilon_t, & \text{if} \quad k^* < t \le n, \end{cases} \tag{20}$$

where $\sigma_{m,t}^2 = \omega_m + \sum_{i=1}^{q} \alpha_{m,i} r_{t-i}^2 + \sum_{j=1}^{p} \beta_{m,j}\sigma_{m,t-j}^2$ for $m = 1, 2$, instead of (9) and (10). That is, the volatility beyond k^* is determined by the observations anterior to k^*. The model (20) seems to be more intuitive modelling of volatility shifts. However, we consider volatility shifts in the sense of (9) and (10) by following the framework of Kokoszka and Leipus (2000) due to its theoretical advantages. Also, our simulation study, not reported here for brevity, shows that both models give almost the same results. It would be an interesting future work to examine theoretical properties of change-point estimator of Kokoszka and Leipus (2000) under the model assumption (20).

7. PROOFS

7.1. Proof of Theorem 1

Throughout this section, the symbol $C > 0$ denotes a generic constant which can take different values from line to line. For a column vector $\boldsymbol{x} \in \mathbb{R}^m, m \in \mathbb{N}$, denote $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}'\boldsymbol{x}}$, and for a matrix A, let $\|A\| := \sup\{\|A\boldsymbol{x}\| : \|\boldsymbol{x}\| \le 1\}$. For a random variable X, $\|X\|_2 = \sqrt{E(X^2)}$.

DOI: 10.1111/jtsa.12098

Let $\sigma_t^2(\theta), \theta \in \Theta$, be the strictly stationary, ergodic and non-anticipative solution of

$$\sigma_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2(\theta), \tag{21}$$

where series $\{r_t\}$ are from (3). Observe also that, as it appears in Francq and Zakoïan (2004), $\sigma_t^2(\theta)$ in (21) can be explicitly rewritten as the first entry of the random vector

$$\sum_{k=0}^{\infty} B^k \boldsymbol{c}_{t-k}(\theta), \tag{22}$$

where

$$\boldsymbol{c}_t(\theta) = \begin{pmatrix} \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad B = B(\theta) = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & & \beta_p \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

Then, $\sigma_t^2(\theta^\circ) = \sigma_t^2$ a.s. implies $\theta = \theta^\circ$ from the proof of Theorem 2.1 in Francq and Zakoïan (2004). We first recall the following two lemmas whose proofs can be found in Lemma 2.3 of Berkes *et al.* (2003) and Francq and Zakoïan (2004) on pages 615-616.

Lemma 1. Assume that **(A2)** holds. There exists $s > 0$ such that

$$\mathbb{E}\left|\sigma_0(\theta^\circ)\right|^s < \infty \quad \text{and} \quad \mathbb{E}|r_0|^s < \infty. \tag{23}$$

Furthermore, let $\rho(B)$ be the spectral radius of B, then

$$\sup_{\theta\in\Theta} \rho(B) < 1. \tag{24}$$

Lemma 2. Under **(A1)-(A2)**, there exist a constant $\rho \in (0, 1)$ and a finite random variable $V \geq 0$ such that for every $t \in \mathbb{N}$,

$$\sup_{\theta\in\Theta} \left|\tilde{\sigma}_t^2(\theta) - \sigma_t^2(\theta)\right| \leq \rho^t V.$$

For matrix A, let $A(i, j)$ denote (i, j)-th entry of A. For two matrices A_1 and A_2 with the same dimension, define $A_1 \leq A_2$ if $A_1(i, j) \leq A_2(i, j)$ for every i, j. For $\eta > 0$, let

$$N_\eta(\theta^\circ) = \left\{\theta \in \mathbb{R}^{p+q+1} : \|\theta - \theta^\circ\| < \eta\right\}$$

be a neighborhood of θ°.

Lemma 3. Suppose that **(A1)-(A2)** hold. Then, there exists $\eta > 0$ such that for every $d \in \mathbb{N}$,

$$\mathbb{E}\left\{\sup_{\theta \in N_\eta(\theta^\circ)} \frac{\sigma_t^2(\theta^\circ)}{\sigma_t^2(\theta)}\right\}^d < \infty, \quad \mathbb{E}\left\{\sup_{\theta \in N_\eta(\theta^\circ)} \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta^\circ)}\right\}^d < \infty.$$

Proof
This is a straightforward adaptation of the proof on page 622 of Francq and Zakoïan (2004), hence omitted for brevity. □

Lemma 4. Suppose that **(A1)-(A2)** hold. Then, there exists $\eta > 0$ such that

$$\mathbb{E}\sup_{\theta \in N_\eta(\theta^\circ)} \left\| \frac{1}{\sigma_t^2(\theta)} \frac{\partial}{\partial \theta'} \sigma_t^2(\theta) \right\|^d < \infty, \quad \mathbb{E}\sup_{\theta \in N_\eta(\theta^\circ)} \left\| \frac{1}{\sigma_t^2(\theta)} \frac{\partial^2}{\partial \theta \partial \theta'} \sigma_t^2(\theta) \right\|^d < \infty,$$

for every $d \in \mathbb{N}$.

Proof
The proof is on page 623 of Francq and Zakoïan (2004). □

Let

$$\hat{\epsilon}_t = \frac{r_t}{\sigma_t\left(\hat{\theta}\right)}, \quad \text{for } t = 1, \ldots, n.$$

The following lemmas indicate that $\hat{\epsilon}_t$ is a good proxy of the residual $\tilde{\epsilon}_t$ in (5).

Lemma 5. Under (4) and **(A1)-(A2)**,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left|\tilde{\epsilon}_t^2 - \hat{\epsilon}_t^2\right| = o_P(1), \tag{25}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\hat{\epsilon}_t - \epsilon_t)^2 = o_P(1). \tag{26}$$

Proof
We will only provide the proof of (25) since (26) can be proved similarly. Note that $\min_{1 \le t \le n} \tilde{\sigma}_t^2\left(\hat{\theta}\right)\sigma_t^2\left(\hat{\theta}\right) \ge \hat{\omega}^2$. Observe also from Lemma 2 that

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left|\tilde{\epsilon}_t^2 - \hat{\epsilon}_t^2\right| &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| \frac{r_t^2}{\tilde{\sigma}_t^2\left(\hat{\theta}\right)} - \frac{r_t^2}{\sigma_t^2\left(\hat{\theta}\right)} \right| = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| \frac{1}{\tilde{\sigma}_t^2\left(\hat{\theta}\right)\sigma_t^2\left(\hat{\theta}\right)} \left\{\tilde{\sigma}_t^2\left(\hat{\theta}\right) - \sigma_t^2\left(\hat{\theta}\right)\right\} \right| r_t^2 \\ &\le \left(\frac{1}{\hat{\omega}}\right)^2 \frac{1}{\sqrt{n}} \sum_{t=1}^{n} r_t^2 \left|\tilde{\sigma}_t^2\left(\hat{\theta}\right) - \sigma_t^2\left(\hat{\theta}\right)\right| \le \left(\frac{\sqrt{V}}{\hat{\omega}}\right)^2 \frac{1}{\sqrt{n}} \sum_{t=1}^{\infty} \rho^t r_t^2.\end{aligned}$$

Thus, (25) follows from the fact that $\hat{\omega} \to \omega^\circ$ in probability and $\sum_{t=1}^{\infty} \rho^t r_t^2$ is finite almost surely due to (23). □

DOI: 10.1111/jtsa.12098

Lemma 6. Under **(A1)**-**(A2)**, it holds that

$$\sup_{0\le u\le 1}\left|\frac{2}{\sqrt{n}}\sum_{t=1}^{[nu]}\epsilon_t\left(\hat{\epsilon}_t-\epsilon_t\right)-\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^\circ)}\right|=o_P(1).$$

Proof
Let

$$I_1(u)=\frac{2}{\sqrt{n}}\sum_{t=1}^{[nu]}\left\{\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t\left(\hat{\theta}\right)\left(\sigma_t(\theta^\circ)+\sigma_t\left(\hat{\theta}\right)\right)}-\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{2\sigma_t^2(\theta^\circ)}\right\}\left\{\epsilon_t^2-1\right\},$$

$$I_2(u)=\frac{2}{\sqrt{n}}\sum_{t=1}^{[nu]}\left\{\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t\left(\hat{\theta}\right)\left(\sigma_t(\theta^\circ)+\sigma_t\left(\hat{\theta}\right)\right)}-\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{2\sigma_t^2(\theta^\circ)}\right\},$$

$$I_3(u)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^\circ)}\left\{\epsilon_t^2-1\right\},\quad I_4(u)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^\circ)}.$$

Then, we can show in a similar manner as in the proof of Lemma 5 that

$$\sup_{0\le u\le 1}|I_1(u)|\vee|I_2(u)|=o_P(1).$$

Observe that

$$\frac{2}{\sqrt{n}}\sum_{t=1}^{[nu]}\epsilon_t\left(\hat{\epsilon}_t-\epsilon_t\right)=\frac{2}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t\left(\hat{\theta}\right)\left(\sigma_t(\theta^\circ)+\sigma_t\left(\hat{\theta}\right)\right)}\epsilon_t^2=\frac{2}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t\left(\hat{\theta}\right)\left(\sigma_t(\theta^\circ)+\sigma_t\left(\hat{\theta}\right)\right)}\left\{\epsilon_t^2-1\right\}$$
$$+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t\left(\hat{\theta}\right)\left(\sigma_t(\theta^\circ)+\sigma_t\left(\hat{\theta}\right)\right)}=I_1(u)+I_2(u)+I_3(u)+I_4(u),$$

hence it leads to

$$\sup_{0\le u\le 1}\left|\frac{2}{\sqrt{n}}\sum_{t=1}^{[nu]}\epsilon_t\left(\hat{\epsilon}_t-\epsilon_t\right)-\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^\circ)}\right|=\sup_{0\le u\le 1}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^\circ)}\left\{\epsilon_t^2-1\right\}\right|+o_P(1).$$

Therefore, to complete the proof, it suffices to prove that the leading term is $o_P(1)$. Applying Taylor expansion yields

$$
\begin{aligned}
&\sup_{0\le u\le 1}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^\circ)-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^\circ)}\left\{\epsilon_t^2-1\right\}\right| \\
&= \sup_{0\le u\le 1}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{1}{\sigma_t^2(\theta^\circ)}\left\{\left(\hat{\theta}-\theta^\circ\right)'\frac{\partial\sigma_t^2(\theta^\circ)}{\partial\theta}+\frac{1}{2}\left(\hat{\theta}-\theta^\circ\right)'\frac{\partial^2\sigma_t^2(\xi)}{\partial\theta\partial\theta'}\left(\hat{\theta}-\theta^\circ\right)\right\}\left\{\epsilon_t^2-1\right\}\right| \\
&\le \sqrt{n}\left\|\hat{\theta}-\theta^\circ\right\|\sup_{0\le u\le 1}\left\|\frac{1}{n}\sum_{t=1}^{[nu]}\frac{1}{\sigma_t^2(\theta^\circ)}\frac{\partial\sigma_t^2(\theta^\circ)}{\partial\theta}\left\{\epsilon_t^2-1\right\}\right\| \\
&\quad+\frac{1}{2}n\left\|\hat{\theta}-\theta^\circ\right\|^2\sup_{0\le u\le 1}\left\|\frac{1}{n^{3/2}}\sum_{t=1}^{[nu]}\frac{1}{\sigma_t^2(\theta^\circ)}\frac{\partial^2\sigma_t^2(\xi)}{\partial\theta\partial\theta'}\left\{\epsilon_t^2-1\right\}\right\| \\
&\le \sqrt{n}\left\|\hat{\theta}-\theta^\circ\right\|\sup_{0\le u\le 1}\left\|\frac{1}{n}\sum_{t=1}^{[nu]}\frac{1}{\sigma_t^2(\theta^\circ)}\frac{\partial\sigma_t^2(\theta^\circ)}{\partial\theta}\left\{\epsilon_t^2-1\right\}\right\| \\
&\quad+\frac{1}{2}n\left\|\hat{\theta}-\theta^\circ\right\|^2\frac{1}{n^{3/2}}\sum_{t=1}^{n}\left\|\frac{1}{\sigma_t^2(\theta^\circ)}\frac{\partial^2\sigma_t^2(\xi)}{\partial\theta\partial\theta'}\right\|\left|\epsilon_t^2-1\right|,
\end{aligned}
$$

where every $\xi=\xi(n,t)$ is an intermediate point between $\hat{\theta}$ and θ°.

From (4), it is shown that

$$
\sqrt{n}\left\|\hat{\theta}-\theta^\circ\right\|\sup_{0\le u\le 1}\left\|\frac{1}{n}\sum_{t=1}^{[nu]}\frac{1}{\sigma_t^2(\theta^\circ)}\frac{\partial\sigma_t^2(\theta^\circ)}{\partial\theta}\left\{\epsilon_t^2-1\right\}\right\|=o_P(1),
$$

since the supremum is $o_P(1)$ by the ergodic theorem. Similarly as in the proof of Lemma 5, it can be shown that

$$
n\left\|\hat{\theta}-\theta^\circ\right\|^2\frac{1}{n^{3/2}}\sum_{t=1}^{n}\left\|\frac{1}{\sigma_t^2(\theta^\circ)}\frac{\partial^2\sigma_t^2(\xi)}{\partial\theta\partial\theta'}\right\|\left|\epsilon_t^2-1\right|=o_P(1),
$$

so it completes the proof. □

Proof of Theorem 1

It can be easily seen that $\hat{\tau}\xrightarrow{p}\tau$. Together with (25), observe that

$$
\begin{aligned}
\tilde{T}_n &= \frac{1}{\sqrt{n}\hat{\tau}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}\tilde{\epsilon}_t^2-\frac{k}{n}\sum_{t=1}^{n}\tilde{\epsilon}_t^2\right|=\frac{1}{\sqrt{n}\hat{\tau}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\tilde{\epsilon}_t^2-u\sum_{t=1}^{n}\tilde{\epsilon}_t^2\right|+\sup_{0\le u\le 1}\left|u-\frac{[nu]}{n}\right|\frac{1}{\sqrt{n}\hat{\tau}}\sum_{t=1}^{n}\tilde{\epsilon}_t^2 \\
&= \frac{1}{\sqrt{n}\tau}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\tilde{\epsilon}_t^2-u\sum_{t=1}^{n}\tilde{\epsilon}_t^2\right|+o_P(1).
\end{aligned}
$$

From the invariance principle for partial sums for an i.i.d. sequence and continuous mapping theorem, it further leads to

$$
\frac{1}{\sqrt{n}\tau}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\epsilon_t^2-u\sum_{t=1}^{n}\epsilon_t^2\right|\xrightarrow{d}\sup_{0\le u\le 1}\left|\mathcal{B}^\circ(u)\right|,\quad\text{as } n\to\infty, \tag{27}
$$

DOI: 10.1111/jtsa.12098

where $\mathcal{B}^{\circ}$ denotes a standard Brownian bridge. Moreover, note that

$$\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\tilde{\epsilon}_t^2-u\sum_{t=1}^{n}\tilde{\epsilon}_t^2-\sum_{t=1}^{[nu]}\epsilon_t^2+u\sum_{t=1}^{n}\epsilon_t^2\right|\le\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\tilde{\epsilon}_t^2-\sum_{t=1}^{[nu]}\epsilon_t^2-\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|$$
$$+\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}-u\sum_{t=1}^{n}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|.$$

Hence, it is sufficient to show that

$$\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\tilde{\epsilon}_t^2-\sum_{t=1}^{[nu]}\epsilon_t^2-\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|=o_P(1), \tag{28}$$

and

$$\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}-u\sum_{t=1}^{n}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|=o_P(1). \tag{29}$$

Observe first from (25) that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\tilde{\epsilon}_t^2-\sum_{t=1}^{[nu]}\epsilon_t^2-\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|\\
&=\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\tilde{\epsilon}_t^2-\sum_{t=1}^{[nu]}\hat{\epsilon}_t^2+\sum_{t=1}^{[nu]}\hat{\epsilon}_t^2-\sum_{t=1}^{[nu]}\epsilon_t^2-\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|\\
&\le\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|\tilde{\epsilon}_t^2-\hat{\epsilon}_t^2\right|+\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\hat{\epsilon}_t^2-\sum_{t=1}^{[nu]}\epsilon_t^2-\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|\\
&=\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[nu]}\left(\hat{\epsilon}_t-\epsilon_t\right)^2+2\sum_{t=1}^{[nu]}\epsilon_t\left(\hat{\epsilon}_t-\epsilon_t\right)-\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|+o_P(1)\\
&\le\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\hat{\epsilon}_t-\epsilon_t\right)^2+\frac{1}{\sqrt{n}}\sup_{0\le u\le 1}\left|2\sum_{t=1}^{[nu]}\epsilon_t\left(\hat{\epsilon}_t-\epsilon_t\right)-\sum_{t=1}^{[nu]}\frac{\sigma_t^2(\theta^{\circ})-\sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^{\circ})}\right|+o_P(1).
\end{aligned}$$

Therefore, (28) is established by (26) and Lemma 6.

DOI: 10.1111/jtsa.12098

Next, there exists intermediate points $\xi = \xi(n,t)$ between $\widehat{\theta}$ and θ° such that

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sup_{0\le u\le 1} \left| \sum_{t=1}^{[nu]} \frac{\sigma_t^2(\theta^\circ) - \sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^\circ)} - u \sum_{t=1}^{n} \frac{\sigma_t^2(\theta^\circ) - \sigma_t^2\left(\hat{\theta}\right)}{\sigma_t^2(\theta^\circ)} \right| \\
&\le \sqrt{n} \left\| \hat{\theta} - \theta^\circ \right\| \cdot \sup_{0\le u\le 1} \left\| \frac{1}{n} \sum_{t=1}^{[nu]} \frac{1}{\sigma_t^2(\theta^\circ)} \frac{\partial \sigma_t^2(\theta^\circ)}{\partial\theta} - \frac{u}{n} \sum_{t=1}^{n} \frac{1}{\sigma_t^2(\theta^\circ)} \frac{\partial \sigma_t^2(\theta^\circ)}{\partial\theta} \right\| \\
&\quad + \sqrt{n} \left\| \hat{\theta} - \theta^\circ \right\|^2 \sup_{0\le u\le 1} \frac{1}{n} \left\| (1-u) \sum_{t=1}^{[nu]} \frac{1}{\sigma_t^2(\theta^\circ)} \frac{\partial^2 \sigma_t^2(\xi)}{\partial\theta\partial\theta'} - u \sum_{t=[nu]+1}^{n} \frac{1}{\sigma_t^2(\theta^\circ)} \frac{\partial^2 \sigma_t^2(\xi)}{\partial\theta\partial\theta'} \right\|,
\end{aligned}$$

where the first term converges in probability to 0 by ergodic theorem and (4). Now, note that the second term is bounded by

$$\frac{1}{\sqrt{n}} n \left\| \hat{\theta} - \theta^\circ \right\|^2 \cdot \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{1}{\sigma_t^2(\theta^\circ)} \frac{\partial^2 \sigma_t^2(\xi)}{\partial\theta\partial\theta'} \right\|.$$

From Lemmas 3 and 4, we can take a neighborhood $N(\theta^\circ)$ of θ° such that

$$\mathbb{E} \sup_{\theta\in N(\theta^\circ)} \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta^\circ)} \sup_{\theta\in N(\theta^\circ)} \left\| \frac{1}{\sigma_t^2(\theta)} \frac{\partial^2 \sigma_t^2(\theta)}{\partial\theta\partial\theta'} \right\| < \infty,$$

which, together with (4) and the ergodic theorem, implies that with probability tending to 1

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \left\| \frac{1}{\sigma_t^2(\theta^\circ)} \frac{\partial^2 \sigma_t^2(\xi)}{\partial\theta\partial\theta'} \right\| &\le \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta\in N(\theta^\circ)} \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta^\circ)} \sup_{\theta\in N(\theta^\circ)} \left\| \frac{1}{\sigma_t^2(\theta)} \frac{\partial^2 \sigma_t^2(\theta)}{\partial\theta\partial\theta'} \right\| \\
&\le \mathbb{E} \sup_{\theta\in N(\theta^\circ)} \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta^\circ)} \sup_{\theta\in N(\theta^\circ)} \left\| \frac{1}{\sigma_t^2(\theta)} \frac{\partial^2 \sigma_t^2(\theta)}{\partial\theta\partial\theta'} \right\|.
\end{aligned}$$

Therefore, the second term converges in probability to 0. This completes the proof. □

7.2. Proof of Theorems 2 and 3

We assume (10), (16), and **(B1)**-**(B3)**. Let $A_M = A_M(n) = \left\{ \left| \hat{k} - k^* \right| \ge M \right\}$. Then, (16) implies that for any given $\epsilon > 0$ there exists $M \in \mathbb{N}$ such that $P(A_M) \le \epsilon$ for sufficiently large n. As in the case of Theorem 1, we define $\sigma_{m,t}^2(\theta_m), \theta_m \in \Theta, m = 1,2$, be the strictly stationary, ergodic and non-anticipative solution of the recursion

$$\sigma_{m,t}^2(\theta_m) = \omega_m + \sum_{i=1}^{q} \alpha_{m,i} r_{m,t-i}^2 + \sum_{j=1}^{p} \beta_{m,j} \sigma_{m,t-j}^2(\theta_m), \quad m = 1,2,$$

DOI: 10.1111/jtsa.12098

where $\{r_{m,t}\}$ satisfy (9). Denote

$$l_{m,t}(\theta_m) = \frac{r_{m,t}^2}{\sigma_{m,t}^2(\theta_m)} + \log \sigma_{m,t}^2(\theta_m), \quad m = 1, 2.$$

We first present some lemmas for the proof of Theorem 2 in the below.

Lemma 7. There exist $\rho \in (0, 1)$ and a strictly stationary sequence $\{V_t\}$ such that for $\bar{k} = k^* + M$ with $M \in \mathbb{N}$,

$$\max_{\bar{k} < t \leq n} \frac{1}{\rho^{t-\bar{k}}} \sup_{\theta_2 \in \Theta} \left| \tilde{\sigma}_t^2(\theta_2) - \sigma_{2,t}^2(\theta_2) \right| \leq V_{\bar{k}}, \tag{30}$$

and

$$\max_{\bar{k} < t \leq n} \frac{1}{\rho^{t-\bar{k}}} \sup_{\theta_2 \in \Theta} \left\| \frac{\partial \tilde{\sigma}_t^2(\theta_2)}{\partial \theta_2} - \frac{\partial \sigma_{2,t}^2(\theta_2)}{\partial \theta_2} \right\| \vee \sup_{\theta_2 \in \Theta} \left\| \frac{\partial^2 \tilde{\sigma}_t^2(\theta_2)}{\partial \theta_2 \partial \theta_2'} - \frac{\partial^2 \sigma_{2,t}^2(\theta_2)}{\partial \theta_2 \partial \theta_2'} \right\| \leq V_{\bar{k}}, \tag{31}$$

when $\hat{k} < \bar{k}$.

Proof
Since $\tilde{\sigma}_t^2(\theta_2) = \tilde{\sigma}_{2,t}^2(\theta_2)$ and $\tilde{r}_{2,t} = r_{2,t}$ for $t \geq \bar{k}$ when $\hat{k} < \bar{k}$, (30) and (31) can be proved in similar ways as in the arguments up to (4.6) and (4.33) in Francq and Zakoïan (2004) respectively. □

Lemma 8. For $m = 1, 2$, there exists a strictly stationary sequence $\{U_t\}$ such that

$$\sup_{\theta_m \in \Theta} \left| \tilde{l}_t(\theta_m) \right| \leq U_t, \tag{32}$$

$$\sup_{\theta_m \in \Theta} \left\| \frac{\partial}{\partial \theta_m} \tilde{l}_t(\theta_m) \right\| \vee \sup_{\theta_m \in \Theta} \left\| \frac{\partial^2}{\partial \theta_m \partial \theta_m'} \tilde{l}_t(\theta_m) \right\| \leq U_t. \tag{33}$$

Proof
The proof of (32) is provided for $m = 1$, since the other case and (33) can be proved in a similar manner. Since $\tilde{\sigma}_t^2(\theta_1) \geq \omega = \min(\omega_1, \omega_2)$, we have that

$$\left| \tilde{l}_t(\theta_1) \right| = \left| \frac{r_t^2}{\tilde{\sigma}_t^2(\theta_1)} + \log \frac{\tilde{\sigma}_t^2(\theta_1)}{\omega} + \log \omega \right| \leq \frac{r_{1,t}^2 \vee r_{2,t}^2}{\omega} + \frac{\tilde{\sigma}_t^2(\theta_1)}{\omega} - 1 + |\log \omega|.$$

Moreover, there exist $C > 0$ and $\rho \in (0, 1)$ such that

$$\sup_{\theta_1 \in \Theta} \tilde{\sigma}_t^2(\theta_1) \leq C \left(1 + \sum_{i=1}^{\infty} \rho^{i-1} \left(r_{1,t-i}^2 \vee r_{2,t-i}^2 \right) \right).$$

Hence, the proof is completed. □

DOI: 10.1111/jtsa.12098

Proof of Theorem 2
We will only prove the case when $m = 1$,

$$\sqrt{n}\left\|\hat{\theta}_1 - \theta_1^\circ\right\| = O_P(1), \tag{34}$$

since the other case, $i = 2$, comes similarly due to Lemmas 7-8. First, we will show that

$$\left\|\hat{\theta}_1 - \theta_1^\circ\right\| = o_P(1). \tag{35}$$

By following the proof of Theorem 2.1 in Francq and Zakoïan (2004), we already have that $l_{1,0}\left(\theta_1^\circ\right)$ is integrable and $\mathbb{E}\left\{l_{1,0}(\theta_1)\right\} > \mathbb{E}\left\{l_{1,0}\left(\theta_1^\circ\right)\right\}$ for every $\theta_1 \neq \theta_1^\circ$. Moreover, we claim that

$$\frac{1}{k^* \wedge \hat{k}} \sum_{t=1}^{k^* \wedge \hat{k}} \tilde{l}_t\left(\theta_1^\circ\right) \xrightarrow{p} \mathbb{E}\left\{l_{1,0}\left(\theta_1^\circ\right)\right\}. \tag{36}$$

Note first that $\left(k^* \wedge \hat{k}\right)/k^* \xrightarrow{p} 1$ and the argument up to (4.7) in Francq and Zakoïan (2004) implies that

$$\frac{1}{k^* \wedge \hat{k}} \sum_{t=1}^{k^* \wedge \hat{k}} \left|\tilde{l}_t\left(\theta_1^\circ\right) - l_{1,t}\left(\theta_1^\circ\right)\right| \xrightarrow{p} 0.$$

From the ergodic theorem, we have that

$$\frac{1}{k^*} \sum_{t=1}^{k^*} l_{1,t}\left(\theta_1^\circ\right) \longrightarrow \mathbb{E}\left\{l_{1,0}\left(\theta_1^\circ\right)\right\}, \quad \text{a.s.}$$

Again, from $\left(k^* \wedge \hat{k}\right)/k^* \xrightarrow{p} 1$, (36) follows from

$$\frac{1}{k^* \wedge \hat{k}} \sum_{t=1}^{k^* \wedge \hat{k}} l_{1,t}\left(\theta_1^\circ\right) \xrightarrow{p} \mathbb{E}\left\{l_{1,0}\left(\theta_1^\circ\right)\right\}.$$

To complete the proof, it suffices to show that for $\theta_1 \neq \theta_1^\circ$, there exist a neighborhood $N(\theta_1)$ of θ_1 such that

$$\inf_{\vartheta \in N(\theta_1)} \frac{1}{\hat{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\vartheta) > \mathbb{E}\left\{l_{1,0}\left(\theta_1^\circ\right)\right\}$$

holds with probability tending to 1, and

$$\frac{1}{\hat{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t\left(\theta_1^\circ\right) \xrightarrow{p} \mathbb{E}\left\{l_{1,0}\left(\theta_1^\circ\right)\right\}. \tag{37}$$

DOI: 10.1111/jtsa.12098

Let $\theta_1 \neq \theta_1^\circ$ and $\tilde{k} = k^* \wedge \hat{k}$. As in the proof of Theorem 2.1 in Francq and Zakoïan (2004), we can take a neighborhood $N(\theta_1)$ such that with probability tending to 1

$$\inf_{\vartheta \in N(\theta_1)} \frac{1}{\tilde{k}} \sum_{t=1}^{\tilde{k}} \tilde{l}_t(\vartheta) > \mathbb{E}\left\{l_{1,0}\left(\theta_1^\circ\right)\right\}.$$

Since $k^*/\hat{k} \to 1$ in probability, observe that

$$\begin{aligned}\inf_{\vartheta \in N(\theta_1)} \frac{1}{\hat{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\vartheta) &= \frac{\tilde{k}}{\hat{k}} \inf_{\vartheta \in N(\theta_1)} \frac{1}{\tilde{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\vartheta) = \frac{\tilde{k}}{\hat{k}} \inf_{\vartheta \in N(\theta_1)} \left\{ \frac{1}{\tilde{k}} \sum_{t=1}^{\tilde{k}} \tilde{l}_t(\vartheta) - \frac{1}{\tilde{k}} \sum_{t=1}^{\tilde{k}} \tilde{l}_t(\vartheta) + \frac{1}{\tilde{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\vartheta) \right\} \\ &\geq \frac{\tilde{k}}{\hat{k}} \inf_{\vartheta \in N(\theta_1)} \frac{1}{\tilde{k}} \sum_{t=1}^{\tilde{k}} \tilde{l}_t(\vartheta) + \frac{\tilde{k}}{\hat{k}} \inf_{\vartheta \in N(\theta_1)} \left\{ \frac{1}{\tilde{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\vartheta) - \frac{1}{\tilde{k}} \sum_{t=1}^{\tilde{k}} \tilde{l}_t(\vartheta) \right\}.\end{aligned}$$

We will show that

$$\sup_{\theta_1 \in \Theta} \left| \frac{1}{\tilde{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\theta_1) - \frac{1}{\tilde{k}} \sum_{t=1}^{\tilde{k}} \tilde{l}_t(\theta_1) \right| = o_P(1). \tag{38}$$

For $\epsilon > 0$ and $M > 0$, we have from Lemma 8 that

$$\begin{aligned}&P\left(\sup_{\theta_1 \in \Theta} \left| \frac{1}{\tilde{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\theta_1) - \frac{1}{\tilde{k}} \sum_{t=1}^{\tilde{k}} \tilde{l}_t(\theta_1) \right| > \epsilon \right) \\ &\leq P\left(\frac{C}{k^*} \sum_{t=k^*+1}^{k^*+M} \sup_{\theta_1 \in \Theta} \left| \tilde{l}_t(\theta_1) \right| > \epsilon \right) + P(A_M) \leq P\left(\frac{C}{k^*} \sum_{t=1}^{M} U_t > \epsilon \right) + P(A_M).\end{aligned}$$

Therefore,

$$\limsup_{n \to \infty} P\left(\sup_{\theta_1 \in \Theta} \left| \frac{1}{\tilde{k}} \sum_{t=1}^{\hat{k}} \tilde{l}_t(\theta_1) - \frac{1}{\tilde{k}} \sum_{t=1}^{\tilde{k}} \tilde{l}_t(\theta_1) \right| > \epsilon \right) \leq \limsup_{n \to \infty} P(A_M),$$

and by letting $M \to \infty$, we obtain (38), which implies (37) by (36). Thus, (35) holds.

Next, to show (34), we will verify asymptotic normality of $\hat{\theta}_1$. Applying Taylor expansion of the score vector at θ_1° gives that

$$\mathbf{0} = \frac{1}{\sqrt{k^*}} \sum_{t=1}^{\hat{k}} \frac{\partial}{\partial \theta_1} \tilde{l}_t\left(\hat{\theta}_1\right) = \frac{1}{\sqrt{k^*}} \sum_{t=1}^{\hat{k}} \frac{\partial}{\partial \theta_1} \tilde{l}_t\left(\theta_1^\circ\right) + \left\{ \frac{1}{k^*} \sum_{t=1}^{\hat{k}} \frac{\partial^2}{\partial \theta_1 \partial \theta_1'} \tilde{l}_t(\xi) \right\} \sqrt{k^*}\left(\hat{\theta}_1 - \theta_1^\circ\right),$$

DOI: 10.1111/jtsa.12098

where ξ is an intermediate point between $\hat{\theta}_1$ and θ_1°. We have

$$\frac{1}{\sqrt{k^*}}\sum_{t=1}^{\hat{k}}\frac{\partial}{\partial\theta_1}\tilde{l}_t\left(\theta_1^\circ\right)-\frac{1}{\sqrt{k^*}}\sum_{t=1}^{\tilde{k}}\frac{\partial}{\partial\theta_1}\tilde{l}_t\left(\theta_1^\circ\right)=o_P(1), \tag{39}$$

$$\frac{1}{k^*}\sum_{t=1}^{\hat{k}}\frac{\partial^2}{\partial\theta_1\partial\theta_1'}\tilde{l}_t(\xi)-\frac{1}{k^*}\sum_{t=1}^{\tilde{k}}\frac{\partial^2}{\partial\theta_1\partial\theta_1'}\tilde{l}_t(\xi)=o_P(1), \tag{40}$$

which can be checked in a similar fashion as in proving (38). Moreover, we claim that

$$\frac{1}{\sqrt{k^*}}\sum_{t=1}^{\tilde{k}}\frac{\partial}{\partial\theta_1}\tilde{l}_t\left(\theta_1^\circ\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\left(\mathbb{E}\epsilon_0^4-1\right)\mathbb{E}\frac{\partial^2}{\partial\theta_1\partial\theta_1'}l_{1,0}\left(\theta_1^\circ\right)\right) \tag{41}$$

and

$$\frac{1}{k^*}\sum_{t=1}^{\tilde{k}}\frac{\partial^2}{\partial\theta_1\partial\theta_1'}\tilde{l}_t(\xi)\xrightarrow{p}\mathbb{E}\frac{\partial^2}{\partial\theta_1\partial\theta_1'}l_{1,0}\left(\theta_1^\circ\right) \tag{42}$$

where the limit is positive definite. First, we present a verification of (41). It follows from the argument up to (4.35) in Francq and Zakoïan (2004) that

$$\frac{1}{\sqrt{k^*}}\sum_{t=1}^{\tilde{k}}\left\|\frac{\partial}{\partial\theta_1}\tilde{l}_t\left(\theta_1^\circ\right)-\frac{\partial}{\partial\theta_1}l_{1,t}\left(\theta_1^\circ\right)\right\|\xrightarrow{p}0. \tag{43}$$

Let $\lambda\in\mathbb{R}^{p+q+1}$ be a column vector. Note that $\left\{\lambda'\frac{\partial}{\partial\theta_1}l_{1,t}\left(\theta_1^\circ\right):t=1,2,\ldots\right\}$ are stationary and square integrable martingale differences with respect to filtration $\{\sigma(\epsilon_t,\epsilon_{t-1},\ldots):t=0,1,\ldots\}$. Thus, from the functional central limit theorem for martingale differences (cf. Theorem 18.2 in Billingsley (1999)), we have that

$$\frac{1}{\sqrt{k^*}}\sum_{t=1}^{[k^*u]}\lambda'\frac{\partial}{\partial\theta_1}l_{1,t}\left(\theta_1^\circ\right)\xrightarrow{d}\sqrt{\upsilon(\lambda)}\,\mathcal{B}(u)\quad\text{in } D[0,1].$$

where $\upsilon(\lambda)=\mathbb{E}\left\{\lambda'\frac{\partial}{\partial\theta_1}l_{1,0}\left(\theta_1^\circ\right)\right\}^2$. Moreover, since $\tilde{k}/k^*\xrightarrow{p}1$, we obtain from Theorem 14.4 in Billingsley (1999) that

$$\frac{1}{\sqrt{k^*}}\sum_{t=1}^{[\tilde{k}u]}\lambda'\frac{\partial}{\partial\theta_1}l_{1,t}\left(\theta_1^\circ\right)\xrightarrow{d}\sqrt{\upsilon(\lambda)}\,\mathcal{B}(u)\quad\text{in } D[0,1],$$

especially,

$$\frac{1}{\sqrt{k^*}}\sum_{t=1}^{\tilde{k}}\lambda'\frac{\partial}{\partial\theta_1}l_{1,t}\left(\theta_1^\circ\right)\xrightarrow{d}\mathcal{N}(0,\upsilon(\lambda))\quad\text{for every } \lambda\in\mathbb{R}^{p+q+1}. \tag{44}$$

Furthermore, since

$$\mathbb{E}\frac{\partial}{\partial\theta_1} l_{1,0}\left(\theta_1^\circ\right)\frac{\partial}{\partial\theta_1'} l_{1,0}\left(\theta_1^\circ\right) = \left(\mathbb{E}\epsilon_0^4 - 1\right)\mathbb{E}\frac{\partial^2}{\partial\theta_1\partial\theta_1'} l_{1,0}\left(\theta_1^\circ\right),$$

(cf. (4.23) in Francq and Zakoïan (2004)), we verify (41) by (43) and (44). We can also verify (42) in a similar fashion to (36), since $k^*/\tilde{k} \xrightarrow{p} 1$ and for any shrinking neighborhood $\mathcal{V}$ of θ_1°,

$$\sup_{\theta_1\in\mathcal{V}}\left\|\frac{1}{k^*}\sum_{t=1}^{k^*}\frac{\partial^2}{\partial\theta_1\partial\theta_1'} l_{1,t}(\theta_1) - \mathbb{E}\frac{\partial^2}{\partial\theta_1\partial\theta_1'} l_{1,0}\left(\theta_1^\circ\right)\right\| \longrightarrow 0 \quad \text{a.s.},$$

(cf. steps (iv) and (vi) in the proof of Theorem 2.2 of Francq and Zakoïan (2004)). Hence,

$$\sqrt{k^*}\left(\hat{\theta}_1 - \theta_1^\circ\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(\mathbb{E}\epsilon_0^4 - 1\right)\left\{\mathbb{E}\frac{\partial^2}{\partial\theta_1\partial\theta_1'} l_{1,0}\left(\theta_1^\circ\right)\right\}^{-1}\right),$$

which completes the proof. □

Proof of Theorem 3

In light of Theorems 1 and 2, once $\hat{k}$ is replaced by the true value k^*, convergence to two independent Brownian bridges is a standard argument from a scaling property of a Brownian motion, that is $c^{-1/2}\mathcal{B}(cu)$ for $0 \le u \le 1$ is again a Brownian motion, and the independent increments of a Brownian motion. Hence, here we will complete the proof by proving the following facts:

$$\frac{1}{\sqrt{k^*}}\max_{1\le k\le\hat{k}}\left|\sum_{t=1}^{k}\tilde{\epsilon}_t^2 - \frac{k}{\hat{k}}\sum_{t=1}^{\hat{k}}\tilde{\epsilon}_t^2\right| = \frac{1}{\sqrt{k^*}}\max_{1\le k\le k^*}\left|\sum_{t=1}^{k}\epsilon_t^2 - \frac{k}{k^*}\sum_{t=1}^{k^*}\epsilon_t^2\right| + o_P(1) \tag{45}$$

and

$$\max_{\hat{k}<k\le n}\left|\sum_{t=\hat{k}+1}^{k}\tilde{\epsilon}_t^2 - \frac{k-\hat{k}}{n-\hat{k}}\sum_{t=\hat{k}+1}^{n}\tilde{\epsilon}_t^2\right| = \max_{k^*<k\le n}\left|\sum_{t=k^*+1}^{k}\epsilon_t^2 - \frac{k-k^*}{n-k^*}\sum_{t=k^*+1}^{n}\epsilon_t^2\right| + o_P\left(\sqrt{n}\right). \tag{46}$$

Note that (45) follows from that

$$\frac{1}{\sqrt{k^*}}\sup_{0\le u\le 1}\left|\sum_{t=1}^{[\hat{k}u]\wedge k^*}\tilde{\epsilon}_t^2 - u\sum_{t=1}^{\hat{k}\wedge k^*}\tilde{\epsilon}_t^2 - \sum_{t=1}^{[\hat{k}u]\wedge k^*}\epsilon_t^2 + u\sum_{t=1}^{\hat{k}\wedge k^*}\epsilon_t^2\right| = o_P(1),$$

which can be verified as in the proof of Theorem 1, and

$$\frac{1}{\sqrt{k^*}}\sup_{0\le u\le 1}\left|\sum_{t=[\hat{k}u]\wedge k^*+1}^{[\hat{k}u]}\tilde{\epsilon}_t^2 - u\sum_{t=\hat{k}\wedge k^*+1}^{\hat{k}}\tilde{\epsilon}_t^2\right| = o_P(1)$$

by applying (16) and

$$\tilde{\epsilon}_t^2 \leq \frac{r_{2,t}^2}{\inf\{\omega : \theta \in \Theta\}} \quad \text{for } t > k^*.$$

The latter relationship (46) can be verified similarly by using Lemma 7. Therefore, the consistency of

$$\hat{\tau}_1 \xrightarrow{p} \tau \quad \text{and} \quad \hat{\tau}_2 \xrightarrow{p} \tau$$

finally provides (18) by continuous mapping theorem. □

ACKNOWLEDGEMENTS

We would like to thank an Associate Editor and anonymous Referees for their valuable comments and suggestions which substantially improved the presentation. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A2006796), and funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1006025).

REFERENCES

Andreou E, Ghysels E. 2002. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* **17**(5): 579–600.

Baek C, Pipiras V. 2012. Statistical tests for a single change in mean against long-range dependence. *Journal of Time Series Analysis* **33**(1): 131–151.

Bai J. 1997. Estimating multiple breaks one at a time. *Econometric Theory* **13**(3): 315–352.

Baillie RT, Bollerslev T, Mikkelsen HO. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**(1): 3–30.

Basrak B, Davis RA, Mikosch T. 2002. Regular variation of GARCH processes. *Stochastic Processes and their Applications* **99**(1): 95–115.

Berkes I, Horváth L, Kokoszka P. 2003. GARCH processes: structure and estimation. *Bernoulli* **9**(2): 201–227.

Berkes I, Horváth L, Kokoszka P. 2005. Near-integrated GARCH sequences. *The Annals of Applied Probability* **15**(1B): 890–913.

Berkes I, Horváth L, Kokoszka P, Shao Q. 2006. On discriminating between long-range dependence and changes in mean. *Annals of Statistics* **34**(3): 1140–1165.

Billingsley P. 1999. *Convergence of Probability Measures*, 2nd edition, Wiley Series in Probability and Statistics: Probability and Statistics. New York: John Wiley & Sons Inc.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3): 307–327.

de Pooter M, van Dijk D. 2004. Testing for changes in volatility in heteroskedastic time series - a further examination. *Econometric Institute Report EI 2004-38*, *Erasmus University Rotterdam, Econometric Institute*.

Diebold FX, Inoue A. 2001. Long memory and regime switching. *Journal of Econometrics* **105**(1): 131–159.

Ding Z, Granger CW. 1996. Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics* **73**(1): 185–215.

Ding Z, Granger CWJ, Engle RF. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1**(1): 83–106.

Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4): 987–1007.

Engle RF, Bollerslev T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* **5**(1): 1–50.

Francq C, Zakoïan JM. 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10** (4): 605–637.

DOI: 10.1111/jtsa.12098

Francq C, Zakoïan JM. 2007. Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and their Applications* **117**(9): 1265–1284.
Giot P, Laurent S. 2004. Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance* **11**(3): 379–398.
Giraitis L, Kokoszka P, Leipus R. 2000a. Stationary ARCH models: dependence structure and Central Limit Theorem. *Econometric Theory* **16**(1): 3–22.
Giraitis L, Leipus R, Surgailis D. 2007. *Recent advances in ARCH modelling*. In *Long Memory in Economics* Springer: Berlin, pp. 3–38.
Giraitis L, Robinson PM, Surgailis D. 2000b. A model for long memory conditional heteroscedasticity. *The Annals of Applied Probability* **10**(3): 1002–1024.
Hartz C, Mittnik S, Paolella M. 2006. Accurate value-at-risk forecasting based on the normal-GARCH model. *Computational Statistics & Data Analysis* **51**(4): 2295–2312.
Hillebrand E. 2005. Neglecting parameter changes in GARCH models. *Journal of Econometrics* **129**(1-2): 121–138.
Jensen AT, Lange T. 2010. On convergence of the QMLE for misspecified GARCH models. *Journal of Time Series Econometrics* **2**(1): 31. Art. 3.
Klemeš V. 1974. The Hurst pheomenon: a puzzle *Water Resources Research* **10**(4): 675–688.
Kokoszka P, Leipus R. 2000. Change-point estimation in ARCH models. *Bernoulli* **6**(3): 513–539.
Kulperger R, Yu H. 2005. High moment partial sum processes of residuals in GARCH models and their applications. *The Annals of Statistics* **33**(5): 2395–2422.
Lee S, Tokutsu Y, Maekawa K. 2004. The cusum test for parameter change in regression models with ARCH errors. *Journal of Japan Statistical Society* **34**(2): 173–188.
Mikosch T, Stărică C. 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics* **86**(1): 378–390.
Paolella M. 2001. Testing the stable Paretian assumption. *Mathematical and Computer Modelling* **34**(9-11): 1095–1112.
Robinson PM. 2001. The memory of stochastic volatility models. *Journal of Econometrics* **101**(2): 195–218.
Surgailis D, Viano MC. 2002. Long memory properties and covariance structure of the EGARCH model. *European Series in Applied and Industrial Mathematics. Probability and Statistics* **6**: 311–329 (electronic).
Taqqu M. 1979. Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* **50**: 53–83.
Teverovsky V, Taqqu MS, Willinger W. 1999. A critical look at Lo's modified R/S statistic. *Journal of Statistical Planning and Inference* **80**(1-2): 211–227.
Vostrikova LJ. 1981. Detecting “disorder" in multidimensional random processes. *Soviet Mathematics Doklady* **24**: 55–59.
Zhang A, Gabrys R, Kokoszka P. 2007. Discriminating between long memory and volatility shifts. *Austrian Journal of Statistics* **36**(4): 253–275.

DOI: 10.1111/jtsa.12098