JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **35:** 189–202 (2014)
Published online 21 December 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12058

ORIGINAL ARTICLE

NON-STATIONARITY AND QUASI-MAXIMUM LIKELIHOOD ESTIMATION ON A DOUBLE AUTOREGRESSIVE MODEL

MIN CHEN[a] DONG LI[b,*] AND SHIQING LING[c]

[a] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*
[b] *Mathematical Sciences Center, Tsinghua University, Beijing, China*
[c] *Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong*

This article first studies the non-stationarity of the first-order double AR model, which is defined by the random recurrence equation $y_t = \phi_0 y_{t-1} + \eta_t \sqrt{\gamma_0 + \alpha_0 y_{t-1}^2}$, where $\gamma_0 > 0$, $\alpha_0 \geq 0$, and $\{\eta_t\}$ is a sequence of i.i.d. symmetric random variables. It is shown that the double AR(1) model is explosive under the condition $\mathbb{E} \log |\phi_0 + \eta_t \sqrt{\alpha_0}| > 0$. Based on this, it is shown that the quasi-maximum likelihood estimator of (ϕ_0, α_0) is consistent and asymptotically normal so that the unit root problem does not exist in the double AR(1) model. Simulation studies are carried out to assess the performance of the quasi-maximum likelihood estimator in finite samples.

Received 10 April 2013; Revised 13 November 2013; Accepted 14 November 2013

Keywords: Double AR model; non-stationarity; quasi-maximum likelihood estimation;
JEL. C13; C22.

1. INTRODUCTION

The autoregressive conditional heteroskedasticity (ARCH) model was first proposed by Engle (1982) for modelling UK inflation rates and was generalized to GARCH model by Bollerslev (1986). Since then, it has been widely used to model the volatility of economic and financial time series and extended in many directions. Now, there are many variants of ARCH/GARCH models. A Wikipedia search reveals a vastly extended family with members carrying acronyms such as EGARCH, fGARCH, GARCH-M, GJR-GARCH, IGARCH, NGARCH, QGARCH, TGARCH, and ARFIMA-GARCH. (The list is probably incomplete.) Each acronym corresponds to an attempt to extend the basic ARCH model to cope with different practical demands. Weiss (1984) introduced a class of ARCH models, defined as

$$y_t = \mu(y_{t-1}, \ldots, y_{t-p}) + \eta_t \, \sigma(y_{t-1}, \ldots, y_{t-q}),$$

where $\mu(\cdot)$ is measurable in its arguments and $\sigma(\cdot)$ is positive and measurable. Such type of models is different from Engle's model. Both conditional mean and volatility functions are an immediate regression on the observed process itself. This model was investigated by many researchers. It is also useful in modelling the volatility of economic and financial time series. One important reason is that it inherits merits of both AR models and ARCH models. Given the past information $\{y_s, s < t\}$, the conditional mean function $\mu(\cdot)$ describes the prediction values, and the conditional variance $\sigma^2(\cdot)$ measures the volatility of the prediction.

Borkovec and Klüppelberg (2001), abbreviated to BK (2001), and Ling (2004) considered the first-order double AR model (DAR(1)), a special Weiss-type model, defined by

* Correspondence to: Dong Li, Mathematical Sciences Center, Tsinghua University, Beijing 100084, China. E-mail: li20072010@gmail.com

$$y_t = \phi_0 y_{t-1} + \eta_t \sqrt{\gamma_0 + \alpha_0 y_{t-1}^2}, \quad t \in \mathbb{N} := \{1, 2, \ldots\}, \tag{1}$$

where $\gamma_0 > 0$, $\alpha_0 \geq 0$, the innovation $\{\eta_t\}$ is a sequence of i.i.d. random variables and η_t is independent of $\{y_s : s < t\}$. The higher-order DAR model was studied by Weiss (1984), Lu (1998), Ling (2007), Zhu and Ling (2013), and others.

For model (1), Guégan and Diebolt (1994) showed that a sufficient condition for weak stationarity is $\phi_0^2 + \alpha_0 < 1$. BK (2001) proved that this condition is necessary for weak stationarity as well and gave a weaker sufficient condition for strict stationarity, which is $\mathbb{E} \log |\phi_0 + \eta_t \sqrt{\alpha_0}| < 0$; see also Cline and Pu (2004). When η_t is normal, Ling (2004) showed that this condition is also almost necessary. However, when η_t is not normal, it is not clear whether BK's condition is (almost) necessary for strict stationarity or not. Our first aim is to study the condition for non-stationarity of the DAR(1) model.

In the stationary case, Weiss (1984) studied the quasi-maximum likelihood estimation (QMLE) of the model (1) and proved its asymptotic normality under $\mathbb{E} y_t^4 < \infty$; see also Tsay (1987). Ling (2004, 2007) showed that the QMLE of $(\phi_0, \gamma_0, \alpha_0)$ is still consistent and asymptotically normal under the weaker condition $\mathbb{E} \log |\phi_0 + \eta_t \sqrt{\alpha_0}| < 0$. Chan and Peng (2005) and Zhu and Ling (2013) discussed the least absolute deviation estimator. All these papers need a minimal requirement that the process $\{y_t\}$ is stationary and ergodic so that laws of large numbers do work. On the other hand, Ling and Li (2008) proved that the QMLE of (ϕ_0, α_0) is consistent and asymptotically normal under $\mathbb{E} \log |\phi_0 + \eta_t \sqrt{\alpha_0}| > 0$ when η_t is normal. However, when η_t is not normal, asymptotic properties of the QMLE in model (1) have not been established up to now in the non-stationary case. In this article, we shall establish the consistency and asymptotic normality of the QMLE of (ϕ_0, α_0) so that the unit root problem does not exist in DAR(1) models. When $\phi_0 = 0$, (1) reduces to an ARCH(1) model and the related results can be available in Nelson (1990), Jensen and Rahbek (2004a, 2004b), and Francq and Zakoïan (2012). As $\alpha_0 = 0$, (1) reduces to an AR(1) process, which is studied by Anderson (1959).

The rest of the article is organized as follows. Section 2 gives a condition for non-stationarity of model (1). Section 3 establishes asymptotic properties of the QMLE. Simulation studies are carried out in Section 4. All proofs of results are given in the Appendix.

2. A CONDITION FOR NON-STATIONARITY OF DAR(1) MODEL

Let η be a generic random variable with the same distribution as η_t. For strict stationarity of model (1), BK (2001) established a sufficient condition, which is $\mathbb{E} \log |\phi_0 + \eta\sqrt{\alpha_0}| < 0$, when η is symmetric and has a positive and continuous density on the real line $\mathbb{R}$ with $\mathbb{E}\eta^2 < \infty$. Here, we will prove that under $\mathbb{E} \log |\phi_0 + \eta\sqrt{\alpha_0}| > 0$ model (1) is explosive for general symmetric innovations, which implies that it is non-stationary.

We first define an auxiliary process $\{x_t\}$ by the random recurrence equation

$$x_t = \left| \phi_0 x_{t-1} + \eta_t \sqrt{\gamma_0 + \alpha_0 x_{t-1}^2} \right|, \tag{2}$$

where the notations are the same as in (1) and the initial value $x_0 = |y_0|$ a.s. Because of the symmetry of $\{\eta_t\}$, the independence of η_t and y_{t-1} in (1) and the homogeneous Markov structures of $\{x_t\}$ and $\{y_t\}$, it is readily seen that $\{x_t, t \in \mathbb{N}\} \stackrel{d}{=} \{|y_t|, t \in \mathbb{N}\}$; see Borkovec (2000), BK (2001), and others.

Let $w_t = \left| \phi_0 x_{t-1} + \eta_t \sqrt{\gamma_0 + \alpha_0 x_{t-1}^2} \right| - |\phi_0 + \eta_t \sqrt{\alpha_0}| x_{t-1}$. From (2), it follows that for $t \geq 1$,

$$\begin{aligned} x_t &= |\phi_0 + \eta_t \sqrt{\alpha_0}| x_{t-1} + w_t \\ &= \cdots \\ &= \sum_{j=0}^{t-1} \left(\prod_{i=0}^{j-1} |\phi_0 + \eta_{t-i} \sqrt{\alpha_0}| \right) w_{t-j} + \left(\prod_{i=0}^{t-1} |\phi_0 + \eta_{t-i} \sqrt{\alpha_0}| \right) x_0 \end{aligned}$$

DOI: 10.1111/jtsa.12058

with the convention $\prod_{i=0}^{-1} \equiv 1$. Thus,

$$\left(\prod_{k=1}^{t} |\phi_0 + \eta_k \sqrt{\alpha_0}|\right)^{-1} x_t = x_0 + \sum_{j=1}^{t} \left(\prod_{k=1}^{j} |\phi_0 + \eta_k \sqrt{\alpha_0}|\right)^{-1} w_j .$$

Note that $|w_t| \leq \sqrt{\gamma_0}|\eta_t|$. If $\mathbb{E}\log|\phi_0 + \eta\sqrt{\alpha_0}| > 0$, then we can show that

$$\begin{aligned}\left(\prod_{k=1}^{t} |\phi_0 + \eta_k \sqrt{\alpha_0}|\right)^{-1} x_t &\to x_0 + \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} |\phi_0 + \eta_k \sqrt{\alpha_0}|\right)^{-1} w_j \\ &:= x_0 + W \quad \text{a.s.,}\end{aligned} \tag{3}$$

as $t \to \infty$. Note that $|W| \leq \sqrt{\gamma_0} \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} |\phi_0 + \eta_k \sqrt{\alpha_0}|\right)^{-1} |\eta_j| < \infty$ a.s., and hence, it is well defined.

Summarizing the aforementioned discussion, we have the following result.

Theorem 1. If $\{\eta_t\}$ is a sequence of i.i.d. symmetric random variables with $\mathbb{E}\log|\phi_0+\eta\sqrt{\alpha_0}| > 0$ and $\mathbb{P}(|y_0|+W=0)=0$, then $\rho^{-t}|y_t| \to \infty$ a.s. for all $1<\rho<\exp\{\mathbb{E}\log|\phi_0+\eta\sqrt{\alpha_0}|\}$, as $t \to \infty$.

Unlike BK (2001) in which η has a positive and continuous density on $\mathbb{R}$ with $\mathbb{E}\eta^2 < \infty$, we weaken this condition and allow it in particular not to have any fractional moments. Here, the unique condition for η is its symmetry. The condition is rather weak, which is met by the most commonly used random variables, for example, normal, double exponential (also sometimes called the Laplace distribution), Student's t_v, and symmetric stable distributions.

Figure 1 exhibits the stationary and non-stationary regions of process (1) when η is $\mathcal{N}(0,1)$, $U[-\sqrt{3},\sqrt{3}]$, the double exponential distribution with the diversity $1/\sqrt{2}$ and standardized Student's t_4-distribution respectively. It

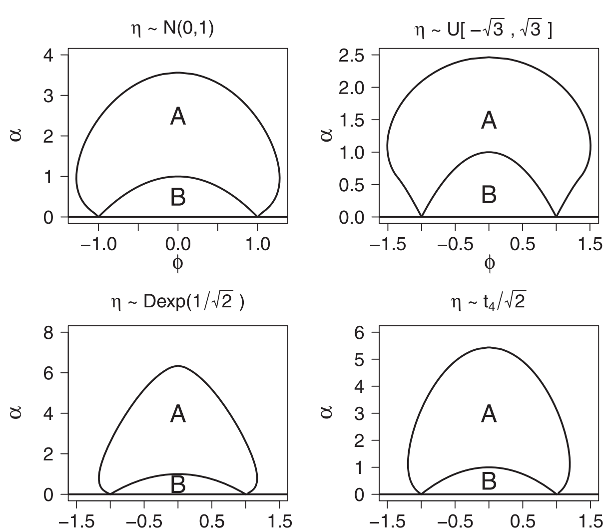

Figure 1. $\mathbb{E}\log|\phi+\eta\sqrt{\alpha}| < 0$ as $(\phi,\alpha) \in A \cup B$ and $\phi^2+\alpha<1$ as $(\phi,\alpha) \in B$ when η is $\mathcal{N}(0,1)$, $U[-\sqrt{3},\sqrt{3}]$, the double exponential distribution with the diversity $1/\sqrt{2}$ and standardized Student's t_4-distribution respectively

DOI: 10.1111/jtsa.12058

is well known that the stationary region of AR(1) model is $|\phi_0| < 1$. From Figure 1, we can see that model (1) may be stationary even if $|\phi_0| \geq 1$.

Furthermore, in Theorem 1, η is allowed to be discrete random variables. This makes a sharp difference with the stationary case studied in Borkovec (2000), BK (2001), and others in literature. Figure 2 shows the non-stationary region of process (1) when η is the Rademacher distribution and a four-point discrete distribution (FPDD), which is defined as

$$\mathbb{P}\left(\eta = \pm\sqrt{2/5}\right) = \mathbb{P}\left(\eta = \pm 2\sqrt{2/5}\right) = 1/4 \tag{4}$$

respectively.

In Theorem 1, we need to choose an initial value y_0 (or x_0) such that $\mathbb{P}(|y_0| + W = 0) = 0$. A similar condition is used in Berkes *et al.* (2009) (Thm 5) when they investigated the growth of a non-stationary random coefficient autoregressive model. Borkovec (2000) also used the assumption that y_0 is large enough when he studied the extremal behaviour of DAR(1) models. From the expression of W, it seems to be very difficult to obtain the closed form of the distribution of W since it heavily depends on the process $\{x_t\}$ itself. However, we can simulate the density or distribution of $|y_0| + W$. Figure 3 shows the density function of $|y_0| + W$ when η is $\mathcal{N}(0, 1)$, $U[-\sqrt{3}, \sqrt{3}]$, Dexp$(1/\sqrt{2})$, standardized Student's t_4-distribution, the Rademacher distribution, and the FPDD defined in (4) respectively. Here, $(\phi_0, \alpha_0, \gamma_0) = (2, 1, 1)$ and the initial value $y_0 = 2$. We use the first 100 terms to approximate W in (3). A total of 20,000 replications are used. From Figure 3, we can see that $|y_0| + W > 0$ holds. Thus, it seems that $\mathbb{P}(|y_0| + W = 0) = 0$ is not unreasonable.

It should be mentioned that, for the case $\mathbb{E}\log|\phi_0 + \eta\sqrt{\alpha_0}| = 0$, we conjecture that model (1) is still non-stationary and its asymptotic behaviour is more complicate. It needs a further study in the future research. A similar technical difficulty was encountered in Francq and Zakoïan (2012, 2013) when they studied the asymptotic behaviour of GARCH model on the boundary curve of the parameter space.

3. QUASI-MAXIMUM LIKELIHOOD ESTIMATION

Assume that the sample $\{y_0, \ldots, y_n\}$ is from model (1). Denote by $(\gamma, \boldsymbol{\theta}')' = (\gamma, \phi, \alpha)'$ the DAR(1) parameter and define the QMLE $(\hat{\gamma}_n, \hat{\boldsymbol{\theta}}_n')'$ as any measurable solution of

$$L_n(\gamma, \boldsymbol{\theta}) = \sum_{t=1}^{n} \ell_t(\gamma, \boldsymbol{\theta}), \quad \ell_t(\gamma, \boldsymbol{\theta}) = -\frac{1}{2}\left\{\log(\gamma + \alpha y_{t-1}^2) + \frac{(y_t - \phi y_{t-1})^2}{\gamma + \alpha y_{t-1}^2}\right\},$$

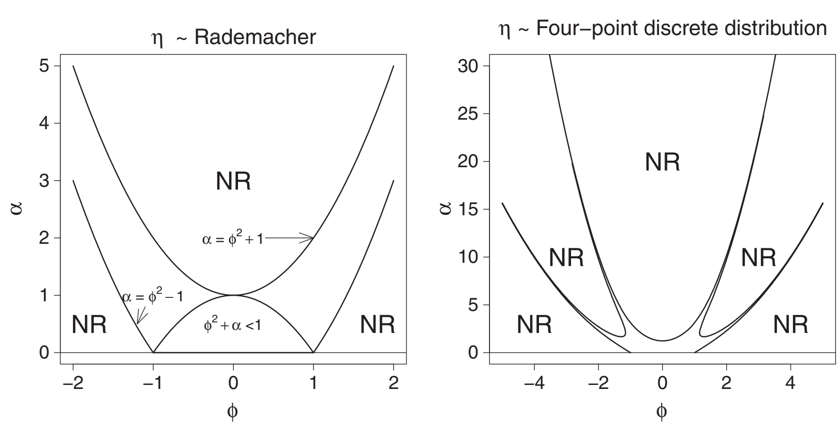

Figure 2. The non-stationary region (NR) with $\mathbb{E}\log|\phi + \eta\sqrt{\alpha}| > 0$ when η is the Rademacher distribution and the four-point discrete distribution in (4) respectively

DOI: 10.1111/jtsa.12058

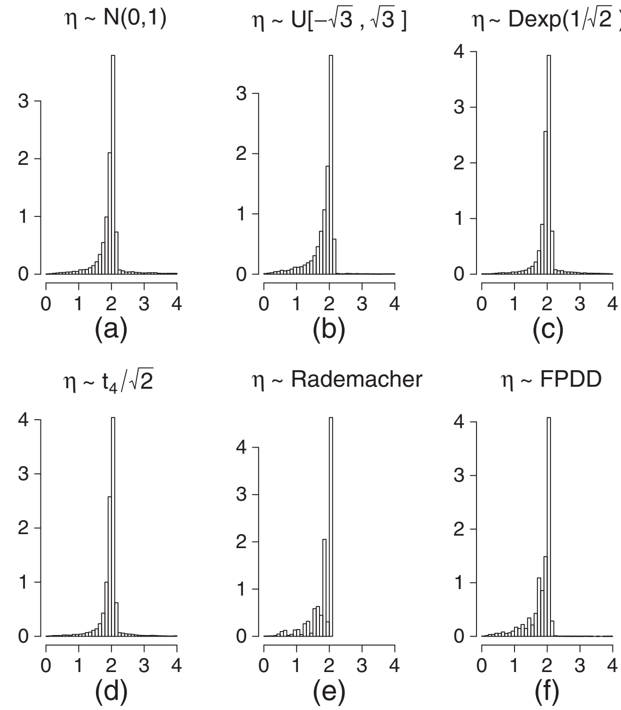

Figure 3. The density function of $|y_0| + W$ for η being $\mathcal{N}(0,1)$, $U[-\sqrt{3}, \sqrt{3}]$, $\text{Dexp}(1/\sqrt{2})$, standardized Student's t_4-distribution, the Rademacher distribution, and the four-point discrete distribution (FPDD) defined in (4) respectively. Here, $(\phi_0, \alpha_0, \gamma_0) = (2, 1, 1)$ and the initial value $x_0 = |y_0| = 2$

that is,

$$(\hat{\gamma}_n, \hat{\boldsymbol{\theta}}_n{}')' = \arg\max_{\Gamma\times\Theta} L_n(\gamma, \boldsymbol{\theta}),$$

where $\Gamma \times \Theta$ is a compact subset of $\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$ with $\mathbb{R}_+ \equiv (0, \infty)$ that contains the true value $(\gamma_0, \boldsymbol{\theta}_0')' = (\gamma_0, \phi_0, \alpha_0)'$. Here, the case $\alpha_0 = 0$ is excluded. Ideally, we should study asymptotic properties of $(\hat{\gamma}_n, \hat{\boldsymbol{\theta}}_n')'$. Unfortunately, as $\{y_t\}$ is explosive, γ is absent from the limit of $L_n(\gamma, \boldsymbol{\theta})/n$; see Lemma 2. A similar phenomenon was observed in Jensen and Rahbek (2004a) when they studied the non-stationary ARCH(1) model and in Ling and Li (2008) when η is normal; see also Francq and Zakoïan (2012). Thus, intuitively, there does not have any consistent estimator of γ_0. This point can be seen from the proof of Theorem 2 that the log-likelihood is completely flat in the direction where $(\phi_0, \alpha_0)'$ is fixed and γ_0 varies. More specifically, for any real sequence λ_n going to zero as $n \to \infty$, we have

$$\Lambda_n \begin{pmatrix} \frac{\partial L_n(\gamma, \boldsymbol{\theta_0})}{\partial \gamma} \\ \frac{\partial L_n(\gamma, \boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \end{pmatrix} \xrightarrow{\mathscr{D}} \mathcal{N}(0, \text{diag}\{0, 1/\alpha_0, (\mathbb{E}\eta^4 - 1)/(4\alpha_0^2)\}),$$

where $\Lambda_n = \text{diag}\{\lambda_n, 1/\sqrt{n}, 1/\sqrt{n}\}$. For deeper insights on asymptotic behaviour of $\hat{\gamma}_n$, one can consult Francq and Zakoïan (2012). The inconsistency of $\hat{\gamma}_n$ is illustrated via simulation studies in Section 4.

For convenient statements, we first give assumptions on $\{\eta_t\}$ and the parameter space:

Assumption 1. The innovation $\{\eta_t\}$ is a sequence of i.i.d. symmetric random variables with zero mean and unit variance.

DOI: 10.1111/jtsa.12058

Assumption 2. The true parameter $\boldsymbol{\theta}_0 \in \Theta$ satisfies $\mathbb{E}\log|\phi_0 + \eta\sqrt{\alpha_0}| > 0$ and $\mathbb{P}(|y_0| + W = 0) = 0$. Both Γ and Θ are compact.

So as to obtain asymptotic properties of $\hat{\boldsymbol{\theta}}_n$, the rate of divergence of $|y_t|$ to infinity is needed, which is guaranteed by Theorem 1. The following theorem states consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_n$.

Theorem 2. If Assumptions 1 and 2 hold, then as $n \to \infty$

(i) $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ in probability;
(ii) If $\kappa = \mathbb{E}\eta^4 < \infty$ and $\boldsymbol{\theta}_0$ is an interior point of Θ, then

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{\mathscr{D}} \mathcal{N}(0, \Sigma),$$

where $\Sigma = \operatorname{diag}\left\{\alpha_0,\ (\kappa - 1)\alpha_0^2\right\}$.

Combining with the results in Ling (2004) for the stationary case produces asymptotic normality of the QMLE of (ϕ_0, α_0) with a root-n rate of convergence for any $\phi_0 \in \mathbb{R}$ and any $\alpha_0 \in \mathbb{R}_+$ with $\mathbb{E}\log|\phi_0 + \eta\sqrt{\alpha_0}| \neq 0$. In the stationary case, the estimator of γ_0 is consistent and asymptotically normal; see Ling (2004). In the explosive case, however, γ_0 is not estimable so that we cannot obtain a consistent estimator, but this does not affect asymptotic properties of the QMLE of (ϕ_0, α_0). Ling and Li (2008) presented an asymptotically unbiased estimator for γ_0, but it is not consistent. Comparing with the results for the classical AR(1) model, the unit root problem does not exist in model (1); see Rmks 1–3 in Ling and Li (2008).

So as to construct confidence intervals for $\boldsymbol{\theta}_0$, estimating κ is needed. To this end, we consider the residual $\{\hat{\eta}_t\}$, where

$$\hat{\eta}_t = \frac{y_t - \hat{\phi}_n y_{t-1}}{\sqrt{\hat{\gamma}_n + \hat{\alpha}_n y_{t-1}^2}},$$

and define the estimator of κ as $\hat{\kappa}_n = \frac{1}{n}\sum_{t=1}^n \hat{\eta}_t^4$. Under Assumptions 1 and 2, we can prove that if $\kappa < \infty$, then $\hat{\kappa}_n = \kappa + o_p(1)$. That is, $\hat{\kappa}_n$ is a consistent estimator of κ.

4. SIMULATION STUDIES

Monte Carlo experiments are carried out to assess the performance of the QMLE of $\boldsymbol{\theta}_0$ in finite samples. When $\eta \sim \mathcal{N}(0, 1)$, the simulation results can be found in Ling and Li (2008). We only report the results when η takes, respectively, the following distributions:

- $U[-\sqrt{3}, \sqrt{3}\,]$.
- Double exponential distribution (Dexp($1/\sqrt{2}$)) with the density

$$f(x) = \frac{1}{\sqrt{2}}\exp\left(-\sqrt{2}|x|\right), \quad x \in \mathbb{R}.$$

- Standardized Student's t_6 ($\sqrt{2/3}\ t_6$) with the density

$$f(x) = \frac{15}{32}\left(1 + \frac{x^2}{4}\right)^{-\frac{7}{2}}, \quad x \in \mathbb{R}.$$

DOI: 10.1111/jtsa.12058

- Rademacher distribution, that is, $\mathbb{P}(\eta = \pm 1) = 1/2$.
- An FPDD defined in (4).

Simulations are carried out with the sample size $n = 100$ and 200. A total of 1000 replications are used. In all simulation studies, we set $\gamma_0 = 1$. Each table includes the empirical mean (EM), empirical standard deviation

Table I. Simulation studies for $\eta \sim \mathrm{U}[-\sqrt{3}, \sqrt{3}\,]$

			$n = 100$			$n = 200$		
ϕ_0	α_0		$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$
1.0	3.0	EM	1.2505	1.0092	2.9555	1.2687	0.9966	2.9793
		ESD	0.9775	0.1811	0.2933	1.0514	0.1252	0.1991
		ASD	—	0.1732	0.2683	—	0.1225	0.1897
1.0	4.0	EM	1.2516	1.0090	3.9477	1.2342	0.9995	3.9831
		ESD	0.8789	0.2024	0.3870	0.9288	0.1466	0.2591
		ASD	—	0.2000	0.3578	—	0.1414	0.2530
2.0	1.0	EM	1.1920	1.9980	0.9844	1.1413	1.9993	0.9912
		ESD	0.7409	0.1028	0.0935	0.7241	0.0725	0.0631
		ASD	—	0.1000	0.0894	—	0.0707	0.0632
2.0	2.0	EM	1.1166	2.0021	1.9749	1.1117	1.9981	1.9903
		ESD	0.6720	0.1502	0.1869	0.6643	0.1053	0.1324
		ASD	—	0.1414	0.1789	—	0.1000	0.1265
−2.0	3.0	EM	1.1645	**−2.0137**	2.9801	1.1771	**−2.0014**	2.9870
		ESD	0.9340	0.1730	0.2711	0.8253	0.1181	0.1949
		ASD	—	0.1732	0.2683	—	0.1225	0.1897
−2.0	4.0	EM	1.1949	**−2.0124**	3.9634	1.1587	**−1.9973**	3.9737
		ESD	0.9149	0.2124	0.3591	0.8380	0.1489	0.2516
		ASD	—	0.2000	0.3578	—	0.1414	0.2530

EM, empirical mean; ESD, empirical standard deviation; ASD, asymptotic standard deviation.

Table II. $\eta \sim \mathrm{Dexp}(1/\sqrt{2})$

			$n = 100$			$n = 200$		
ϕ_0	α_0		$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$
1.0	4.0	EM	2.4800	1.0074	3.9618	2.3243	0.9986	3.9693
		ESD	9.1838	0.2154	0.9665	12.574	0.1424	0.6738
		ASD	—	0.2000	0.8944	—	0.1414	0.6325
1.0	6.0	EM	5.3166	0.9912	5.8731	7.2252	1.0040	5.9428
		ESD	67.395	0.2579	1.4068	138.20	0.1765	0.9688
		ASD	—	0.2449	1.3416	—	0.1732	0.9487
2.0	2.0	EM	1453.6	2.0013	2.0152	2.0402	2.0044	1.9763
		ESD	45876	0.1481	0.4605	6.5248	0.1011	0.3069
		ASD	—	0.1414	0.4472	—	0.1000	0.3162
2.0	4.0	EM	3.4803	1.9996	3.9670	2.1050	1.9987	3.9427
		ESD	25.507	0.1995	0.9270	7.2692	0.1427	0.5915
		ASD	—	0.2000	0.8944	—	0.1414	0.6325
−1.0	6.0	EM	4.6524	**−1.0095**	5.8352	4.4432	**−1.0000**	5.9190
		ESD	46.197	0.2409	1.3419	28.842	0.1720	0.9289
		ASD	—	0.2449	1.3416	—	0.1732	0.9487
−2.0	4.0	EM	2.8904	**−2.0038**	3.9575	8.1028	**−2.0039**	3.9853
		ESD	16.777	0.2013	0.9155	178.67	0.1399	0.6501
		ASD	—	0.2000	0.8944	—	0.1414	0.6325

EM, empirical mean; ESD, empirical standard deviation; ASD, asymptotic standard deviation.

DOI: 10.1111/jtsa.12058

(ESD), and asymptotic standard deviation (ASD) of the QMLE of $(\gamma_0, \phi_0, \alpha_0)$. The ASD is obtained from the asymptotic covariance matrix Σ in Theorem 2. Since γ_0 is inestimable, its ASDs are not available in all tables. The simulation results are reported in Tables I–V.

From each table, we see that all biases of the estimates of (ϕ_0, α_0) are very small and the differences between the ESDs and ASDs are very small. The larger the sample size n, the smaller the ESDs. For the same sample size n, the larger α_0, the larger all ESDs and ASDs. All results show that the QMLE of (ϕ_0, α_0) performs well in the finite samples. However, the QMLE of γ_0 has a larger bias and a quite larger empirical standard deviation for some cases, for example, $\eta \sim \mathrm{Dexp}(1/\sqrt{2})$. This is caused by the inestimability of γ_0.

Table III. $\eta \sim$ standardized Student's t_6

				$n = 100$			$n = 200$	
ϕ_0	α_0		$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$
1.0	5.0	EM	3.6877	1.0078	4.9581	2.6544	1.0012	4.9886
		ESD	35.719	0.2338	1.2439	14.339	0.1649	0.8822
		ASD	—	0.2236	1.1180	—	0.1581	0.7906
1.0	6.0	EM	3.7139	1.0092	5.9565	4.1232	1.0008	5.9121
		ESD	36.030	0.1792	0.9822	41.095	0.1738	0.9106
		ASD	—	0.2449	1.3416	—	0.1732	0.9487
2.0	2.0	EM	2.7431	2.0048	1.9667	2.7161	2.0029	1.9673
		ESD	16.908	0.1402	0.4351	16.233	0.0983	0.3110
		ASD	—	0.1414	0.4472	—	0.1000	0.3162
2.0	4.0	EM	36.942	2.0001	3.9605	2.6050	2.0013	4.0021
		ESD	1082.2	0.2009	0.8896	13.976	0.1393	0.6721
		ASD	—	0.2000	0.8944	—	0.1414	0.6325
−2.0	7.0	EM	2.8831	−2.0093	6.9523	6.6976	−2.0085	6.9397
		ESD	24.584	0.2678	1.6144	94.661	0.1938	1.1228
		ASD	—	0.2646	1.5652	—	0.1871	1.1068
−2.0	5.0	EM	8.7821	−2.0043	4.9297	6.0257	−1.9934	4.9773
		ESD	212.81	0.2315	1.1189	97.151	0.1579	0.7618
		ASD	—	0.2236	1.1180	—	0.1581	0.7906

EM, empirical mean; ESD, empirical standard deviation; ASD, asymptotic standard deviation.

Table IV. $\eta \sim$ Rademacher distribution

				$n = 100$			$n = 200$	
ϕ_0	α_0		$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$
1.0	4.0	EM	1.0074	0.9974	3.9543	0.9994	0.9970	3.9809
		ESD	0.0501	0.2140	0.0628	0.0281	0.1384	0.0261
		ASD	—	0.2000	0.0000	—	0.1414	0.0000
2.0	1.0	EM	1.0025	2.0064	0.9887	0.9968	1.9991	0.9948
		ESD	0.0504	0.1056	0.0151	0.0359	0.0712	0.0070
		ASD	—	0.1000	0.0000	—	0.0707	0.0000
−1.0	5.0	EM	1.0025	−0.9970	4.9478	1.0016	−1.0088	4.9738
		ESD	0.0367	0.2286	0.0672	0.0250	0.1617	0.0356
		ASD	—	0.2236	0.0000	—	0.1581	0.0000
−2.0	1.0	EM	1.0039	−1.9981	0.9893	0.9981	−2.0004	0.9949
		ESD	0.0503	0.1029	0.0155	0.0358	0.0700	0.0070
		ASD	—	0.1000	0.0000	—	0.0707	0.0000

EM, empirical mean; ESD, empirical standard deviation; ASD, asymptotic standard deviation.

  J. Time Ser. Anal. **35:** 189–202 (2014)
DOI: 10.1111/jtsa.12058

Table V. $\eta \sim$ four-point discrete distribution defined in (4)

			$n=100$			$n=200$		
ϕ_0	α_0		$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\phi}_n$	$\hat{\alpha}_n$
4.0	3.0	EM	1.0842	3.9947	2.9793	1.0620	3.9976	2.9813
		ESD	0.5951	0.1783	0.1837	0.5970	0.1188	0.1295
		ASD	—	0.1732	0.1800	—	0.1225	0.1273
2.0	5.0	EM	1.0690	1.9843	4.9338	1.0500	1.9958	4.9781
		ESD	0.5721	0.2183	0.3142	0.5671	0.1593	0.2191
		ASD	—	0.2236	0.3000	—	0.1581	0.2121
0.0	3.0	EM	1.0568	−0.0003	2.9637	1.0475	−0.0043	2.9815
		ESD	0.5940	0.1778	0.1903	0.5924	0.1210	0.1268
		ASD	—	0.1732	0.1800	—	0.1225	0.1273
−2.0	4.0	EM	1.0557	−1.9941	3.9533	1.0510	−1.9880	3.9803
		ESD	0.5615	0.2072	0.2552	0.5505	0.1452	0.1715
		ASD	—	0.2000	0.2400	—	0.1414	0.1697
−4.0	2.0	EM	1.0315	−3.9937	1.9768	1.0846	−3.9978	1.9903
		ESD	0.5917	0.1435	0.1292	0.5754	0.1013	0.0879
		ASD	—	0.1414	0.1200	—	0.1000	0.0849

EM, empirical mean; ESD, empirical standard deviation; ASD, asymptotic standard deviation.

ACKNOWLEDGEMENTS

The authors thank the Editor, the Associate Editor, and two anonymous referees for their helpful comments that improved the presentation. The second author also thanks Dr. Tao CAI at Sun Yat-sen Univeristy for his great help in plotting Figure 1. Chen's research is supported in part by the National Natural Science Foundation of China under the grant No.10990012 and No.10721101, CAS No. 2008DP173182. Ling's research is supported in part by Hong Kong RGC Grants numbered HKUST641912 and 603413.

REFERENCES

Anderson TW. 1959. On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* **30**: 676–687.

Berkes I, Horváth L, Ling S. 2009. Estimation in nonstationary random coefficient autoregressive models. *Journal of Time Series Analysis* **30**: 395–416.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Eonometrics* **31**: 307–327.

Borkovec M. 2000. Extremal behavior of the autoregressive process with ARCH(1) errors. *Stochastic Processes and their Applications* **85**: 189–207.

Borkovec M, Klüppelberg C. 2001. The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Annals of Applied Probability* **11**: 1220–1241.

Brown BM. 1971. Martingale central limit theorems. *Annals of Mathematical Statistics* **42**: 59–66.

Chan NH, Peng L. 2005. Weighted least absolute deviation estimation for an AR(1) process with ARCH(1) errors. *Biometrika* **92**: 477–484.

Cline DBH, Pu HH. 2004. Stability and the Lyapounov exponent of threshold AR-ARCH models. *Annals of Applied Probability* **14**: 1920–1949.

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

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

Francq C, Zakoïan J-M. 2013. Inference in nonstationary asymmetric GARCH models. *Annals of Statistics* **41**: 1970–1998.

Guégan D, Diebolt J. 1994. Probabilistic properties of the β-ARCH-model. *Statistica Sinica* **4**: 71–87.

Jensen ST, Rahbek A. 2004a. Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* **72**: 641–646.

DOI: 10.1111/jtsa.12058

Jensen ST, Rahbek A. 2004b. Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**: 1203–1226.
Ling S. 2004. Estimation and testing stationarity for double autoregressive models. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **66**: 63–78.
Ling S. 2007. A double AR(p) model: structure and estimation. *Statistica Sinica* **17**: 161–175.
Ling S, Li D. 2008. Asymptotic inference for a nonstationary double AR(1) model. *Biometrika* **95**: 257–263.
Lu Z. 1998. On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica* **8**: 1205–1217.
Nelson DB. 1990. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**: 318–334.
Tsay RS. 1987. Conditional heteroscedastic time series models. *Journal of the American Statistical Association* **82**: 590–604.
Weiss AA. 1984. ARMA models with ARCH errors. *Journal of Time Series Analysis* **5**: 129–143.
Zhu K, Ling S. 2013. Quasi-maximum exponential likelihood estimators for a double AR (p) model. *Statistica Sinica* **23**: 251–270.

APPENDIX: PROOF OF THEOREM 2

Before the proof, we first give three lemmas.

Lemma 1. If Assumptions 1 and 2 hold, then

(a)

$$\sup_{\boldsymbol{\theta}\in\Theta}\sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\ell_t(\gamma,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| = O_p(1);$$

(b)

$$\sup_{\boldsymbol{\theta}\in\Theta}\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^3\ell_t(\gamma,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| = O_p(1),\ \text{where } \theta_1=\phi, \theta_2=\alpha, i,j,k=1,2.$$

Proof

(a) Since the parameter space $\Gamma\times\Theta$ is compact, there exist constants $0<\underline{\alpha}<\bar{\alpha}<\infty$, $0<\underline{\gamma}<\bar{\gamma}<\infty$ and $0<M<\infty$ such that $|\phi|\le M$, $\gamma\in[\underline{\gamma},\bar{\gamma}]$ and $\alpha\in[\underline{\alpha},\bar{\alpha}]$. By the expression of $\ell_t(\gamma,\boldsymbol{\theta})$, we have

$$\frac{\partial\ell_t(\gamma,\boldsymbol{\theta})}{\partial\phi} = \frac{y_{t-1}(y_t-\phi y_{t-1})}{\gamma+\alpha y_{t-1}^2}.$$

By Theorem 1 and the law of large numbers, it follows that

$$\sup_{\boldsymbol{\theta}\in\Theta}\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\ell_t(\gamma,\boldsymbol{\theta})}{\partial\phi}\right| \le \frac{2M}{\underline{\alpha}}+\frac{1}{\sqrt{\underline{\alpha}}}\sqrt{\frac{\max\{\gamma_0,\alpha_0\}}{\min\{\underline{\gamma},\underline{\alpha}\}}}\,\frac{1}{n}\sum_{t=1}^{n}|\eta_t| = O_p(1).$$

Similarly, we can show that

$$\sup_{\boldsymbol{\theta}\in\Theta}\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\frac{\partial L_n(\gamma,\boldsymbol{\theta})}{\partial\alpha}\right| = O_p(1).$$

Thus, (a) holds.

DOI: 10.1111/jtsa.12058

(b) We consider $\partial^3 \ell_t(\gamma, \boldsymbol{\theta})/\partial\alpha^3$. By a calculation, it follows that

$$\frac{\partial^3 \ell_t(\gamma, \boldsymbol{\theta})}{\partial \alpha^3} = \frac{3y_{t-1}^6 (y_t - \phi y_{t-1})^2}{(\gamma + \alpha y_{t-1}^2)^4} - \frac{y_{t-1}^6}{(\gamma + \alpha y_{t-1}^2)^3}.$$

By Theorem 1 and the law of large numbers, we have

$$\sup_{\boldsymbol{\theta}\in\Theta} \sup_{\gamma\in\Gamma} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^3 \ell_t(\gamma, \boldsymbol{\theta})}{\partial \alpha^3} \right| \le \frac{24M^2}{\underline{\alpha}^4} + \frac{6}{\underline{\alpha}^3} \frac{\max\{\gamma_0, \alpha_0\}}{\min\{\underline{\gamma}, \underline{\alpha}\}} \frac{1}{n} \sum_{t=1}^{n} \eta_t^2 = O_p(1).$$

Similarly, we can show that other third-order mixed partial derivatives of $\frac{1}{n}\sum_{t=1}^{n} \ell_t(\gamma, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are $O_p(1)$. Thus, (b) holds. □

Lemma 2. Let $f(\boldsymbol{\theta}) = \{\log(\alpha_0/\alpha) + 1 - (\alpha_0/\alpha) - (\phi - \phi_0)^2/\alpha\}/2$. If Assumptions 1 and 2 hold, then

$$\sup_{\boldsymbol{\theta}\in\Theta} \sup_{\gamma\in\Gamma} \left| \frac{1}{n} \sum_{t=1}^{n} [\ell_t(\gamma, \boldsymbol{\theta}) - \ell_t(\gamma_0, \boldsymbol{\theta_0})] - f(\boldsymbol{\theta}) \right| = o_p(1).$$

Proof
By the expression of $\ell_t(\gamma, \boldsymbol{\theta})$ and Theorem 1, we have

$$\sup_{\gamma\in\Gamma} \frac{1}{n} \sum_{t=1}^{n} |\ell_t(\gamma, \boldsymbol{\theta}_0) - \ell_t(\gamma_0, \boldsymbol{\theta}_0)| \to 0 \quad \text{a.s. as } n \to \infty.$$

Thus, it suffices to show

$$\sup_{\boldsymbol{\theta}\in\Theta} \sup_{\gamma\in\Gamma} \left| \frac{1}{n} \sum_{t=1}^{n} [\ell_t(\gamma, \boldsymbol{\theta}) - \ell_t(\gamma, \boldsymbol{\theta}_0)] - f(\boldsymbol{\theta}) \right| = o_p(1).$$

First, for each $\boldsymbol{\theta} \in \Theta$, we have

$$\begin{aligned}
&\sup_{\gamma\in\Gamma} \left| \frac{1}{n} \sum_{t=1}^{n} [\ell_t(\gamma, \boldsymbol{\theta}) - \ell_t(\gamma, \boldsymbol{\theta}_0)] - f(\boldsymbol{\theta}) \right| \\
&\le \sup_{\gamma\in\Gamma} \left| \frac{1}{2n} \sum_{t=1}^{n} \left[\log \frac{\gamma + \alpha_0 y_{t-1}^2}{\gamma + \alpha y_{t-1}^2} - \log \frac{\alpha_0}{\alpha} \right] \right| + \sup_{\gamma\in\Gamma} \left| \frac{1}{2n} \sum_{t=1}^{n} \left[\eta_t^2 \frac{\gamma_0 + \alpha_0 y_{t-1}^2}{\gamma + \alpha y_{t-1}^2} - \frac{\alpha_0}{\alpha} \right] \right| \\
&+ \sup_{\gamma\in\Gamma} \left| \frac{1}{2n} \sum_{t=1}^{n} \left[\eta_t^2 \frac{\gamma_0 + \alpha_0 y_{t-1}^2}{\gamma + \alpha_0 y_{t-1}^2} - 1 \right] \right| + \sup_{\gamma\in\Gamma} \left| \frac{1}{2n} \sum_{t=1}^{n} \left[\frac{(\phi_0 - \phi)^2 y_{t-1}^2}{\gamma + \alpha y_{t-1}^2} - \frac{(\phi - \phi_0)^2}{\alpha} \right] \right| \\
&+ \sup_{\gamma\in\Gamma} \left| \frac{1}{2n} \sum_{t=1}^{n} \frac{2(\phi_0 - \phi)\eta_t y_{t-1} \sqrt{\gamma_0 + \alpha_0 y_{t-1}^2}}{\gamma + \alpha y_{t-1}^2} \right| \\
&:= I_1 + I_2 + I_3 + I_4 + I_5.
\end{aligned}$$

DOI: 10.1111/jtsa.12058

For I_1, using the mean value theorem and Theorem 1,

$$I_1 \le \frac{\bar{\alpha}\bar{\gamma}}{2\underline{\alpha}}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\underline{\gamma}+\underline{\alpha}y_{t-1}^2} = o_p(1).$$

As for I_2, by Theorem 1, we have

$$\begin{aligned}
I_2 &\le \left|\frac{1}{2n}\sum_{t=1}^{n}\left(\eta_t^2-1\right)\frac{\alpha_0}{\alpha}\right| + \sup_{\gamma\in\Gamma}\left|\frac{1}{2n}\sum_{t=1}^{n}\eta_t^2\frac{\alpha\gamma_0-\alpha_0\gamma}{\alpha(\gamma+\alpha y_{t-1}^2)}\right| \\
&\le o_p(1) + \sup_{\gamma\in\Gamma}\left|\frac{1}{2n}\sum_{t=1}^{n}\eta_t^2\frac{\alpha\gamma_0-\alpha_0\gamma}{\alpha(\gamma+\alpha y_{t-1}^2)}\right| \\
&\le o_p(1) + M\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\eta_t^2}{\underline{\gamma}+\underline{\alpha}y_{t-1}^2}\right) = o_p(1).
\end{aligned}$$

Similarly, we can show that $I_3 = I_4 = o_p(1)$. For I_5, we have

$$\begin{aligned}
I_5 &\le M\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\eta_t y_{t-1}\sqrt{\gamma_0+\alpha_0 y_{t-1}^2}}{\underline{\gamma}+\alpha y_{t-1}^2}\right| \\
&\quad + M\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}\eta_t y_{t-1}\sqrt{\gamma_0+\alpha_0 y_{t-1}^2}\left\{\frac{1}{\gamma+\alpha y_{t-1}^2}-\frac{1}{\underline{\gamma}+\alpha y_{t-1}^2}\right\}\right| \\
&\le o_p(1) + M\left(\frac{1}{n}\sum_{t=1}^{n}\frac{|\eta_t|}{\underline{\gamma}+\underline{\alpha}y_{t-1}^2}\right) = o_p(1)
\end{aligned}$$

by Theorem 1. Thus, for each $\boldsymbol{\theta} \in \Theta$,

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta})-\ell_t(\gamma,\boldsymbol{\theta}_0)] - f(\boldsymbol{\theta})\right| = o_p(1). \tag{A.1}$$

We next show that (A.1) holds uniformly in $\boldsymbol{\theta} \in \Theta$. For any $\varepsilon > 0$, there exists a $\delta > 0$ such that $|f(\boldsymbol{\theta}_1) - f(\boldsymbol{\theta}_2)| < \varepsilon/3$ for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ with $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < \delta$ because $f(\boldsymbol{\theta})$ is continuous uniformly on the compact set Θ. For any $\eta \in (0, \delta)$, let $V_j = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_j\| < \eta\}$ for some $\boldsymbol{\theta}_j \in \Theta$. Then, $\{V_j : \boldsymbol{\theta}_j \in \Theta\}$ forms an open covering of Θ. We can choose a finite number of $\{V_j, \boldsymbol{\theta}_j \in \Theta\}$ covering Θ, say, $\{V_j\}_{j=1}^K$. By Lemma 1(a), we have

DOI: 10.1111/jtsa.12058

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta}\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta})-\ell_t(\gamma,\boldsymbol{\theta}_0)]-f(\boldsymbol{\theta})\right| \\
&\leq \max_{1\leq j\leq K}\sup_{\boldsymbol{\theta}\in V_j\cap\Theta}\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta})-\ell_t(\gamma,\boldsymbol{\theta}_j)]\right. \\
&\quad\left.+\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta}_j)-\ell_t(\gamma,\boldsymbol{\theta}_0)-f(\boldsymbol{\theta}_j)]+\frac{1}{n}\sum_{t=1}^{n}[f(\boldsymbol{\theta}_j)-f(\boldsymbol{\theta})]\right| \\
&\leq \max_{1\leq j\leq K}\left\{\sup_{\boldsymbol{\theta}\in V_j\cap\Theta}\|\boldsymbol{\theta}-\boldsymbol{\theta}_j\|\sup_{\boldsymbol{\theta}\in V_j\cap\Theta}\sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\ell_t(\gamma,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\right. \\
&\quad\left.+\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta}_j)-\ell_t(\gamma,\boldsymbol{\theta}_0)-f(\boldsymbol{\theta}_j)]\right|+\sup_{\boldsymbol{\theta}\in V_j\cap\Theta}|f(\boldsymbol{\theta}_j)-f(\boldsymbol{\theta})|\right\} \\
&\leq \eta O_p(1)+\max_{1\leq j\leq K}\left\{\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta}_j)-\ell_t(\gamma,\boldsymbol{\theta}_0)-f(\boldsymbol{\theta}_j)]\right|\right\}+\frac{\varepsilon}{3}.
\end{aligned}
$$

We first take $\eta \in (0,\delta)$ small enough such that $\mathbb{P}(\eta|O_p(1)| > \varepsilon/3) < \varepsilon/2$ uniformly in n. For this η, K is fixed and hence, by (A.1), the second term is $o_p(1)$ as $n \to \infty$. Thus, we can take an N large enough such that

$$
\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta}\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta})-\ell_t(\gamma,\boldsymbol{\theta}_0)]-f(\boldsymbol{\theta})\right|>\varepsilon\right)<\varepsilon,
$$

as $n > N$, that is, the conclusion holds. □

Lemma 3. If Assumptions 1 and 2 hold and $\kappa < \infty$, then

(a)

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\ell_t(\gamma_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\xrightarrow{\mathscr{D}}\mathcal{N}(0,\Omega_1),
$$

(b)

$$
\sup_{\gamma\in\Gamma}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{\partial\ell_t(\gamma,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{\partial\ell_t(\gamma_0,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\right\|=o_p(1);
$$

where $\Omega_1 = \mathrm{diag}(\alpha_0^{-1}, (\kappa-1)/(4\alpha_0^2))$.

Proof
(a) For any $\mathbf{c} = (c_1, c_2)' \in \mathbb{R}^2$, by Theorem 1, we can show that

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\{\left[\mathbf{c}'\frac{\partial\ell_t(\gamma_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right]^2|\mathcal{F}_{t-1}\right\} &= \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{c_1^2y_{t-1}^2}{\gamma_0+\alpha_0y_{t-1}^2}+\frac{c_2^2(\kappa-1)y_{t-1}^4}{4(\gamma_0+\alpha_0y_{t-1}^2)^2}\right\} \\
&= \frac{c_1^2}{\alpha_0}+\frac{c_2^2(\kappa-1)}{4\alpha_0^2}+o_p(1) \\
&= \mathbf{c}'\Omega_1\mathbf{c}+o_p(1).
\end{aligned}
$$

By the Cramér–Wold device and the martingale central limit theorem in Brown (1971), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \ell_t(\gamma_0,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{\mathscr{D}} \mathcal{N}(0,\Omega_1).$$

Thus, (a) holds.

By Theorem 1, we can show that

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\sup_{\boldsymbol{\theta}\in\Theta}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\frac{\partial \ell_t(\gamma,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}-\frac{\partial \ell_t(\gamma_0,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|\right\}\\
&\quad\le \frac{M}{\sqrt{n}}\sum_{t=1}^{n}(1+\eta_t^2)\left(\frac{1}{\sqrt{\underline{\gamma}+\underline{\alpha}\,y_{t-1}^2}}+\frac{1}{\underline{\gamma}+\underline{\alpha}\,y_{t-1}^2}\right)=o_p(1).
\end{aligned}$$

Thus, (b) holds. □

Proof of Theorem 2

(i) Note that $f(\boldsymbol{\theta})$ has a unique maximizer at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ on Θ. For any $\varepsilon>0$, we have $c\equiv\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\ge\varepsilon} f(\boldsymbol{\theta})<f(\boldsymbol{\theta}_0)=0$. Thus, by Lemma 2,

$$\begin{aligned}
&\mathbb{P}\left(\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|\ge\varepsilon\right)\\
&=\mathbb{P}\left(\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|\ge\varepsilon,\ \frac{1}{n}\sum_{t=1}^{n}[\ell_t(\hat{\gamma}_n,\hat{\boldsymbol{\theta}}_n)-\ell_t(\gamma_0,\boldsymbol{\theta}_0)]\ge 0\right)\\
&\le\mathbb{P}\left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\ge\varepsilon}\ \sup_{\gamma\in\Gamma}\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta})-\ell_t(\gamma_0,\boldsymbol{\theta}_0)-f(\boldsymbol{\theta})]+c\ge 0\right)\\
&\le\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta}\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}[\ell_t(\gamma,\boldsymbol{\theta})-\ell_t(\gamma_0,\boldsymbol{\theta}_0)-f(\boldsymbol{\theta})]\right|\ge -c\right)\to 0.
\end{aligned}$$

Thus, (i) holds.

(ii) By Taylor's expansion, Lemmas 1(b) and 3(b), it follows that

$$\begin{aligned}
0&=\frac{\partial L_n(\hat{\gamma}_n,\hat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\\
&=\frac{\partial L_n(\gamma_0,\hat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}+\left(\frac{\partial L_n(\hat{\gamma}_n,\hat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}-\frac{\partial L_n(\gamma_0,\hat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\right)\\
&=\frac{\partial L_n(\gamma_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}+\left[\frac{\partial^2 L_n(\gamma_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}+n\left\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right\|O_p(1)\right](\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+o_p(\sqrt{n}).
\end{aligned}$$

Thus, (ii) holds by (i) of this theorem, Lemma 3(a), and

$$\frac{1}{n}\frac{\partial^2 L_n(\gamma_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}=-\mathrm{diag}(\alpha_0^{-1},(2\alpha_0^2)^{-1})+o_p(1).$$

□

J. Time Ser. Anal. 35: 189–202 (2014)
DOI: 10.1111/jtsa.12058