



ELSEVIER

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Asymptotic inference in multiple-threshold double autoregressive models

Dong Li^a, Shiqing Ling^b, Jean-Michel Zakoïan^{c,*}

^a Yau Mathematical Sciences Center and Center for Statistical Science, Tsinghua University, Beijing 100084, China

^b Hong Kong University of Science and Technology, Hong Kong

^c University Lille 3 and CREST, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France

ARTICLE INFO

Article history:

Available online xxxx

JEL classification:

C13

C22

Keywords:

Compound Poisson process

Ergodicity

Quasi-maximum likelihood estimation

Strict stationarity

MTDAR model

Score test

ABSTRACT

This paper investigates a class of multiple-threshold models, called Multiple Threshold Double AR (MTDAR) models. A sufficient condition is obtained for the existence and uniqueness of a strictly stationary and ergodic solution to the first-order MTDAR model. We study the Quasi-Maximum Likelihood Estimator (QMLE) of the MTDAR model. The estimated thresholds are shown to be n -consistent, asymptotically independent, and to converge weakly to the smallest minimizer of a two-sided compound Poisson process. The remaining parameters are \sqrt{n} -consistent and asymptotically multivariate normal. In particular, these results apply to the multiple threshold ARCH model, with or without AR part, and to the multiple threshold AR models with ARCH errors. A score-based test is also presented to determine the number of thresholds in MTDAR models. The limiting distribution is shown to be distribution-free and is easy to implement in practice. Simulation studies are conducted to assess the performance of the QMLE and our score-based test in finite samples. The results are illustrated with an application to the quarterly US real GNP data over the period 1947–2013.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Tong's (1978) threshold autoregressive (TAR) models have been extensively investigated in the literature and are arguably the most popular class of nonlinear time series models for the conditional mean. For financial time series, however, the conditional mean modeling has to be completed by a specification of the conditional variance. Indeed, typical effects such as the volatility clustering or the leverage effect have been widely documented in the empirical finance literature and such effects cannot be captured with independent innovations. For the conditional variance, the popularity of GARCH-type models, both in applied and theoretical works, has always increased since their introduction by Engle (1982). See, for example, Francq and Zakoïan (2010) for an overview on GARCH models. When the GARCH model is not directly applied to observations, but rather to the innovations of linear or nonlinear time series model, it can be more natural and convenient to specify the volatility as a function of the past observations rather than the past innovations. An example of such model

is the double AR model introduced by Weiss (1984) and studied by Ling (2004, 2007).

In this article, we study the probabilistic properties and the estimation of a Multiple Threshold Double AR (MTDAR) model. More precisely, the model we consider in this article is the MTDAR($m; p$) defined by

$$y_t = \sum_{i=1}^m \left\{ c_i + \sum_{j=1}^p \phi_{ij} y_{t-j} + \eta_t \left(\omega_i + \sum_{j=1}^p \alpha_{ij} y_{t-j}^2 \right)^{1/2} \right\} \times I\{y_{t-d} \in \mathcal{R}_i\}, \quad (1.1)$$

where m, p and d are positive integers, $c_i, \phi_{ij} \in \mathbb{R}$, $\omega_i > 0$, $\alpha_{ij} \geq 0$, the m sets $\mathcal{R}_i = (r_{i-1}, r_i]$ constitute a partition of the real line, $-\infty = r_0 < r_1 < \dots < r_{m-1} < r_m = +\infty$, $I\{B\}$ denotes the indicator function of some event B , and $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with zero mean. The standard ARCH(p) model can be obtained as a particular case by taking $c_i = \phi_{ij} = 0$ and $\alpha_{ij} = \alpha_j$ for all i and j , while a version of TAR(p) model is obtained by canceling the α_{ij} 's.

The first aim of this paper is to study the stability properties of the MTDAR($m; 1$) model. The probabilistic structure of TAR models was studied by Chan et al. (1985), Chan and Tong (1985) and Tong (1990). Relying on the approach developed in the

* Corresponding author.

E-mail addresses: dongli@math.tsinghua.edu.cn (D. Li), maling@ust.hk (S. Ling), zakoian@ensae.fr (J.-M. Zakoïan).

<http://dx.doi.org/10.1016/j.jeconom.2015.03.033>

0304-4076/© 2015 Elsevier B.V. All rights reserved.

book by [Meyn and Tweedie \(1996\)](#), we obtain explicit ergodicity conditions depending on the parameters of the extremal regimes (when $y_{t-d} \in \mathcal{R}_1$ or $y_{t-d} \in \mathcal{R}_m$) and the innovations distribution. A different approach was used by [Cline and Pu \(2004\)](#) who established sharp ergodicity conditions for a general class of threshold AR-ARCH models under assumptions we will discuss further.

The second aim of this article is to study the asymptotic properties of the Gaussian Quasi-Maximum Likelihood (QML) estimator of the vector of parameters, including the double-AR coefficients and the thresholds. The third aim of this article is to develop a score-based test to determine the number of thresholds in MTDAR models.

The literature on the estimation of threshold time-series models is vast. To cite but a few of such articles, let us mention [Chan \(1993\)](#), [Hansen \(2000\)](#), and more recently, [Li and Ling \(2012\)](#), [Li et al. \(2013a\)](#). These articles study the asymptotic properties of least-squares estimators (LSE) in threshold linear (AR or MA) models. Our framework is that of a threshold nonlinear¹ time-series model, for which a QML criterion allows to simultaneously estimate the conditional mean and variance. Simultaneous QML estimation of the conditional mean and variance was studied by [Francq and Zakoian \(2004\)](#) in the case of ARMA-GARCH, and by [Meitz and Saikkonen \(2011\)](#) for a general class of nonlinear AR-GARCH(1,1) models. A difference with these papers is that the conditional variance in Model (1.1) is specified in function of the observations rather than the innovations. [Bardet and Wintenberger \(2009\)](#) proved the asymptotic properties of the Gaussian QMLE for a general class of multidimensional causal processes, in which both the conditional mean and variance are specified as functions of the observations. However, their conditions for consistency and asymptotic normality require the existence of moments of orders 2 and 4, respectively, which we do not need for the class (1.1). Moreover, their assumptions rule out the possibility of thresholds in the parameter vector. To our knowledge, asymptotic results for estimation of nonlinear multiple threshold time series models had not yet been established in the literature.

The article is organized as follows. In Section 2, we study the existence of a strictly stationary and geometrically ergodic solution. In Section 3, we derive the asymptotic properties of the QML estimator. Some special cases of MTDAR models are analyzed in Section 4. Section 5 develops a score-based test to determine the number of thresholds in MTDAR models. Section 6 reports simulation results on the QMLE and the score test in finite samples. An empirical application is proposed in Section 7. All proofs of Theorems are displayed in the [Appendix](#).

2. Stability properties of the MTDAR model

2.1. First-order model

We focus on the MTDAR(m ; 1) model

$$y_t = \sum_{i=1}^m (c_i + \phi_i y_{t-1} + \eta_t \sqrt{\omega_i + \alpha_i y_{t-1}^2}) I\{y_{t-1} \in \mathcal{R}_i\}. \quad (2.1)$$

Without loss of generality, we assume in this section that $r_1 \leq 0 \leq r_{m-1}$. The aim of this section is to establish conditions for the existence of a strictly stationary and nonanticipative solution² to (2.1). We make the following assumption.

A0: The distribution of η_t has a positive density f over \mathbb{R} . Moreover, $E|\eta_t|^s < \infty$ for some $s > 0$.

Let

$$\begin{aligned} \mu_1 &= E \log |\phi_1 - \eta_t \sqrt{\alpha_1}|, & p_1 &= P(\eta_t < \phi_1 / \sqrt{\alpha_1}), \\ \mu_m &= E \log |\phi_m + \eta_t \sqrt{\alpha_m}|, & p_m &= P(\eta_t > -\phi_m / \sqrt{\alpha_m}). \end{aligned}$$

By convention, $P(\eta_0 < a/b) = I\{a > 0\}$ if $b = 0$. Under **A0**, μ_1 and μ_m are well-defined but may be equal to $-\infty$ when $\phi_1 = 0$ or $\phi_m = 0$. We will prove the following result, using the approach developed by [Meyn and Tweedie \(1996\)](#) for establishing the geometric ergodicity of Markov chains.

Theorem 2.1. *Let Assumption A0 hold and assume*

$$\begin{aligned} \gamma &:= \max\{(1 + p_1)\mu_1 + (1 - p_1)\mu_m, \\ &\quad (1 - p_m)\mu_1 + (1 + p_m)\mu_m\} < 0. \end{aligned} \quad (2.2)$$

Then there exists a strictly stationary, nonanticipative solution $\{y_t\}$ to the MTDAR(m ;1) Model (2.1) and the solution is unique and geometrically ergodic with $E|y_t|^u < \infty$ for some $u > 0$.

Remark 2.1. A simple sufficient condition for (2.2) is $\mu_1 < 0$ and $\mu_m < 0$. Note that the strict stationarity condition only depends on the coefficients of the two extremal regimes. This remarkable feature was obtained in the first-order multiple threshold AR model by [Chan et al. \(1985\)](#).

Remark 2.2. When the model is the multiple-threshold AR(1) model (or simply, when $\alpha_1 = \alpha_m = 0$), condition (2.2) reduces to

$$(0 < \phi_1 < 1, 0 < \phi_m < 1) \quad \text{or} \quad (\phi_1 < 0, \phi_m < 0, \phi_1 \phi_m < 1)$$

or

$$(-1 < \phi_1 \phi_m < 0, \phi_1 < 1, \phi_m < 1),$$

which is slightly stronger, when $\phi_1 \phi_m < 0$, than the necessary and sufficient condition ($\phi_1 < 1, \phi_m < 1$ and $\phi_1 \phi_m < 1$) established by [Chan et al. \(1985\)](#). For a standard AR(1) model ($\phi_1 = \phi_m$) we obtain the standard stationarity constraint $|\phi_1| < 1$. [Fig. 1](#) gives the regions of (ϕ_1, ϕ_m) when $(\alpha_1, \alpha_m) = (0.1, 0.5), (1, 0.5), (1, 1)$ and $(1, 3)$ with $\eta_t \sim N(0, 1)$. We can see that ϕ_1, ϕ_m and $\phi_1 \phi_m$ may be greater than 1, the upper boundary given by [Chan et al. \(1985\)](#) for TAR(1, m) models.

Remark 2.3. When Model (2.1) reduces to a multiple threshold ARCH model, at least in its extremal regimes (i.e. $\phi_1 = \phi_m = 0$) with a symmetric density f , the condition (2.2) reduces to

$$\max(\alpha_1 \alpha_m^3, \alpha_1^3 \alpha_m) < \exp\{-4E \log \eta_t^2\}.$$

In particular, if $\alpha_1 = \alpha_m$, we retrieve the standard ARCH(1) condition: $\alpha_1 < \exp\{-E \log \eta_t^2\}$. When $\alpha_1 = \alpha_m = \alpha$ and $\phi_1 = \phi_m = \phi$ (which is in particular the case when $m = 1$), the condition (2.2) reduces to $\mu_1 + \mu_m < 0$, that is, $E \log |\phi^2 - \alpha \eta_t^2| < 0$. Moreover, if the distribution of η_t is symmetric, we then get the condition $\mu_1 = \mu_m < 0$, that is, $E \log |\phi - \eta_t \sqrt{\alpha}| < 0$. This is the necessary and sufficient strict stationarity condition obtained by [Ling \(2004\)](#) and [Ling and Li \(2008\)](#) for the double AR(1) model when η_t is normally distributed.

Remark 2.4. [Cline and Pu \(2004\)](#) obtained sharp ergodicity conditions for a general class of models encompassing Model (2.1), by an alternative approach called the *piggyback* method. From their Example 4.1, a condition for geometric ergodicity for our model is $(1 - p_m)\mu_1 + (1 - p_1)\mu_m < 0$ in our notations, which is in general a bit less restrictive than our condition (2.2). On the other hand, they need the assumption that $\sup_{x \in \mathbb{R}} \{(1 + |x|)f(x)\} < \infty$ which we do not require.

¹ nonlinearity having two causes: the thresholds and the presence of a volatility.

² A solution (y_t) is called nonanticipative if y_t can be written as a measurable function of $\{\eta_j : j \leq t\}$.

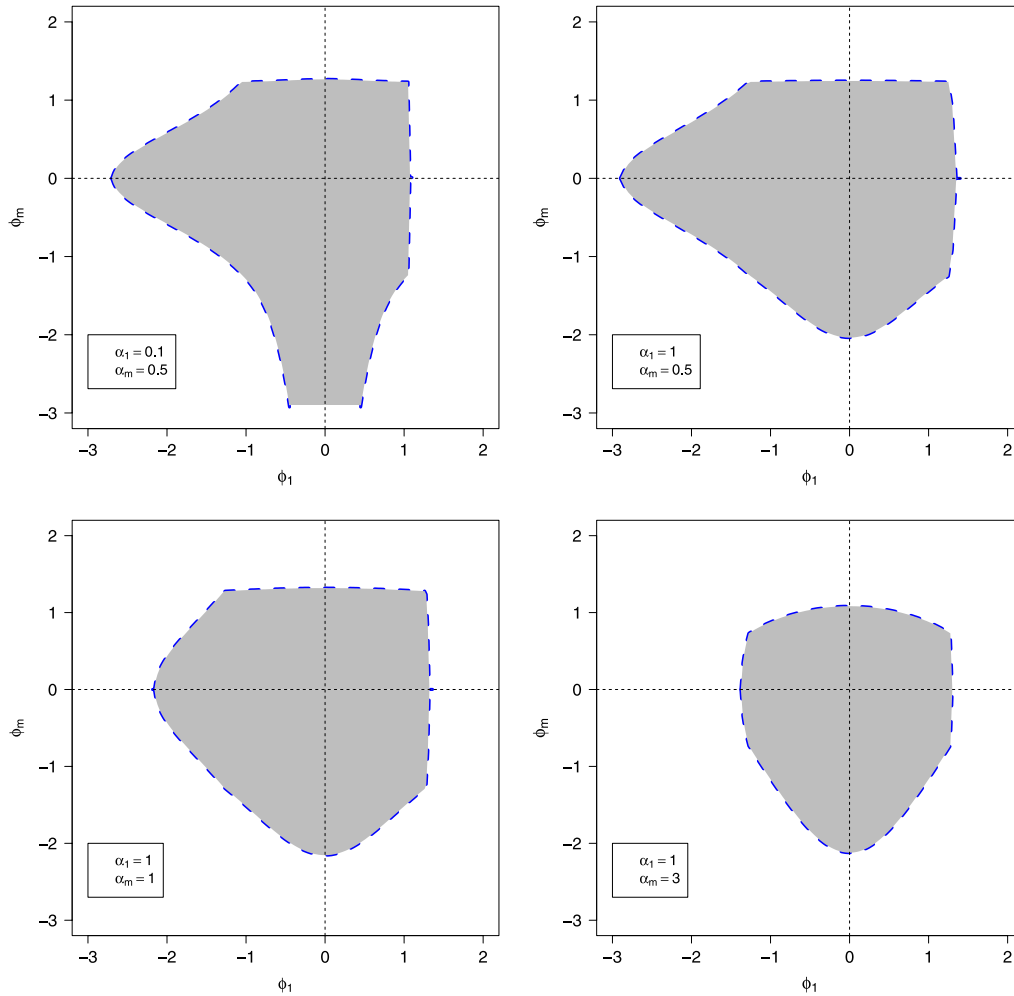


Fig. 1. Region of (ϕ_1, ϕ_m) with $(\alpha_1, \alpha_m) = (0.1, 0.5), (1, 0.5), (1, 1)$ and $(1, 3)$.

Remark 2.5. In the proof, we show that, for some constants $u, K > 0$ and $\rho \in (0, 1)$, we have

$$E(V(y_2) | y_0 = y) \leq K + \rho V(y) \quad \text{for all } y,$$

where $V(y) = 1 + |y|^u$. This condition entails that (y_t) is V -uniformly ergodic: with $\rho_1 \in (\rho, 1)$, there exists a constant $M > 0$ such that

$$\sup_A |P(y_k \in A | y_0 = y) - P(y_k \in A)| \leq M \rho_1^k V(y) \quad \text{for all } y \in \mathbb{R}, k \geq 1, \quad (2.3)$$

see [Meyn and Tweedie \(1996, Theorem 16.0.1\)](#). Thus, (2.2) is a sufficient condition for Assumption A.6 in the next section for MTDAR($m, 1$) models.

The following theorem provides a sufficient second-order stationarity condition. Let

$$\beta_1 = \frac{E[(\phi_1 - \eta_t \sqrt{\alpha_1})^2 I\{\phi_1 - \eta_t \sqrt{\alpha_1} > 0\}]}{\phi_1^2 + \alpha_1},$$

$$\beta_m = \frac{E[(\phi_m + \eta_t \sqrt{\alpha_m})^2 I\{\phi_m + \eta_t \sqrt{\alpha_m} > 0\}]}{\phi_m^2 + \alpha_m},$$

with, by convention, $\beta_i = 0$ if $\phi_i = \alpha_i = 0$.

Theorem 2.2. Let Assumption A0 hold and assume $E\eta_t^2 < \infty$ and

$$\begin{cases} (\phi_1^2 + \alpha_1)\{(\phi_1^2 + \alpha_1)\beta_1 + (\phi_m^2 + \alpha_m)(1 - \beta_1)\} < 1, \\ (\phi_m^2 + \alpha_m)\{(\phi_1^2 + \alpha_1)(1 - \beta_m) + (\phi_m^2 + \alpha_m)\beta_m\} < 1. \end{cases} \quad (2.4)$$

Then, there exists a strictly stationary, nonanticipative solution (y_t) with $E y_t^2 < \infty$.

Note that a simple sufficient condition for (2.4) to hold is that $\phi_1^2 + \alpha_1 < 1$ and $\phi_m^2 + \alpha_m < 1$. However, in (2.4) the volatility coefficients α_1, α_m are not both constrained to be less than 1: for instance when $\phi_1 = \phi_m = 0$ and η_t has a symmetric distribution with $E\eta_t^2 = 1$, the second-order stationarity condition becomes: $(\alpha_1 + \alpha_m) \max(\alpha_1, \alpha_m) < 2$.

2.2. Higher-order model

For the general p th order Model (1.1), a simple sufficient geometric ergodicity condition can be obtained by applying [Cline and Pu \(2004, Corollary 2.2\)](#). Under Assumption A0, and assuming that f is locally bounded away from 0 and satisfies $\sup_{x \in \mathbb{R}} \{(1 + |x|)f(x)\} < \infty$, then Model (1.1) admits a strictly stationary, geometrically ergodic solution if one of the following conditions holds:

1. $\sum_{j=1}^p \left(\sup_{1 \leq i \leq m} |\phi_{ij}|^r + \sup_{1 \leq i \leq m} (\alpha_{ij}^{r/2}) E|\eta_t|^r \right) < 1$, $r \in (0, 1)$;
2. $\left(\sum_{j=1}^p \sup_{1 \leq i \leq m} |\phi_{ij}| \right)^r + \sum_{j=1}^p \sup_{1 \leq i \leq m} (\alpha_{ij}^{r/2}) E|\eta_t|^r < 1$ with either $r \in (1, 2)$, or $r = 2$ and f symmetric.

Such conditions are simple but are over restrictive for the first-order model when compared to those of [Theorem 2.1](#). Obtaining stability conditions that only depend on the extreme regime coefficients for the p th order model with $p > 1$ seems difficult and is left for further research.

3. QML estimation of MTDAR($m; p$) models

We now turn to the estimation of the MTDAR($m; p$) model by the Gaussian QML method, assuming that the orders m and p are known positive integers. This method, which uses a Gaussian likelihood without assuming that the errors are Gaussian, has proven useful for GARCH-type models because it provides a consistent and asymptotically normal estimator under mild assumptions. In particular, it does not require high-order moments of the observed process.

The parameter, consisting of the AR and volatility coefficients and the thresholds, is denoted $\vartheta = (\lambda', \mathbf{r}') \equiv (\phi_1', \alpha_1', \dots, \phi_m', \alpha_m', \mathbf{r}')$, with $\phi_i = (c_i, \phi_{i1}, \dots, \phi_{ip})'$ and $\alpha_i = (\omega_i, \alpha_{i1}, \dots, \alpha_{ip})'$, $i = 1, \dots, m$, and $\mathbf{r} = (r_1, \dots, r_{m-1})'$. The true parameter value is denoted $\vartheta_0 = (\lambda_0', \mathbf{r}_0')$.

Assume that a sample $\{y_1, \dots, y_n\}$ is generated from Model (2.1). Given initial values $\{y_{1-p}, \dots, y_0\}$, the conditional log-likelihood function (omitting a constant) is defined as

$$L_n(\vartheta) = \sum_{t=1}^n l_t(\vartheta) \quad \text{with}$$

$$l_t(\vartheta) = -\frac{1}{2} \sum_{i=1}^m \left\{ \log(\alpha_i' \mathbf{X}_{t-1}) + \frac{(y_t - \phi_i' \mathbf{Y}_{t-1})^2}{\alpha_i' \mathbf{X}_{t-1}} \right\} \\ \times I\{r_{i-1} < y_{t-d} \leq r_i\},$$

where $\mathbf{Y}_{t-1} = (1, y_{t-1}, \dots, y_{t-p})'$, $\mathbf{X}_{t-1} = (1, y_{t-1}^2, \dots, y_{t-p}^2)'$, $r_0 = -\infty$, $r_m = \infty$. It can be shown that the choice of initial values does not matter for the asymptotic properties of the QML estimator. To save space the proof will be omitted (see Berkes et al. (2003), Francq and Zakoian (2004) for a proof in the case of GARCH and ARMA-GARCH models). In practice, d is unknown and can be estimated following the lines of proof of Chan (1993), Li and Ling (2012) among others. For simplicity, we assume that d is known and $1 \leq d \leq \max(p, 1)$.

Let $\Theta_r = \{\mathbf{r} = (r_1, \dots, r_{m-1})' \in [-\Gamma, \Gamma]^{m-1} : r_{i+1} - r_i \geq \delta\}$ for some constants $\delta > 0$ and $\Gamma > 0$. The parameter space is $\Theta = \Theta_\lambda \times \Theta_r$, where Θ_λ is a compact subset of $\mathbb{R}^{2m(p+1)}$ with $\omega_i \geq \underline{\omega}$ for some constant $\underline{\omega} > 0$ and $\alpha_{ij} \geq \underline{\alpha}$ for some constant $\underline{\alpha} > 0$. The maximizer of $L_n(\vartheta)$ is denoted by $\hat{\vartheta}_n$, i.e.,

$$\hat{\vartheta}_n = \arg \max_{\vartheta} L_n(\vartheta).$$

Since $L_n(\vartheta)$ is not continuous in \mathbf{r} , one can use two steps to find $\hat{\vartheta}_n$:

- For each fixed \mathbf{r} , maximize $L_n(\vartheta)$ over Θ_λ and get its maximizer $\hat{\lambda}_n(\mathbf{r})$.
- Since the profile log-likelihood $L_n^*(\mathbf{r}) \equiv L_n(\hat{\lambda}_n(\mathbf{r}), \mathbf{r})$ is a piecewise constant function over \mathbb{R}^{m-1} and only takes a finite number of possible values, one can get the maximizer $\hat{\mathbf{r}}_n$ of $L_n^*(\mathbf{r})$ by the enumeration approach and finally obtain the estimator $\hat{\vartheta}_n = (\hat{\lambda}_n'(\hat{\mathbf{r}}_n), \hat{\mathbf{r}}_n')$ by a plug-in method.

Let $\{y_{(1)}, \dots, y_{(n)}\}$ denote the order statistics of the sample $\{y_1, \dots, y_n\}$. If $(y_{(j_1)}, \dots, y_{(j_{m-1})})'$ is an estimate of \mathbf{r}_0 for some $j_1 < \dots < j_{m-1}$, then $L_n^*(\mathbf{r})$ is a constant over the $(m-1)$ -dimensional cube \mathcal{A} defined by $\mathcal{A} = \{\mathbf{r} = (r_1, \dots, r_{m-1})' : r_i \in [y_{(j_i)}, y_{(j_{i+1})}], i = 1, \dots, m-1\}$. Thus, there exist infinitely many \mathbf{r} such that $L_n(\cdot)$ can achieve its global maximum and each $\mathbf{r} \in \mathcal{A}$ can be considered as an estimate of \mathbf{r}_0 . In this case, we generally take $\hat{\mathbf{r}}_n = (y_{(j_1)}, \dots, y_{(j_{m-1})})'$ as a QMLE of \mathbf{r}_0 .

The strong consistency of the QMLE $\hat{\vartheta}_n$ of ϑ_0 relies on the following assumptions.

- A1:** The true value ϑ_0 belongs to Θ and $(\phi_{i0}', \alpha_{i0}') \neq (\phi_{i+1,0}', \alpha_{i+1,0}')$ for $i = 1, \dots, m-1$.

- A2:** (η_t) is a sequence of i.i.d. random variables with zero mean and unit variance, and η_1 has a positive density over \mathbb{R} .

- A3:** The process $\{y_t\}$ is a strictly stationary, nonanticipative and ergodic solution of Model (1.1) such that $E|y_t|^v < \infty$ for some $v > 0$.

Theorem 3.1. Assume that **A1–A3** hold. Then $\hat{\vartheta}_n \rightarrow \vartheta_0$ a.s.

To obtain the asymptotic normality of $\hat{\lambda}_n$, we make the following additional assumptions.

- A4:** ϑ_0 is an interior point of Θ .

- A5:** $E\eta_0^4 < \infty$.

To obtain the convergence rate of $\hat{\mathbf{r}}_n$, we also require two additional assumptions.

- A6:** The process (y_t) is V -uniformly ergodic.

- A7:** There exist nonrandom vectors $\mathbf{w}_i^* = (w_{i1}, \dots, w_{ip})'$ with $w_{id} = r_{i0}$ and $\mathbf{W}_i^* = (1, W_{i1}, \dots, W_{ip})'$ with $W_{id} = r_{i0}^2$ such that for $i = 1, \dots, m-1$,

$$\{(\phi_{i0} - \phi_{i+1,0})' \mathbf{w}_i^*\}^2 + \{(\alpha_{i0} - \alpha_{i+1,0})' \mathbf{W}_i^*\}^2 > 0.$$

Remark 3.1. Assumption **A7** is similar to the CONDITION 4 in Chan (1993) and implies that either the conditional mean function or volatility function in Model (2.1) is discontinuous at each threshold r_{i0} . It is a necessary condition for the n -convergence rate of $\hat{\mathbf{r}}_n$. If $\alpha_{i0} = \alpha_{j0}$ for all i and j , then Assumption **A7** is equivalent to $(\phi_{i0} - \phi_{i+1,0})' \mathbf{w}_i^* \neq 0$ for all i , which is Assumption 3.4 in Li and Ling (2012) that the conditional mean function is discontinuous at each threshold r_{i0} .

The next theorem gives the convergence rate of $\hat{\mathbf{r}}_n$ and shows that the asymptotic distribution of the estimators of the AR and ARCH coefficients is the same as if the thresholds were known.

Theorem 3.2. If the assumptions **A1–A7** hold, then

- $n(\hat{\mathbf{r}}_n - \mathbf{r}_0) = O_p(1)$;
- $\sqrt{n} \sup_{\|\mathbf{r} - \mathbf{r}_0\| \leq B/n} \|\hat{\lambda}_n(\mathbf{r}) - \hat{\lambda}_n(\mathbf{r}_0)\| = o_p(1)$ for any fixed constant $0 < B < \infty$. Further, it follows that

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) = \sqrt{n}(\hat{\lambda}_n(\mathbf{r}_0) - \lambda_0) + o_p(1) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{J}),$$

where $\mathbf{J} = \text{diag}\{J_1, \dots, J_m\}$ and $J_i = \Sigma_i^{-1} \Omega_i \Sigma_i^{-1}$ with

$$\Omega_i = E \begin{pmatrix} \frac{\mathbf{Y}_{t-1} \mathbf{Y}_{t-1}' I_{it}}{\alpha_{i0}' \mathbf{X}_{t-1}} & \frac{\kappa_3}{2} \frac{\mathbf{Y}_{t-1} \mathbf{X}_{t-1}' I_{it}}{(\alpha_{i0}' \mathbf{X}_{t-1})^{3/2}} \\ \frac{\kappa_3}{2} \frac{\mathbf{X}_{t-1} \mathbf{Y}_{t-1}' I_{it}}{(\alpha_{i0}' \mathbf{X}_{t-1})^{3/2}} & \frac{\kappa_4 - 1}{4} \frac{\mathbf{X}_{t-1} \mathbf{X}_{t-1}' I_{it}}{(\alpha_{i0}' \mathbf{X}_{t-1})^2} \end{pmatrix},$$

$$\Sigma_i = E \left\{ \text{diag} \left(\frac{\mathbf{Y}_{t-1} \mathbf{Y}_{t-1}' I_{it}}{\alpha_{i0}' \mathbf{X}_{t-1}}, \frac{1}{2} \frac{\mathbf{X}_{t-1} \mathbf{X}_{t-1}' I_{it}}{(\alpha_{i0}' \mathbf{X}_{t-1})^2} \right) \right\},$$

$I_{it} = I\{r_{i-1,0} < y_{t-d} \leq r_{i0}\}$ and $\kappa_j = E\eta_t^j$ for $i = 1, \dots, m$ and $j = 3, 4$.

To construct confidence intervals for λ_0 , we need to estimate the matrices Σ_i and Ω_i by their sample counterparts. For example, we can estimate Σ_i by

$$\hat{\Sigma}_i = \text{diag} \left(\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{Y}_{t-1} \mathbf{Y}_{t-1}' I\{\hat{r}_{i-1,n} < y_{t-d} \leq \hat{r}_{in}\}}{\hat{\alpha}_{in}' \mathbf{X}_{t-1}}, \right. \\ \left. \frac{1}{2n} \sum_{t=1}^n \frac{\mathbf{X}_{t-1} \mathbf{X}_{t-1}' I\{\hat{r}_{i-1,n} < y_{t-d} \leq \hat{r}_{in}\}}{(\hat{\alpha}_{in}' \mathbf{X}_{t-1})^2} \right).$$

It can be shown that $\hat{\Sigma}_i$ is a consistent estimator of Σ_i by using Theorem 3.2.

To study the limiting distribution of $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$, we first define $(m-1)$ independent 1-dimensional two-sided compound Poisson processes $\{\mathcal{P}_i(z), z \in \mathbb{R}\}$ as

$$\mathcal{P}_i(z) = I\{z < 0\} \sum_{k=1}^{N_1^{(i)}(|z|)} Y_k^{(i,i+1)} + I\{z \geq 0\} \sum_{k=1}^{N_2^{(i)}(z)} Z_k^{(i+1,i)}, \quad z \in \mathbb{R}, \quad (3.1)$$

for $i = 1, \dots, m-1$, where $\{N_1^{(i)}(z), z \geq 0\}$ and $\{N_2^{(i)}(z), z \geq 0\}$ are two independent Poisson processes with $N_1^{(i)}(0) = N_2^{(i)}(0) = 0$ a.s. and with the same jump rate $\pi(r_{i0})$, where $\pi(\cdot)$ is the density function of y_1 . The sequences of variables $\{Y_k^{(i,i+1)} : k \geq 1\}$ and $\{Z_k^{(i+1,i)} : k \geq 1\}$ are i.i.d., mutually independent, and are distributed as $F_{(i,i+1)}(\cdot|r_{i0})$ and $F_{(i+1,i)}(\cdot|r_{i0})$, respectively, where $F_{(i,j)}(\cdot|r_{k0})$ is the conditional distribution function of $\xi_{d+1}^{(i,j)}$ given $y_1 = r_{k0}$, and

$$\xi_t^{(i,j)} = \log \frac{\alpha'_{j0} \mathbf{x}_{t-1}}{\alpha'_{j0} \mathbf{x}_{t-1}} + \frac{\left\{ (\phi_{i0} - \phi_{j0})' \mathbf{y}_{t-1} + \eta_t \sqrt{\alpha'_{j0} \mathbf{x}_{t-1}} \right\}^2}{\alpha'_{j0} \mathbf{x}_{t-1}} - \eta_t^2.$$

Here, we work with the left continuous version for $N_1^{(i)}(\cdot)$ and the right continuous version for $N_2^{(i)}(\cdot)$ for $i = 1, \dots, m-1$.

We further define a spatial compound Poisson process $\wp(\mathbf{s})$ as follows,

$$\wp(\mathbf{s}) = \sum_{i=1}^{m-1} \mathcal{P}_i(s_i), \quad \mathbf{s} = (s_1, \dots, s_{m-1})' \in \mathbb{R}^{m-1}. \quad (3.2)$$

Clearly, $\wp(\mathbf{s})$ goes to ∞ a.s. when $\|\mathbf{s}\| \rightarrow \infty$ since $\mathbb{E}Y_1^{(i,i+1)} > 0$ and $\mathbb{E}Z_1^{(i+1,i)} > 0$ by a conditional argument for $i = 1, \dots, m-1$. Therefore, there exists a unique random $(m-1)$ -dimensional cube $[\mathbf{M}_-, \mathbf{M}_+] \equiv [M_-^{(1)}, M_+^{(1)}] \times \dots \times [M_-^{(m-1)}, M_+^{(m-1)}]$ on which the process $\wp(\mathbf{s})$ attains its global minimum a.s. That is,

$$[\mathbf{M}_-, \mathbf{M}_+] = \arg \min_{\mathbf{s} \in \mathbb{R}^{m-1}} \wp(\mathbf{s}),$$

which is equivalent to

$$[M_-^{(i)}, M_+^{(i)}] = \arg \min_{z \in \mathbb{R}} \mathcal{P}_i(z), \quad i = 1, \dots, m-1.$$

Note that, the processes $\{\mathcal{P}_i(z) : i = 1, \dots, m-1\}$ being independent, so are $\{M_-^{(i)} : i = 1, \dots, m-1\}$. By a technique similar to that used in the proof of Theorem 3.3 in Li and Ling (2012), we can show that $\wp_n(\mathbf{s}) \Rightarrow \wp(\mathbf{s})$ as $n \rightarrow \infty$, and we deduce the following result. The proof is omitted.

Theorem 3.3. *If Assumptions A1–A7 hold, then $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$ converges weakly to \mathbf{M}_- and its components are asymptotically independent as $n \rightarrow \infty$. Furthermore, $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$ is asymptotically independent of $\sqrt{n}(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0)$.*

In practice, it is difficult to use $M_-^{(i)}$ to construct confidence intervals for r_{i0} since $M_-^{(i)}$ does not have a closed form. A simulation procedure is needed to obtain the percentiles of $M_-^{(i)}$. Such a procedure is exactly the same as that given in Li et al. (forthcoming) for the MTDAR(1, p) model and hence the details are omitted, see also Li and Ling (2012) and Gonzalo and Wolf (2005).

4. Special cases

This section applies the results in Section 3 to several popular conditionally heteroscedastic threshold models. Asymptotic results for such models are new in the literature, to our knowledge.

4.1. Multiple threshold ARCH models

The multiple threshold ARCH (MTARCH) model is defined as

$$y_t = \eta_t \sqrt{h_t} \quad \text{and} \quad h_t = \sum_{i=1}^m \left(\omega_i + \sum_{j=1}^p \alpha_{ij} y_{t-j}^2 \right) I\{r_{i-1} < y_{t-d} \leq r_i\}.$$

When threshold parameters are known, variants of this class of MTARCH models were studied by Gouriéroux and Monfort (1992), Rabemananjara and Zakoian (1993), Zakoian (1994), Li and Li (1996) among others. Recently, Chan et al. (2014) considered a special MTARCH model in which the volatility function h_t is piecewise constant, and studied the asymptotic properties of the QMLE of the parameter when the thresholds are unknown.

Let $\boldsymbol{\lambda} \equiv (\alpha'_1, \dots, \alpha'_m)'$ and its QMLE be $\hat{\boldsymbol{\lambda}}_n$. By Theorem 3.2, we have

Theorem 4.1. *If the assumptions A1–A7 with $\phi_{10} = \dots = \phi_{m0} = 0$ hold, then*

$$n(\hat{\mathbf{r}}_n - \mathbf{r}_0) = O_p(1) \quad \text{and} \quad \sqrt{n}(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{J}),$$

where $\mathbf{J} = (\kappa_4 - 1) \text{diag}\{J_1, \dots, J_m\}$ with

$$J_i = \left\{ E \left(\frac{\mathbf{x}_{t-1} \mathbf{x}'_{t-1} I_{it}}{(\alpha'_{i0} \mathbf{x}_{t-1})^2} \right) \right\}^{-1}.$$

The limiting distribution of $\hat{\mathbf{r}}_n$ is the same as that in Theorem 3.3, except that the jump distributions in the corresponding two-sided compound Poisson processes are replaced by the conditional distribution of

$$\xi_t^{(i,j)} = \log \frac{\alpha'_{j0} \mathbf{x}_{t-1}}{\alpha'_{j0} \mathbf{x}_{t-1}} + \frac{\eta_t^2 (\alpha_{i0} - \alpha_{j0})' \mathbf{x}_{t-1}}{\alpha'_{j0} \mathbf{x}_{t-1}}, \quad \text{given } y_{t-d}.$$

4.2. Multiple threshold AR models with conditionally heteroscedastic errors

We consider the class of multiple threshold AR model with conditionally heteroscedastic errors defined as follows:

$$\begin{cases} y_t = \sum_{i=1}^m \left(c_i + \sum_{j=1}^p \phi_{ij} y_{t-j} \right) I\{r_{i-1} < y_{t-d} \leq r_i\} + \varepsilon_t, \\ \varepsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \omega + \sum_{j=1}^p \alpha_j y_{t-j}^2. \end{cases}$$

This model generalizes Ling's (2007) double AR model by considering a threshold effect in the mean part. Zhang et al. (2011) investigated asymptotic properties of the QMLE when the threshold is known. However, when the threshold is unknown, the asymptotic properties of the QMLE are not available in literature.

Let $\boldsymbol{\lambda} \equiv (\phi'_1, \dots, \phi'_m, \alpha')'$ with $\boldsymbol{\alpha} = (\omega, \alpha_1, \dots, \alpha_p)'$ and its QMLE be $\hat{\boldsymbol{\lambda}}_n$. We have the theorem:

Theorem 4.2. *If the assumptions A1–A7 with $\alpha_{10} = \dots = \alpha_{m0} = 0$ hold, then*

$$n(\hat{\mathbf{r}}_n - \mathbf{r}_0) = O_p(1) \quad \text{and} \quad \sqrt{n}(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega} \boldsymbol{\Sigma}^{-1}),$$

with

$$\boldsymbol{\Sigma} = E \left\{ \text{diag} \left(\frac{\mathbf{y}_{t-1} \mathbf{y}'_{t-1} I_{1t}}{\alpha'_{0} \mathbf{x}_{t-1}}, \dots, \frac{\mathbf{y}_{t-1} \mathbf{y}'_{t-1} I_{mt}}{\alpha'_{0} \mathbf{x}_{t-1}}, \frac{1}{2} \frac{\mathbf{x}_{t-1} \mathbf{x}'_{t-1}}{(\alpha'_{0} \mathbf{x}_{t-1})^2} \right) \right\},$$

$$\Omega = E \left(\begin{array}{c|c} \tilde{\Omega} & \begin{array}{c} \frac{\kappa_3}{2} \frac{\mathbf{Y}_{t-1}\mathbf{X}'_{t-1}I_{1t}}{(\alpha'_0\mathbf{X}_{t-1})^{3/2}} \\ \vdots \\ \frac{\kappa_3}{2} \frac{\mathbf{Y}_{t-1}\mathbf{X}'_{t-1}I_{mt}}{(\alpha'_0\mathbf{X}_{t-1})^{3/2}} \end{array} \\ \hline \begin{array}{c} \frac{\kappa_3}{2} \frac{\mathbf{X}_{t-1}\mathbf{Y}'_{t-1}I_{1t}}{(\alpha'_0\mathbf{X}_{t-1})^{3/2}} \quad \dots \quad \frac{\kappa_3}{2} \frac{\mathbf{X}_{t-1}\mathbf{Y}'_{t-1}I_{mt}}{(\alpha'_0\mathbf{X}_{t-1})^{3/2}} \end{array} & \begin{array}{c} \frac{\kappa_4-1}{4} \frac{\mathbf{X}_{t-1}\mathbf{X}'_{t-1}}{(\alpha'_0\mathbf{X}_{t-1})^2} \end{array} \end{array} \right),$$

$$\tilde{\Omega} = \text{diag}(I_{1t}, \dots, I_{mt}) \otimes \left(\frac{\mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}}{\alpha'_0\mathbf{X}_{t-1}} \right),$$

where \otimes is the Kronecker product.

The limiting distribution of $\hat{\mathbf{r}}_n$ is the same as that in Theorem 3.3, except that the jump distributions in the corresponding two-sided compound Poisson processes are replaced by the conditional distribution of

$$\xi_t^{(i,j)} = \frac{\{(\phi_{i0} - \phi_{j0})'\mathbf{Y}_{t-1}\}^2}{\alpha'_0\mathbf{X}_{t-1}} + \frac{2\eta_t(\phi_{i0} - \phi_{j0})'\mathbf{Y}_{t-1}}{\sqrt{\alpha'_0\mathbf{X}_{t-1}}}, \quad \text{given } y_{t-d}.$$

4.3. AR models with multiple-threshold conditionally heteroscedastic errors

In linear AR models, the asymptotic properties of estimators or tests on the AR coefficients are generally derived under three types of assumptions on the errors process: i.i.d., martingale difference or uncorrelated sequence (see for instance Francq et al. (2005)). To further specify the errors dependence structure, various heteroscedasticity models were introduced, see Li et al. (2002), Francq and Zakoian (2010) for an overview. A special case of Model (1.1), in which the thresholds are only present in the volatility, is defined as follows:

$$\begin{cases} y_t = \sum_{j=1}^p \phi_j y_{t-j} + \varepsilon_t, \\ \varepsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \sum_{i=1}^m \left(\omega_i + \sum_{j=1}^p \alpha_{ij} y_{t-j}^2 \right) I\{r_{i-1} < y_{t-d} \leq r_i\}. \end{cases}$$

Let $\boldsymbol{\lambda} \equiv (\phi', \alpha'_1, \dots, \alpha'_m)'$ and its QMLE be $\hat{\boldsymbol{\lambda}}_n$. We have the theorem:

Theorem 4.3. *If the assumptions A1–A7 with $\phi_{10} = \dots = \phi_{m0} = \phi_0$ hold, then*

$$n(\hat{\mathbf{r}}_n - \mathbf{r}_0) = O_p(1) \quad \text{and} \quad \sqrt{n}(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega} \boldsymbol{\Sigma}^{-1}),$$

where

$$\boldsymbol{\Sigma} = E \left\{ \text{diag} \left(\sum_{i=1}^m \frac{\mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}I_{it}}{\alpha'_0\mathbf{X}_{t-1}}, \frac{1}{2} \frac{\mathbf{X}_{t-1}\mathbf{X}'_{t-1}I_{1t}}{(\alpha'_0\mathbf{X}_{t-1})^2}, \dots, \frac{1}{2} \frac{\mathbf{X}_{t-1}\mathbf{X}'_{t-1}I_{mt}}{(\alpha'_0\mathbf{X}_{t-1})^2} \right) \right\},$$

$$\boldsymbol{\Omega} = E \left(\begin{array}{c|c} \sum_{i=1}^m \frac{\mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}I_{it}}{\alpha'_0\mathbf{X}_{t-1}} & \begin{array}{c} \frac{\kappa_3}{2} \frac{\mathbf{Y}_{t-1}\mathbf{X}'_{t-1}I_{1t}}{(\alpha'_0\mathbf{X}_{t-1})^{3/2}} \quad \dots \quad \frac{\kappa_3}{2} \frac{\mathbf{Y}_{t-1}\mathbf{X}'_{t-1}I_{mt}}{(\alpha'_0\mathbf{X}_{t-1})^{3/2}} \end{array} \\ \hline \begin{array}{c} \frac{\kappa_3}{2} \frac{\mathbf{X}_{t-1}\mathbf{Y}'_{t-1}I_{1t}}{(\alpha'_0\mathbf{X}_{t-1})^{3/2}} \\ \vdots \\ \frac{\kappa_3}{2} \frac{\mathbf{X}_{t-1}\mathbf{Y}'_{t-1}I_{mt}}{(\alpha'_0\mathbf{X}_{t-1})^{3/2}} \end{array} & \tilde{\Omega} \end{array} \right),$$

$$\tilde{\Omega} = \frac{\kappa_4-1}{4} \text{diag} \{ (\alpha'_0\mathbf{X}_{t-1})^{-2}I_{1t}, \dots, (\alpha'_0\mathbf{X}_{t-1})^{-2}I_{mt} \} \otimes (\mathbf{X}_{t-1}\mathbf{X}'_{t-1}).$$

The limiting distribution of $\hat{\mathbf{r}}_n$ is the same as that in Theorem 3.3 with the form of $\xi_t^{(i,j)}$ defined as in Theorem 4.1, that is, the AR part does not affect the function of $\xi_t^{(i,j)}$ in terms of $\{y_t\}$.

5. Determination of the number of thresholds in MTDAR models

It is always an important issue to determine the number of thresholds in threshold models. In this section, we will develop a score-based test as in Ling and Tong (2011) and Li et al. (forthcoming), which is asymptotically distribution-free and is easy to implement in practice. The relevant critical values are available without bootstrap, unlike the likelihood ratio test in Chan (1990), Chan and Tong (1990), Wong and Li (1997, 2000), among others.

Under the null H_0 , we suppose that $\{y_t\}$ follows a MTDAR($m; p$) model:

$$y_t = \sum_{i=1}^m \left(\phi'_i \mathbf{Y}_{t-1} + \eta_t \sqrt{\alpha'_i \mathbf{X}_{t-1}} \right) I\{r_{i-1} < y_{t-d} \leq r_i\}. \quad (5.1)$$

Here and throughout we use the notations in Section 3. Denote by $\hat{\boldsymbol{\vartheta}}_n$ the QMLE of $\boldsymbol{\vartheta}_0$. By Theorem 3.2, it follows that

$$\sqrt{n}(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) = \boldsymbol{\Sigma}^{-1} n^{-1/2} \sum_{t=1}^n D_t(\boldsymbol{\vartheta}_0) + o_p(1),$$

where $\boldsymbol{\Sigma} = \text{diag}\{\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m\}$, $D_t(\boldsymbol{\vartheta}) = (D'_{1t}(\boldsymbol{\vartheta}), \dots, D'_{mt}(\boldsymbol{\vartheta}))'$ with

$$D_{it}(\boldsymbol{\vartheta}) = \left(\frac{(y_t - \phi'_i \mathbf{Y}_{t-1})\mathbf{Y}'_{t-1}}{\alpha'_i \mathbf{X}_{t-1}}, \frac{[(y_t - \phi'_i \mathbf{Y}_{t-1})^2 - (\alpha'_i \mathbf{X}_{t-1})]\mathbf{X}'_{t-1}}{2(\alpha'_i \mathbf{X}_{t-1})^2} \right)' \times I\{r_{i-1} < y_{t-d} \leq r_i\}.$$

Following Ling and Tong (2011), we first define a score-marked empirical process as follows:

$$T_{in}(x, \boldsymbol{\vartheta}) = n^{-1/2} \sum_{t=1}^n \mathbf{U}^{-1} D_{it}(\boldsymbol{\vartheta}) I\{r_{i-1} < y_{t-d} \leq x\}, \quad x \in \mathbb{R}, \quad (5.2)$$

for $i = 1, \dots, m$, where $\mathbf{U} = \text{diag}\{\mathbf{I}_{p+1}, \sqrt{(\kappa_4-1)/2}\mathbf{I}_{p+1}\}$ with \mathbf{I}_{p+1} being the identity matrix and $I(r_{i-1} < y_{t-d} \leq x) = 0$ if $x \leq r_{i-1}$ by convention. Let $\mathbb{D}[-\infty, \infty]$ denote the space of functions on $[-\infty, \infty]$ which are right continuous and have left-hand limits, equipped with the Skorokhod topology as in Billingsley (1968). By Theorem A.1 in Li et al. (forthcoming), Theorem 3.2 and Taylor's expansion, we can get the asymptotic property of $\{T_{in}(x, \hat{\boldsymbol{\vartheta}}_n)\}$.

Theorem 5.1. *Under the null H_0 that $\{y_t\}$ follows Model (5.1) with the true value $\boldsymbol{\vartheta}_0$, if the assumptions A1–A7 hold and η_t is symmetrically distributed, then, for $i = 1, \dots, m$,*

$$\sup_{x \in \mathbb{R}} \|T_{in}(x, \hat{\boldsymbol{\vartheta}}_n) - T_{in}(x, \boldsymbol{\vartheta}_0) + \mathbf{U}^{-1} \boldsymbol{\Sigma}_{ix} \boldsymbol{\Sigma}_{i,\infty}^{-1} n^{-1/2} \sum_{t=1}^n D_{it}(\boldsymbol{\vartheta}_0)\| = o_p(1),$$

and $T_{in}(x, \hat{\boldsymbol{\vartheta}}_n)$ converges weakly to $G_i(x)$ in $\mathbb{D}[-\infty, \infty]$, where all $G_i(x)$'s are independent and $G_i(x)$ is a $2(p+1)$ -dimensional Gaussian process with mean zero and covariance kernel $\mathbf{K}_{i,xy} = \boldsymbol{\Sigma}_{i,x \wedge y} - \boldsymbol{\Sigma}_{ix} \boldsymbol{\Sigma}_{i,\infty}^{-1} \boldsymbol{\Sigma}_{iy}$ for $x, y \in (r_{i-1,0}, r_{i0}]$, where

$$\boldsymbol{\Sigma}_{ix} = E \left[\text{diag} \left\{ \frac{\mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}}{\alpha'_{i0} \mathbf{X}_{t-1}}, \frac{\mathbf{X}_{t-1}\mathbf{X}'_{t-1}}{2(\alpha'_{i0} \mathbf{X}_{t-1})^2} \right\} \times I\{r_{i-1,0} \wedge x < y_{t-d} \leq r_{i0} \wedge x\} \right].$$

Almost all paths of $G_i(x)$ are continuous in $x \in (r_{i-1,0}, r_{i0}]$.

We first estimate Σ_{ix} by

$$\hat{\Sigma}_{ix} = \frac{1}{n} \sum_{t=1}^n \text{diag} \left\{ \frac{\mathbf{Y}_{t-1} \mathbf{Y}'_{t-1}}{\hat{\alpha}'_{in} \mathbf{X}_{t-1}}, \frac{\mathbf{X}_{t-1} \mathbf{X}'_{t-1}}{2(\hat{\alpha}'_{in} \mathbf{X}_{t-1})^2} \right\} \\ \times I\{\hat{r}_{i-1,n} \wedge x < y_{t-d} \leq \hat{r}_{in} \wedge x\},$$

where $i = 1, \dots, m$, $\hat{r}_{0n} = -\infty$ and $\hat{r}_{mn} = \infty$. It is not hard to see that $\hat{\Sigma}_{ix}$ is a consistent estimator of Σ_{ix} uniformly in $x \in [-\infty, \infty]$. Now, we define the test statistic

$$S_{in} = \max_{x \in [a_i, \hat{r}_{in}]} \frac{[\beta' \hat{\Sigma}_{ix}^{-1} T_{in}(x, \hat{\theta}_n)]^2}{\beta' (\hat{\Sigma}_{ia_i}^{-1} - \hat{\Sigma}_{i\hat{r}_{in}}^{-1}) \beta}, \quad (5.3)$$

where $a_i \in (\hat{r}_{i-1,n}, \hat{r}_{in})$ and β is a nonzero $2(p+1) \times 1$ constant vector. We need to choose a_i such that $\hat{\Sigma}_{ia_i}^{-1}$ exists in practice. S_{in} is to test if the linear AR model is adequate in the regime $(r_{i-1}, r_i]$. Since we do not have a specific alternative, it is considered to be a portmanteau test as called in Ling and Tong (2011). However, when the alternative is a TAR model, i.e., there is another threshold in $(r_{i-1}, r_i]$, this test is more powerful than that under other alternatives, see the simulation study in Ling and Tong (2011). Furthermore, when $p = 1$, it is equivalent to the log-likelihood ratio test in Chan (1991). Thus, it is expected that S_{in} can be used to determine the number of threshold in Model (5.1).

By Theorem 5.1 and the continuous mapping theorem, we have the following result.

Theorem 5.2. *If the assumptions in Theorem 5.1 hold, then, for any $2(p+1) \times 1$ nonzero constant vector β , it follows that*

$$\lim_{n \rightarrow \infty} P(S_{in} \leq y) = P\left(\max_{\tau \in [0,1]} B_i^2(\tau) \leq y\right)$$

for any $y \in \mathbb{R}$ and $B_i(\tau)$ is a standard Brownian motion on $\mathbb{C}[0, 1]$.

The limiting distribution of S_{in} , unlike the test of Chan (1991), does not depend on the choice of a_i since the weight function cancels out the related component, see Ling and Tong (2011). From the formula in Shorack and Wellner (1986, p.34)

$$P\left(\max_{\tau \in [0,1]} B_i^2(\tau) \leq x\right) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{(2k+1)^2 \pi^2}{8x}\right), \\ x \geq 0,$$

we can choose an approximate critical value C_α such that $P(\max_{\tau \in [0,1]} B_i^2(\tau) \geq C_\alpha) = \alpha$ for rejecting the null H_0 at different significance levels α . Numerous simulation studies show that $\beta = (1, \dots, 1)'$, with a_i around the 15% quantile of data $\{y_t : \hat{r}_{i-1,n} < y_{t-d} \leq \hat{r}_{in}\}$, produces a good power.

When the alternative is a MTDAR($m+1, p$) model, it is not hard to see that there are $(m-1)$ tests which have less power among $\{S_{1n}, \dots, S_{mn}\}$. Thus, it is not a good idea to use a single test, say S_{1n} , to test the number of thresholds in Model (5.1). Testing the null Model (5.1) is equivalent to the joint testing that the linear AR models are adequate in all the regimes. Thus, a natural test statistic for this is

$$S_n = \max_{1 \leq i \leq m} \{S_{in}\}. \quad (5.4)$$

By Theorem 5.1, all the $G_i(x)$, $i = 1, \dots, m$, are independent and so are the $B_i(x)$, $i = 1, \dots, m$. Thus, it follows that, under the assumptions of Theorem 5.1,

$$\lim_{n \rightarrow \infty} P(S_n \leq x) = P\left(\max_{1 \leq i \leq m} \max_{\tau \in [0,1]} B_i^2(\tau) \leq x\right) \\ = \left\{ \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{(2k+1)^2 \pi^2}{8x}\right) \right\}^m, \quad x \geq 0.$$

Table 1 provides approximate critical values of S_n for rejecting the null H_0 at the significance levels $\alpha = 1\%, 5\%, 10\%$ and $m = 1, \dots, 4$.

Table 1

Approximate critical values of S_n at the significance level α .

$m \setminus \alpha$	10%	5%	1%
1	3.841	5.024	7.879
2	4.979	6.216	9.136
3	5.670	6.930	9.878
4	6.169	7.442	10.408

6. Simulation studies

To assess the performance of the QMLE in finite samples, we use the sample sizes $n = 300, 600$ and 900 , each with 1000 replications of the following model

$$y_t = \begin{cases} \phi_1 y_{t-1} + \eta_t \sqrt{\omega_1 + \alpha_1 y_{t-1}^2}, & \text{if } y_{t-1} \leq r_1, \\ \phi_2 y_{t-1} + \eta_t \sqrt{\omega_2 + \alpha_2 y_{t-1}^2}, & \text{if } r_1 < y_{t-1} \leq r_2, \\ \phi_3 y_{t-1} + \eta_t \sqrt{\omega_3 + \alpha_3 y_{t-1}^2}, & \text{if } y_{t-1} > r_2, \end{cases} \quad (6.1)$$

with true value $(\phi_1, \omega_1, \alpha_1; \phi_2, \omega_2, \alpha_2; \phi_3, \omega_3, \alpha_3; r_1, r_2) = (0.5, 1, 0.3; 1, 0.5, 3; -0.7, 1, 0.5; -1, 0)$ and $\eta_t \stackrel{i.i.d.}{\sim} N(0, 1)$. For Model (6.1), we have $E \log |\phi_1 + \eta_t \sqrt{\alpha_1}| = -0.8725$ and $E \log |\phi_3 + \eta_t \sqrt{\alpha_3}| = -0.5623$. Thus, model (6.1) is strictly stationary, although $E \log |\phi_2 + \eta_t \sqrt{\alpha_2}| = 0.0719 > 0$.

Table 2 reports the empirical means (EM), empirical standard deviations (ESD) and asymptotic standard deviations (ASD) of the QMLE in Model (6.1). Here, the ASD of $\hat{\lambda}_n$ and \hat{r}_n are computed by using Σ and Ω in Theorem 3.2 and by simulating the compound Poisson processes defined in (3.1), respectively. From Table 2, we see that both ESDs and ASDs generally become smaller and closer to each other when the sample size n increases. We also see that the values of the ESD for \hat{r}_n are approximately halved whenever the value of n is doubled. This partially illustrates the n -consistency of \hat{r}_n . To see the overall feature of \hat{r}_n , Fig. 2 displays the histograms of $n(\hat{r}_{1n} - r_{10})$ and $n(\hat{r}_{2n} - r_{20})$, respectively, when $n = 600$. The important variability of the estimated thresholds can be explained by the fact that their limiting distributions depend on the jump sizes $Y_k^{(i,i+1)}$ and $Z_k^{(i+1,i)}$. According to our experiments, the smaller the jump sizes, the wider the range of the limiting distribution of the estimated thresholds.

For Model (6.1), by Theorem 2.1, we know that $\phi_2 = 1$ does not entail a unit-root behavior. This fact can also be seen from Fig. 3(a). From Ling and Li (2008), we know that ω_2 is not identifiable for the double AR model $y_t = \phi_2 y_{t-1} + \eta_t \sqrt{\omega_2 + \alpha_2 y_{t-1}^2}$ with true value $(\phi_2, \omega_2, \alpha_2) = (1, 0.5, 3)$. However, if such a double AR model is regarded as the middle regime of Model (6.1), then ω_2 remains identifiable and its estimator is asymptotically normal. This is illustrated in Fig. 3(b). We also did some simulation studies when $n = 250$ but the results were not much different from those obtained when $n = 300$.

To examine the performance of our score-based tests in finite samples, we take $\beta = (1, \dots, 1)'$ and use 1000 replications. Under the null, H_0 , $\{y_t\}$ follows a MTDAR(2;1) model, i.e.,

$$y_t = \begin{cases} 0.1 y_{t-1} + \eta_t \sqrt{0.5 + 0.6 y_{t-1}^2}, & \text{if } y_{t-1} \leq 0, \\ 0.4 y_{t-1} + \eta_t \sqrt{0.3 + 0.2 y_{t-1}^2}, & \text{if } y_{t-1} > 0, \end{cases}$$

where $\eta_t \stackrel{i.i.d.}{\sim} N(0, 1)$. At the significance level 0.05, the empirical sizes of the tests based on the statistics S_{1n} , S_{2n} and S_n are 0.071, 0.030 and 0.073 when $n = 100$, and 0.070, 0.032 and 0.053 when $n = 200$, respectively. This shows that the size of S_n gets closer to

Table 2
Simulation results for Model (6.1).

n		ϕ_1	ω_1	α_1	ϕ_2	ω_2	α_2	ϕ_3	ω_3	α_3	r_1	r_2
		0.5	1	0.3	1	0.5	3	-0.7	1	0.5	-1	0
300	EM	0.4915	0.9476	0.2977	1.0125	0.4534	3.1094	-0.7081	1.0302	0.4540	-0.9884	0.0158
	ESD	0.0710	0.4307	0.1128	0.2904	0.1908	0.8413	0.1462	0.2553	0.2772	0.0631	0.0791
	ASD	0.0686	0.2799	0.0744	0.2596	0.1336	0.7785	0.1471	0.1851	0.2130	0.0900	0.0923
600	EM	0.4942	0.9778	0.2980	1.0144	0.4675	3.0777	-0.7010	1.0320	0.4631	-0.9933	0.0092
	ESD	0.0506	0.2801	0.0737	0.1858	0.1172	0.6135	0.1070	0.1596	0.1832	0.0301	0.0364
	ASD	0.0485	0.1979	0.0526	0.1836	0.0945	0.5505	0.1040	0.1309	0.1506	0.0450	0.0461
900	EM	0.4978	0.9778	0.3023	1.0028	0.4798	3.0872	-0.6977	1.0162	0.4802	-0.9949	0.0044
	ESD	0.0420	0.2444	0.0645	0.1540	0.0884	0.5094	0.0858	0.1197	0.1456	0.0215	0.0243
	ASD	0.0396	0.1616	0.0430	0.1499	0.0771	0.4494	0.0849	0.1069	0.1230	0.0300	0.0308

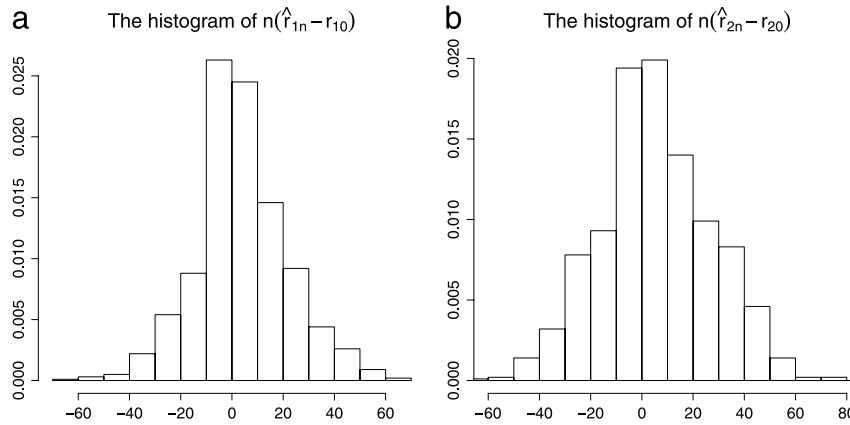


Fig. 2. The histograms of $n(\hat{r}_{1n} - r_{10})$ and $n(\hat{r}_{2n} - r_{20})$ when $n = 600$.

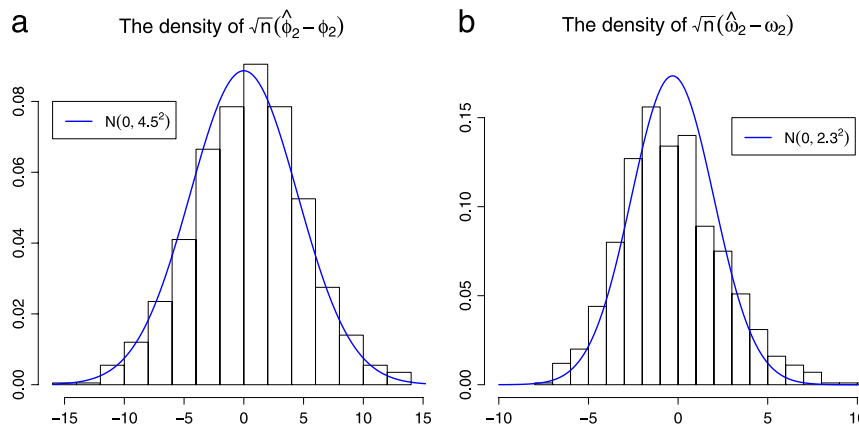


Fig. 3. Empirical and asymptotic densities of $\sqrt{n}(\hat{\phi}_2 - \phi_2)$ and $\sqrt{n}(\hat{\omega}_2 - \omega_2)$ when $n = 600$. The asymptotic variances are computed by Theorem 3.2.

its nominal value than those of the two other tests as the sample size increases. The alternative H_1 is a MTDAR(3;1) model, i.e.,

$$y_t = \begin{cases} 0.1y_{t-1} + \eta_t\sqrt{0.5 + 0.6y_{t-1}^2}, & \text{if } y_{t-1} \leq 0, \\ (0.4 + \lambda)y_{t-1} + \eta_t\sqrt{0.3 + (0.2 + |\lambda|)y_{t-1}^2}, & \text{if } 0 < y_{t-1} \leq 2, \\ 0.4y_{t-1} + \eta_t\sqrt{0.3 + 0.2y_{t-1}^2}, & \text{if } y_{t-1} > 2. \end{cases}$$

Fig. 4 illustrates the power of the tests S_{1n} and S_{2n} defined in (5.3) when λ varies from -4 to 2 . Even if the range of λ is large, the alternative model is still stationary and ergodic by Theorem 2.1. In this alternative model, a third threshold is introduced in the second regime. It is expected that S_{1n} has a less power. However, when $\lambda > 0.5$, its power is significantly increasing. This is because the process $\{y_t\}$ returns to the first regime $(-\infty, 0]$ much less frequently than to the second regime $(0, \infty)$ in this case. Also for this reason, S_{2n} has no power when $\lambda > 0.5$. Fig. 5 gives the plot

of the power of S_n . It can be seen that the power of S_n increases when $|\lambda|$ increases or when the sample size n increases from 100 to 200. However, as a referee pointed out, the power exceeds 20% only when $|\lambda| > 0.5$. This may be because the second regime in the alternative model is irrelevant to its stationarity and ergodicity and hence its effect on the behavior of the whole model is small. According to our simulation study, the test based on S_n should be useful for determining the number of thresholds in the MTDAR models.

7. An empirical example

In modern macroeconomics, the US Gross National Product (GNP) is perhaps the most examined univariate time series, see Potter (1995) and the references therein. Many researchers pointed out that the US GNP sequence contains nonlinearity and asymmetric effects causing the contraction (or recession)

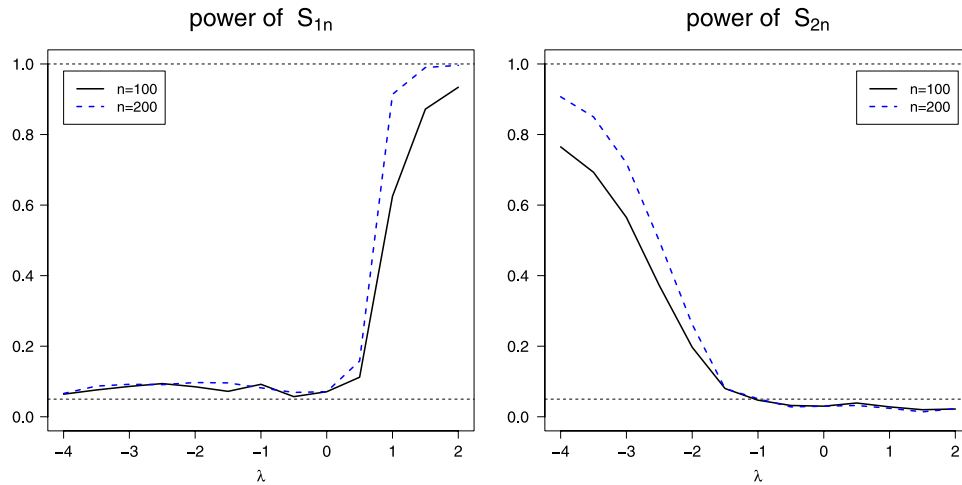


Fig. 4. Powers of the test statistics S_{1n} and S_{2n} at the significance level 5%, based on 1000 simulations of the MTDAR(3;1) model.

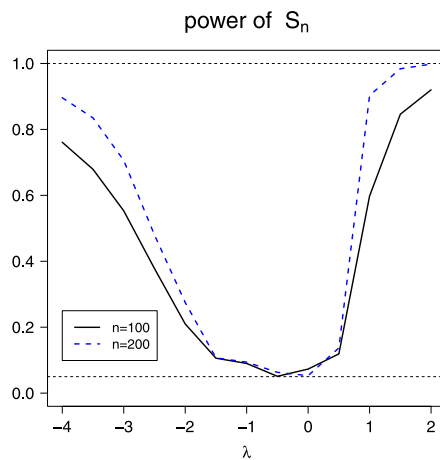


Fig. 5. Power of the test statistic S_{η} at the significance level 5%, based on 1000 simulations of the MTDAR(3;1) model.

and expansion regimes. To characterize such nonlinearity and asymmetric effects, [Tiao and Tsay \(1994\)](#) suggested that two-regime threshold models may be appropriate for contraction and expansion, see also [Potter \(1995\)](#). It is also reasonable to model “bad times”, “good times” and “normal times” of a given time series, see [Koop and Potter \(1999\)](#). Based on this idea, [Li and Ling \(2012\)](#) used a three-regime TAR model to fit the growth rate of the quarterly US real GNP data over the period 1947–2009.

In this section, we will use the score test of Section 5 to reanalyze the quarterly US real GNP data over the period 1947–06/2013 with a total of 266 observations. Let x_1, \dots, x_{266} denote the original data. We define the growth rate series as $y_t = 100(\log x_t - \log x_{t-1})$, $t = 2, \dots, 266$. The data $\{x_t\}$ and the growth rate series $\{y_t\}$ are plotted in [Fig. 6](#).

We first used a two-regime TDAR model to fit the data $\{y_t\}$. Based on the AIC and BIC, we selected the following model:

$$y_t = \begin{cases} 0.177 + 0.424y_{t-1} - 0.306y_{t-2} \\ (0.202) \quad (0.128) \quad (0.210) \\ + \eta_t \sqrt{0.690 + 0.631y_{t-3}^2}, & \text{if } y_{t-2} \leq 0.244, \\ (0.452) \quad (0.538) \\ 0.421 + 0.331y_{t-1} + 0.213y_{t-2} - 0.105y_{t-3} \\ (0.116) \quad (0.069) \quad (0.087) \quad (0.065) \\ + \eta_t \sqrt{0.601 + 0.022y_{t-1}^2}, & \text{if } y_{t-2} > 0.244, \\ (0.144) \quad (0.058) \quad (0.169) \end{cases} \quad (7.1)$$

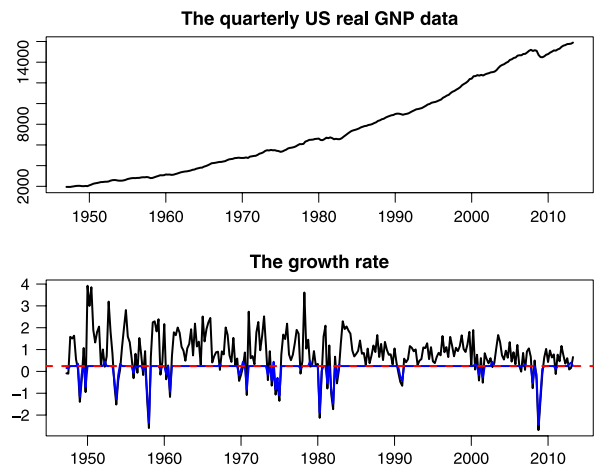


Fig. 6. The original data and the growth rate. The dash line below is the frontier between the contraction regime and the expansion regime.

where the estimated value of d is 2, $\hat{\tau}_n = 0.244$, the AIC and BIC’s values are 207.246 and 257.362, respectively. The 95% confidence interval of the threshold is $[-0.173, 0.534]$ by the simulation method in [Li et al. \(forthcoming\)](#), where 2000 replications were used. The p -values of the Ljung–Box test statistic $Q(M)$ and the McLeod–Li test statistic $Q^2(M)$ (see [Li and Li, 1996](#); [McLeod and Li, 1983](#)) suggest that Model (7.1) is adequate for $\{y_t\}$, see [Fig. 7](#). Moreover, the value of the score test S_{η} in (5.4) is 0.767, which indicates that the two-regime TDAR model (7.1) is sufficient for $\{y_t\}$. The first regime (i.e., $y \leq 0.244$) and the second one in Model (7.1) characterize the dynamic behaviors of the contraction and the expansion, respectively. The number of observations in the contraction and expansion regimes are 56 and 209, respectively. This means that the US total economic activity was most of time in expansion since the World War II. By computing the roots of AR polynomials in the two regimes, we find that the expected durations of the contraction and the expansion periods are approximately 5.32 quarters and 5.95 quarters, respectively. The average length of the stochastic cycles is totally 11.27 quarters, which is about 3 years. This result is similar to that in [Tsay \(2010, p. 42\)](#). Finally, note that the first regime is characterized by a more important volatility than the second regime, with also a larger delay of response (Y_{t-3}^2 instead of Y_{t-1}^2).

Acknowledgments

The authors would like to thank the Editor and two anonymous referees for their helpful comments that improved the

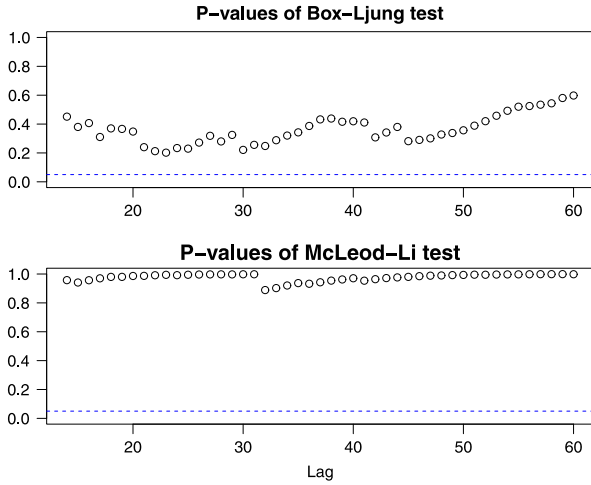


Fig. 7. The p-values of the Ljung–Box test statistic and the McLeod–Li test statistic.

presentation. The research was partially supported by the Start-up Fund of Tsinghua University (No. 553310013) and the National Natural Science Foundation of China (NSFC) under Grant (No. 11401337) (to Li), and the Hong Kong Research Grants Commission (Grant HKUST641912 and 603413) (to Ling) and the Labex ECODEC and the IDR ACP “Regulation and systemic risks” (to Zakoian).

Appendix

Let $K > 0$ and $0 < \rho < 1$ denote generic constants, whose values can change throughout the proofs.

A.1. Proof of Theorem 2.1

Proof. We will verify the following criterion, which is straightforwardly deduced from Meyn and Tweedie (1996, Theorem 19.1.3): if $\{y_t\}$ is a homogeneous Markov chain on $E \subset \mathbb{R}$ which is Feller, aperiodic, μ -irreducible, where μ is a σ -finite measure whose support has a non-empty interior, and if there exist a compact set C , an integer $s \geq 1$ and a function $V : \mathbb{R} \rightarrow \mathbb{R}^+$ such that

$$V(y) \geq 1, \quad \forall y \in C \quad (\text{A.1})$$

and for some $\delta > 0$

$$E[V(y_t)|Y_{t-s} = y] \leq (1 - \delta)V(y), \quad \forall y \notin C \quad (\text{A.2})$$

then $\{y_t\}$ is geometrically ergodic and $E[V(y_t)] < \infty$.

It is clear that $\{y_t\}$ defined by (2.1), with initial value y_0 , is an homogeneous Markov chain on \mathbb{R} endowed with its Borel σ -field $\mathcal{B}(\mathbb{R})$. Denote by λ the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The transition probabilities of $\{y_t\}$ are given, for $y \in \mathbb{R}$, $B \in \mathcal{B}(\mathbb{R})$, by

$$P(y, B) = P(y_t \in B | y_{t-1} = y) = \sum_{i=1}^m \left(\int_{B_i(y)} f(x) dx \right) I\{y \in \mathcal{R}_i\},$$

where $B_i(y)$ is the set $\frac{B - \phi_i y - c_i}{\sqrt{\omega_i + \alpha_i y^2}}$. Since $P(\cdot, B)$ is continuous, for any $B \in \mathcal{B}(\mathbb{R})$, the chain $\{y_t\}$ has the Feller property. Now, because the density f is positive over \mathbb{R} , we have $P(y, B) > 0$ whenever $\lambda(B) > 0$. Thus the chain is λ -irreducible. It can also be shown that $P^k(y, B) = P(y_t \in B | y_{t-k} = y) > 0$ for any integer $k \geq 1$, whenever $\lambda(B) > 0$, which establishes the aperiodicity of the chain.

Let $\sigma_{it} = (\omega_i + \alpha_i y_{t-1}^2)^{1/2}$. We have

$$y_t = (\phi_1 y_{t-1} + \sigma_{1t} \eta_t) I\{y_{t-1} < r_1\}$$

$$+ (\phi_m y_{t-1} + \sigma_{mt} \eta_t) I\{y_{t-1} > r_{m-1}\} + z(y_{t-1}, \eta_t),$$

where $|z(y_{t-1}, \eta_t)| \leq K(1 + |\eta_t|)$. Write

$$\begin{aligned} \sigma_{1t} I\{y_{t-1} < r_1\} &= -\sqrt{\alpha_1} y_{t-1} I\{y_{t-1} < r_1\} + u_1(y_{t-1}), \\ \sigma_{mt} I\{y_{t-1} > r_{m-1}\} &= \sqrt{\alpha_m} y_{t-1} I\{y_{t-1} > r_{m-1}\} + u_m(y_{t-1}), \end{aligned}$$

where $u_1(y_{t-1})$ and $u_m(y_{t-1})$ are bounded random terms. It follows that

$$\begin{aligned} y_t &= [a_1(\eta_t) I\{y_{t-1} < r_1\} + a_m(\eta_t) I\{y_{t-1} > r_{m-1}\}] y_{t-1} \\ &\quad + \{u_1(y_{t-1}) + u_m(y_{t-1})\} \eta_t + z(y_{t-1}, \eta_t), \end{aligned} \quad (\text{A.3})$$

where $a_1(x) = \phi_1 - x\sqrt{\alpha_1}$ and $a_m(x) = \phi_m + x\sqrt{\alpha_m}$. Let

$$b(\eta, y) = a_1(\eta) I\{y < r_1\} + a_m(\eta) I\{y > r_{m-1}\}.$$

Expanding (A.3) we find that if $y_{t-2} = y$,

$$y_t = b(\eta_t, y_{t-1}) b(\eta_{t-1}, y) y + v(y_{t-1}, \eta_t, y), \quad (\text{A.4})$$

where

$$\begin{aligned} v(y_{t-1}, \eta_t, y) &= b(\eta_t, y_{t-1}) [\{u_1(y) + u_m(y)\} \eta_{t-1} + z(y, \eta_{t-1})] \\ &\quad + \{u_1(y_{t-1}) + u_m(y_{t-1})\} \eta_t + z(y_{t-1}, \eta_t). \end{aligned}$$

Suppose that $y > r_{m-1}$. Then

$$\begin{aligned} I\{y_{t-1} < r_1\} &= I\left\{\eta_{t-1} < \frac{r_1 - \phi_m y}{\sqrt{\omega_m + \alpha_m y^2}}\right\} \\ &= I\left\{\eta_{t-1} < -\frac{\phi_m}{\sqrt{\alpha_m}}\right\} + \varepsilon_1 I\{\eta_{t-1} \in A_1(y)\}, \end{aligned} \quad (\text{A.5})$$

and

$$\begin{aligned} I\{y_{t-1} > r_{m-1}\} &= I\left\{\eta_{t-1} > \frac{r_{m-1} - \phi_m y}{\sqrt{\omega_m + \alpha_m y^2}}\right\} \\ &= I\left\{\eta_{t-1} > -\frac{\phi_m}{\sqrt{\alpha_m}}\right\} + \varepsilon_2 I\{\eta_{t-1} \in A_2(y)\}, \end{aligned} \quad (\text{A.6})$$

with, by convention, $I\{\eta < a/b\} = 1 - I\{\eta > a/b\} = I\{a > 0\}$ if $b = 0$, for some sets $A_i(y)$, with $\varepsilon_i = 0, \pm 1$, $i = 1, 2$. Similarly, if $y < r_1$,

$$I\{y_{t-1} < r_1\} = I\left\{\eta_{t-1} < \frac{\phi_1}{\sqrt{\alpha_1}}\right\} + \varepsilon_3 I\{\eta_{t-1} \in A_3(y)\}, \quad (\text{A.7})$$

and

$$I\{y_{t-1} > r_{m-1}\} = I\left\{\eta_{t-1} > \frac{\phi_1}{\sqrt{\alpha_1}}\right\} + \varepsilon_4 I\{\eta_{t-1} \in A_4(y)\}. \quad (\text{A.8})$$

Let

$$\begin{aligned} H_1(\eta_0, \eta_1) &= \left\{ a_1(\eta_0) I\left\{\eta_1 < \frac{\phi_1}{\sqrt{\alpha_1}}\right\} \right. \\ &\quad \left. + a_m(\eta_0) I\left\{\eta_1 > \frac{\phi_1}{\sqrt{\alpha_1}}\right\} \right\} a_1(\eta_1), \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} H_m(\eta_0, \eta_1) &= \left\{ a_1(\eta_0) I\left\{\eta_1 < -\frac{\phi_m}{\sqrt{\alpha_m}}\right\} \right. \\ &\quad \left. + a_m(\eta_0) I\left\{\eta_1 > -\frac{\phi_m}{\sqrt{\alpha_m}}\right\} \right\} a_m(\eta_1). \end{aligned} \quad (\text{A.10})$$

We thus have, in view of (A.4)–(A.8), for $y > r_{m-1}$

$$y_t = \left\{ a_1(\eta_t) I\left\{\eta_{t-1} < -\frac{\phi_m}{\sqrt{\alpha_m}}\right\} + a_m(\eta_t) I\left\{\eta_{t-1} > -\frac{\phi_m}{\sqrt{\alpha_m}}\right\} \right\}$$

$$\begin{aligned} & \times a_m(\eta_{t-1})y + v(y_{t-1}, \eta_t, y) \\ & + [a_1(\eta_t)\varepsilon_1 I\{\eta_{t-1} \in A_1(y)\} + a_m(\eta_t)\varepsilon_2 I\{\eta_{t-1} \in A_2(y)\}] \\ & \times a_m(\eta_{t-1})y \end{aligned}$$

$$:= H_m(\eta_t, \eta_{t-1})y + R_m(\eta_t, \eta_{t-1}, y), \quad (\text{A.11})$$

and for $y < r_1$,

$$y_t = H_1(\eta_t, \eta_{t-1})y + R_1(\eta_t, \eta_{t-1}, y) \quad (\text{A.12})$$

where $R_1(\eta_t, \eta_{t-1}, y)$ is defined similarly to $R_m(\eta_t, \eta_{t-1}, y)$. Note that (2.2) can be equivalently written as

$$\gamma = \max \{E \log |H_1(\eta_0, \eta_1)|, E \log |H_m(\eta_0, \eta_1)|\} < 0. \quad (\text{A.13})$$

Because $E \log |H_1(\eta_0, \eta_1)| < 0$, and $E|H_1(\eta_0, \eta_1)|^s < \infty$, there exists $u \in (0, 1)$ such that

$$\rho \equiv \max \{E|H_1(\eta_0, \eta_1)|^u, E|H_m(\eta_0, \eta_1)|^u\} < 1$$

(see for instance Francq and Zakoian (2010, Lemma 2.2)). Let $V : y \rightarrow V(y) = 1 + |y|^u$. Using the elementary equality $(a + b)^u \leq a^u + b^u$ for $a, b \in \mathbb{R}$, we have, for $y > r_{m-1}$, by (A.11)

$$\begin{aligned} E[|y_t|^u | y_{t-2} = y] - |y|^u \\ \leq \{\rho - 1\} |y|^u + E\{R_m(\eta_t, \eta_{t-1}, y)\}^u. \end{aligned} \quad (\text{A.14})$$

Now

$$\begin{aligned} E\{R_m(\eta_t, \eta_{t-1}, y)\}^u & \leq E\{v(y_{t-1}, \eta_t, y)\}^u \\ & + E\{[a_1(\eta_t)^u I\{\eta_{t-1} \in A_1(y)\} + a_m(\eta_t)^u I\{\eta_{t-1} \in A_2(y)\}]\} \\ & \times a_m(\eta_{t-1})^u |y|^u. \end{aligned}$$

It can be seen that $|a_1(\eta_t)| \leq K(1 + |\eta_t|)$, $|a_m(\eta_t)| \leq K(1 + |\eta_t|)$ and $|v(y_{t-1}, \eta_t, y)| \leq K(1 + |\eta_t|)(1 + |\eta_{t-1}|)$. Moreover, for $i = 1, \dots, 4$, $\lambda(A_i(y)) \leq K/y$. It follows, by the Schwarz inequality, that for any integer p such that $u/s < 1/p < 1 - u$, with u sufficiently small,

$$\begin{aligned} EI\{\eta_{t-1} \in A_1(y)\} \{a_m(\eta_{t-1})\}^u |y|^u \\ \leq K[\lambda(A_1(y))]^{1-1/p} [E\{a_m(\eta_{t-1})\}^{up}]^{1/p} |y|^u \leq K. \end{aligned}$$

The same inequalities hold with a_1 replaced by a_m . Thus, by (A.14), we have, for $y > r_{m-1}$,

$$E[|y_t|^u | y_{t-2} = y] - |y|^u \leq \{\rho - 1\} |y|^u + K.$$

The same arguments show that the same inequality holds for $y < r_1$.

Let $0 < \delta < 1 - \rho$ and let

$$C = \{y \in \mathbb{R}; (\rho - 1 + \delta)|y|^u + K \geq 0\} \cup [r_1, r_{m-1}].$$

The set C is a non-empty compact set such that $\lambda(C) > 0$. Moreover (A.1) and (A.2) hold with $s = 2$. Thus, there exists a unique geometrically ergodic solution with $E|y_t|^u < \infty$ to Model (2.1). This completes the proof. \square

A.2. Proof of Theorem 2.2

Condition (2.4) means that $\rho_1 = \max\{EH_0^2(\eta_t, \eta_{t-1}), EH_m^2(\eta_t, \eta_{t-1})\} < 1$, with the notation introduced in (A.9)–(A.10). Starting from (A.11) we have, for $y > r_{m-1}$,

$$\begin{aligned} E[y_t^2 | y_{t-2} = y] - y^2 \\ = \{EH_m^2(\eta_t, \eta_{t-1}) - 1\} y^2 + 2yEH_m(\eta_t, \eta_{t-1})R_m(\eta_t, \eta_{t-1}, y) \\ + E\{R_m(\eta_t, \eta_{t-1}, y)\}^2. \end{aligned}$$

Since $a_m(\eta_{t-1})^2$ is bounded over the set $A_1(y)$, we have, for $y > r_{m-1}$,

$$EI\{\eta_{t-1} \in A_1(y)\} a_m(\eta_{t-1})^2 y^2 \leq EI\{\eta_{t-1} \in A_1(y)\} y^2 \leq Ky.$$

Treating in the same way the other terms involved in $R_m(\eta_t, \eta_{t-1}, y)$ it follows that $E\{R_m(\eta_t, \eta_{t-1}, y)\}^2 \leq Ky$. Similarly we have $|EH_m(\eta_t, \eta_{t-1})R_m(\eta_t, \eta_{t-1}, y)| \leq Ky$ and thus, for $y > r_{m-1}$

$$E[y_t^2 | y_{t-2} = y] - y^2 \leq \{\rho_1 - 1\} y^2 + K|y|.$$

The same inequality holds for $y < r_1$. Letting $0 < \delta < 1 - \rho_1$ and introducing the compact set

$$C = \{y \in \mathbb{R}; (\rho_1 - 1 + \delta)y^2 + K|y| \geq 0\} \cup [r_1, r_{m-1}],$$

we conclude that (A.1) and (A.2) hold with $s = 2$, and $V(y) = 1 + y^2$. The proof is complete. \square

A.3. Proof of Theorem 3.1

Before proving Theorem 3.1, we first prove several intermediate results.

Lemma A.1. Under the assumptions of Theorem 3.1, $El_t(\vartheta)$ has a unique maximum at ϑ_0 .

Proof. Let $\beta(\vartheta) = -2E\{l_t(\vartheta) - l_t(\vartheta_0)\}$, $A_i = I\{r_{i-1} < y_{t-d} \leq r_i\}$, $A_{i0} = I\{r_{i-1,0} < y_{t-d} \leq r_{i0}\}$ and

$$\Gamma_{i,j0} = \log\left(\frac{\alpha'_i \mathbf{X}_{t-1}}{\alpha'_{j0} \mathbf{X}_{t-1}}\right) + \frac{\alpha'_{j0} \mathbf{X}_{t-1}}{\alpha'_i \mathbf{X}_{t-1}} - 1 + \frac{(\phi_{j0} - \phi_i)' \mathbf{Y}_{t-1}}{\alpha'_i \mathbf{X}_{t-1}}.$$

Clearly, $\Gamma_{i,j0} \geq 0$ by the elementary inequality $\log x + \frac{1}{x} - 1 \geq 0$ for all $x > 0$ and the equality holds if and only if $x = 1$. Then, $\beta(\vartheta) = \sum_{i=1}^m \sum_{j=1}^m E\{\Gamma_{i,j0} A_i A_{j0}\} \geq 0$.

Next, we show that $\beta(\vartheta) = 0$ holds if and only if $\vartheta = \vartheta_0$. Assume $\beta(\vartheta) = 0$. We first prove $r_1 = r_{10}$. If $r_1 > r_{10}$, then $\Gamma_{1,10} A_1 A_{10} = \Gamma_{1,10} A_{10} = 0$ and $\Gamma_{1,20} A_1 A_{20} = \Gamma_{1,20} I\{r_{10} < y_{t-d} \leq r_1 \wedge r_{20}\} = 0$. Since the density of y_t is positive and continuous by A2, it follows that $\Gamma_{1,10} = \Gamma_{1,20} = 0$, which implies that $(\phi'_{10}, \alpha'_{10})' = (\phi'_1, \alpha'_1)' = (\phi'_{20}, \alpha'_{20})'$. This contradicts A1. Hence, $r_1 \leq r_{10}$. If $r_1 < r_{10}$, using the same technique, we can get $(\phi'_{10}, \alpha'_{10})' = (\phi'_2, \alpha'_2)' = (\phi'_{20}, \alpha'_{20})'$ if $r_2 \geq r_{10}$, which results in $r_1 < r_2 < r_{10}$. Repeating the above procedure, we can get $r_1 < \dots < r_m < r_{10}$, which is a contradiction with the partition $\mathbb{R} = \cup_{i=1}^m A_i$. Finally, we get $r_1 = r_{10}$ and then in turn $(\phi'_1, \alpha'_1)' = (\phi'_{10}, \alpha'_{10})'$ implied by $\Gamma_{1,10} A_{10} = 0$. Similarly, we have $r_i = r_{i0}$ for $i = 2, \dots, m-1$ and then in turn $(\phi'_j, \alpha'_j)' = (\phi'_{j0}, \alpha'_{j0})'$ for $j = 2, \dots, m$. Thus, $\vartheta = \vartheta_0$ and $El_t(\vartheta)$ is uniquely maximized at ϑ_0 . \square

Lemma A.2. Under the assumptions of Theorem 3.1, for any $\eta > 0$,

$$\lim_{l \rightarrow \infty} P\left(\max_{l \leq n < \infty} \sup_{\|\vartheta - \vartheta_0\| > \eta} \sum_{t=1}^n [l_t(\vartheta) - l_t(\vartheta_0)] \geq 0\right) = 0.$$

Proof. Let $V_{\tilde{\eta}} = \{\tilde{\vartheta} : \|\tilde{\vartheta} - \vartheta\| \leq \tilde{\eta}\}$ and $X_t(\tilde{\eta}) = \sup_{\vartheta \in \Theta} \sup_{V_{\tilde{\eta}}} |l_t(\tilde{\vartheta}) - l_t(\vartheta)|$. Since η_t has a density function, we can show that

$$EX_t(\tilde{\eta}) \rightarrow 0 \quad (\text{A.15})$$

as $\tilde{\eta} \rightarrow 0$. Thus, for any $\epsilon > 0$, there is $\tilde{\eta} > 0$ such that $EX_t(\tilde{\eta}) < \epsilon/2$. Since $X_t(\tilde{\eta})$ is asymptotically strictly stationary and ergodic, by Lemma 1 in Chow and Teicher (1978, p. 66) and the ergodic theorem, for any $\epsilon_1 > 0$, we have

$$P\left(\max_{l \leq n < \infty} \frac{1}{n} \left| \sum_{t=1}^n [X_t(\tilde{\eta}) - EX_t(\tilde{\eta})] \right| \geq \frac{\epsilon}{2}\right) < \epsilon_1,$$

as l is large enough. Thus, for any $\epsilon, \epsilon_1 > 0$, there exists a constant $\tilde{\eta} > 0$ such that

$$\begin{aligned} P\left(\max_{l \leq n < \infty} \frac{1}{n} \sum_{t=1}^n X_t(\tilde{\eta}) \geq \epsilon\right) \\ \leq P\left(\max_{l \leq n < \infty} \frac{1}{n} \left| \sum_{t=1}^n [X_t(\tilde{\eta}) - EX_t(\tilde{\eta})] \right| \geq \frac{\epsilon}{2}\right) < \epsilon_1. \end{aligned} \quad (\text{A.16})$$

By the ergodic theorem, for each $\vartheta \in \Theta$ and any $\epsilon > 0$,

$$\lim_{l \rightarrow \infty} P\left(\max_{l \leq n < \infty} \left| \frac{1}{n} \sum_{t=1}^n [l_t(\vartheta) - El_t(\vartheta)] \right| \geq \epsilon\right) = 0. \quad (\text{A.17})$$

Since Θ is compact, we can choose a collection of balls of radius $\Delta > 0$ covering Θ , and the number of such balls is a finite integer N . In the i^{th} ball, we take a point ξ_i and denote this ball by $V(\xi_i)$. By (A.15)–(A.17), for any $\epsilon > 0$, we have

$$\begin{aligned} P\left(\max_{l \leq n < \infty} \frac{1}{n} \sup_{\vartheta} \left| \sum_{t=1}^n [l_t(\vartheta) - El_t(\vartheta)] \right| \geq \epsilon\right) \\ \leq P\left(\max_{1 \leq j \leq N} \sup_{\vartheta \in V(\xi_j)} \max_{l \leq n < \infty} \left| \frac{1}{n} \sum_{t=1}^n [l_t(\vartheta) - l_t(\xi_j)] \right| \geq \frac{\epsilon}{3}\right) \\ + P\left(\max_{1 \leq j \leq N} \sup_{\vartheta \in V(\xi_j)} \left| El_t(\vartheta) - l_t(\xi_j) \right| \geq \frac{\epsilon}{3}\right) \\ + P\left(\max_{1 \leq j \leq N} \max_{l \leq n < \infty} \left| \frac{1}{n} \sum_{t=1}^n [l_t(\xi_j) - El_t(\xi_j)] \right| \geq \frac{\epsilon}{3}\right) \\ \leq P\left(\sup_{\xi_i \in \Theta} \sup_{\vartheta \in V(\xi_i)} \max_{l \leq n < \infty} \left| \frac{1}{n} \sum_{t=1}^n [l_t(\vartheta) - l_t(\xi_j)] \right| \geq \frac{\epsilon}{3}\right) \\ + \sum_{j=1}^N P\left(\sup_{\vartheta \in V(\xi_j)} \left| El_t(\vartheta) - l_t(\xi_j) \right| \geq \frac{\epsilon}{3}\right) \\ + \sum_{j=1}^N P\left(\max_{l \leq n < \infty} \left| \frac{1}{n} \sum_{t=1}^n [l_t(\xi_j) - El_t(\xi_j)] \right| \geq \frac{\epsilon}{3}\right) < \epsilon, \end{aligned} \quad (\text{A.18})$$

as l is large enough and Δ is small enough.

Since $El_t(\vartheta)$ has a unique maximum at ϑ_0 , Θ is compact, and $El_t(\vartheta)$ is continuous, there exists a constant $c > 0$, such that

$$\max_{\|\vartheta - \vartheta_0\| > \eta} E[l_t(\vartheta) - l_t(\vartheta_0)] \leq -c \quad (\text{A.19})$$

for any $\eta > 0$. By (A.18)–(A.19), it follows that

$$\begin{aligned} P\left(\max_{l \leq n < \infty} \sup_{\|\vartheta - \vartheta_0\| > \eta} \left\{ \sum_{t=1}^n [l_t(\vartheta) - l_t(\vartheta_0)] + \frac{cn}{2} \right\} > 0\right) \\ = P\left(\max_{l \leq n < \infty} \sup_{\|\vartheta - \vartheta_0\| > \eta} \left\{ \sum_{t=1}^n [l_t(\vartheta) - El_t(\vartheta)] \right. \right. \\ \left. \left. - \sum_{t=1}^n [l_t(\vartheta_0) - El_t(\vartheta_0)] + n[El_t(\vartheta) - El_t(\vartheta_0)] + \frac{cn}{2} \right\} > 0\right) \\ \leq P\left(\max_{l \leq n < \infty} \sup_{\vartheta} \left\{ 2 \left| \sum_{t=1}^n [l_t(\vartheta) - El_t(\vartheta)] \right| - cn + \frac{cn}{2} \right\} > 0\right) \\ \leq P\left(\max_{l \leq n < \infty} \sup_{\vartheta} \left| \frac{1}{n} \sum_{t=1}^n [l_t(\vartheta) - El_t(\vartheta)] \right| > \frac{c}{4}\right) \rightarrow 0, \end{aligned} \quad (\text{A.20})$$

as $l \rightarrow \infty$. \square

Proof of Theorem 3.1. By Lemma A.2, for any $\epsilon > 0$, we have

$$\begin{aligned} \lim_{l \rightarrow \infty} P\left(\max_{l \leq n < \infty} \|\hat{\vartheta}_n - \vartheta_0\| > \epsilon\right) \\ = \lim_{l \rightarrow \infty} P\left\{ \max_{l \leq n < \infty} \|\hat{\vartheta}_n - \vartheta_0\| > \epsilon, \right. \\ \left. \max_{l \leq n < \infty} \sum_{t=1}^n [l_t(\hat{\vartheta}_n) - l_t(\vartheta_0)] \geq 0 \right\} \\ \leq \lim_{l \rightarrow \infty} P\left\{ \max_{l \leq n < \infty} \sup_{\|\vartheta - \vartheta_0\| > \epsilon} \sum_{t=1}^n [l_t(\vartheta) - l_t(\vartheta_0)] \geq 0 \right\} = 0. \end{aligned}$$

Thus, the conclusion holds. \square

A.4. Proof of Theorem 3.2

Since $\hat{\vartheta}_n$ is consistent by Theorem 3.1, we restrict the parameter space to an open neighborhood of ϑ_0 . To this end, define $V_\delta = \{\vartheta \in \Theta : \|\lambda - \lambda_0\| < \delta, |r_i - r_{i0}| < \delta, i = 1, \dots, m-1\}$ for some $0 < \delta < 1$ to be determined later. Choose δ small enough so that $\{r : |r - r_{i-1,0}| < \delta\} \cap \{r : |r - r_{i0}| < \delta\} = \emptyset$ for $i = 2, \dots, m-1$. Note that

$$L_n(\lambda, r) - L_n(\lambda, r_0) = \sum_{i=1}^{m-1} L_n^{(i)}(\lambda, r_i),$$

where

$$\begin{aligned} L_n^{(i)}(\lambda, r_i) \\ = -\frac{1}{2} \sum_{t=1}^n \left[\log\left(\frac{\alpha'_i \mathbf{X}_{t-1}}{\alpha'_{i+1} \mathbf{X}_{t-1}}\right) + \frac{(y_t - \phi'_i \mathbf{Y}_{t-1})^2}{\alpha'_i \mathbf{X}_{t-1}} - \frac{(y_t - \phi'_{i+1} \mathbf{Y}_{t-1})^2}{\alpha'_{i+1} \mathbf{X}_{t-1}} \right] \\ \times \text{sign}(r_i - r_{i0}) I\{r_i \wedge r_{i0} < y_{t-d} \leq r_i \vee r_{i0}\}. \end{aligned}$$

For (i), it is equivalent to prove $n|\hat{r}_{in} - r_{i0}| = O_p(1)$ for each $i = 1, \dots, m$. To obtain $n|\hat{r}_{in} - r_{i0}| = O_p(1)$, it suffices to prove that there exist constants $B > 0$ and $\gamma > 0$ such that, for any $\epsilon > 0$,

$$P\left(\sup_{\substack{B/n < |r_i - r_{i0}| \leq \delta \\ \vartheta \in V_\delta}} \frac{L_n^{(i)}(\lambda, r_i) - L_n^{(i)}(\lambda, r_{i0})}{nG_i(|r_i - r_{i0}|)} < -\gamma\right) > 1 - \epsilon \quad (\text{A.21})$$

for $i = 1, \dots, m$, as n is large enough, where $G_i(u) = P(r_{i0} < y_0 \leq r_{i0} + u)$.

We now show (A.21) holds for the case $p = 1$ and $i = 1$. The proof of the general case would go through by the same technique used in Chan (1993, p. 529). Here, we only treat the case $r_1 > r_{10}$. The proof for the case $r_1 \leq r_{10}$ is similar. Writing $r_1 = r_{10} + u$ for some $u \geq 0$. By a simple calculation, it follows that

$$\begin{aligned} \frac{2\{L_n^{(1)}(\lambda, r_1) - L_n^{(1)}(\lambda, r_{10})\}}{nG_1(u)} \\ = -K_1 \frac{G_{1n}(u)}{G_1(u)} + K_2 \frac{\sum_{t=1}^n \eta_t I(r_{10} < y_{t-1} \leq r_{10} + u)}{nG_1(u)} \\ + K_3 \frac{\sum_{t=1}^n (\eta_t^2 - 1) I(r_{10} < y_{t-1} \leq r_{10} + u)}{nG_1(u)} + O_p(\sqrt{\delta}), \end{aligned}$$

where $G_{1n}(u) = n^{-1} \sum_{t=1}^n I(r_{10} < y_{t-1} \leq r_{10} + u)$,

$$K_1 = \log \frac{\alpha'_{10} \mathbf{X}}{\alpha'_{20} \mathbf{X}} + \frac{\alpha'_{20} \mathbf{X}}{\alpha'_{10} \mathbf{X}} - 1 + \frac{(\phi_{20} - \phi_{10})r_{10}}{\alpha'_{10} \mathbf{X}},$$

$$K_2 = \frac{2\{(\phi_{10} - \phi_{20})r_{10}\}\sqrt{\alpha'_{20} \mathbf{X}}}{\alpha'_{10} \mathbf{X}} \quad \text{and} \quad K_3 = \frac{(\alpha_{10} - \alpha_{20})' \mathbf{X}}{\alpha'_{10} \mathbf{X}}$$

with $\mathbf{X} = (1, r_{10}^2)'$. Similar to Claim 2 in Chan (1993), for any $\varepsilon > 0$ and $\epsilon > 0$, there exists a positive constant B such that as n is large enough

$$P\left(\sup_{B/n < u \leq \delta} \left| \frac{G_{1n}(u)}{G_1(u)} - 1 \right| < \epsilon\right) > 1 - \varepsilon,$$

$$P\left(\sup_{B/n < u \leq \delta} \left| \frac{\sum_{t=1}^n \eta_t I(r_{10} < y_{t-1} \leq r_{10} + u)}{nG_1(u)} \right| < \epsilon\right) > 1 - \varepsilon,$$

$$P\left(\sup_{B/n < u \leq \delta} \left| \frac{\sum_{t=1}^n (\eta_t^2 - 1) I(r_{10} < y_{t-1} \leq r_{10} + u)}{nG_1(u)} \right| < \epsilon\right) > 1 - \varepsilon.$$

Note that $K_1 > 0$ by A7. Choosing δ small enough and $\gamma = K_1/4$, (A.21) holds and so does (i). \square

(ii). Similar to the proof of Theorem 3.1, it is not hard to show that $\sup_{\|\mathbf{r}-\mathbf{r}_0\| \leq B/n} \|\hat{\lambda}_n(\mathbf{r}) - \lambda_0\| = o_p(1)$. After a simple calculation, we have

$$\sup_{\|\mathbf{r}-\mathbf{r}_0\| \leq B/n} \left\| \frac{1}{n} \frac{\partial L_n(\lambda_0, \mathbf{r})}{\partial \lambda} - \frac{1}{n} \frac{\partial L_n(\lambda_0, \mathbf{r}_0)}{\partial \lambda} \right\| = O_p(n^{-1}),$$

$$\sup_{\|\lambda - \lambda_0\| < \epsilon} \sup_{\mathbf{r} \in \mathcal{C}_r} \left\| \frac{1}{n} \frac{\partial^2 L_n(\lambda, \mathbf{r})}{\partial \lambda \partial \lambda'} - \frac{1}{n} \frac{\partial^2 L_n(\lambda_0, \mathbf{r})}{\partial \lambda \partial \lambda'} \right\| = O_p(\epsilon), \quad (\text{A.22})$$

$$\sup_{\|\mathbf{r}-\mathbf{r}_0\| \leq B/n} \left\| \frac{1}{n} \frac{\partial^2 L_n(\lambda_0, \mathbf{r})}{\partial \lambda \partial \lambda'} - \frac{1}{n} \frac{\partial^2 L_n(\lambda_0, \mathbf{r}_0)}{\partial \lambda \partial \lambda'} \right\| = O_p(n^{-1}).$$

By Taylor's expansion of $\partial L_n(\lambda, \mathbf{r})/\partial \lambda$, it follows that

$$0 = \frac{1}{n} \frac{\partial L_n(\hat{\lambda}_n(\mathbf{r}), \mathbf{r})}{\partial \lambda}$$

$$= \frac{1}{n} \frac{\partial L_n(\lambda_0, \mathbf{r})}{\partial \lambda} + \frac{1}{n} \frac{\partial^2 L_n(\bar{\lambda}, \mathbf{r})}{\partial \lambda \partial \lambda'} [\hat{\lambda}_n(\mathbf{r}) - \lambda_0], \quad (\text{A.23})$$

where $\bar{\lambda}$ lies in the ball $\mathcal{B}(\lambda_0, \|\hat{\lambda}_n(\mathbf{r}) - \lambda_0\|)$. By the ergodic theorem, we have

$$-\frac{1}{n} \frac{\partial^2 L_n(\vartheta_0)}{\partial \lambda \partial \lambda'} \longrightarrow \Sigma := \text{diag}(\Sigma_1, \dots, \Sigma_m), \quad \text{a.s.}$$

as $n \rightarrow \infty$. Thus, by (A.22) and (A.23),

$$\sup_{\|\mathbf{r}-\mathbf{r}_0\| \leq B/n} \left\| \sqrt{n} [\hat{\lambda}_n(\mathbf{r}) - \lambda_0] - \Sigma^{-1} \frac{1}{\sqrt{n}} \frac{\partial L_n(\vartheta_0)}{\partial \lambda} \right\| = o_p(1),$$

which implies that

$$\sqrt{n} \sup_{\|\mathbf{r}-\mathbf{r}_0\| \leq B/n} \|\hat{\lambda}_n(\mathbf{r}) - \hat{\lambda}_n(\mathbf{r}_0)\|$$

$$\leq \sup_{\|\mathbf{r}-\mathbf{r}_0\| \leq B/n} \left\| \sqrt{n} [\hat{\lambda}_n(\mathbf{r}) - \lambda_0] - \Sigma^{-1} \frac{1}{\sqrt{n}} \frac{\partial L_n(\vartheta_0)}{\partial \lambda} \right\|$$

$$+ \left\| \sqrt{n} [\hat{\lambda}_n(\mathbf{r}_0) - \lambda_0] - \Sigma^{-1} \frac{1}{\sqrt{n}} \frac{\partial L_n(\vartheta_0)}{\partial \lambda} \right\| = o_p(1).$$

By the martingale central limit theorem, we can show that the conclusion holds. \square

References

- Bardet, J.-M., Wintenberger, O., 2009. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.* 37, 2730–2759.
- Berkes, I., Horváth, L., Kokoszka, P., 2003. GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

- Billingsley, P., 1968. *Convergence of Probability Measures*. Wiley, New York.
- Chan, K.S., 1990. Testing for threshold autoregression. *Ann. Statist.* 18, 1886–1894.
- Chan, K.S., 1991. Percentage points of likelihood ratio tests for threshold autoregression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 53, 691–696.
- Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* 21, 520–533.
- Chan, K.S., Li, D., Ling, S., Tong, H., 2014. On conditional heteroscedastic AR models with thresholds. *Statist. Sinica* 24, 625–652.
- Chan, K.S., Petrucci, J.D., Tong, H., Woolford, S.W., 1985. A multiple-threshold AR(1) model. *J. Appl. Probab.* 22, 267–279.
- Chan, K.S., Tong, H., 1985. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.* 17, 666–678.
- Chan, K.S., Tong, H., 1990. On likelihood ratio tests for threshold autoregression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 52, 469–476.
- Chow, Y.S., Teicher, H., 1978. *Probability Theory: Independence, Interchangeability, Martingales*. Springer-Verlag.
- Cline, D.B.H., Pu, H.-M., 2004. Stability and the Lyapunov exponent of threshold AR-ARCH models. *Ann. Appl. Probab.* 14, 1920–1949.
- Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.
- Francq, C., Roy, R., Zakoian, J.-M., 2005. Diagnostic checking in ARMA models with uncorrelated errors. *J. Amer. Statist. Assoc.* 100, 532–544.
- Francq, C., Zakoian, J.-M., 2004. Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli* 10, 605–637.
- Francq, C., Zakoian, J.-M., 2010. *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley.
- Gonzalo, J., Wolf, M., 2005. Subsampling inference in threshold autoregressive models. *J. Econometrics* 127, 201–224.
- Gouriéroux, C., Monfort, A., 1992. Qualitative threshold ARCH models. *J. Econometrics* 52, 159–199.
- Hansen, B.E., 2000. Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
- Koop, G., Potter, S.M., 1999. Dynamic asymmetries in US unemployment. *J. Bus. Econom. Statist.* 17, 298–312.
- Li, C.W., Li, W.K., 1996. On a double-threshold autoregressive heteroscedastic time series model. *J. Appl. Econometrics* 11, 253–274.
- Li, D., Ling, S., 2012. On the least squares estimation of multiple-regime threshold autoregressive models. *J. Econometrics* 167, 240–253.
- Li, D., Ling, S., Li, W.K., 2013a. Asymptotic theory on the least squares estimation of threshold moving-average models. *Econometric Theory* 29, 482–516.
- Li, D., Ling, S., Zhang, R.M., 2013b. On a Threshold Double Autoregressive Model. *J. Bus. Econom. Statist.* (forthcoming).
- Li, W.K., Ling, S., McAleer, M., 2002. Recent theoretical results for time series models with GARCH errors. *J. Econ. Surv.* 16, 245–269.
- Ling, S., 2004. Estimation and testing stationarity for double-autoregressive models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66, 63–78.
- Ling, S., 2007. A double AR(p) model: structure and estimation. *Statist. Sinica* 17, 161–175.
- Ling, S., Li, D., 2008. Asymptotic inference for a non-stationary double AR(1) model. *Biometrika* 95, 257–263.
- Ling, S., Tong, H., 2011. Score based goodness-of-fit tests for time series. *Statist. Sinica* 21, 1807–1829.
- McLeod, A.I., Li, W.K., 1983. Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Series Anal.* 4, 269–273.
- Meitz, M., Saikkonen, P., 2011. Parameter estimation in nonlinear AR-GARCH models. *Econometric Theory* 27, 1236–1278.
- Meyn, S.P., Tweedie, R.L., 1996. *Markov chains and stochastic stability*. Springer-Verlag, New-York.
- Potter, S.M., 1995. A nonlinear approach to US GNP. *J. Appl. Econometrics* 10, 109–125.
- Rabemananjara, R., Zakoian, J.M., 1993. Threshold ARCH models and asymmetries in volatility. *J. Appl. Econometrics* 8, 31–49.
- Shorack, G.R., Wellner, J.A., 1986. *Empirical Processes with Applications to Statistics*. Wiley.
- Tiao, G.C., Tsay, R.S., 1994. Some advances in nonlinear and adaptive modeling in time series. *J. Forecast.* 13, 109–131.
- Tong, H., 1978. On a threshold model. In: Chen, C.H. (Ed.), *Pattern Recognition and Signal Processing*. Sijthoff and Noordhoff, Amsterdam.
- Tong, H., 1990. *Nonlinear Time Series. A Dynamical System Approach*. Oxford University Press, New York.
- Tsay, R.S., 2010. *Analysis of Financial Time Series*, third ed.. John Wiley & Sons.
- Weiss, A.A., 1984. ARMA models with ARCH errors. *J. Time Series Anal.* 5, 129–143.
- Wong, C.S., Li, W.K., 1997. Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika* 84, 407–418.
- Wong, C.S., Li, W.K., 2000. Testing for double threshold autoregressive conditional heteroscedastic model. *Statist. Sinica* 10, 173–189.
- Zakoian, J.-M., 1994. Threshold heteroskedastic models. *J. Econom. Dynam. Control* 18, 931–955.
- Zhang, X., Wong, H., Li, Y., Ip, W.-C., 2011. A class of threshold autoregressive conditional heteroscedastic models. *Stat. Interface* 4, 149–157.